This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3028967, IEEE Access

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version XXXX 00, 0000.

Digital Object Identifier XXX

A Cooperative Multi-Agent System for Crowd Sensing Based Estimation in Smart Cities

DAVIDE ANDREA GUASTELLA [1,2], VALÉRIE CAMPS[2], AND MARIE-PIERRE GLEIZES[2]
[1]Institut de Recherche en Informatique de Toulouse, Université Toulouse III - Paul Sabatier, France (e-mail: {davide.guastella, valerie.camps, marie-pierre.gleizes}@irit.fr)
[2]Università degli Studi di Palermo, Italy

Corresponding author: Davide Andrea Guastella (e-mail: davide.guastella@irit.fr).

ABSTRACT The concept of *Smart City* has spread as a solution to ensure better access to information and services to citizens, but also as a means to reduce the environmental footprint of cities. To this end, a continuous and wide observation of the environment is necessary to analyze information that enables government bodies to act on the environment appropriately. Moreover, a diffused acquisition of information requires adequate infrastructure and proper devices, which results in relevant installation and maintenance costs. Our proposal enables reducing the number of necessary sensors to be deployed while ensuring that information is available at any time and anywhere. We present the HybridIoT system to cope with the lack of environmental information in the urban context through an estimation technique that integrates heterogeneous data acquired from some different sensors. HybridIoT can be deployed in large-scale contexts and ensures data accessibility even if devices enter or leave the system at any time and everywhere. We compare the results to those obtained by state-of-the-art techniques to assess the validity of our proposal, in particular concerning the properties of openness, large-scale, and heterogeneity, of primary importance in the context of the development of systems to be deployed in the smart city context.

INDEX TERMS Smart city, Cooperative Multi-Agent Systems, Missing Information Estimation, Heterogeneous Data Integration

I. INTRODUCTION

The increasing diffusion and accessibility of the *Internet Of Things* (IoT) sensors enabled cities to become urban sensing platforms [1]. Data acquisition, through these platforms, enables cities to become "Smart", using environmental information collected in a participatory way to improve services and reduce their ecological footprint [2]. Environmental information includes information about air, water, soil, land, flora and fauna, energy, noise, waste and emissions, but also information about decisions, policies and activities that affect the environment [3]. Wide dissemination of environmental information such as temperature, humidity, CO_2 enables government agencies to map cities and regions to provide accurate information at any time and everywhere in parts of the environment that are not sufficiently covered by sensors. This could help to assess important information such as pollution [4], hydrological forecasting [5] or traffic estimation [6]. For these reasons, it is necessary to conceive effective solutions that enable leveraging the potential of data perceived through a wide range of sensing devices. However, in large scale contexts is difficult to deploy a large number of devices so as to enable sufficient informative coverage through the urban environment; this is due to high costs of maintenance and installation of the network infrastructure [6].

Thanks to their increasing computational power and accessibility, smart devices can be exploited to make the data acquisition in cities a participatory activity. This is the key concept of *Mobile Crowd Sensing* (MCS), which leverages device mobility and sensing capabilities, as well as human collaboration and intelligence to distributively perform tasks and provide cost-efficient applications and services [7]. MCS enables integrating different types of smart devices into a large scale sensing infrastructure. However, smart devices can embed a limited set of sensors: in this case, it is necessary to compensate for the lack of data through a mechanism for estimating the missing information.

This paper presents the HybridIoT system, based on a

Multi-Agent System (MAS), that enables coping with the lack of environmental information at large scale. MAS are systems composed of multiple interacting and autonomous entities known as *agents*, each one acting and sensing within a common environment. Agents have a partial view of the environment, they act jointly to produce a result for a goal that cannot be achieved individually. Due to the distribution of tasks within the agents composing a MAS and the possibility to decentralize control and decision, MAS are more suitable to model and simulate complex systems than traditional approaches [8]. MAS offer a framework to model, study, and control complex systems with a bottom-up approach by focusing on the entities and their interactions to solve a wide variety of problems [9].

The novelty of our contribution lies in:

- a technique for estimating missing environmental information at large scale whereas *ad hoc* sensors (whose sole purpose is to perceive the environment) are not available;
- limiting the installation of a large number of *ad hoc* sensors by using virtual sensors;
- propose a technique to estimate missing data by integrating heterogeneous information.

The difficulty of conceiving such system is to address the following three properties:

- **openness**: refers to the capacity of handling new input at any time without the need for any reconfiguration;
- **large scale**: the system can be deployed in a large, urban context and ensure correct operation with a significative number of devices;
- **heterogeneity**: the system handles different types of information without any *a priori* configuration.

HybridIoT leverages two types of estimation: **endogenous**, in which historical data of a device is used whereas this is not capable of providing any information from the direct observation of the environment; **exogenous**, estimating missing information by integrating heterogeneous information perceived from different data sources. To motivate the use of exogenous estimation, consider an urban sensing platform used to monitor the environment. The platform can exploit a multitude of sensors capable of perceiving environmental information on a large scale. Sensors can be of different types, acquire heterogeneous information or data of the same type using different units of measurement. In such a context, we hypothesize that the abundance of different information is crucial for both AI and data analysis techniques to extract useful knowledge from the urban environment to achieve the smartness goal of a city. The exogenous estimation is beneficial in the smart city context for two aspects:

- enables integrating heterogeneous information to estimate missing values and to overcome the lack of sensors of the same type;
- the integration of heterogeneous information enables defining a coherent representation of the environment to monitor and to understand its evolution.

Exogenous estimation enables to leverage the large amount of information perceived by heterogeneous devices to provide wide information coverage. This ensures that users can access accurate and timely information on the state of the environment, while experts can use the information to improve services offered to citizens.

Our previous works are part of the neOCampus project [10] and focus on the estimation of missing values in the smart city to reduce the number of sensors deploy and ensure data availability to both citizens and experts. In [11], we propose a MAS to estimate missing information by exploiting the historical values perceived by individual sensors; the proposed solution performs an endogenous estimation. In [12], we propose a MAS to estimate missing information through a mechanism of cooperation between agents associated with devices perceiving homogeneous information (of the same type) and in a local environment. This paper, compared to our previous works, focuses on a new technique for estimating missing information by integrating heterogeneous information whose type is not known in advance.

The rest of this paper is organized as follows: section II discusses some state of the art propositions to address the data estimation and integration in urban contexts according to openness, heterogeneity and large scale properties. Section III presents the HybridIoT system, describing how it is capable of estimating missing information without the need to deploy a large number of *ad hoc* sensing devices. Section IV presents an experimental case and the results obtained from both endogenous and exogenous estimations. In Section VI, we conclude our work and point out some future perspectives.

II. BACKGROUND AND MOTIVATION

This section presents different state of the art techniques to address data estimation and integration in urban context.

This paper proposes a method to estimate environmental information in a distributed way. As the estimation mechanism is the core of our contribution, the sensing platforms [13] that use mobile devices such as smartphones are not detailed in this work. However, these platform tools are complementary to our proposal: an adequate sensing platform enables developing a network infrastructure necessary to accomplish distributed estimation of missing values. The described state of the art methods rather focus on the exploitation of data in urban contexts; these methods are discussed according to openness, heterogeneity and large scale properties.

Ali et al. [14] proposed a data fusion technique to identify anomalous behaviors based on the integration of heterogeneous data: credit card data, loyalty card data, GPS, age and gender-based on face images. The anomaly is detected from the credit card data, loyalty card and GPS data. Images are processed by a pre-trained deep learning network; the classification is done using *Support Vector*

Machine (SVM). Anomaly and image classification results are combined through a data fusion technique. The result of the fusion technique is then ranked in ascending order to identify the anomaly. Finally, the expert verifies the analysis at the final step and produces the final decision based on highly intelligible information, which is a much smaller amount than the raw data. This technique uses a pre-defined set of heterogeneous information. Consequently, it is not possible to enter new types of information into the system without a proper configuration. Deep learning and SVM, used to detect anomalies and pursue classification, provide high performance but require a pre-configuration using the available data. Also, the input of new data may require additional system configuration that cannot be anticipated. Moreover, the scalability problem is not addressed.

Zhu et al. [15] proposed a city-wide air quality estimation technique that uses a limited number of monitoring stations that are geographically sparse. The prediction of air quality index is divided into three stages: the first stage interacts with input data flows, from both online and historical data; the second stage deals with non-causality detection, which is based on the spatiotemporal extended Granger causality model; the last stage, the online prediction stage, is based on the use of neural networks. A fine-grained air quality map is generated after inferring the air quality index. Authors do not consider the openness, therefore it is not possible to introduce data acquired from new sensors during the operation of the system. Moreover, the proposed solution focuses only on the estimation of the air quality index, thus the heterogeneity is not addressed.

Shan et al. [16] proposed a multi-sensor data fusion technique based on a *Multiple Linear Regression* (MLR) model to improve the accuracy of traffic state estimation. The contribution is twofold: firstly, MLR is used to analyze large sets of GPS data obtained by monitoring taxies' position to extract the inherent spatiotemporal correlations of traffic states for specific road segments. Secondly, authors proposed an information fusion technique to extract correlations from incomplete data to increase the performance of traffic state estimation.

Tomaras et al. [17] proposed HELIoS (HEalthy LIving Smart), a framework that combines heterogeneous information such as urban traffic and pollution data to diagnose the health state of urban areas in a smart city. Data has been acquired hourly from 3402 heterogeneous fixed sensors deployed in different urban areas. The method uses three regression models to predict the health state of urban areas:

- **Support Vector Regression** (SVR): a regression variant of the SVM classifier;
- **Random forest**: a set of multiple decision trees where each tree is trained using a different subset of the training set;
- **Gaussian process**: a non-linear and non parametric model that is an extension of the multivariate Gaussian distribution for an infinite collection of real value variables.

The technique is not designed to be deployed at large scale and is not able to face unpredictable events such as sensors malfunction or unavailability.

Table I lists the described methods along with their strengths and weaknesses according to the following properties: openness, large-scale, heterogeneity. HybridIoT is different to the presented technique because it addresses all the three properties at the same time. Moreover, our proposal is based on the AMAS approach which allows distributing the computation among several sensing devices, therefore our technique is scalable.

We used four indicators to depict strengths and weaknesses of each described method: (++) a challenge has been discussed and authors describe a precise method to address it, (+) a challenge has been discussed and addressed but authors did not provide a precise explanation of the solution, (–) the challenge has been mentioned but not addressed, (– –) the challenge was neither mentioned nor discussed.

TABLE I: Comparison of state of the art approaches for heterogeneous data integration and estimation.

	Openness	Heterogeneity	Large-scale
Ali et al. (2018)	– –	++	– –
Zhu et al. (2015)	– –	++	+
Shan et al. (2016)	– –	++	+
Tomaras et al. (2018)	– –	++	– –

Section III discusses the HybridIoT system and how the properties described previously are addressed.

III. PROPOSED SOLUTION

This section presents HybridIoT, a MAS based technique for estimating missing environmental information at large scale environments.

MAS are proven to be an effective solution for estimating missing information at large scale environments [12]. The MAS paradigm offers an intuitive and natural way to solve a complex problem by means of a distributed computation between interacting autonomous agents that have specific and limited tasks. An agent can be defined as a computer system that is situated in some environment, and that is capable of autonomous action in this environment in order to meet its design objectives [9]. An agent has a partial view of the environment in which it is located. As a result, it has a limited control, it can only influence a part of its environment (including other agents) through its actions. An autonomous agent does not need external interventions to modify its behavior. Agents operate jointly to achieve a global objective that cannot be pursued individually [18]. Agents have different functionalities, they can enter or leave the environment at any time, they can interact with other agents. When interacting, agents autonomously constitute organizations in which they operate jointly. The activity of agents in emerging organizations enables MAS to be an effective solution to solve complex problems, where the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3028967, IEEE Access

solution to a problem results from the interactions between the subsystems with which the agents are associated [19].

Using the MAS approach enables IoT devices to include collective intelligence mechanisms: the computation of IoT devices is no more individual, rather a joint and collective activity between devices to address the complexity of the physical environment. The interaction between the agents allows solving problems such as momentary unavailability or lack of sensors. Openness and dynamics are sources of unexpected events and an open system plunged into a dynamic environment needs to be able to adapt to these changes, to self-organize [20]. In the AMAS approach, agents can adapt themselves and change their behavior according to the evolution of the environment in which they are situated.

To further motivate the choice of the MAS approach, consider a centralized distributed system for estimating missing information using a large number of IoT sensing devices in an urban context. The system has the following tasks:

- communicate with all devices available in the environment,
- acquire the information from the available mobile devices,
- evaluate estimates whereas sensors are not available or able to perceive the environment. The system should evaluate a subset of sensors whose values must be used to estimate the missing information.

