CIMSA 2003 - International Symposium on
Computational Intelligence for Measurement Systems and Applications
Lugano, Switzerland, 29-31 July 2003

Analysis and metrological characterization of the state of atmospheric pollution by means of mathematical models and agents.

A. Amato[1], G. Andria[2], T. Delvecchio[1], V. Di Lecce[2], A. Guerriero[2], A.M.L. Lanzolla[2], C. Pasquale[1], V. Piuri[3]

[1] II Engineering Faculty of Politecnico di Bari, V. del Turismo n.8, 74100 Taranto, Italy
[2] DEE Politecnico di Bari, Via Re David 200, 70121 Bari, Italy
[3] DTI University of Milan, Via Bramante 65, 26013 Crema (CR), Italy

Abstract ***This work deals with the development of an expert system based on agents for analysis, validation and characterization of environmental data relevant to monitoring of the level of atmospheric pollution in an urban area with high road traffic level. A mathematical model to identify the local pollutant concentrations and to forecast their temporal dynamics is here proposed. In this context, suitable "agency" to control, to validate and to interpret environmental data, has been developed and some correlations existing between the polluting values measured in various monitoring stations have been carried out.***

I. INTRODUCTION

Air quality is one of the main environmental problems for people, because of its direct influence on the state of human health; in this application both the interface and the computation complexity required the use of new and powerful computer paradigms.

In [1] and [2] the use of expert systems and agents has supplied a valid solution to improve complex acquiring-tasks about suitable parameters.

Agents are now being widely discussed by researchers in mainstream computer science [3]. In monitoring environment applications, an "agent" is usually seen as a self-directed program object having its own software structure and able to solve particular subtasks independently and to communicate its solution to a more complex problem-solving process, either on its own initiative or on a specific request from other agents [4].

As well known, the interest and the attention devoted to the environmental thematic, to the monitoring and to control activities about air quality, are growing quickly. Environmental National Agencies are imposing the implementation of environmental monitoring stations located in a wide geographic area, in order to:

supply representative real data related to the atmospheric pollution processes;

characterize the causes determining the pollution phenomena;

document the time-behavior of air quality in the analyzed territories;

forecast the eventual emergency situations.

The choice of a monitoring process for air pollutants depends on the typology and relevant proprieties to the substance that must be acquired, and on the kind of information to be gained from the measured values jointly with the relevant uncertainly level. Then, to reach an effective control of the air quality, the monitoring stations must be able to measure and record periodically a lot of the polluting substances, by using appropriate and reliable instrumentations [5]. In this way, it is possible to evaluate the behavior and dynamics of the measured phenomena with suitable accuracy, in order to identify long-term pollutant's biological effects [6].

As well known, pollutant concentrations depend on emissive sources, geographic conformations, and meteorological conditions (particularly wind and rain), able to modify the atmospheric dispersion of pollutants. Therefore to evaluate properly the air quality, it is necessary to work out a sophisticated monitoring system to measure chemical substances and meteorological quantities [7].

The analysis of environmental data, recorded in several monitoring stations installed in Taranto, Italy, represent an example of a real case. This town has been declared a *territory with elevated environmental risk,* because of many environmental high-impact factories (cement works, steel mills, oil refineries etc.) located near the urban area.

The monitoring stations are distributed in various typologies of urban sites (parks, residential areas, high road traffic area, suburban zones), according to the national prescriptions. Each station consists of:

chemical equipments to measure the concentrations of polluting substances: sulfur dioxide, nitrogen oxides, particulate matter (PM10), methane and not methane carbon ozone, monoxide, hydrocarbons and polyaromatic hydrocarbons;

meteorological devices to measure wind speed, wind direction, atmospheric temperature, solar radiation, relative humidity, atmospheric pressure and rain.

The measurement network is characterized by 6 fixed stations. All the equipments carry out approximately 50 measures/minute (raw data), that are subsequently elaborated to generate hourly medium values (valid data). These values are considered valid if at least 75% of the raw data are right. Then, the valid data are transmitted to a hub, where all

Under grants of MIUR Research projects: 1) "APE: Agencies for PErception in environmental monitoring", Research Unit of Università di Milano & Politecnico di Bari, Italy, 2) "Metrologia per la caratterizzazione di strumentazione digitale adoperata nel Controllo Ambientale" Research Unit of Politecnico di Bari, Italy.

