Pre-print version
V. Di Lecce, C. Pasquale, V. Piuri, "Agent-based communication for an environmental monitoring application", in Proc. of the 2004 International Conference on Computer, Theory and Application (ICCTA 2004), Alexandria, Egypt, September, 2004

Agent-Based Communication for an Environmental Monitoring Application

V. Di Lecce[1], C. Pasquale[2] and V. Piuri[2]
[1] DEE – Politecnico di Bari – Via Re David 200 – 70121 Bari – Italy
[2] DTI – University of Milan – Via Bramante 65, 26013 Crema (CR) – Italy
[email: dilecce@poliba.it , pasqualeciro@libero.it, piuri@dti.unimi.it]

Abstract –This paper describes a system based on agents to analyze, validate and characterize environmental data relevant to monitor atmospheric pollution in an urban area with high pollution level. The proposed MAS architecture is a multi-agent system developed to support distributed workflow management environments where distributed components are used to implement each of the workflow steps. Indeed air quality monitoring systems are typically used by experts that are able to manage complex data and to understand the effect of single pollutant, but common users are unable to do it. Using an adequate ontological approach, it is possible to define a system capable of doing data mining and giving information to unskilled users too.

Key-Words: Agent, multi-agent system, ontology, data communication, data mining.

I. INTRODUCTION

The choice of a monitoring process for air pollutants depends on the typology and relevant properties of the substance that must be acquired, even if it has crucial relevance to identify the kind of information that is gained from the measured values jointly with the relevant uncertainly level too. 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 [1]. 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 [2].

As it is well known, pollutant concentrations depend on emissive sources, geographic conformations, and meteorological conditions (particularly wind and rain), capable of modifying atmospheric dispersion of pollutants. Therefore to evaluate the air quality properly, it is necessary to work out a sophisticated monitoring system to measure chemical substances and meteorological quantities [3].

The analysis of environmental data, recorded in several monitoring stations installed in a city (Taranto, Italy), are an example of a real case.

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.

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 medium values hourly (valid data). These values are considered valid if at least 75% of the raw data are right.

In this paper, we present an experience in environmental data treatment, using agent techniques to analyze and validate data acquired by the presented monitoring net. The software agent technology can be an interesting tool for the creation of new models for complex software systems. In a project based on distributed computer system, software agents are basically designed to cooperate (either with other agents or with humans) in a seeming intelligent way.

This work introduces a model of Multi Agent System (MAS), layered in five levels, suitable to supply answer to a query characterized by a high semantic level. This is possible using progressive interpreting/multiplying techniques of a complex query in simple queries according to well known compilers and O.S. Theories.

With this approach, a large number of agents will be generated and in this regard they will lead to many problems, such as the server loading, the scalability problems and communication process among agents. For this reason an appropriate understanding of what agents will communicate through the messages is indispensable

An environment, in which agents are speaking/working, is like a public square in which men directly communicate among themselves. For example if two men are speaking about a clearly definite argument, and there is somebody that knows the discussion object and he is interested to share his knowledge, he answers to the questions of the conversation between two men without being directly questioned.

Human language is characterized by a high semantic level; instead the language used among communication processes is often characterized by a low level of implementation. A notable effort is essential in the work to develop software capable of representing this sort of communication. As described in other works [4], KQML has been used for agent communication.

The Foundation for Intelligent Physical Agents (FIPA) has specifications for interaction protocols, communicative acts, and content messages for agent communication. KQML is a subset of both the complexity and completeness of these specifications. In this project, the decision was made to use KQML protocols. In the context of this project, KQML protocols

Under grants of MIUR Research projects: "APE: Agencies for PErception in environmental monitoring", Research Unit of Università di Milano & Politecnico di Bari, Italy.
*Contacting author: V. Di Lecce, DEE - Politecnico di Bari - Via Re David 200 - 70125 Bari - Italy, dilecce@poliba.it, tel.+390805963309, fax+390805963410

are sufficient to support the workflow-based communication. It is necessary to use a high-level language according to Chomsky theories to simulate this example of spoken communication among intelligent agents.

Aid in this study is the definition of a correct ontology, that allows us to define messages and communication acts for a specific application with a concise meaning and without ambiguity.