On a large scale context, several thousand devices could face unpredictable situations in which information has to be estimated and provided to both users and experts. A centralized implementation of such a system requires a significative computational power.

HybridIoT exploits both homogeneous and heterogeneous information to provide accurate estimates whereas devices are not available for an area of the environment not covered by *ad hoc* sensors. A group of information is characterized as homogeneous if it is composed of information of the same type. Contrarily, a group of heterogeneous information contains information of different types, not necessarily correlated.

The rest of the section is organized as follows: Section III-A provides some definitions required for the presentation of HybridIoT; Section III-B presents the general operation of the proposed system and an example to motivate its use at large scale contexts; Section III-C describes how HybridIoT estimates missing information by an endogenous process; Section III-D describes how our system estimates missing information by an exogenous process that is subsequent to the endogenous one.

A. DEFINITIONS

The following definitions provide an overview of the main features of HybridIoT. The definitions are valid for both endogenous and exogenous estimation schemes, presented afterward.

*Definition 1 (**Ambient Context Window**):* an *Ambient Context Window* (ACW) C_t contains homogeneous environmental information perceived in a discrete time interval $T = [t-\delta, t],\ t-\delta < t$. An ACW has $|C_t| = |T|$ homogeneous context entries, one for each time instant. Each ACW C_t is associated with an index t that corresponds to the time instant in which the information at time t has been perceived.

*Definition 2 (**Context Entry**):* a *Context Entry* $E^i_t \in \mathbb{R}$ is a numerical information perceived at time $t \in T$, where T is the time window of the C_i. The value of a context entry can be any type of environmental information such as temperature, humidity, lightness etc.

*Definition 3 (**ACW Distance**):* the distance between two ACWs is defined as the absolute difference in time between the context entries divided by the number of entries γ of the two ACWs. The smaller the difference is, the more similar two ACWs are. The context distance between two ACWs C_t and C_k is defined by the following formula:

$$d\left(C_t, C_k\right) = \frac{\sum_{l \in [1,\gamma]} |E^t_l - E^k_l|}{\gamma} \tag{1}$$

where $\gamma = |C_t| = |C_k|$, l is an index in the range $[1, \gamma]$, E^t_l and E^k_l are the context entries of index l in the ACW C_t and C_k respectively.

The (1) is generic as it does not consider the unit of the information used, therefore it can be used for any type of numerical environmental information.

The distance d satisfies the following properties:

- $d(C_t, C_k) \geq 0$,
- $d(C_t, C_k) = 0 \iff C_t = C_k$,
- $d(C_t, C_k) = d(C_k, C_t)$,
- $d(C_t, C_p) \leq d(C_t, C_k) + d(C_k, C_p)$

where C_t, C_k, C_p are ACWs for information times t, k, p respectively. Therefore, d is a metric.

*Definition 4 (**Ambient Context Agent**):* an *Ambient Context Agent* (ACA) is associated with a sensing device; its goal is to provide environmental information in a local part of the environment. Thanks to cooperative behavior, ACAs are capable of providing estimates even if *ad hoc* sensors are unavailable. Each ACA is characterized by a knowledge base containing ACWs. Each ACW belongs to a unique ACA.

*Definition 5 (**Sensing Device**):* a sensing device is any physical instrumentation that embeds sensors capable of detecting events and changes in its environment. Sensing devices can be either fixed or mobile: fixed sensing devices are installed in specific positions and enable continuous monitoring of the environment. They acquire information continuously and store these in dedicated databases. Mobile sensing devices can embed a variable number of sensors: the information can be acquired on-demand by the user (the owner of the device) or using a fixed schedule.

B. GENERAL OPERATION

This section presents the general operation of the HybridIoT system. Fig. 1 shows the main steps of the proposed system.

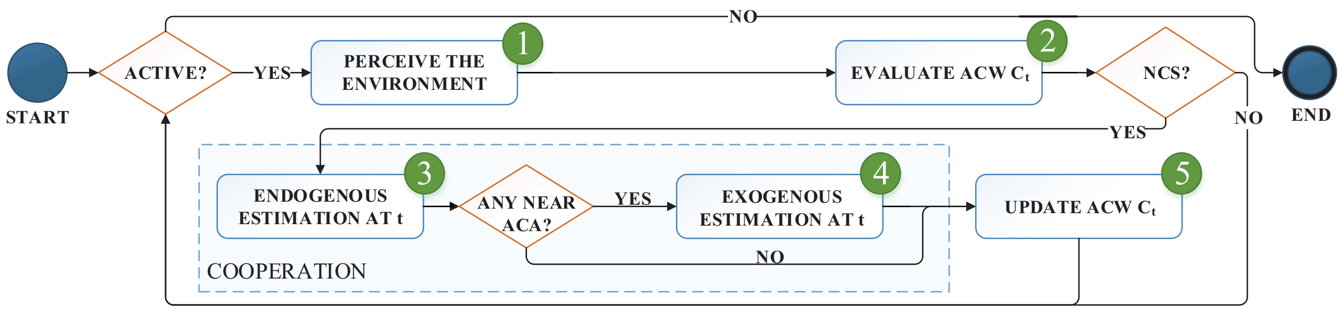


FIGURE 1: Main steps of the proposed system.

Let ACA_i be the ACA associated to the i-th sensing device in the environment. If the associated sensing device is available, the associated ACA_i is able to perceive the environment, thus a new information perceived at time t is presented to the agent (step ❶). Following, a new ACW_t is created and associated with the perceived information (step ❷). The new ACW is added to the set of ACWs of the ACA_i. Each ACW contains a variable number of entries; Section III-B1 describes the process pursued by agents for determining ACWs of variable sizes.

If the information at time t is not available due to the unavailability of the sensing device, the ACA_i generates an exception as it is not able to provide any information. This exception is solved by exploiting the *Adaptive Multi-Agent System Approach* (AMAS) [19]. In this approach, an exception is considered as a *Non Cooperative Situation* (NCS) that has to be solved in a local and cooperative way. In HybridIoT, ACAs can encounter two types of NCS:

- **incompetence NCS**: an ACA cannot provide an environmental information (because no *ad hoc* sensor is available or it encountered a problem);
- **uselessness NCS**: an ACA cannot help the agent that is faced to an incompetence NCS because the distance between their ACW is not sufficiently low.

ACAs solve the incompetence NCS by pursuing two types of estimation processes:

- **endogenous estimation**: the ACAs cooperate with agents that perceive the same type of information as the ACA that encountered an incompetence NCS (step ❸);
- **exogenous estimation**: the ACAs cooperate with agents that perceive heterogeneous information (step ❹).

The idea behind the cooperative behavior is that the ACA_i cooperates with other agents whose ACWs, independently from their type, can be used to estimate the missing information.

The comparison between ACWs allows ACAs to discover similar dynamics in the evolution of the information previously acquired. The ACA_i that has encountered a NCS, identifies the ACAs whose ACWs are similar to C_t and can be used to estimate the missing information at time t.

The estimation of the missing information considers the last perceived information weighted by a value w_t obtained from a cooperative process among ACAs. This weight represents the variation that the estimated information at time t assumes with respect to the last information perceived at time $t-1$. The value w_t is added to the last information perceived by ACA_i at time $t-1$; we assume that consecutive environmental information does not vary significantly unless there is noise or unpredictable environmental factors.

Once the information at time t has been estimated, the ACW C_t is updated by the ACA_i (step ❺).

To motivate the application of exogenous estimation, let us consider the urban context representation in Fig. 2. The temperature sensor ⑤ encounters an incompetence NCS as it is not able, at a given time, to provide a temperature value by directly observing its environment. The only homogeneous ACAs with which the ACA ⑤ could cooperate are the ACAs ⑥ and ⑦. To estimate the missing value, ACA ⑤ searches in its history for ACWs that are sufficiently similar to the last observed ACW containing the information to be estimated. Following, ACA ⑤ cooperates with agents ⑥ and ⑦ (step ❸) to estimate the missing value.

As shown in Fig. 2, ACAs ⑥ and ⑦ may be outside the part of the environment where the dynamics of information are similar to those observed by ACA ⑤. In the neighborhood of ACA ⑤ there are four ACAs, respectively ①, ②, ③ and ④. These ACAs perceive humidity and CO_2 information. Although the information perceived by these sensors cannot be used directly to estimate a temperature value for ACA ⑤, they cooperate with ACA ⑤ to indicate how the last perceived information of the respective sensors has varied over time.

This indication is at the basis of the exogenous estimation process, which enables calculating estimates using different types of information. The idea behind the exogenous estimation process is the following: the ACA ⑤ that encountered an issue in providing information, searches for similar ACWs in its history: each ACW is relative to a precise time instant. Let ξ be the set of ACWs similar to C_t (containing the information to be estimated), which similarity is calculated by (1). The ACA that encountered a problem starts a cooperative process with the other ACAs providing them with set ξ (step ❹), independently of the information perceived.

The other ACAs calculate a weight obtained by comparing

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3028967, IEEE Access

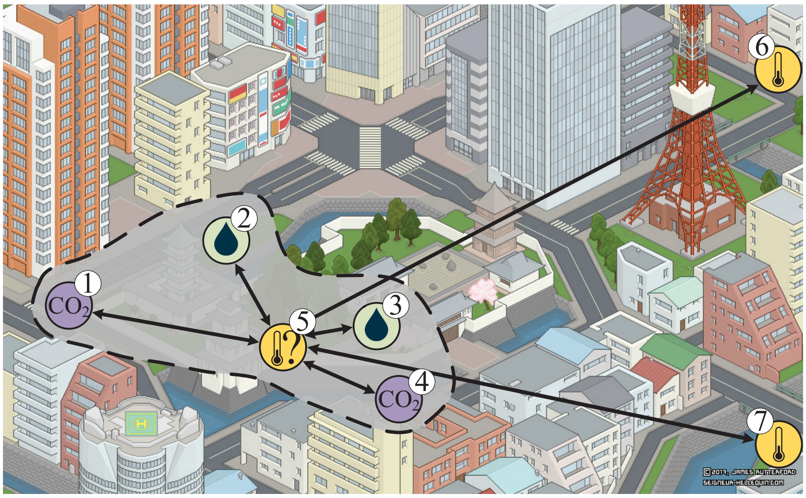


FIGURE 2: Example of exogenous estimation: temperature sensor ⑤ cooperates with heterogeneous (①-④) and homogeneous (⑥ and ⑦) sensors in order to provide an accurate estimate of the missing value. The dashed area represents the part of the environment that groups sensors pertinent for helping sensor ⑤ to estimate a missing information. Image source: seigneur-hellequin.com.

their ACWs at the time instants indicated by the set ξ. The ACA that encountered a NCS uses the weights provided cooperatively by other agents to adjust the weight that must be added to the last perceived information (at $t-1$). Although the weights have been obtained from agents that perceive different types of information, exogenous estimation allows exploiting a large quantity of information that may be available through sensing devices deployed in the environment.

Thanks to the exogenous estimation, agents cooperate not only with ACAs that perceive different information but also with those perceiving information of the same type with different units or scales. This is possible because each ACA compares the ACWs in its knowledge base. The information exchanged between agents refers to comparisons between ACWs of the same type, containing the same type of information. Let us consider that an agent ACA_c perceives temperature in degree Celsius and ACA_f in degree Fahrenheit. If ACA_c has to calculate an estimate for missing information, it cooperates with ACA_f to estimate the missing value. ACA_f compares its ACWs, containing temperature values in degree Fahrenheit, and calculates a weight that represents the variation that the next value (that is, the estimate) must assume with respect to the last perceived information. Because the weight is a variation value, it does not depend on the unit and can be used by ACA_c to estimate the missing value in degree Celsius.

The use of a distance measure d defined in (1) that considers ACW of the same size has two advantages:

- an ACA retrieves more quickly the ACWs, in its knowledge base, similar to the one containing the information to be estimated;
- ACWs containing similar information dynamics are more likely to have the same size. The distance measure d, therefore, allows ACAs to calculate estimates of missing information by observing information dynamics similar to those of the ACW containing the information to be estimated.

Fig. 3 shows an example of how cooperation takes place between ACA ⑤ and other ACAs. The ACA ⑤ has encountered an incompetence NCS as it is not able to provide information. The ACA ⑤ performs an endogenous estimation by cooperating with ACAs ⑥ and ⑦ (step ❸). Following, an exogenous estimation is performed by cooperating with ACAs ①, ②, ③ and ④ (step ❹), that perceive information of different types with respect to ACA ⑤. Once the ACA ⑤ obtained the weights through cooperation, some ACAs can encounter a uselessness NCS as their cooperation is not helpful to provide an estimate. This is because the ACA ⑤ discriminates the obtained weights to be used in order to provide an accurate estimate.