0-7803-7783-4/03/$17.00 ©2003 IEEE

received data are automatically recorded (using CDR, for legal and long term preservation) and arranged in a suitable graphic-table for reading and interpretation.

In this paper we present an experience in environmental data treatment, using the agent concept for analysis and validation of the data acquired by the presented monitoring net.

The paper is so structured: in the following the problem of data validation is discussed. After the agent/agency implementation is presented. Conclusions and future work close the paper.

II. DATA STRUCTURES AND VALIDATION

The data validation process is a crucial phase for all the scientific applications, especially in the environmental field where physical and biological data are strongly related.

The high complexity of the measurement stations (in particular for chemical equipments) implies, unfortunately, that data are characterized by:

missing data, caused by periodic calibrations of the instruments and/or by working periods of the equipment;

wrong values, due to incorrect calibrations of instruments and/or to faulty-working of the same instruments;

wrong records of data, (presence of uninterpretable symbols) due to errors in the transmission data phase and/or recording process.

In order to obtain a suitable data set, it is necessary to implement some filters to identify incorrect and/or inconsistent data, based on the elimination of:

discontinuous data,

values outside of the admissible ranges,

values not in accordance with the expected behavior of the monitored quantities.

There are some interpolation and validation techniques, which can be used to approach and to solve these problems.

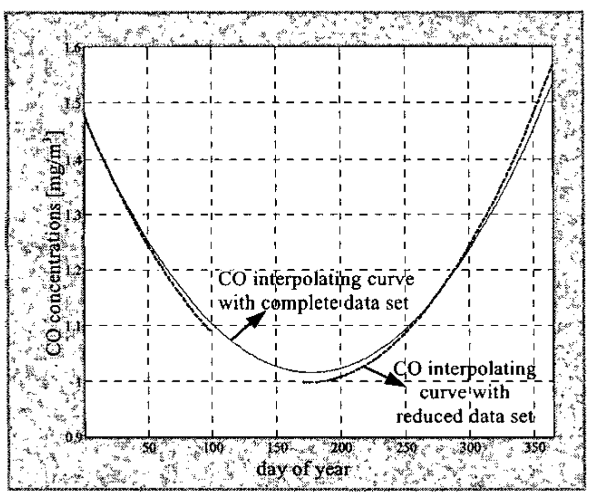

Fig. 1. Behaviour of fitting polynomial curves representing the daily average CO concentrations relevant year 2001, calculated by considering the complete set of data (continuous line) and the reduced one (dashed line).

In particular, some filters to eliminate incorrect data, based on the elimination of both evident discontinuity in the collected data and their values outside the admissible ranges, have been developed.

To reduce the problem of missing data some reconstruction techniques, based on both the continuity hypothesis of pollution values and interpolation methods, can be carried out. By using these techniques, it is possible to work with a continuous and valid set of data and to identify a mathematic model, based on a suitable polynomial interpolation, able to describe the behavior of daily average concentrations of pollutants during the days of the year.

This technique is useful in this case to validate data coming from sensors on the instant t_0. The acquired data is compared with its trend. This was computed by its historical dataset for a possible margin of error. There are a lot of methodology, known in literature like A.R. (auto regressive method), A.R.M.A. (auto regressive moving average), to compute the trend [20]. In this paper the approach, shown in Fig. 1, extracting the fitting polynomial curves is used. The agency architecture is independent from validation-task: different validation methodology, such as the aforesaid, can be used.

To evaluate the reliability of the developed model, several tests have been carried out, by eliminating a subset of data in a random way, from the complete time series. By recalculating the interpolation polynomial obtained using the so reduced data set, and by comparing the results carried out using the original data set, a measurement error lower then 4% has been obtained. An example is that shown in Fig. 1, about the daily average concentrations relevant to the year 2001, in which some data are considered missing [8]. Fig. 2 shows the architecture data-flow.