Ontology is a term used to refer to the common sense of some domains of interest. Ontology is a collection of concepts and their interrelationship, which can collectively provide an abstract view of an application domain [5, 6].

For example many researchers investigate the correct ontology to extract and access information within the Web[7, 8]. But there are two distinct problems/tasks for an ontology-based model: one is the extraction of semantic concepts from the keywords and the other one is the actual construction of the ontology for a Multi Agent System (MAS) architecture [9]. With regard to the first problem, the key issue is to identify appropriate concepts for problem domains. For the second problem, we would like to generate ontology, in order to make our approach scalable, as shown in other works [10].

In the presented analysis the manual extraction of a (specific) ontology is performed in environmental application. It is necessary to define a basic ontology, that defines a mutual relationship between some types of general vision, measured environmental parameters and common sense about air quality conditions.

The paper is structured as follows the second section discusses the problem of data validation. The third section gives an overview of agents communication (approaches and innovations). The fourth section describes system architecture implementation. Test and conclusions close the paper.

II. DATA VALIDATION TECHNIQUES

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 not interpretable 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, discordant values with the expected behaviour of the monitored quantities.

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

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

To reduce the problem of missing data some reconstruction techniques, based both on 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, founded on a suitable polynomial interpolation, that is able to describe the behaviour 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 are compared with its trend. This was computed by its historical dataset for a possible margin of error. There are a lot of methodologies, known in literature as A.R. (auto regressive method), A.R.M.A. (auto regressive moving average), to compute the trend [11]. This paper describes how the approach, shown in Fig. 1, of 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.

Several tests have been carried out, by eliminating a subset of data in a random way, from the complete time series, in order to evaluate the reliability of the developed model. 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 than 4% has been obtained. A typical example is shown in Fig. 1, and describes the daily average concentrations relevant to the year 2001, in which some data are considered as missing [12].

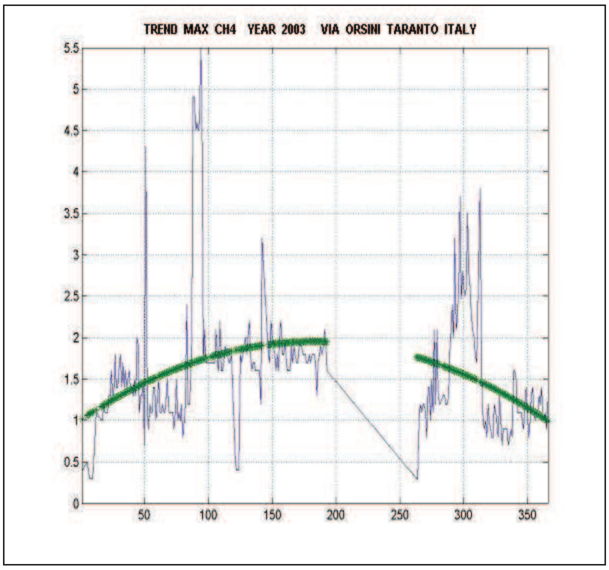

Figure 1 Interpolation trend of a monitoring station about CH_4

III. AGENT COMMUNICATION

Although agents have been used in a wide range of applications, they do not represent a universal solution. There are many applications for which conventional software development paradigms are far more appropriate.

According to Wooldridge and Jennings [13], every agent has been designed with five fundamental blocks: local data , processing unit and control unit, a buffer and an interface. The local data are a necessary memory to use and acquire the knowledge for software agent. In this way the agents act as sellers and buyers using a bilateral communication. The processing unit includes qualities that allow or facilitate achievement or accomplishment of agent’s goal. The control unit allows agents to interact with each other by making reciprocal commitments and carry out particular tasks. A buffer is necessary to handle processing data. An interface characterizes every agent for the management of communication protocol.

Analyzing the communication problems, when a user asks a query to the MAS, in an open environment where autonomous agents can contract, we must design a defined architectural model, that has an important role to determine the rules of interaction, between who interacts and the method of interacting [14].

This section gives a brief overview of the Agent Communication Language (ACL) and supporting protocols, which have been the methodology of choice for the implementation of agent communications, traditionally. Therefore it explains how this approach is implemented using UML.