1) Evaluating Dynamic Size ACWs

ACAs learn the environmental dynamics through the definition of ACWs describing the evolution of information in discrete time windows of limited size. The information on environmental dynamics is fundamental in estimating missing values; in fact, the estimation mechanism and the mechanism for defining ACWs are interdependent: the former requires a sufficient number of ACWs to estimate missing information, the latter needs information (either real or estimated) to define ACWs.

Sensors perceive information at frequencies not known in advance. On the one hand, windows containing a small amount of information with no significant variations have no particular semantic significance, so their use for estimating missing value would be irrelevant. On the other hand, too much information makes the ACWs no specific of the time

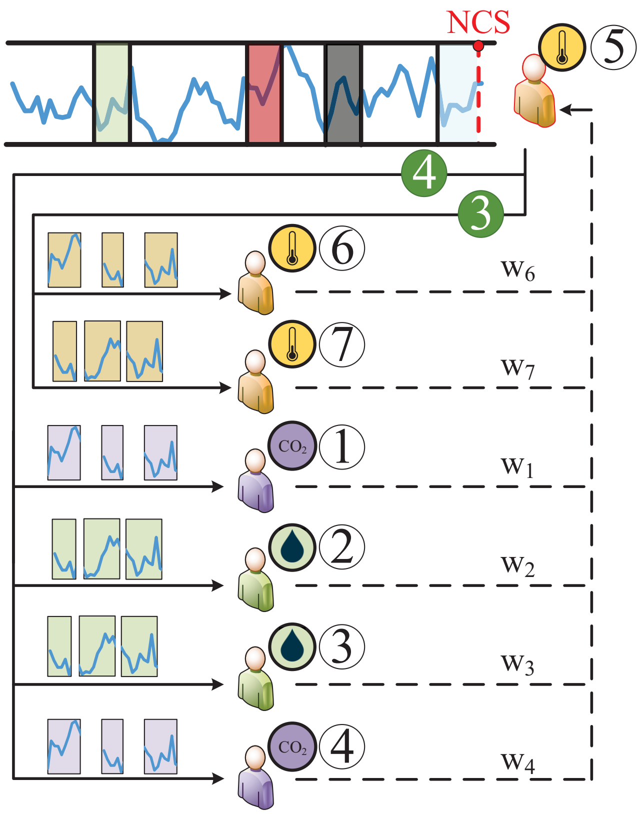


FIGURE 3: ACA ⑤ cooperates with other agents by performing both endogenous and exogenous estimations. The weights obtained cooperatively are used to provide an estimate of the missing value. Steps ❸ and ❹ are the same as in Fig. 1.

interval they describe. The mechanism to determine ACWs of variable size allows overcoming this problem; agents autonomously determine the size that each ACW must have so that it is useful for estimating missing information in the presence of similar information dynamics [11].

This section describes how agents determine ACWs with a variable number of entries.

Consider two ACWs referring to the same time instant t, but of different sizes, $\mathrm{C}_t^{(4)}$ and $\mathrm{C}_t^{(7)}$, containing respectively 4 and 7 entries in the time interval $[t-4, t]$ and $[t-7, t]$. Here we suppose that the agent has previously collected enough information to calculate both ACWs. The agent evaluates the estimate of information t by simulating its absence using both $\mathrm{C}_t^{(4)}$, $\mathrm{C}_t^{(7)}$. This results in two different estimates that are compared to the real value. The ACW minimizing the bias, that is the discrepancy between the estimate and the real value, is defined as representative of the information at t.

In HybridIoT, the agents calculate to 15 different ACWs for determining the one to be associated with specific information. Let Σ_t be the set of 15 most similar ACWs to C_t for an arbitrary ACA. The ACA calculates as many estimates as the number of ACW in Σ_t. Because the ACAs compare ACW containing the same number of entries, each $C \in \Sigma_t$ is compared to different disjoint sets of ACWs; this implies that for each ACW in Σ_t, the technique calculates different estimates. After calculating one estimate for each ACW in Σ_t, the agent chooses the context window that minimizes the bias, that is, the one that provides an estimate which value is the closest to the perceived information.

After experimentations in our previous work, we observed that the amount of information depends on the data acquisition frequency of the sensor: a high frequency (e.g. every 10 seconds) leads to ACWs containing a significative amount of information; on the other hand, a low frequency (e.g. daily) leads to the formation of ACWs containing a small amount of information. In the first case, it is necessary to have a sufficient amount of information as the variability between consecutive information is low. In the second case, consecutive information has high variability, therefore ACWs containing much information would lead to inaccurate estimates.

Following Section III-C presents the endogenous estimation process carried out by ACAs.

C. ENDOGENOUS ESTIMATION

This section describes the endogenous estimation scheme where ACAs rely on homogenous data to provide estimates for missing information.

The endogenous estimation of missing information is divided into two steps:

1) **Cooperative weights evaluation**: ACAs evaluate the weights to be used for estimating missing information;
2) **Missing information calculation**: ACAs estimate missing information using the calculated weights.

1) Cooperative Weights Evaluation

When ACA_i is not able to perceive the environment through a sensing device, an incompetence NCS occurs. Let t be the time instant at which the information needs to be estimated by ACA_i. For the sake of simplicity, we assume that ACA_i has assembled the ACWs for the information before t.

The ACA_i evaluates a subset $\xi^{(t)}$ containing the ACWs that minimize the distance $d(C_t, C_k), \forall C_k \in \xi, k \neq t$ where C_t is the ACW containing the information to be estimated at time t and $C_t \notin \xi^{(t)}$.

The ACA_i evaluates a weight w_t obtained as a weighted average of the difference between the last two values of each context $C_k \in \xi^{(t)}$. The distances between C_t and the ACWs in ξ are used to calculate the weight w_t: the smaller the distance between an ACW $C_k \in \xi^{(t)}$ and the ACW C_t, the more likely the missing information at time t will be similar to the last value of the ACW C_k.

The calculation of w_t is the average of the differences of the last two context entries for each ACW $C_k \in \xi$ by using the distance $d(C_k, C_t), \forall C_k \in \xi$. The value w_t is calculated in such a way to be independent from the sampling rate of the data. The weight w_t is computed as follow:

$$w_t = \frac{\sum_{C_k \in \xi^{(t)}} \left(E_\ell^k - E_{\ell-1}^k\right) \cdot d\left(C_t, C_k\right)}{\sum_{C_k \in \xi^{(t)}} d\left(C_t, C_k\right)} \tag{2}$$

where C_t is the ACW containing the information to be estimated at time t, $C_k \in \xi^{(t)}$, $|\xi^{(t)}| = 10$, is the k-th most similar ACW to C_t and E_ℓ^k and $E_{\ell-1}^k$ are respectively the ℓ^{th} and $(\ell-1)^{th}$ context entries of the ACW $C_k \in \xi^{(t)}$, namely the last two context entries of C_k.

The set $\xi^{(t)}$ contains a maximum of 10 ACWs. We observed through experiments that HybridIoT is capable of estimating accurate information even with a limited number (10) of ACWs. Moreover, the use of a limited number of ACWs is advantageous because it enables avoiding noise in the estimation process.

2) Missing Information Calculation

Let C_j be the ACW of the ACA$_i$ containing the information to be estimated at time t. The estimated context entry E_t^j is computed as follows:

$$E_t^j = E_{t-1}^j + w_t \tag{3}$$

where E_t^j is the estimate for the missing information at time t, $E_{t-1}^j \in C_j$ is the last information perceived by ACA$_i$, w_t the weight obtained through a cooperation behavior between ACA$_i$ and the other available agents. The weight w_t is calculated by (2). Equation (3) considers the last information acquired because we assume that environmental information perceived in consecutive temporal instants do not present relevant changes. If there is no other agent in the neighborhood of the ACA, then E_t^j is the estimated value for the missing information at time t. Finally, the ACW C_j is updated by adding the context entry E_t^j.

This section presented the endogenous estimation scheme carried out by ACAs. The MAS approach enables the estimation of accurate environmental information without the need to install a large number of sensors in an urban context. The endogenous estimation scheme enables estimating missing information instantaneously by using a multitude of information acquired through the available sensing devices.

Section III-D presents the exogenous estimation scheme for estimating missing information using heterogeneous information.

D. EXOGENOUS ESTIMATION

The exogenous estimation is pursued after the endogenous process and makes use of heterogeneous data to provide accurate estimates. Exogenous estimation addresses the lack of a sufficient number of sensing devices providing homogeneous information.

The exogenous estimation process is divided into three steps:

1) **Evaluation of the set of cooperating ACAs**: an ACA determines the set of ACAs with which cooperates to estimate missing information;
2) **Cooperative evaluation of the set of weights**: the ACAs cooperatively evaluate the weights to be used for estimating the missing information;
3) **Evaluation of the estimate**: the estimate for the missing information is calculated using the obtained weights.

1) Evaluation of the Set of Cooperating ACAs

The Exogenous estimation is based on a cooperative process between different ACAs. However, a large number of ACAs can be present in the environment, some of which are not relevant for the calculation of the estimate. An ACA uses one of the following criteria to choose the ACAs to cooperate with:

- **Nearest ACAs**: the ACA cooperates with the nearest ones in the environment (the Euclidean distance is used);
- **Most confident ACAs**: the ACA cooperates with those which have the highest degree of confidence.

The Nearest ACAs criterion enables to choose ACAs that are in the immediate proximity of the ACA that has encountered a NCS. The use of this criterion can be very efficient in open environments where no barriers are present (such as weather stations). For example, this criterion enables to choose ACAs both inside and outside a building because of their proximity, even if they could perceive information that differs significantly.

The most confident criterion enables selecting a set of ACAs that do not necessarily have to be in the immediate proximity: ACAs are chosen according to the similarity between the information perceived by the agents.

2) Cooperative Evaluation of the Set of Weights

Let t be the time instant at which the ACA$_i$, associated with the i-th sensing device, has to estimate the information, and $\Gamma^{(t)}$ be the set of agents the ACA$_i$ cooperates with, evaluated according to one of the previous criteria.

Let $\xi^{(t)}$ be the set of ACWs chosen by the ACA$_i$ to estimate the missing information at time t, $\Sigma^{(t)} = \{\sigma_1, \sigma_2, ..., \sigma_k\}$ with $k = |\xi^{(t)}|$ the indexes of the ACWs chosen by ACA$_i$, where $\sigma_i \in \Sigma^{(t)}$ is the index of the ${\sigma_i}^{th}$ ACW in $\xi^{(t)}$. In the exogenous estimation process, the set $\Sigma^{(t)}$ is communicated to the other ACAs in $\Gamma^{(t)}$. In this way, the ACA$_i$ provides to each ACA in $\Gamma^{(t)}$ an indication of which are the temporal instants associated with ACWs that are similar to C_t, that is, the ACW containing the information to estimate at time t. Each ACA in $\Gamma^{(t)}$, therefore, evaluates the distance between the ACW observed at the time instant t and the ACWs whose indices are indicated by the set $\Sigma^{(t)}$. Each ACA calculates a weight by using the (2).

The cooperation between ACAs, therefore, yields a set W of weights, one for each ACA in $\Gamma^{(t)}$. The weights in W are calculated by ACAs that perceive both homogeneous and heterogeneous information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3028967, IEEE Access

3) Evaluation of the Estimate

Once the ACA$_i$ obtains a set W of weights through cooperation with other agents, it evaluates one estimate for each weight $w \in W$, resulting in a set of estimates. To avoid taking into account information that is not relevant to the output estimate, the ACA$_i$ evaluates a histogram H obtained from an empirical cumulative distribution of the estimates. The histogram H includes the estimates obtained from both endogenous and exogenous estimation. The ACA$_i$ selects the bin which average value is closest to the estimate obtained by the endogenous process. Then the agent calculates the average value of the elements within the bin. The result of the average operation is returned as the estimate for the missing information at time t.

Let E_{end} be the estimate obtained by ACA$_i$ through the endogenous estimation process (calculated by using (3)). Let f be a function that returns the mean of the values contained in a bin H_k of the histogram H such that the difference between H_k and the endogenous estimate E_{end} is minimized:

$$f(H_k) = |\overline{H_k} - E_{end}| \tag{4}$$

where $\overline{H_k}$ is the average of the values in the k-th bin of the histogram of frequencies H. The ACA$_i$ evaluates the estimate E_{exo} as follow:

$$E_{exo} = \frac{\text{argmin}_k \, f(H_k) + E_{end}}{2}. \tag{5}$$

The average operator in (5) enables to weigh equally the results obtained from endogenous and heterogeneous estimation.