III. AGENTS AND AGENCY

In this work the focus is on the multi-agent architecture and on the used agents features. An important capacity of an agent in multi-agent environments is the ability to decide its own actions based on its own goals. However, when we consider the relationships between individual and community goals, two issues arise: how the satisfaction of individual goals affects the satisfaction of community goals, and how the satisfaction of community goals leads to the satisfaction of the individual goals [12]. In the proposed architecture, the satisfaction of community goals is guaranteed, because data are collected by the hub and processed by an independent unit in order to generate a continuous, valid, credible data set, by means of two different phases:

Generation of a suitable mathematical model, based on interpolation techniques, in order to optimally characterize the time and space-behavior of pollutants and to determine its uncertainty;

Trying to find out the causes of eventual numerical differences between stored values and those to be attended before defining completely the invalid data. The factors, that are considered in order to explain the odd values, vary meaningfully, according to the type of pollutant in question and to the environment characteristics (for example, peculiar characteristic of the monitoring place, season, typology of instrumentation, human and meteorological factors and so on).

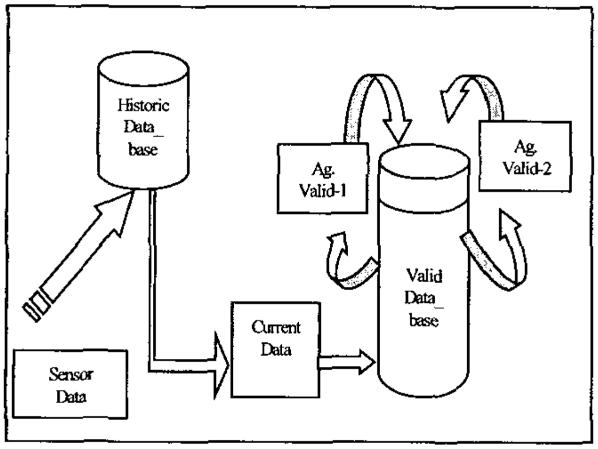

Fig. 2: Block diagram of the application data-flow

If the validation operations have not managed to find a cause for the odd-looking data, or if a malfunctioning of the instrumentation has been identified as the cause, the data are effectively invalidated [1].

The described monitoring system for urban pollution can be structured as an "agency" [9,10], in which it must execute and coordinate various measurement activities such as acquisition, transmission, analysis, elaboration and multilevel data validation.

In the following the basic elements of the multi-agent frameworks can be distinguished and presented: a common environment, agents, and interactions between agents and the environment.

A. Common Environnent

In the classical approach the environment is defined as a world in which the agent was designed to work. In particular a dynamic environment is characterized to be susceptible to change, especially when change is not due to an agent's control. In our approach the environment is represented by the database in which all data (raw data from sensor, agents treaded data and valid data) are stored, the database continuously evolves, so that it must be considered as a dynamic environment.

For the stations used in Taranto the data amount is about 15 Mbytes/year for each one. To environment perception all the agents have to read the database, this operation is time-expensive.

For this reason a *communication room* is added in the environment and used to speed up the agents awareness of the environment.

B. Agents

To define precisely what agents are, is a difficult task. In our case the Temporal Agent and Spatial Agent are define as referring to a component of software which is capable of acting exactingly in order to accomplish tasks on behalf of its user. They are several dimensions to classify existing software agents.

Agents may be classified by their mobility (static or mobile agents); they may be classed as either deliberative or reactive. And more the agents may be classified along several ideal and primary attributes which agents should exhibit (autonomy, learning, and cooperation).

The developed agents can be classified:

- Static or mobile,
- Reactive, because they are sensible at the variation of database;
- Cooperation and autonomy (collaborative agent), because the Temporal Agent and Spatial Agent can co-operate on their own without the need for human guidance. Hence agents have individual internal states and goals, and they act in such a manner as to meet its goals on behalf of its user.
- If Temporal Agent and Spatial Agent acquire the characteristic of learn they can be classified as smart agent [11].

The temporal validation is based on a temporal models (as in section II).

In this way the agent, compare acquired and expected data (extract applying a forecasting model). If the data are verified in the expected range, it is labeled as "valid" and valid for processing, otherwise is labeled as "suspected", this situation is reported in the communication room. If an agent with "suspected" data is recognized in the communication room the second agent analyzes the correlation between the "suspect" data and the other collected data from the other stations using a suitable proximity matrix (weighting function generated according to the monitoring stations relevant to near positions).

