Data Fusion and Inference Systems for Environmental Decision Support

D.A. Sarigiannis

European Commission - Joint Research Centre
Institute for Systems, Informatics and Safety
Systems Analysis and Information Assessment Unit
T.P. 210, 21020 Ispra (VA), Italy
e-mail: denis.sarigiannis@jrc.it

Abstract

Integrated environmental management in urban areas is nowadays considered as a sine qua non objective of Community and national environmental and development policies. Large amounts of scientific information on the state of the environment are now available from a large pool of data sources. This work presents the on-going research efforts at the Joint Research Centre of the European Commission towards integration of these data sources and effective coupling of the environmental information with appropriate models and decision-support tools. State-of-the-art data fusion techniques and inference systems are used in an integrated environment in support of multi-criteria environmental assessment and policy-making.

Introduction

Integrated environmental management in urban areas requires the simultaneous consideration of a multitude of information classes, including environmental quality data, impact pathway models, economic analyses, the respective regulatory framework, and the priorities of the stakeholders involved in (or affected by) the decision-making process. Assimilation of these types of information in a comprehensive yet functional framework is necessary for the development of user-friendly decision support tools which foster public dialogue and enhanced stakeholder participation in urban policy-making. (Fedra 1990). Integrated air pollution assessment in particular includes accurate representation of the pollution distribution in time and space (environmental pressure), identification of the main emission sources (causes), and evaluation of the possible alternatives for coping with the observed environmental burden.

The information base for urban environmental management decision-making should stem from measurements that ought to be:

(a) reliable, objective and accurate at the local level, and

(b) representative and comparable at higher (regional, national, and international) levels.

Site selection of ground-level measurements determines the accuracy and representativeness of point measurements; it is thus of fundamental importance for spatial air quality comparisons and for the determination of the effectiveness of air pollution abatement measures and policies at the local and regional levels. Furthermore, the choice and siting of the measuring network represents a factor of significant economic relevance for policy-makers.

The actual siting of measuring stations should theoretically be based on criteria of population density, historical data on air emissions inventories, location of industrial and traffic emissions, environmental conditions and meteorology. In practice, these variables are most often weighted in an empirical way and are highly site-specific. Moreover, the optimal design of the measuring network requires initial information regarding the geographic patterns of the ambient air concentration. Existing modelling tools require validation and bench marking in a variety of urban environments limiting thus their reliability.

Earth observation from satellites offers a new opportunity to integrate spatial information on pollution dispersion. This can be provided on the one hand by the assessment of the phyto-sanitary state of natural ecosystems and on the other by the estimation of the rate of atmospheric turbidity due to aerosols, namely particulate optical thickness. Atmospheric turbidity shows the particulate matter load but it does not provide information on the chemical nature of the substances present in the atmosphere. However, particulate matter is a tracer for pollutants dispersion, and turbidity is an appropriate pollution indicator due to its strong correlation with the concentration of small particles, that have recently been proven to be most harmful to human health (Lipfert 1994; Pope *et al* 1995). To date, Earth observation-derived data are used extensively in atmospheric studies at the global scale. They provide significant aid to better understanding of both natural processes and anthropogenic impact on the environment (CEOS 1992). There are, on the contrary, very few examples of use of Earth observation data in urban or regional air pollution investigations, where spatially resolved information needs to be derived (Sifakis and Soulakellis 1996).

Features of environmental systems

Environmental systems and, consequently, environmental management problems have several attributes which render their formal representation different from that of other, say, industrial systems. Some of the distinctive features of environmental systems are highlighted below:

- Dynamics. Environmental systems evolve with time. Many environmental processes are best treated as continuous for modelling purposes. Other processes, however, are better described as single events requiring discrete event models.
- Spatial coverage. Environmental systems involve physical processes, which take place in a two or three-dimensional space. Modelling such processes rigorously usually needs the solution of sets of partial differential equations linking spatially referenced variables with time referenced ones. Data are stored in spatial databases often via a geographic information system (GIS) to assist in spatial analysis.
- Complexity. Environmental systems are complex, usually involving interactions between physico-chemical and biological processes with social ones. Models of such systems require multi-disciplinary approaches, the success of which depends heavily on the establishment of common grounds for communication among different scientific disciplines and various classes of stakeholders.
- Randomness. Many environmental processes are stochastic. Parameter uncertainty characterises the models representing them, and it is common to define the parameter space only approximately. Statistical analysis and qualitative interpretation of model equations are therefore warranted.
- Periodicity. Many environmental processes are periodic in time, adding thus a further degree of complexity in parameter calibration and validation. Data storage can also be affected by the periodicity of input data.
- Heterogeneity and scale. The media in which environmental processes take place are not homogeneous and cannot easily be characterised by measurable parameters. Processes in different media may have quite varying characteristic time and space scales.
- Paucity of information. Observational data on environmental systems, particularly on internal states, typically suffice only for the characterisation of simply parameterised models.

Users of environmental management decision support systems vary, including scientists, managers (decision-makers) and concerned stakeholders. Each of the user categories has different needs, knowledge acquisition and communication standards, and final objectives.

Integrated solutions to environmental problems are inevitably at the interface between the domain of observation, studies and scientific action, which allow the emergence of knowledge, and the domain of political strategies and social pressure. These domains have their own indicators, their concepts, their culture and behaviour, as well as their sources of bias and misunderstanding.

The main feature of our work is the identification of the need and the development of a state of the art methodology and of the respective tools for integrating the variety of information classes affecting decision-making on environmental problems. First all relevant stakeholders and the actual decision-makers are identified, and the fundamental question a decision is called to answer is well formulated. Second, the relevant scientific/technical information is organised and put in an easy-to-manage form. A series of indicators correlating the decision-makers' question and the scientific information (data) is created. Then, the sequence of the use of the different tool is as follows:

a) Simulation of technical/environmental system operation
b) Calculation of impacts
c) Calculation of values of indicators
d) Modification of the system or operational assumptions
e) Revision of the process to evaluate the impact of the newly selected set of assumptions on the values of the indicators

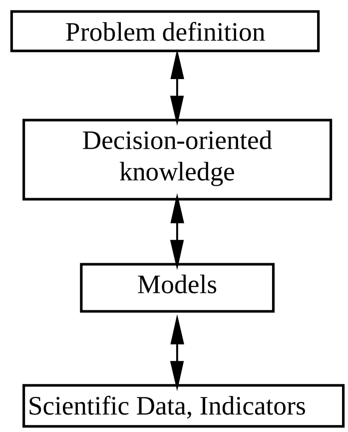

Fig. 1: Flow chart of relevant information classes among stakeholders levels

The information requirements to support the complete definition of an environmental problem and a structured approach towards participatory decision-making are identified and the flow of different information classes among stakeholders is facilitated. This is the first step toward decision analysis. The flow between the different levels of information necessary for integrated environmental management is given in figure 1.

Inference models may play a double role in this process:

(a) they may be used to complete the information necessary to define well (in the mathematical sense) the environmental problem, and
(b) they may be used to support the actual decision-making by predicting the behaviour of the environmental

system based on the historical knowledge base. In this paper I present the methodology and some results of the application of inference models to air pollution distribution assessment and to the optimisation of air quality monitoring networks.

Methodology

Environmental Data Fusion

Fusing environmental data coming from a variety of information sources together with other data classes, such as socio-economic indicators is necessary for effective environmental management and decision-making. In the case of air pollution assessment, the so far state of the art involved the coupled use of atmospheric dispersal models supported by ground-based measurements to calculate continuous in space and time pollution profiles (Günther, Radermarcher and Rieckert 1995). Recent developments in satellite image processing allow the integration of data coming from Earth observation satellite missions in air quality monitoring (Sifakis, 1995). The effective integration of EO-derived data with ground measurements of air quality and atmospheric pollution transport models necessitates, however, the development of an integrating computational environment that permits the assimilation of these diverse data types. For this operation to be user-friendly this information technology tool be based on an object-oriented platform. Object-oriented data management and modelling technology is the most appropriate to use for building the ICAROS® tool kit.