Fig. 4 resumes the cooperative estimation between ACA$_i$ and the other agents. For the sake of simplicity, we show one lane for all the agent with which the ACA$_i$ cooperates, as the cooperative behavior is identical among the agents.

E. SYNTHESIS

This section presented the HybridIoT system for estimating missing information in urban contexts through both endogenous and exogenous estimations.

Thanks to cooperative behavior, the ACAs estimate missing data using a large amount of data acquired from sensors in the urban context. HybridIoT does not use any specific data fusion technique for heterogeneous data; this is an advantage because the system can use information whose type is not known in advance. Nevertheless, heterogeneous information cannot be used directly for estimating a missing value. For example, it is not possible to estimate wind speed information using only humidity values. To estimate missing information using heterogeneous values, agents indicate the trend that the missing information should assume with respect to the last perceived value. A multitude of indications, resulting from the presence of numerous heterogeneous sensors, allows the ACA that encountered a NCS to accurately estimate the missing values.

Section IV discusses the evaluation and the results obtained by HybridIoT. We pursue multiple experiments to validate both endogenous and exogenous estimations carried out by ACAs on a real dataset.

IV. EXPERIMENTAL RESULTS

After discussing the proposed technique implemented by ACAs for estimating missing information, this section presents the results obtained by HybridIoT on a real weather dataset. Section IV-A discusses the dataset, describing the type of information, its size, the frequency of acquisition and where data has been acquired. Following, the section describes the k-fold technique used to validate the obtained results. Then, we provide some details of the machine used to validate the proposed solution. Two experimental cases are presented: endogenous (Section IV-C) and exogenous (Section IV-D) estimations. In the former case, we evaluate our proposal using only one type of information (for instance, the temperature). Although HybridIoT can estimate missing information from continuous data of any type, we focus on temperature values to evaluate the accuracy of the proposed technique. We compared the results obtained by endogenous estimation to those of state of the art techniques. In the second experimental case, we evaluate the exogenous estimation using the same dataset but estimating missing temperature values using different types of information. Section IV-E discusses the obtained results.

A. DATASET

We used a dataset containing real environmental information acquired in the Emilia Romagna region in Italy [21]. In

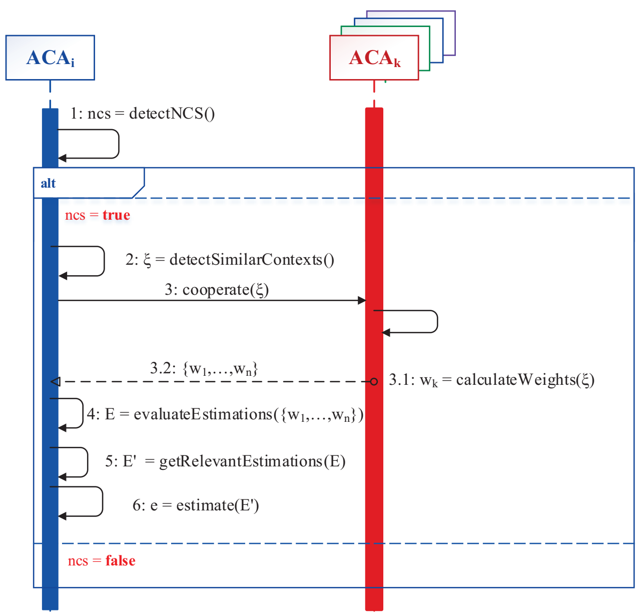


FIGURE 4: Steps of the cooperation between the ACA that detected a NCS and the subset of neighbors agents with which the ACA cooperates.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3028967, IEEE Access

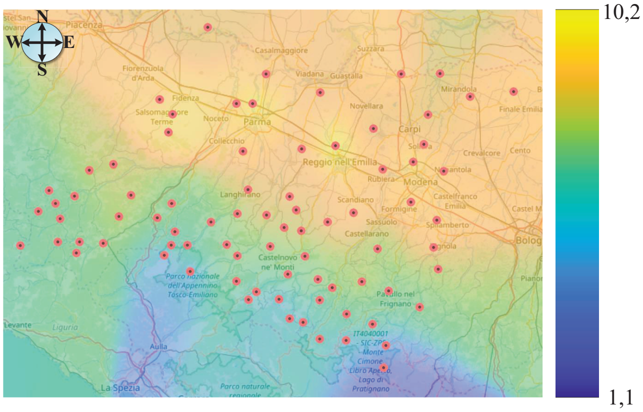


FIGURE 5: Distribution of the temperature values (degree Celsius) on the Emilia-Romagna region (obtained through OpenStreetMap [22]) obtained from the dataset of 80 sensors. The x-axis reports the longitude, the y-axis the latitude.

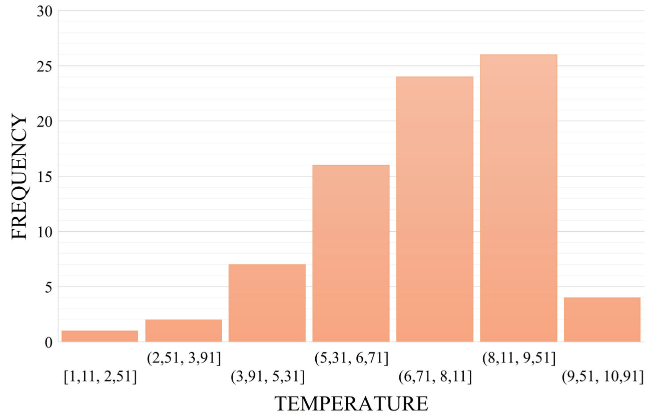


FIGURE 6: Histogram of frequencies for the temperature values from the dataset of 80 weather sensors.

Emilia Romagna region, the prevailing climate is temperate subcontinental, with hot and humid summers followed by cold and harsh winters. Orographically the region is divided almost symmetrically between the Po Valley and the hills and mountains of the northern Apennines.

We considered the average daily air temperatures at 2 meters of altitude, the average daily solar irradiance, the average daily wind speed at 10 meters of altitude and the average daily air humidity at 2 meters of altitude. Data have been collected in 196 days from September 8 2017 to April 25 2018.The days when some sensors were not operational are not considered. We used two versions of this dataset for the experimentations: (*i*) a version containing all the types of information previously listed, that consists in an array of 8112 numerical values acquired from 9 weather sensors; (*ii*) a version containing only temperature values, which results in an array of 15680 numerical values acquired from 80 weather sensors. This difference in terms of the number of samples and sensors, between the two datasets versions, is because for different days not all the information has been acquired correctly by each sensor.

Fig. 5 shows the distribution of mean temperature values for the first version of the dataset, obtained through a standard normalized convolution: this is a non-direct methodology widely used for filtering incomplete or uncertain data which is based on the separation of both data and operator into a signal part and a certainty part [21]. Fig. 5 shows that temperatures are uniformly distributed over the territory except in urban areas (center, north-west) where the temperatures are more intense than in the rest of the region, especially in the south-east part, closer to the Mediterranean sea.

Fig. 6 shows the histogram of temperature values frequencies for the first version of the dataset, containing information acquired by the 80 sensors. The histogram shows that on average the temperature perceived by the sensors in the region is between 6.72°C and 9.51°C.

Fig. 7 shows the distribution of radiation, relative humidity, wind speed, and temperature for the dataset obtained from 9 weather sensors. Here, the wind speed is more intense in the south-east part of the region, which is closer to the Mediterranean sea, and as in Fig. 5, the temperatures are higher in urban areas.

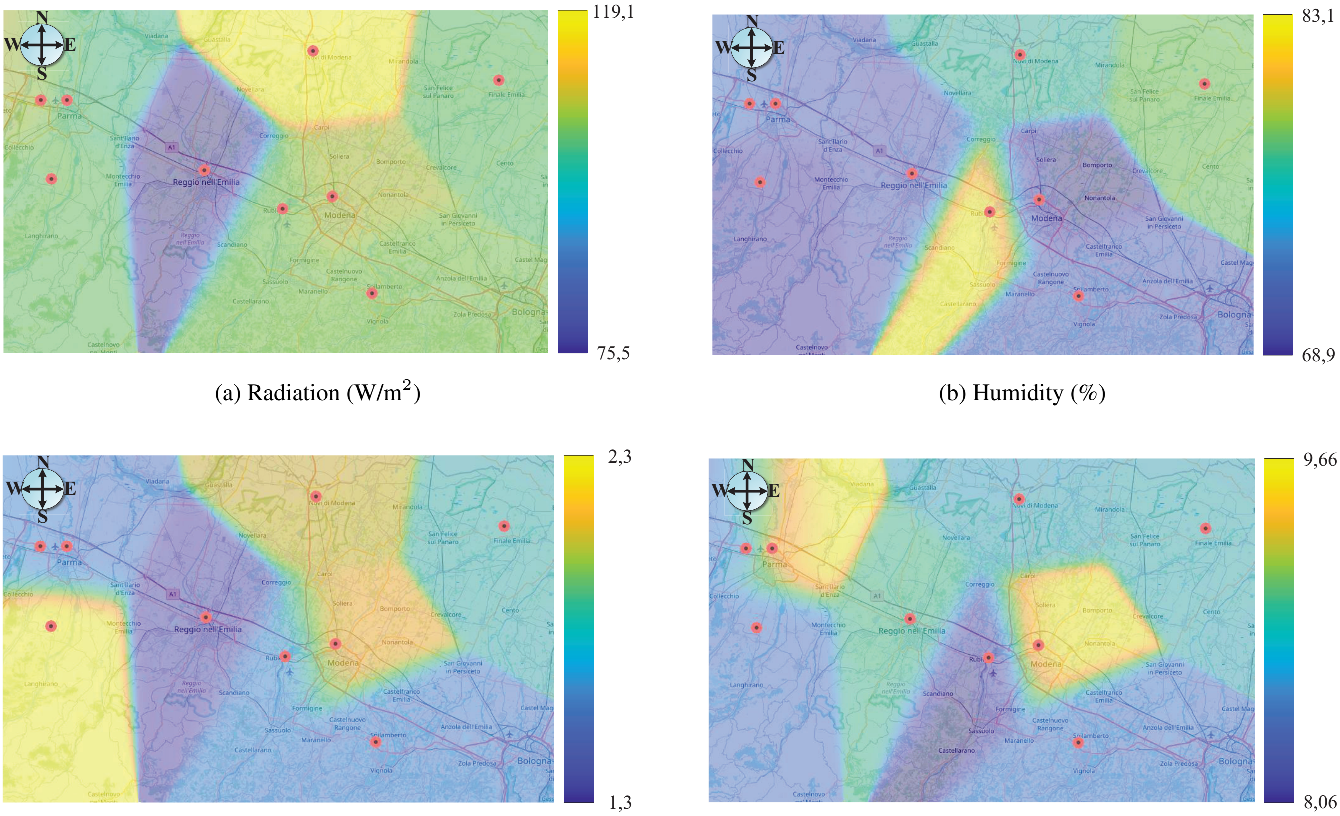
We do not report the histograms of the values for the dataset obtained from the 9 sensors because the information content is limited and the graphs would be superfluous.

B. EVALUATION METHOD

We used the k-fold cross-validation to evaluate the accuracy of the obtained results. This validation technique partitions the original sample in k subsamples. Among the k subsamples, the k-th one is retained as the validation data, for which estimated data is compared to the real information; the remaining $k-1$ subsamples are used as training data. During the training phase, the ACAs assemble the ACWs for the available data. During the test phase, the information from the k-th partition is estimated, thus simulating the unavailability of the sensor. The test phase is then repeated k times, with each of the k subsamples used exactly once as the test data.