The validation-agents must be able to receive information about the environment. Eighteen actions, private and communicative, are defined for all used agent, as in Tab.1, a classified in the follow:

- Private actions: look about communication room (perceive), read database, update database, update communication room (confirm), request temporal validation using the method proposed in section 2, request spatial validation, validate data, no-validate data, lock and unlock records.
- Communicative actions: send a decision to a specified station (inform), and request an action to another station (request), accepted, rejected and not-understood a communication proposal.

In the first case agent updates database and it informs "communication room" about its state, in the second case agent defines "suspect" data, so the first agents produce new sub-goals that another agent can solve. The suspected data is

Private	Communicative
Perceive	Inform
Read	Request
Update	Accept communication
Confirm	Reject communication
Request temporal-valid	(Not-Understood)
Request spatial-valid	(Repeat)
Valid	
No-valid	
Lock record	
Unlock record	

Table 1. Set of actions (ontology)

analysed by a spatial validation agent, which accept or reject data using a proximity matrix. The validation methodology compares the suspect data with the values obtained from the near monitoring-stations, that are displace in Taranto. If the measured values are as good as extract value on the same instant from the other stations, the agent update the database, and the suspect data is acetated, otherwise it considers data no-valid and its acquisition is due to malfunctioning of the instrumentation. Furthermore, there is the case in which more than one data of near stations are over the range of tolerability, so these data aren't wrong, but spatial agent accept all acquired data, because there is a anomalous local environmental reason to validates data.

C. Interaction Between Agents And Environment

The emergence of a number of prototypical *agent languages* [13][14][15] is one sign that agent technology is more widely used, and that many more agent-based applications are likely to be developed in the near future.

One such language is Shoham's AGENT0 system, which is the inspiration of this work.

In this language, an agent is specified in terms of sets of:
1. Agent's *capabilities:* actions the agent is able to perform;
2. Initial *beliefs* and *commitments*: logical statements about the world, that the agent believes to be true or false;
3. *Commitment Rules* consist of antecedent conditions that are matched against incoming messages and the agent's internal states.

The key component, which determines how the agent acts, is the commitment rule set. Each commitment rule contains a *message condition*, a *mental condition*, and an *action* [12]. To determine whether any rule fires, the message condition is matched against the messages the agent has received; the mental condition is matched against the beliefs of the agent. If the rule fires, then the agent becomes committed to the action. Actions may be *private*, corresponding to an internally executed subroutine, or *communicative*, as described in B subsection.

When operating in multi-agent environments, intelligent agents must generally co-ordinate their actions; they must communicate the proper knowledge and information. To communication content identification is not an easy task [16], because it requires every agent to have a message model of a recipient and to infer the impact of a message on the recipient based on that model. In this work, using a homogenous class of agents and a simple ontology, the communication primitives can be represented of the publishing of the single agent cognitive structure. Each agent generates an 18-bit length word representing its status and its message to the agency. A small bit matrix, called communication room, reports all the agent actions and not the data. In this way the data can be read, from the database only for the bare necessities.

Communication between agents could present sequences of commitments and timing problems appear when this communication takes longer than usual.

An agent sees communication room, in this way the agent, like to look about the environment and when invoked it requests to initialise a transaction with database. Several agents can request the records on the same time.

The resolutions of conflictive situations are solved using the operative system primitives: when the spatial validation agent tries to access to the same records used by a temporal validation agent on the same time, the operative system lock-unlock records [18].

IV. STEP-BY-STEP REASONING

The reasoning procedure to control this tool has been accomplished by using a language like AUML (Agent unified modelling language), in particular using AIP (Agent interaction protocol) [19]. Figure 3 describes agency component works, in function of time, step-by-step.

In the first step of the procedure, sensors store acquired data on the database, in the second step temporal validation agent look about communication room and then it requests to read and lock some/new records that it must be temporal validated.

In the next step it updates communication room and it changes its state. In the following step the agent runs validation-task and it updates the database. After validation, it un-lock the records. Moreover it informs the environment, communicating the results of validation and its new state. It had defined if the data is *valid* or *no-valid* for its knowledge. In the first case the next step of the agent wait for another data. Otherwise agent it produces a sub-goal that the spatial validation agent can solve.