The motivation for ACL was to formalize a method by which agents can communicate effectively and efficiently. The message format supplies the agent with knowledge of which the agent it is communicating to, a protocol for establishing dialogue, the language by which agents are communicating, terms by which other agents will interpret expressions, and exception handling. A lot of researchers [15, 16] use the well known KQML to do it.

It is not the scope of this paper to cover the KQML specification entirely but we want to test the use of a

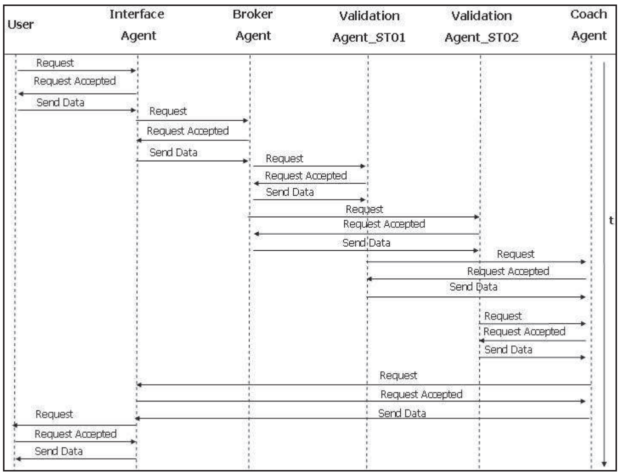

Figure 2 UML chart of a typical MAS communication

well known protocol as SMTP (Simple Mail Transfer Protocol), to carry KQML.

Analyzing the action of changing message among agents into the MAS when *User* makes a request, it is possible to observe no overlap. Figure 2 shows an example of UML chart of a typical MAS communication.

According to the same Saeb's motivation our architectural approach is structured as follows: the design of agents in a way that it can be implemented using the SMTP and TCP/IP protocols; agents are taken as subclass of each one. This method simplifies the coding and provides distributed task with low processing time [17].

IV. THE SYSTEM COMMUNICATION ARCHITECTURE

A multi agent system can be defined as an organization composed of autonomous agents and proactive agents, that interact with each other to achieve common goals. There are a lot of goals that a MAS can satisfy, but to do it, the agents need to know information about the environment in order to take decisions [9].

Basic idea is that all users (humans or computer based systems), using WEB, can formulate a query to a system to obtain environmental information. For this reason a MAS, based on classes of agents, is used. A class identifies agents that have same features and common goals, but they are able to satisfy their own scope with different approaches and with different results. In the proposed system the researchers have planned a multitude of agents with the same features (capable of sending and receiving messages in Internet). Engineering the MAS, three macro classes are identified: First one is Interface area (the goal is to interpret external queries according with services offered by the MAS), the second one is Brokerage area (agents with the one goal: to manage information of traffic flows) and the third one is Analysis area (agents with specific features to satisfy environmental applications: to validate and to forecast pollutant values). To implement the system with more scalability, five sub classes are introduced: *Interface (one agent), Broker (one agent), Validation (more agents), Forecast (more agents) and Coach (one agent).*

Interface agent translates the message from human/natural language into an Agent Communication Language (ACL) [18]. In order to make the integration easier in Internet, the first step is to use a drop-down menu. In this way a user can formulate a query using an assigned form.

Broker agent decides which agents can satisfy the requirement. Broker analyzes a local database in which services offered by MAS are stored. And starting from one query, it produces as many messages as the request needs. The language, that *Broker* uses to communicate, is generated using understandable and common language for all agents.

Validation agents try to validate data, using the knowledge of the geographic position in which sensors are placed.

Forecast agents have their own knowledge, based on a forecasting model, known in literature. Each forecast agent applies its analysis method, in this way it is possible to define which is the best forecast method among those ones that are used.

Coach is a complementary agent of the *Broker*, because it is able to assemble - information contained in the messages, that the *Validation* and *Forecast* agents have sent. In the end of the process the *Coach* sends a message to the *Interface*, that translates the answer from ACL into a human language.