Fig. 8 shows the pipeline used to evaluate the results obtained by HybridIoT. The first step of the validation consists in using the training partitions to define the ACWs to be used for estimating the missing data (step ❶). The k-th partition is being validated by estimating its data (step ❷) to simulate the unavailability of the sensor. During this step, the agent pursues an endogenous estimation. Then, the ACA determines the set of the other agents with which it can cooperate to provide an accurate estimate (step ❸). Finally, the ACA pursues an exogenous estimation by cooperating with the available agents (step ❹). The steps ❶ to ❹ are repeated for each partition. The estimates provided for all the partitions are merged to compare the results to the real values (step ❺). The average error is calculated by comparing

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3028967, IEEE Access



(a) Radiation (W/m^2)

(b) Humidity (%)

(c) Wind speed (m/s)

(d) Temperature (°C)

FIGURE 7: Distribution of radiation, humidity, wind speed and temperature values in the Emilia-Romagna region (obtained through OpenStreetMap [22]). The red markers represent the position of the weather sensors. The x-axis reports the longitude, the y-axis the latitude.

the estimated and the real values for each sensor of the dataset (step ❻). In our experiments, we used a k value of 5, 10 and 15. For the dataset containing only homogeneous information, the values of k chosen (5, 10, 15) yield to test partitions containing respectively 20%, 9% and 6% of data with respect to the size of the dataset. No significant differences in terms of error were observed using different values of k.

All the experiments were carried out on a machine equipped with i7 − 7820HQ, 32GB RAM and Windows 10. With such a configuration, estimating a missing value is practically instantaneous. The proposed solution has been developed in Java language and simulates a multi-agent system distributed in a large scale environment. To evaluate the performance of the solution with respect to the quality of the estimation of missing data, we did not consider additional computational overhead such as communication costs between agents. Our implementation does not make use of any particular agent-based technology.

The following subsection presents the results obtained by HybridIoT using the endogenous estimation method on the introduced dataset.

C. ENDOGENOUS ESTIMATION

We carried out both cooperative and non cooperative evaluation: in the first case, the ACAs estimate missing information by cooperating with the other agents, while in the second case the agents use only their ACWs to estimate missing information. When using cooperation, the ACAs choose the ACAs with which it cooperates according to two criteria:

- *Nearest Agents*: ACAs cooperate with the nearest with respect to their position;
- *Most Confident Agents*: ACAs cooperate with those that have a high confidence value.

In order to assess the effectiveness of the cooperation, three evaluations have been carried out using different percentages of ACAs involved in the cooperative process: 25%, 50% and 75% respectively. The results for the cooperative scheme are calculated as the mean of the results obtained by these three evaluations.

Fig. 9 shows the average error, in degree Celsius, obtained by estimating temperature values through non cooperative and cooperative cases using both agents selection criteria described. For the non-cooperative case, the average error among the considered sensors is 0.074°C, the standard deviation is 1.865°C. For the cooperative case, the average

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3028967, IEEE Access

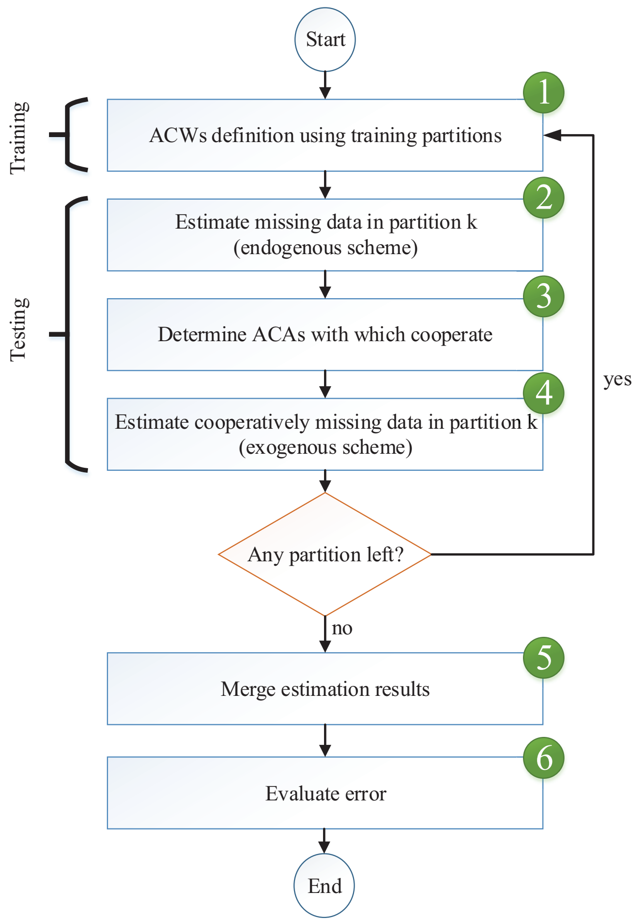
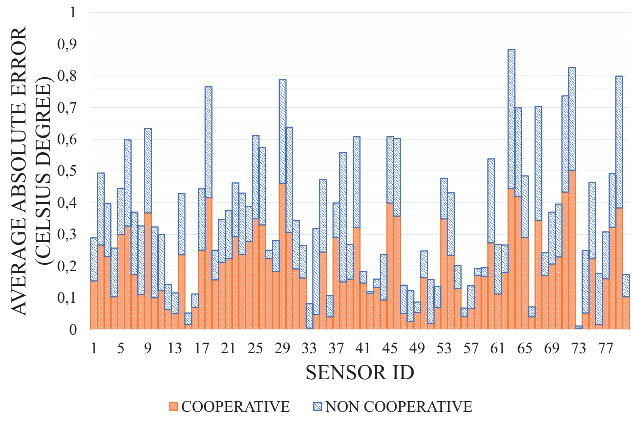
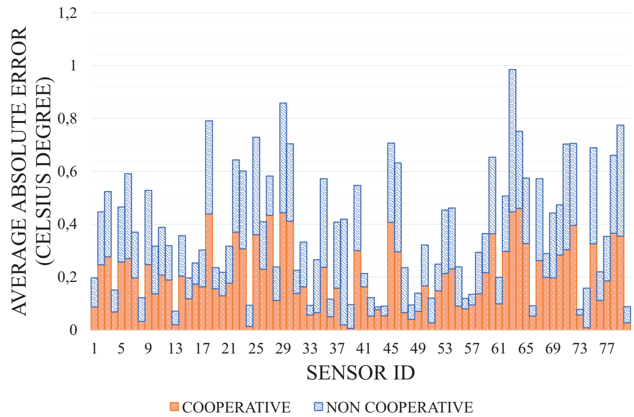


FIGURE 8: Main steps of the pipeline used to evaluate the proposed solution.



(a) Nearest neighbors criterion case.



(b) Most confident neighbors criterion case.

FIGURE 9: Average absolute error (in degree Celsius) obtained from endogenous estimation, for both cooperative and non-cooperative cases. For the cooperative case, both nearest neighbors criterion (Fig. 9a) and most confident neighbors criterion (Fig. 9b) are used to evaluate the agents with which cooperate.

error is 0.060°C, the standard deviation is 1.970°C using the most confident agents criterion. By using the nearest agents criteria, the average error is 0.081°C, the standard deviation is 1.968°C. Fig. 9 shows that the results obtained by the cooperative method are comparable to those obtained by the endogenous (individual) estimation; this proves that the joint operation of a collective of autonomous agents can accurately estimate missing information. Moreover, the most confident criterion gives better results than the nearest criterion: cooperation between agents observing similar environmental dynamics enable agents to provide accurate estimates compared to the use of physically close sensors. This result proves that the proposed technique constitutes a step forward towards the development of large-scale ubiquitous systems for the estimation of missing values, where the distributed computation and the mutual interaction between devices allow obtaining significant results and addressing the challenges of openness and highly dynamic environment.

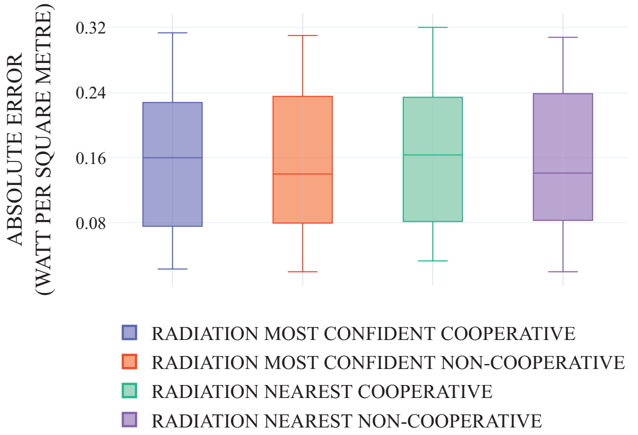
We evaluated the endogenous estimation scheme separately on different types of information to prove that the proposed technique is able to estimate information of different types without any configuration. Fig. 10 shows the absolute error obtained from estimating solar radiation, average daily wind speed, and relative humidity. The error obtained is fairly similar among the different information because of the limited amount of information available and the limited number of variations in the observed data.

TABLE II: Average absolute error obtained by endogenous estimation.

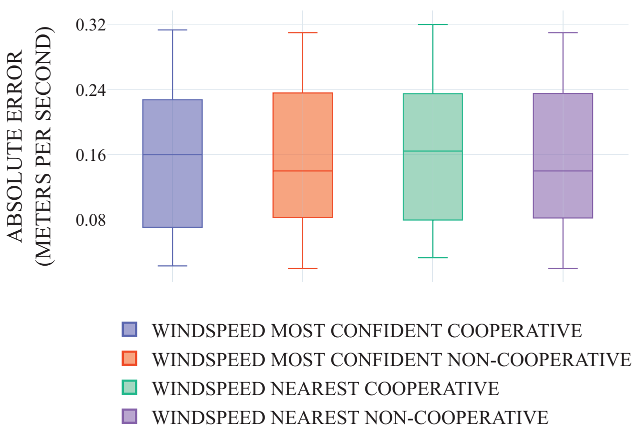
	Average Absolute Error	
	Most confident agent strategy	**Nearest agent strategy**
Humidity (%)	0.16320	0.16240
Solar radiation (W/m^2)	0.15851	0.16265
Wind speed (m/s)	0.15888	0.16141

Table II summarizes the average absolute error obtained by using endogenous estimation on solar radiation, wind speed, and relative humidity.

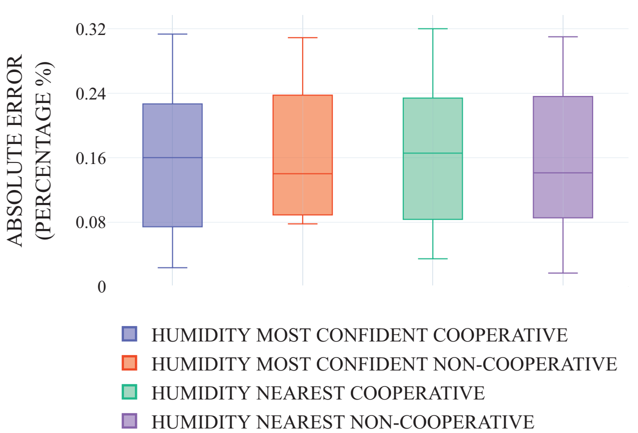
Although the information used is of a different type, with different ranges of values and different distributions,



(a) Radiation



(b) Wind Speed



(c) Humidity

FIGURE 10: Box plot of absolute error obtained through the endogenous estimation, for both cooperative and non cooperative cases, using solar radiation, wind speed, and relative humidity. We used both the nearest neighbor criterion and most confident criterion.

our technique enables obtaining a low error without any configuration: it can intercept and learn the environmental dynamics to accurately estimate missing information, independently from its type. Two main consequences can be discerned from the obtained results: the technique (*i*) makes it possible to obtain precise estimates for different types of environmental information, and (*ii*) allows estimating of environmental information using information which type is not known *a priori*; this suggests that the user does not have to provide, in advance, any estimation or configuration mechanism depending on the type of data and new types of information can be introduced without any particular modification.

1) Comparison to the State of the Art

The results obtained by the endogenous estimation have been compared to state of the art techniques using the KNIME analytic platform. This free software provides a graphical drag-and-drop environment where pipelines can be assembled by connecting nodes that perform data analysis tasks [23]. In KNIME, nodes are components represented as boxes having input and output ports. Each node transforms and processes data according to specific functionalities. Input/output connections ports allow data to flow through the pipeline. The KNIME platform was chosen for the experiments because of its availability, ease of use and easy reproducibility of experiments that do not require any programming languages. Moreover, in KNIME a large number of regression techniques are available, which enables a more exhaustive comparison with our proposal. The following regression techniques have been used for estimating missing information, the nodes being available on KNIME: linear regression, polynomial regression, random forest regression [24], fuzzy rules [25], gradient boost trees regression [26], Autoregressive Integrated Moving Average (ARIMA) [27], Pace regression [28], Radial Basis Function (RBF) [29] and isotonic regression [30]. Pace regression, RBF and isotonic regression nodes are available through the Weka data mining framework [31], which can be integrated with KNIME. The comparison with the state of the art does not consider the methods described in Section II as authors do not provide any tool to realize a comparison with the dataset used in this paper.