Subsequently, the steps of spatial agent are the same of the temporal validation agents so:

- look about communication room (perceive);
- read and lock dataset;
- update communication room (confirm);
- run its own task to valid data in function of proximity matrix;
- update and un-lock dataset;
- update communication room (confirm);
- wait for another data.

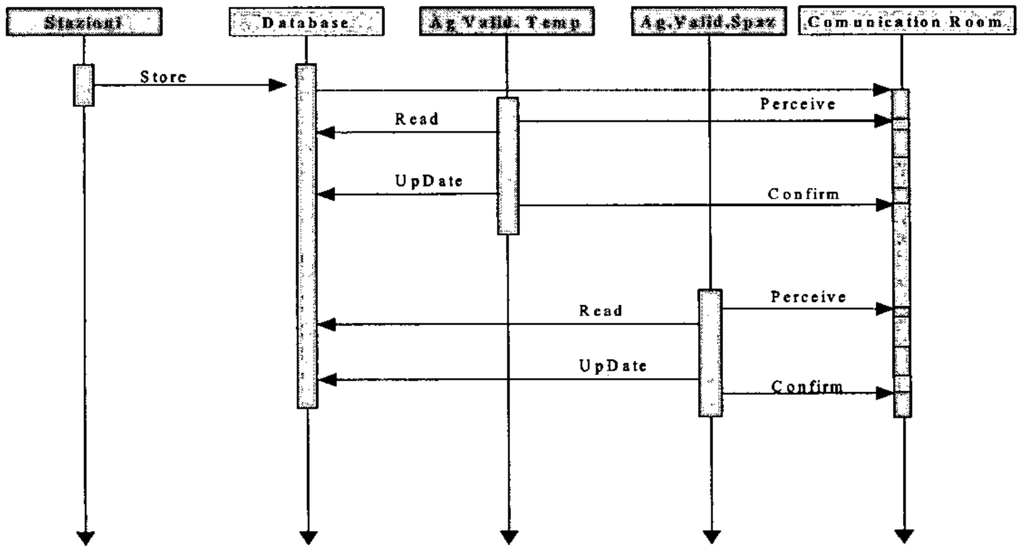

Fig. 3: Diagram about interaction between agents and environment

V. IMPLEMENTATION WITH MATLAB

The implementation of agency whit Matlab/Simulink is possible, because graphical tools provided Simulink allows the design of agents architecture. Really, we can program as a sequence of block diagram.

A main reason to use Matlab/Simulink software, yield on the fact that one of its facilities, brings the opportunity to generate a C-code ready to be compiled and executed even with real time features.

Moreover, this environment contains further advanced and specialised methods of message passing between agents, as for example: agent *actions* take place as part of State flow diagram execution. The action can be executed as part of a transition from one state to another, or depending on the activity status of communication room.

Furthermore, the rules for interaction between the different agent *can be* described with a language that can include all the available elements: to inform, request, accept, reject and each other.

The declarative knowledge (rules of reasoning) of agents has to be implemented using Matlab/Simulink which does not include object-oriented programming capabilities. Therefore, differences between both paradigms (agents and simulation/ control) in the analysis phase are solved by splitting objects and developing reasoning in a sequential order.

VI. CONCLUSIONS

The proposed approach allows us to investigate on data validation techniques based on the use of agents/agency. The results confirm the utility of A.I. applied to environmental data management, usually characterized by:

- high complexity of the monitoring stations;
- presence of missing and/or incorrect data with high volume of the acquired data that often involves the competences of chemistry and environmental expert to a suitable data-validation process.

In the proposed approach two agents family, arranged in an agency, have been provided to validate the data in asynchronous way and to generate a valid predictive model of the pollution.

Subsequently, the authors will propose the development of a web-based service implemented in Taranto, in order to provide a suitable system for town-dweller and for public administrator.

Therefore, this could be a first step to a general application of agents in the world of environmental control and monitoring system by using tools that are common in the former area. The drawback is that object oriented programming are not extended enough and agents have problems to be applied in practical implementations. This paper shows a possibly first success in practical application/implementation of agents in environmental control.