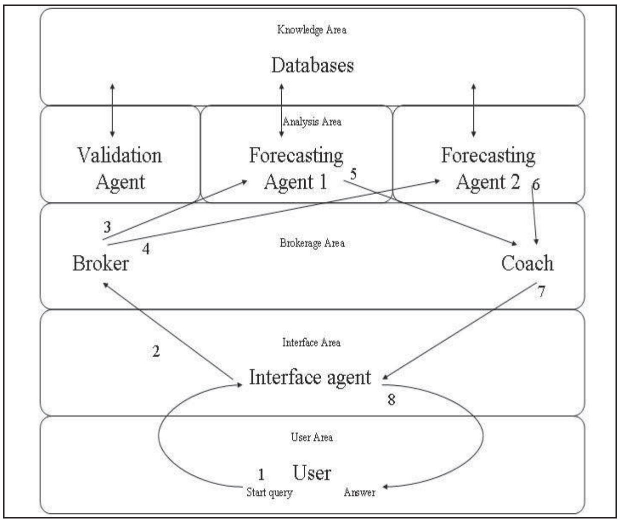

Figure 3 Dataflow of Multi Agent System architecture for environmental monitoring system

V. ONTOLOGY AND COMMUNICATION ACTS

There is no “correct” way or methodology for developing ontology. We also need to remember that ontology is a model representing reality of the world and the concepts of ontology must reflect this reality. Among different solutions, we have defined an initial version of the ontology, used in start up and debug of environmental monitoring applications.

For example a user can effect a query compiling a form through a drop-down menu.

Action	Pollutants	Place	Range	Relia-bility	Time	Historical dataset

The “Action”, that a user requires, is a possible application that the MAS can satisfy.

In the study for the analysis of the environmental data the possible “Actions” are:

1. To validate a datum (in the space),
2. To foresee a datum,
3. To foresee a datum with technique # 1,
4. To foresee a datum with technique # 2,
5. To foresee a datum with technique # 3,
6. To query database on the acquired historical data,
7. To calculate the average,
8. To calculate the maximum one,
9. To calculate the least one,
10. To calculate the most reliable one.

The described actions must be completed by the other fields that identify and clarify the type of question introduced by the user. If one of the fields is not inserted, it is compiled with default value.

The “Pollutants” and every weather parameter are necessary elements of the ontology to define all the queries. In the proposed menu the cardinality of this field is 29. The first ones are 23 standard chemical and weather parameters: $S0_2$, NO, $N0_2$, $N0_x$, CO $C0_2$, CH_4, NMHC, THC, Benzene, Toluene, Xilene, Polycyclic Aromatic Hydrocarbons, 03, (Ozone), Radiation, Total Suspended Dusts, PM_{10}, Temperature, Wind Direction and Wind Speed, Rain, Damp, Relative Pressure.

The remaining 5 indicators are linear combinations among different parameters to carry out: industrial pollution, car traffic pollution, combustion pollution, general air quality level, meteor-climatic condition.

Industrial pollution is characterized by an indicator that is average of the following products of combustion analyzed by fixed stations: SO_2, NO, NO_2, NO_x, CO, CO_2, CH_4, PM_{10}. The partial value is weighed in relation to speed and direction of the wind. The indicator introduces in output 5 different results: low, middle-low, middle, middle-tall, tall.

Car traffic pollution is characterized by the average of Benzene, Toluene, Xilene and Polycyclic Aromatic Hydrocarbons.

Pollution due to domestic combustion (boilers and kitchen) is set by the average weighed of CO, CO_2 and CH_4, the result is weighed in relation to the temperature.

Air quality level is average of the last three parameters valuating parameters of temperature, pressure and wind (speed and direction).

Meteor-climatic conditions is specified by a chart that evaluates temperature, damp, atmospheric pressure and wind in speed and direction.

“Place” is another necessary parameter to identify the query in relation to spatial concentration of a defined pollutant.

The forecast applications have been chosen among six stations located on the territory that acquire and store data. Besides we have the possibility to define some global results such as the average of all the stations: “town condition”, the average of three stations in the urban centre, the average among three stations in elevated traffic zones, the average weighted in operation of the distance among 6 stations to calculate the condition of residential zones of southern Taranto.

“Range” is an auxiliary parameter in the use of the techniques of spatial validation. The measure acquired by a station is compared to those neighbours using a proximity chart. The validation task finds relations among the measured datum and the values of the other stations, using technique of weighed average applied to some stations that are near (next) to the station that has provided the value in examination. The validation value is provided by the near stations. The weight is the function of the distance. We can choose to use a little set: 3 stations, middle: 4 stations, or wide: 6 stations.