The evaluation using the state of art techniques has been carried out using the default configuration for each node. Table III summarizes the default parameters used to configure the nodes in KNIME.

The k-fold cross-validation and an auto-regressive model are used to evaluate the accuracy of the state of the art techniques; the required nodes are available in KNIME, using a k value of 5, 10 and 15. An auto-regressive model assumes that the previously observed samples can be used to predict accurately the value at the next time step. For instance, we used 4 samples to implement the auto-regressive model using KNIME. There is no formal rule for choosing the value of k to use for cross-validation. However, as k gets larger, the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3028967, IEEE Access

TABLE III: Configuration of state of art techniques nodes available in KNIME.

Technique	Properties
Linear Regression	No properties
Polynomial Regression	Maximum polynomial degree: 3
Random Forest Regression	No properties
Fuzzy Rules	Missing Values: Best Guess
Gradient Boost Trees Regression	Missing Value Handling: XGBoost Alpha: 1.0
ARIMA	AR/I/MA order: 1 Estimation method: conditional likelihood [32]
Pace Regression	Estimator: ordinary least squares Threshold value: 2
RBF	Number of Gaussian basis functions: 2 Ridge factor for quadratic penalty on output weights: 0.01 Tolerance parameter for delta values: 1.0e-6 Scale optimization option: one scale per unit Use conjugate gradient descent: true Use normalized basis functions: true Size of the thread pool: 1 Number of threads to use: 1 Use random number seed: true
Gaussian Processes	Level of Gaussian Noise with respect to transformed target: 1 Kernel used: polynomial
Isotonic Regression	No properties

difference in size between the training set and the resampling subsets gets smaller. Consequently, the difference between the estimated and real values becomes smaller [33].

We used the 15 sensors that gave the worst results using the endogenous estimation in HybridIoT to evaluate the state of the art techniques. Fig. 11 shows the box plots that depict the absolute error obtained by the state of art techniques, calculated as the average among the considered folds (5, 10 and 15).

The results in Fig. 11 are obtained from the dataset containing only temperature values. Although the average error obtained for each sensor is less than 1°C, the proposed cooperative approach outperforms the state-of-the-art techniques. In fact, cooperation enables ACAs to use not only information in their historic, but also information coming from a multitude of agents in the environment.

We compared the results of the endogenous estimation scheme to a pipeline of standard techniques using the same dataset and cross-validation [21]. This pipeline includes Voronoi tessellation to determine the geographical area to which sensors belong, hierarchical clustering to group the sensors that perceive similar information and normalized convolution to estimate missing information. The areas of two or more sensors are merged if their Voronoi areas are adjacent and they are grouped by the clustering process. Finally, for a given point, a normalized convolution is used to estimate missing environmental information using the sensors in the corresponding Voronoi region. Normalized convolution is a standard method used to reconstruct incomplete or uncertain data from a spatio-temporal signal widely used in geo-statistical applications [34]. We used the dataset of only temperature values for this evaluation.

Fig. 12 shows the comparison of results obtained by this pipeline and HybridIoT using respectively the nearest agents and most confident agents criteria for the cooperative case. HybridIoT shows better results with respect to the pipeline using both agents' selection criteria (near agents and most confident agents). Moreover, we outline some fundamental differences from HybridIoT: (*i*) HybridIoT can learn environmental dynamics and estimate missing information instantaneously in areas of the environment not sufficiently covered by sensors; (*ii*) HybridIoT enables introducing new sensing devices at any time; (*iii*) using the pipeline, a device that needs to estimate missing information must be located in a region where at least one working sensor must be present. In HybridIoT, an ACA can estimate missing information through cooperation but also using previously perceived values, thus overcoming the lack of sensors in the proximity of the ACA.

We presented the results obtained by the endogenous estimation scheme. We compared the results to those obtained by state of the art regression techniques to assess the accuracy of the results obtained. Also, we carried out a comparison using a pipeline of standard techniques including Voronoi tessellation and hierarchical clustering, on the same dataset. In this case, HybridIoT outperforms the results obtained by the pipeline.

The following section presents the results obtained by HybridIoT using the exogenous estimation scheme.

D. EXOGENOUS ESTIMATION

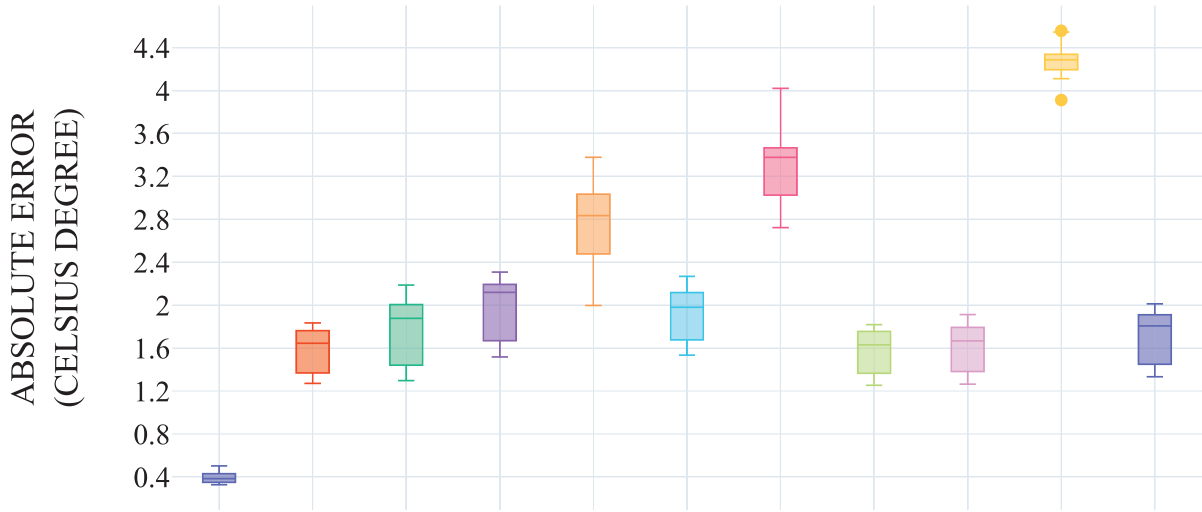
As far as we know, nowadays no solution is available to estimate missing information by integrating heterogeneous information from different data sources. Because we cannot compare the results obtained through the exogenous estimation to any specific technique, we show that the obtained are on a par with those obtained using the endogenous scheme.

Exogenous estimation is addressed through cooperation between ACAs that perceive different types of information. In our experiments, we used the proposed technique to estimate temperature values using ACWs containing solar irradiance, humidity and wind speed.

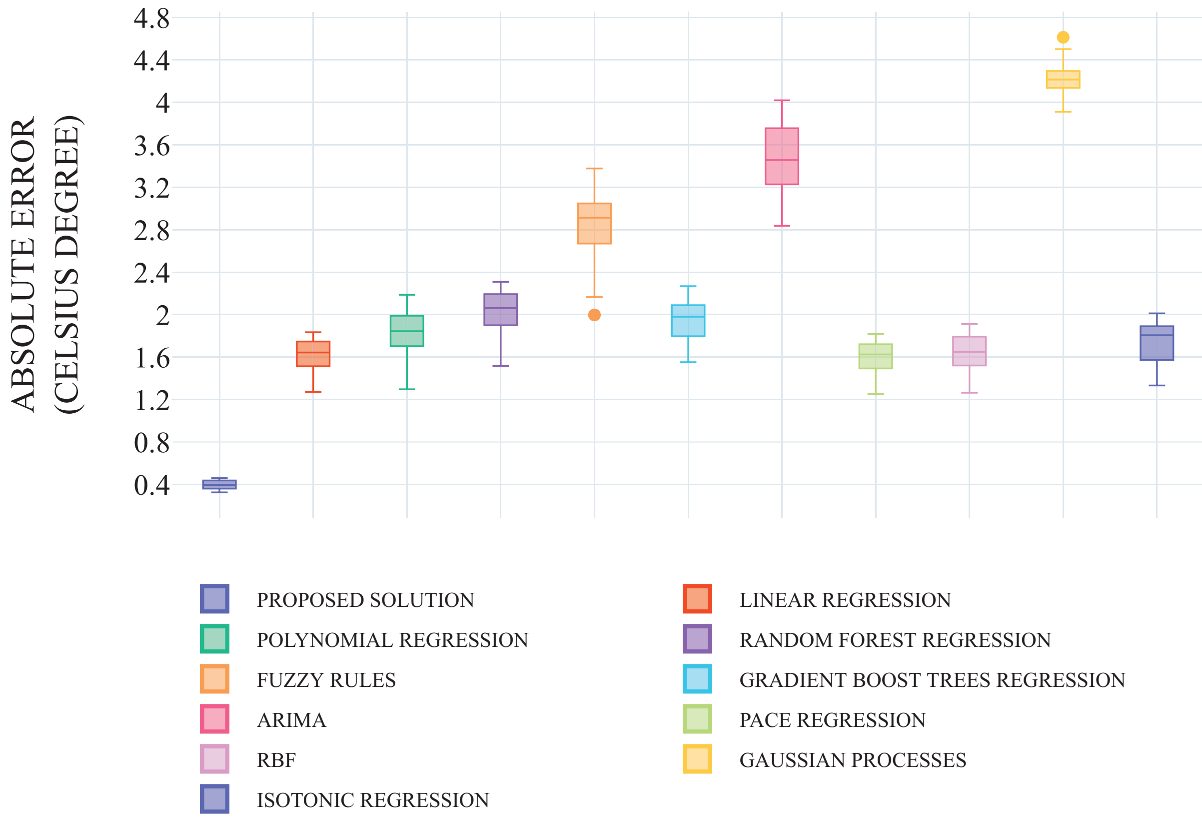
As described in the previous paragraph, an ACA that encountered a NCS evaluates different estimations for each weight obtained from a cooperative process with other agents. Then, the ACA evaluates a histogram of the estimation values to exclude those that are not relevant. In our experiments the number of bins has been set to 10: this value was obtained by using the Freedman-Diaconis rule [35], which allows calculating the size of the classes of a histogram. The number of bins was calculated as the median of the overall numbers of bins obtained by applying the Freedman-Diaconis rule for all the sensors.

Fig. 13 shows the average error for the regional temperature dataset using heterogeneous ACWs. As for endogenous estimation, the estimation of a missing value using the exogenous technique is almost instantaneous.

Table IV summarizes the average absolute error and the standard deviation obtained by estimating the temperature using heterogeneous ACWs. The results show that ACAs can provide accurate estimates through heterogeneous information without using any specific data fusion technique. The error resulting from exogenous estimation is low despite the used ACWs contains values that are semantically



(a) Nearest agents criterion case



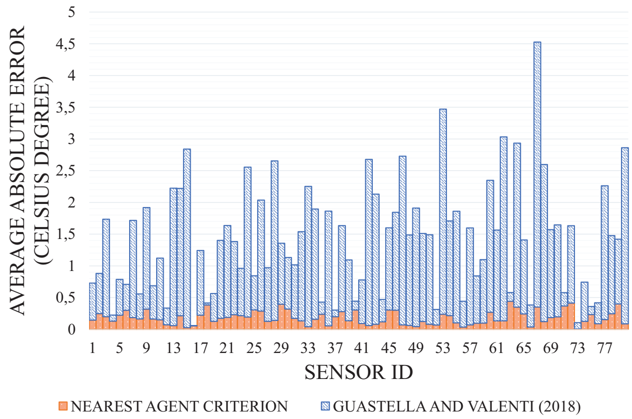
(b) Most confident agents criterion case

FIGURE 11: Box plot of absolute error (in degree Celsius) obtained by the state of the art techniques on the 15 sensors that gave the worst results using the proposed solution. Fig. 11a compares the results of our solution using the nearest agents criterion, Fig. 11b compares the results of our solution using the most confident agents criterion.

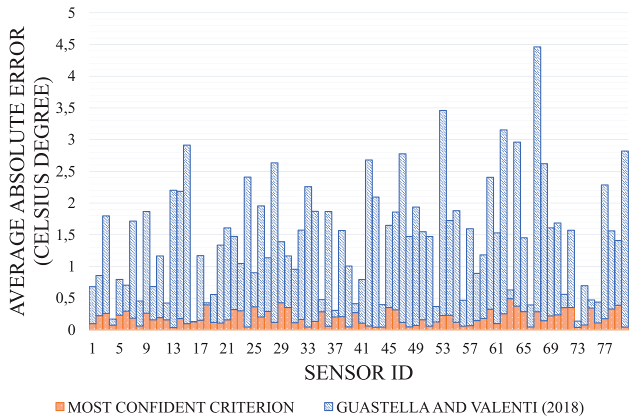
different from the temperature values. The information used in this estimation scheme does not have significant variation compared to the temperature values. Despite the differences in the variations and information between ACWs, the results in Table IV show that the proposed method can accurately estimate missing information through heterogeneous values.