Environmental information processing imposes specific requirements on data management platforms. Among the most important are the efficient modelling and storage of complex objects and the related access operations. Storing all components of an object together increases the efficiency of processing search queries and consequently

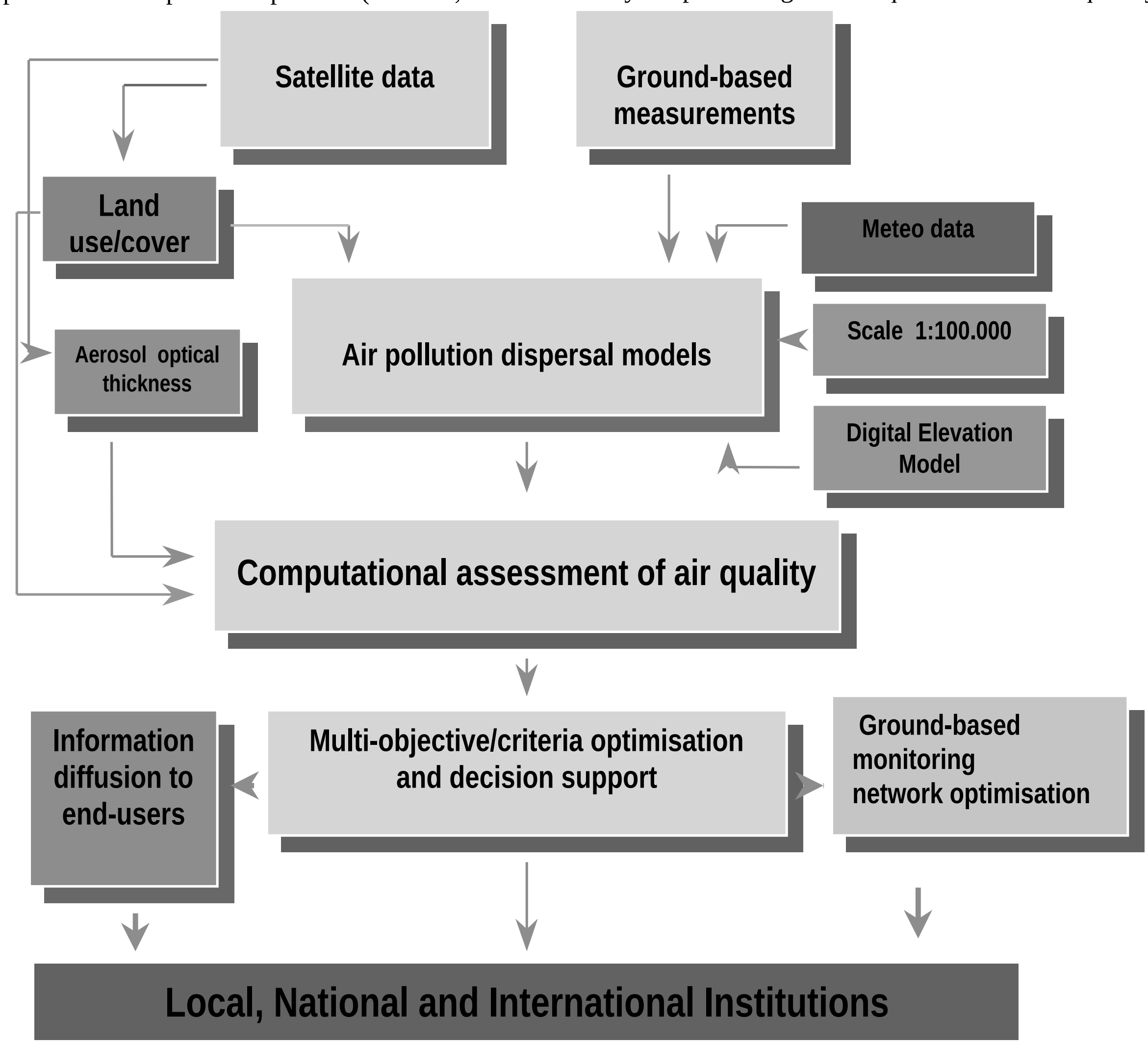

Fig. 2: Information flow in the integrated ICAROS environmental data assimilation system

the use of environmental data for decision-making support. At the same time, every component of the object can still be accessed directly and without restrictions. Object-oriented data management systems allow extensions in the form of application-specific objects and operations. Thus, they enable the management of long and unstructured documents or even pixel graphics. They are not limited to fixed length attributes, as is normally the case with relational databases. On the contrary, they allow processing of very large data objects effectively supporting the different elements impinging upon environmental decision-making.