The degree of “Reliability” is a further choosing parameter, that can be selected, if the datum is valid. The reliability of the measure is a parameter defined by laws, but it can be selected by the user too. Cardinality of this field has been defined as equal to ten. The number of choices is given by the degree of reliability and it will have a 99% variation at the least to foresee values from the Italian normative to make the datum valid. The criteria of evaluation indicators about air quality have been modified with the D.M. 2 April 2002 ns. 60, that it was acquired by Directives 1999/30/CE and 2000/69/CE.

"Time" of the validation is a necessary parameter tightly to ask for: a spatial validation, a forecast or an examination about a pollutant or an environmental parameter.

The "Historical dataset" allows the user to choose what dataset an agent has to use for calculating the forecast. In our application, the cardinality of this field has been defined as equal to 10. The possibilities of choice are: 1 month, 3 months, 6 months, 1 year, 2 years, 3 years, 4 years, 5 years, 6 years, and 7 years.

Using a sequential model, the user sends the query to *Interface* agent, that translates the message into a different message using an ontology that is able to be understood or interpreted by the MAS agent. Then *Interface* agent replies the message to *Broker*, that breaks down the goal of query into elementary tasks or essential objectives. Broker sends to *Validation* and *Forecast* agents as many messages as the sub-goals are. The agents reply the answers to the *Coach*. It waits for all messages and then it assembles the information and it replies to *Interface* agent that converts the final message to the user.

VI. TESTBED

This section provides the simulation results of the proposed architecture, that satisfy the expectative when there are only five types of agent in the environment. Applying the architecture to six station in which there are a "Spatial Validation" agent and three different "Forecast" agents for each station, moreover an "Interface" agent a "Broker" and a "Coach". We are testing the QoS (Quality of Services) of this complex system as shown in [19]. The QoS (Quality of Service) of the MAS during interchange messages communication must be measured. It will be verified that the processing time is greater than communication time $T_p > (t_1 + t_2)/2$.

We are testing this approach, to evaluate the performance and the limits of our proposed approach, increasing the number of agents in a common and heterogeneous environment.

VII. CONCLUSION

In this work the authors propose ontology for a multi agent system in environmental application that shows how agents, that satisfy simple objectives, can solve more complex problems. For doing it, it is necessary to introduce only two classes of agents: the first agent is able to convert to/from a natural language, the second one is able to manage (disassemble and gather up) services offered by agents placed in a common environment.

Combining these elements with an approach arranged in layer it is possible to carry out a scalable system that is usable in many other applications.

The proposed approach shows as a query: - *How is air quality in tomorrow*?- is divided in a lot of simplex queries. The combination of information given by different agents to the *Coach* and the communication *Coach-Interface-User* solve the problem of the system apparent intelligence. Some results show the time and the way that a message needs to be generated by a user. Then it is decomposed in other simple messages that MAS agents can satisfy.

An example of communication is shown in figure 4: a user asks for a query to know a forecast of industrial pollution in Taranto. In the MAS the agents generate as many messages as there are necessary to satisfy the query of the user.

Communication example

IP TABLE	
Translator:	1
Broker:	2
Validation Ag.:	3
Forecasting Ag#1:	41
Forecasting Ag#2:	42
Forecasting Ag#3:	43
Coach:	5