This section presented the results obtained through the exogenous estimation scheme carried out by agents in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3028967, IEEE Access



(a) Nearest neighbors criterion case.



(b) Most confident neighbors criterion case.

FIGURE 12: Average absolute error (in degree Celsius) obtained from the pipeline that uses Voronoi tessellation and Hierarchical clustering and HybridIoT, using nearest agent (Fig. 12a) and most confident criterion (Fig. 12b).

TABLE IV: Average absolute error and standard deviation (in degree Celsius) obtained by exogenous estimation.

	Average Error	
	Most Confident agent strategy	Nearest agent strategy
Temp. from solar radiation	0.02716	0.03796
Temp. from humidity	0.02654	0.03679
Temp. from wind speed	0.02531	0.03914
Temp. from all info	0.07	0.05935
	Standard Deviation	
	Most Confident agent strategy	**Nearest agent strategy**
Temp. from solar radiation	1.639136	1.645494
Temp. from humidity	1.646173	1.647901
Temp. from wind speed	1.64716	1.651358
Temp. from all info	1.65435	1.62805

HybridIoT. The following section discusses HybridIoT with respect to the results obtained by both endogenous and exogenous estimation schemes. We show that the results obtained by exogenous estimation are on a par with those obtained by the endogenous scheme.

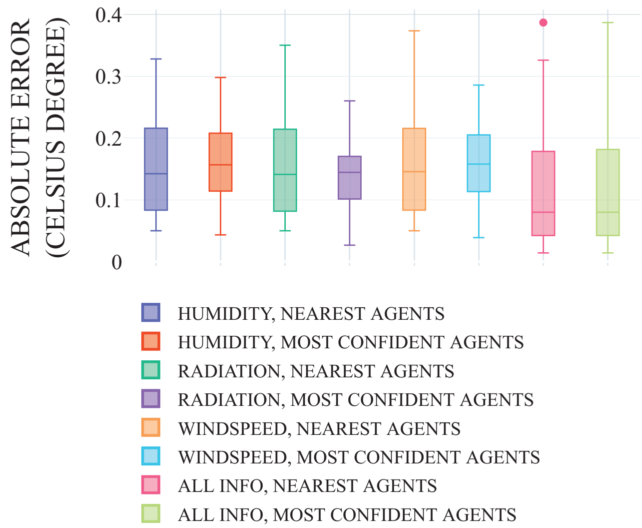


FIGURE 13: Box plot of absolute error (in degree Celsius) obtained through exogenous estimation. Each box depicts the error obtained using different types of ACWs and agents' selection criteria.

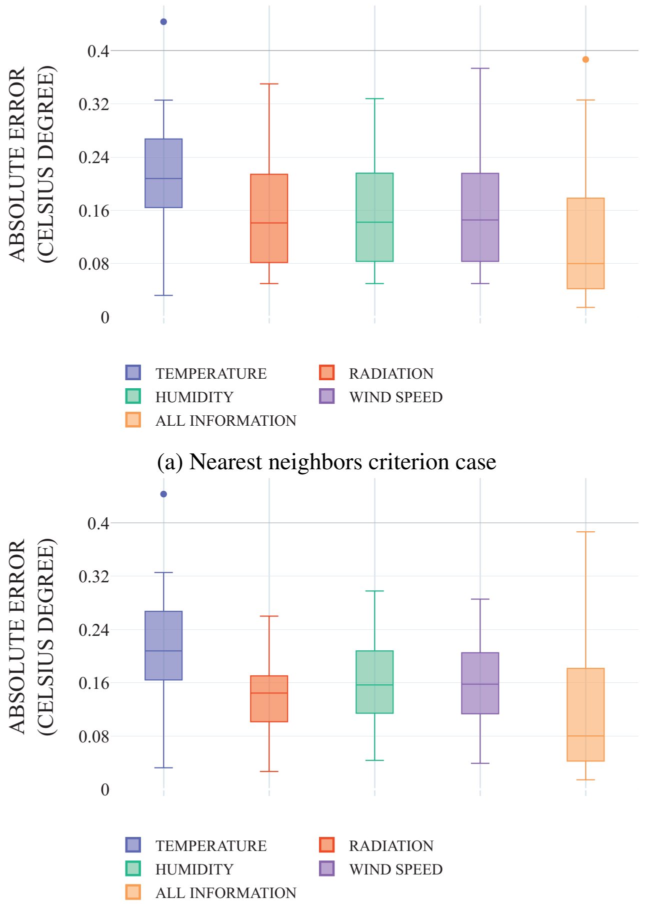
E. DISCUSSION

The obtained results assess the validity of HybridIoT for estimating missing heterogeneous information. With respect to the state of the art techniques, our method does not require any *a priori* configuration and the estimation of missing value is almost instantaneous.

To prove the effectiveness of the exogenous estimation, we show that the related results are on a par with those obtained by the endogenous estimation, using the same dataset. Fig. 14 shows the box plots of error, in degree Celsius, obtained by comparing the real temperature values with the estimates calculated through both endogenous and exogenous estimations, this last using radiation, humidity and wind speed. Fig. 14a shows the results obtained by using the most confident agents criterion. Fig. 14b shows the results obtained by using the nearest agents criterion. The sensors used for this experiment are those that acquired correctly all the available information (temperature, humidity, wind speed, solar radiation) within the considered temporal period. We used both the nearest agents and most confident criteria for the cooperative scheme: in the two cases, the results show that the exogenous estimation outperforms the endogenous one.

The results in Fig. 14 show the presence of outliers in both endogenous estimations using temperature and exogenous estimations using all the information. The reason is that temperature information is more subject to variations with respect to the other information employed (for instance humidity, radiance, wind speed). When estimating a missing value using different types of information that do not vary significantly in time, the weights generated by the estimation process lead to a negligible variation of the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3028967, IEEE Access



(a) Nearest neighbors criterion case

(b) Most confident neighbors criterion case

FIGURE 14: Box plot of absolute error (in degree Celsius) obtained by comparing the results of endogenous and exogenous estimations. Fig. 14a shows the results obtained by using the most confident agents criterion, Fig. 14b the results obtained by using the nearest agents criterion.

last perceived value. Despite the two selection criteria, the estimation process may include ACWs containing environmental dynamics significantly different from the one containing the information to be estimated. Therefore, some estimates can be considered as outliers because they are significantly different from the real value.

Applying exogenous estimation using all types of information available is effective for estimating missing information. Although applying exogenous estimation using one type of different information (with respect to the one that must be estimated) allows to obtain good results, the error obtained by using all types of available information is considerably low. This proves that using different types of ACWs allows ACAs to evaluate accurate estimates by exploiting multiple types of information perceived by sensors available in the environment.

Thanks to HybridIoT, the ACAs can evaluate precise estimates thanks to a cooperative behavior of ACAs. Agents evaluate estimates instantaneously using information perceived by the available sensors. Because each agent pursues an autonomous computation and has a partial view of the environment, the technique can be employed in large scale urban contexts. Moreover, there is no need for any particular configuration depending on the information used or the sensing devices. Our proposal enables addressing the property of openness thanks to the cooperative behavior between agents: sensors can enter or leave the system at any time without compromising its operation.

V. DEPLOYMENT IN URBAN CONTEXT

Today smart devices capable of sensing the surrounding environment are becoming more and more affordable; low prices lead to an increasing diffusion of these devices and consequently to a rise in their production. The diffusion of harmful gases for both the environment and citizens arises from the massive production of these devices [36]. Using already available devices such as smartphones or connected vehicles could limit the number of sensors to deploy in an urban context, lowering the demand for new devices and thus their production. The use of these available devices should not replace existing sensors; sensing infrastructure should be capable of integrating smartphones and connected vehicles with existing sensors to make the urban sensing participatory. This integration is advantageous for the sustainable development of cities: sensing infrastructures integrate a large number of devices capable of perceiving geo-localized information on a large-scale that allow for a local and precise description of the environmental dynamics. Integrating heterogeneous information can be beneficial to experts for defining new services for citizens or improving existing ones to achieve the smartness objective of a city. Nevertheless, although the limited number of devices capable of sensing the environment available on the market, we assume that the low cost and miniaturization can lead to the integration of sensors into devices such as smartphones or connected vehicles.

The deployment of HybridIoT on a real context, considering data acquired from existing sensors and mobile devices such as smartphones, has no consequence on the nominal operation of the system:

- operates regardless of the number of devices, these can enter or leave the system without any external intervention;
- delays in the transmission of information between devices do not affect the functioning of the system: agents evaluate estimates using the available information which is geo-localized and provided with time-stamps;
- mobility does not affect the functioning of the system: agents perform a local computation in the part of the environment where they are situated.

In the experiments conducted, the intermittence of sensors enables simulating both openness and mobility properties. Each ACA has a boolean variable which value is 1 if the associated sensor is turned on, 0 otherwise. By controlling this variable during the agent's operation is possible to simulate the intermittence of the associated sensor, not affecting the functioning of the entire system. In cross-validation, the value is set to 0 when evaluating the estimates for a given test partition. In using this variable, it is possible to simulate the malfunctioning of devices, therefore assessing that the system addresses the property of openness as agents cooperate only with available ACAs. Although the results discussed are obtained from fixed devices, in the demonstration of the proposed system [37], the sensors can be relocated during the demonstration, thus showing that HybridIoT is able to consider mobile sensors. This functionality does not affect the operation of individual agents.

Currently, we plan to deploy the proposed system on the campus of the University of Toulouse III – Paul Sabatier and to use devices and smartphones for collecting information. The following factors are crucial for a reliable implementation of the system on a real context:

- developing an application for mobile devices. Different operating systems must be considered and therefore the effort could be considerable. The mobile application must ensure the privacy of users when gathering and treating information perceived by personal devices. Users should also agree to use the designed application and participate in the experiments;
- implementing a communication protocol between devices. In a personal device, several applications can degrade the device performance. This can be overcome by using lightweight application protocols that improve smartphone performance in terms of bandwidth consumption, battery lifetime, and communication latency [38];
- managing communication delays due to the transmission of remote information: the communication on smartphones is bandwidth-limited and relatively expensive, especially when access to the network is achieved via cellular connection [39];
- verifying that the data acquired by the devices are correct and that there are no anomalies that could prevent the proper functioning of the system. In this way, having data been perceived from devices functioning properly, it is possible to validate the results obtained by the proposed technique;
- building a consistent database for evaluating the results obtained from the system deployed in a real context: this requires continuous monitoring of the environment from both *ad hoc* devices and smartphones. However, smartphones functioning depends on the will of its owner: the device is turned on/off according to non-controllable patterns, therefore it is not possible to exert strict control [39].

These factors are independent of the functioning of the system, have no direct consequences in its operation and parameters.

VI. CONCLUSION AND PERSPECTIVES

The development of sustainable smart cities requires the deployment of ICT to ensure better services and available information at any time and everywhere. As IoT devices become more powerful and low-cost, the implementation of an extensive sensors network for an urban context can be expensive and resource-consuming. Moreover, the production of IoT devices themselves leads to the production of harmful gases into the environment.

This paper presented HybridIoT, a MAS based technique to estimate missing heterogeneous information addressing the properties of openness and large scale computation. Using HybridIoT it is possible to avoid the installation of a large number of sensors by using virtual devices able to provide accurate information estimates where *ad hoc* sensors are not available. This leads to relevant cost savings due to the low number of sensors to be installed and the management costs related to the sensor network. In this way, it is possible to reach a significant compromise between an extensive instrumentation of cities and the availability of information at any time and everywhere.

The advantages of HybridIoT over the state of the art are the heterogeneity, openness and scalability. Heterogeneity enables to integrate different types of devices that could perceive different types of information. Openness enables sensing devices to enter or leave the system without the need for any re-configuration. Cooperation enables agents to exploit the data acquired from a multitude of devices to estimate missing information.

We carried out experimentations using both homogeneous and heterogeneous information of a real weather dataset, and compared the results of endogenous estimation to state of the art techniques. To the best of our knowledge there is no technique whose results can be compared to those obtained by our solution in case of exogenous estimation. The results obtained by endogenous and exogenous estimations are compared in order to show that the latter are on a par with those obtained by the endogenous estimation. The results assess the validity of our proposal in estimating environmental information in large scale urban settings using a limited number of heterogeneous devices.