REFERENCES

[1] S. Bandini, D. Bogni, S. Manzoni, "Knowledge based Environmental data Validation, Proceedings" *IEMSS 2002 Integrate Assessment and Decision Support 24-27 June 2002 Lugano*

[2] Emil M. Petriu, Gilles G. Patry, Thom E. Whalen, Abdul Al-Dhaher, and Voicu Z. Groza "Intelligent Robotic Sensor Agents for Environment Monitoring" *VIMS2002 International Symposium on Virtual and Intelligent Measurement Systems Mt. Alyeska Resort, AK, USA, 18-20 May 2002*

[3] Nils J. Nilsson "Intelligenza Artificiale", *ISBN 88-7303-746-1, APOGEO, 98*

[4] Bing Qiao, Jianying Zhu "Agent-Based Intelligent Manufacturing System for the 21st Century", *International Forum for Graduates and Young Researchers at EXPO 2000 HANNOVER*

[5] Gunter Baumbach, "Air Quality Control", ISBN 3-540-57992-3 Springer- Verlag Berlin Heidelberg New York

[6] Verein Deutscher Ingenieure, "Guideline VDI 2310, part 19: Maximum Immission Values Referring to Human Health. Maximum Immission Concentrations for Suspended Particulate Matter", *Berlin: Beuth April 1992*

[7] Platt, U., and D. Perner, "Direct measurement of atmospheric HCHO, HNO_2, O_3, NO_2 and SO_2 by differential optical absorption spectroscopy", *J. of Geophys. Res.* 85, *1980*

[8] Andria G., Di Lecce V., Lanzolla A., "Mathematical model for measurement and characterization of air pollution in areas with high road-traffic levels", *XVII Imeko World Congress, June 22-27, 2003, Dubrovnik, Croatia*

[9] Amigoni, F., Somalvico, M., Zanisi, D. "A Theoretical Framework for the Conception of Agency. Int. Journal of Intelligent Systems", *1999, p. 449-474*

[10] Amigoni, F., Schiaffonati, V., Somalvico, M. "A Multilevel Architecture of Creative Dynamic Agency", *Foundations of Science, Kluwer Academic Publisher, 2001*

[11] Hyacinth S. Nwana. "Software Agents: An Overview". *Knowledge Engineering Review, Vol. 11, No 3, pp. 1-40, Sept 1996. Cambridge University Press, 1996.*

[12] J. Ll. Dela Rosa, A. Oller, J. Vehi, J. Puyol, "Soccer Team based on Agent-Oriented Programming", *Robotics and Autonomous Systems, 21, pp.167-176, 1997*

[13] J.LL. de la Rosa, J. Aguilar, I.Serra, "Euristics for Cooperation of Export Systems. Application to Process Control", Ed. PIAR, 1994. ISBN 84-605-0725-9.

[14] M. Wooldridge, N.R. Jennings, "Intelligent Agents : Theory and Practice", *Submitted to Knowledge Engineering Review, October 1994, Revised January 1995.*

[15] "COORDINATION: Linguistic Support for Multiple Cooperating Agents". ESPRIT Project No 9102 in the Basic Research Action. Work Area: Basic Aspects of Multiple Computing Systems. Dates: 1.2.94 - 31.1.96.

[16] P.J.Gmytrasiewicz, E.H.Durfee, D.K.Wehe. "The Utility of Communication in Coordinating Intelligent Agents". *Proceedings of the 9th National Conference on Artificial Intelligence, 1991.*

[17] Y.Shoham. "Agent-oriented programming", *Artificial Intelligence, 60 (1993) 51-92.*

[18] Andrews S.Tannenbaum, Albert S. Woodhull, "Sistemi Operativi. Progetto ed implementazione" *Utet Libreria ISBN 88-7750-424-2*

[19] J. Odell, H. Van Dyke Parunak, Bernhard Bauer, "Extending UML for Agent", *AGENT-ORIENTED INFORMATION SYSTEMS (AOIS-2000) Conference, 5-6 June 2000, Stockholm (Sweden)*

[20] Broersen M. T. Piet, "The Quality of Models for ARMA Processes", *IEEE Transactions on Signal Processing, Vol. 46, N. 6, 1998*