				Action	Pollutant	Place	Range	Reliability	Time	Historical dataset			
user2translator				forecast	Industrial pollution	Taranto			Today				
	IP From	IP To	Protocol number	Action	Pollutant	Place	Range	Reliability	Time	Historical dataset	Answer	Coach goal	# of replied messages
translator2broker	1	2	xxx	forecast	CO	Taranto			11/01/2004				
broker2validation_ag	2	41 (diverso da 3)	xxx	forecast	CO	StAction #1			11/01/2004			Average	3
broker2forecast.#1	2	41	xxx	forecast	CO	StAction #1			11/01/2004			Average	3
broker2forecast.#2	2	42	xxx	forecast	CO	StAction #1			11/01/2004			Average	3
broker2forecast.#3	2	43	xxx	forecast	CO	StAction #1			11/01/2004			Average	3
forecast.#1_2coach	41	5	xxx	forecast	CO	StAction #1		default	11/01/2004	5 years (default)	10,00	Average	3
forecast.#2_2coach	42	5	xxx	forecast	CO	StAction #1		default	11/01/2004	5 years (default)	11,00	Average	3
forecast.#3_2coach	43	5	xxx	forecast	CO	StAction #1		default	11/01/2004	5 years (default)	11,00	Average	3
coach2translator	5	1	xxx	forecast	CO	StAction #1		default	11/01/2004	5 years (default)	10,67		
translator2user					Industrial pollution	Taranto		default	11/01/2004	5 years (default)	10,89		

Figure 4. Exchanged message among agents in Multi Agent System architecture

REFERENCES

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

[2] 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.

[3] Platt, U., and D. Perner, "Direct measurement of atmospheric HCHO, HNO2, O3, NO2 and SO2 by differential optical absorption spectroscopy", J. of Geophys. Res. 85, 1980.

[4] M. B. Blake, "Using Agent Control and Communication in a Distributed Workflow Information System, Cooperative Information Systems", Lecture Notes in Computer Science, Springer-Verlag, (Proc. of the 10th Int. Conf. on Cooperative Information Systems (CoopIS 2002), Irvine, California, (November 2002) 163-177.

[5] T. R. Gruber, "Toward Principles for the design of Ontlogies used for Knowledge Sharing", in Proc. of International Workshop on Formal Ontology, March 1993.

[6] N. f. Noy, D. L. McGuinness, "Ontology Development 101: A Guide to Creating Your First Ontology " Technikai report KSL-01-05. Knowledge Systems Laboratory, Stanford, USA, 2001.

[7] Latifur Khan, Feng Luo, "Ontology Construction for Information Selection", Proc. of the 14th IEEE International Conference on Tools with Artificial Intelligence ICTAI, 2002.

[8] Jane Hunter, "Enhancing the Semantic Interoperability ofMultimedia Through a Core Ontology", IEEE Transaction on Circuit and System for video technology, Vol. 13, No.1, January 2003.

[9] Sthephane Faulkner, Manuel Kolp, "Ontological Basis for Agent ADL", The 15th Conference on Advanced Information Systems Engineering, CAISE, Velden, Austria, 16 - 20 June,2003.

[10] N. Guarino, C. Masolo, G. Vetere, "OntoSeek: Content-Based Access to the Web", IEEE Intelligent System, Vol. 14 No 3, pp.70-80, 1999.

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

[12] 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.

[13] M. Wooldridge and N. R. Jennings " Pitfalls of agent- oriented development," Proceeding of the second international conference on Autonomous agents, pages 385–391, 1998.

[14] V. Di Lecce, C. Pasquale, V. Piuri, "Information and knowledge sharing in a distributed system: application to environmental monitoring system", The 46th IEEE International Midwest Symposium on Circuits and Systems Nile Hilton Hotel, Cairo – Egypt, 27th - 30th of December 2003.

[15] M.B. Blake, KOJAC: Implementing KQML with Jini to Support Agent-Based Communications in Emarkets, Proc. AAAI-2000 Workshop on Knowledge-based Electronic Markets (KBEM2000) (AAAI Press, Technical Report WS-00-04) Austin, TX (August 2000) 1-8

[16] Y. Labrou and T. Finin, A semantics approach for KQML – a general purpose communication language for software agents. Proc. 3rd Int Conf. on Information and Knowledge Management (CIKM-94), Gaithersburg, MD (December 1994) 447-455.

[17] M. Saeb, C. Fathy "A modified Diffusion Load Banlance Algorithm Employing Mobile Agents", WSEAS International Conference onCircuit, Systems & Computers, Corfu, Greece, July, 2003.

[18] Jason Jen-Yen Chen, Shin-Wei Su, "AgentGateway: a communication tool for multi agent system". International Journal of Information Science, 2002, Elsevier Science Inc.

[19] X. Xiao and L. M. Ni, *Internet QoS: a big picture*, IEEE Network, vol. 13, no. 2, pp. 8-18, Man-Apr. 1999.