HybridIoT does not require any input parameters or configuration. The system can operate in open, dynamic environments such as cities, where devices can appear or disappear without any prior notification. Agents are able to provide estimates in almost instantaneous time. The state of the art techniques used to compare our proposal are among the most common and powerful methods to carry out regression over time series. However, these techniques do not appear to be able to operate in open environments where devices may appear or disappear unexpectedly and do not make use of heterogeneous information which type is not known *a priori*.

Further studies will focus on the use of HybridIoT in different contexts, notably at building scale to improve the energy consumption of appliances and at city scale using the Floating Car Data (FCD) to estimate in real-time the traffic conditions by avoiding the installation of intrusive technology. The estimated information can be used, in both cases, to improve the quality of life of citizens. Possible future developments of the proposed solution include a mechanism to limit the number of weights to be used. This could be done by introducing a threshold measure to avoid the use of weights that lead to significant changes in the estimated value since the last perceived information.

REFERENCES

[1] C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos. Sensing as a service model for smart cities supported by internet of things. Transactions on Emerging Telecommunications Technologies, 25(1):81–93, 2014. 10.1002/ett.2704.

[2] J. Dutta, C. Chowdhury, S. Roy, A. I. Middya, and F. Gazi. Towards smart city: Sensing air quality in city based on opportunistic crowd-sensing. In Proceedings of the 18th International Conference on Distributed Computing and Networking, ICDCN '17, New York, NY, USA, 2017. ACM. doi: 10.1145/3007748.3018286.

[3] Guide to the environmental information regulations. Available at https://ico.org.uk/for-organisations/guide-to-the-environmental-information-regulations/, 2017.

[4] A. R. Honarvar and A. Sami. Towards sustainable smart city by particulate matter prediction using urban big data, excluding expensive air pollution infrastructures. Big Data Research, 17:56 – 65, 2019. doi: 10.1016/j.bdr.2018.05.006.

[5] S. K. Sood, R. Sandhu, K. Singla, and V. Chang. IoT, big data and HPC based smart flood management framework. Sustainable Computing: Informatics and Systems, 20:102 – 117, 2018. doi: 10.1016/j.suscom.2017.12.001.

[6] J. Guerrero-Ibáñez, S. Zeadally, and J. Contreras-Castillo. Sensor technologies for intelligent transportation systems. Sensors, 18(4):1212, 2018. doi: 10.3390/s18041212.

[7] X. Kong, X. Liu, B. Jedari, M. Li, L. Wan, and F. Xia. Mobile crowdsourcing in smart cities: Technologies, applications, and future challenges. IEEE Internet of Things Journal, 6(5):8095–8113, 2019.

[8] V. Noël and F. Zambonelli. Following the problem organisation: A design strategy for engineering emergence. In D. Camacho, L. Braubach, S. Venticinque, and C. Badica, editors, Intelligent Distributed Computing VIII, pages 311–317. Springer, 2015.

[9] M. Wooldridge. An Introduction to MultiAgent Systems. Wiley, 2 edition, 2009.

[10] M.-P. Gleizes, J. Boes, B. Lartigue, and F. Thiébolt. neOCampus: A demonstrator of connected, innovative, intelligent and sustainable campus. In G. De Pietro, L. Gallo, R. J. Howlett, and L. C. Jain, editors, Intelligent Interactive Multimedia Systems and Services 2017, volume 76, pages 482–491. Springer, 2018.

[11] D. A. Guastella, V. Camps, and M.-P. Gleizes. Estimating missing environmental information by contextual data cooperation. In M. Baldoni, M. Dastani, B. Liao, Y. Sakurai, and R. Zalila Wenkstern, editors, PRIMA 2019: Principles and Practice of Multi-Agent Systems, pages 523–531. Springer, 2019. doi: 10.1007/978-3-030-33792-6_37.

[12] D. Guastella, V. Camps, and M.-P. Gleizes. Multi-agent systems for estimating missing information in smart cities. In Proceedings of the 11th International Conference on Agents and Artificial Intelligence - Volume 2: ICAART,, pages 214–223. SciTePress, 2019. doi: 10.5220/0007381902140223.

[13] O. Alvear, C. Calafate, J.-C. Cano, and P. Manzoni. Crowdsensing in smart cities: Overview, platforms, and environment sensing issues. Sensors, 18(2):460, 2018. doi: 10.3390/s18020460.

[14] A. M. Ali and P. Angelov. Anomalous behaviour detection based on heterogeneous data and data fusion. Soft Computing, 22(10):3187–3201, 2018. doi: 10.1007/s00500-017-2989-5.

[15] J. Y. Zhu, C. Sun, and V. O. K. Li. Granger-causality-based air quality estimation with spatio-temporal (s-t) heterogeneous big data. In 2015 IEEE Conference on Computer Communications Workshops, pages 612–617, 2015. doi: 10.1109/INFCOMW.2015.7179453.

[16] Z. Shan, Y. Xia, P. Hou, and J. He. Fusing incomplete multisensor heterogeneous data to estimate urban traffic. IEEE MultiMedia, 23(3):56–63, 2016. doi: 10.1109/MMUL.2016.37.

[17] D. Tomaras, V. Kalogeraki, N. Zvgouras, N. Panagiotou, and D. Gunopulos. Evaluating the health state of urban areas using multi-source heterogeneous data. In 2018 IEEE 19th International Symposium on "A World of Wireless, Mobile and Multimedia Networks", pages 14–22, 2018. doi: 10.1109/WoWMoM.2018.8449761.

[18] A. Dorri, S. S. Kanhere, and R. Jurdak. Multi-agent systems: A survey. IEEE Access, 6:28573–28593, 2018. doi: 10.1109/ACCESS.2018.2831228.

[19] J.-P. Georgé, M.-P. Gleizes, and V. Camps. Cooperation. In G. Di Marzo Serugendo, M.-P. Gleizes, and A. Karageorgos, editors, Self-organising Software: From Natural to Artificial Adaptation, pages 193–226. Springer, 2011. doi: 10.1007/978-3-642-17348-6.

[20] C. Bernon, V. Camps, M.-P. Gleizes, and G. Picard. Engineering adaptive multi-agent systems. In Agent-Oriented Methodologies, pages 172–202. IGI Global, 2005.

[21] D. A. Guastella and C. Valenti. Estimating missing information by cluster analysis and normalized convolution. In 2018 IEEE 4th International Forum on Research and Technology for Society and Industry (RTSI), pages 1–6, 2018. doi: 10.1109/RTSI.2018.8548454.

[22] OpenStreetMap contributors. Planet dump retrieved from https://planet.osm.org. https://www.openstreetmap.org, 2017.

[23] M. R. Berthold, N. Cebron, F. Dill, T. R. Gabriel, T. Kötter, T. Meinl, P. Ohl, K. Thiel, and B. Wiswedel. KNIME - the konstanz information miner: Version 2.0 and beyond. ACM SIGKDD Explorations Newsletter, 11(1):26–31, 2009. doi: 10.1145/1656274.1656280.

[24] T. K. Ho. Random decision forests. In Proceedings of 3rd International Conference on Document Analysis and Recognition, volume 1, pages 278–282 vol.1, 1995. doi: 10.1109/ICDAR.1995.598994.

[25] M. R. Berthold and K.-P. Huber. Missing values and learning of fuzzy rules. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 06(2):171–178, 1998. doi: 10.1142/S021848859800015X.

[26] J. H. Friedman. Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5):1189–1232, 2001. doi: 10.1214/aos/1013203451.

[27] J. Junttila. Structural breaks, ARIMA model and finnish inflation forecasts. International Journal of Forecasting, 17(2):203 – 230, 2001. doi: 10.1016/S0169-2070(00)00080-7.

[28] Y. Wang and I. H. Witten. Modeling for optimal probability prediction. In Proc. Nineteenth International Conference on Machine Learning, pages 650–657. Morgan Kaufmann, 2002.

[29] B. Walczak and D.L. Massart. The radial basis functions - partial least squares approach as a flexible non-linear regression technique. Analytica Chimica Acta, 331(3):177–185, 1996. doi: 10.1016/0003-2670(96)00202-4.

[30] W. B. Wu, M. Woodroofe, and G. Mentz. Isotonic regression: Another look at the changepoint problem. Biometrika, 88(3):793–804, 10 2001. doi: 10.1093/biomet/88.3.793.

[31] I. H. Witten, F. Eibe, and M. A. Hall. Data Mining: Practical Machine Learning Tools and Techniques. Elsevier, 2011. doi: 10.1016/C2009-0-19715-5.

[32] J. D. Hamilton. Time Series Analysis. Princeton University Press, 1994.

[33] M. Kuhn and K. Johnson. Applied Predictive Modeling. Springer, 2013. doi: 10.1007/978-1-4614-6849-3.

[34] R. T. Lemos and B. Sansó. Spatio-temporal variability of ocean temperature in the portugal current system. J. Geophys. Res. (AGU), 111(C4):C04010, 2006.

[35] D. Freedman and P. Diaconis. On the histogram as a density estimator:l2 theory. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete (Springer), 57(4):453–476, 1981. doi: 10.1007/BF01025868.

[36] The Shift Project. Lean ICT: Towards Digital Sobriety. Available at https://theshiftproject.org/wp-content/uploads/2019/03/Lean-ICT-Report_The-Shift-Project_2019.pdf.

[37] Davide A. Guastella. HybridIoT Demo. https://youtu.be/f1276rOa6NI, 2020. [Online].

[38] N. De Caro, W. Colitti, K. Steenhaut, G. Mangino, and G. Reali. Comparison of two lightweight protocols for smartphone-based sensing. In 2013 IEEE 20th Symposium on Communications and Vehicular Technology in the Benelux (SCVT), pages 1–6, 2013. ISSN: 2373-0854.

[39] A. Faggiani, E. Gregori, L. Lenzini, V. Luconi, and A. Vecchio. Smartphone-based crowdsourcing for network monitoring: Opportunities,

challenges, and a case study. IEEE Communications Magazine, 52(1):106–113, 2014.

DAVIDE A. GUASTELLA is a double degree Ph.D. candidate in artificial intelligence at the Toulouse Institute of computer science research (Institut de Recherche en Informatique de Toulouse, IRIT) and the University of Palermo, Italy. Davide is a member of the Cooperative Multi-Agent Systems team (Systèmes Multi-Agent Coopératifs, SMAC). He focuses his research on cooperative multi-agent systems and their application in the smart city to optimize energy consumption in large-scale contexts through continuous, non-invasive monitoring of the environment.

Davide A. Guastella received its BSc at the University of Palermo and a joint double-degree MSc from the University of Palermo and the University of Paris Est Marne la Vallée.

VALÉRIE CAMPS is Assistant Professor at Paul Sabatier University and a member of the SMAC team at the IRIT (Institut de Recherche en Informatique de Toulouse, France, irit.fr) laboratory. She has been working on adaptive multi-agent systems for the past 20 years. Her interests deal with cooperation, self-organization, and emergent behaviors in multi-agent systems. She has contributed to the definition of the AMAS (Adaptive Multi-Agent Systems) approach and has participated to the ADELFE (toolkit to develop software with emergent functionality) methodology definition.

Valérie Camps has been also applying the AMAS principles to the user's personalization, to the construction of evolving ontologies and to the modeling of dynamic contexts in ambient systems and smart cities. In particular, she has been involved in the French ANR projects "DYNAMO" (DYNAMic Ontology) and "INCOME" (Software Infrastructure of Multi-scale Context Management for the Internet of Objects) and is currently participating in the C2C region project (Cooperation between connected Car-Cars and human acceptability). She has co-advised 9 Ph.D. students (including three currently) in these domains. She has participated in more than one hundred international PCs on Multi-Agent Systems and more generally in Artificial Intelligence.

MARIE-PIERRE GLEIZES is full Professor since 2006 at the University of Toulouse III Paul Sabatier and researcher at IRIT (Institute of Computer Science in Toulouse – France). She is the leader of the SMAC (Cooperative Multi-Agent Systems) research team composed of 19 permanent researchers and 20 Ph.D. and post-docs. At the level of the university, she is also in charge of the scientific project neOCampus which aims at designing a smart, innovative, sustainable campus, in which 11 research laboratories participate. His main research concerns the Multi-Agent Systems essentially around self-adaptation, complex systems, self-organization and emergence. She focuses his research on the ability of a system to learn in a supervised way but also in an agnostic way in using endogenous feed-backs. She also works on the design of local agent behavior to enable the collective of agents to solve optimization and constraint-based problems.