The fusion of different data sources within ICAROS® is organised in a modular way in order to allow for variations in the time and space scales of reference related to the input data and to the information content required by identified end-users (see figure 2). Each data source (Earth observation data, ground-level measurements, and atmospheric transport models) is considered as a distinct but comprehensive object. The information contained in these objects includes air quality data (pollutants spatial cover and concentration, diffusion pathways, etc.), meteorological data (cloud cover and time evolution) (see figure 3), data on land use and emissions intensity.

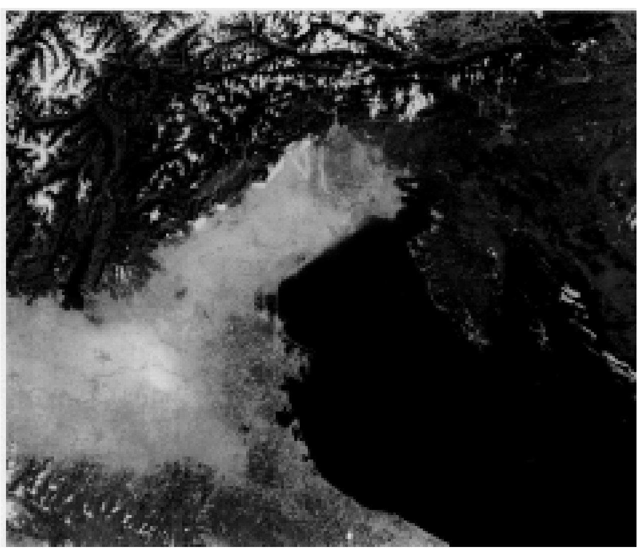

Fig. 3: Pollution cloud in the Po valley (N. Italy)

Furthermore, it includes meta-information allowing the user to identify the origin of the environmental data, their uncertainty levels, the spatial and temporal scales of reference, and the format requirements for communication with the other data sources available in the system. In this way, users have the option to use the ICAROS® data assimilation environment with environmental data acquisition techniques and atmospheric transport models that satisfy the needs of the specific local applications.

The following urban air quality-related issues can be affronted with state-of-the-art methodologies and techniques for ground-level remote sensing measurements:

- Monitoring of power plant emissions and other warm exhausts (industrial stacks, flares) by remote sensing
- Monitoring of heterogeneous and diffuse emissions by remote sensing: waste / garbage deposits, industrial areas, gas filling stations and gasoline tanks, highways, agriculture
- Investigation of sources of particulate matter and optical visibility near the ground
- Investigation of air quality by path integrated trace gas measurements
- Investigation of ozone distribution by mobile ozone lidar / Doppler sodar system

The ICAROS® integrating environment is able to support computationally the state-of-the-art techniques used for the resolution of the above issues. ICAROS® is currently being implemented in two pilot projects aiming at the best use of varied environmental stress information to optimise the air quality monitoring network in the larger area of Brescia (urban conglomerate) and the area of Cayenne, the capital of French Guiana.

Statistical processing of the values of aerosol optical thickness derived from Earth observation via satellite with regard to the concentration of principal urban pollutants has so far indicated that there is significant correlation between EO-derived air quality indicators and measured concentrations of pollutants. Such an attempt has recently been made successfully by the author and his colleagues using data from the greater Athens area in Greece and the area surrounding the province of Brescia in Northern Italy.

System and User Requirements

The ICAROS® computational environment is mainly addressed to environmental and transport policy- and decision-makers at the urban and regional levels. The typical user is the technical service of such decision-making bodies, whereas the final recipients of the information generated by ICAROS® are the actual decision-makers. Decisions in the area of environmental protection in general, and in particular in air pollution control and abatement need to take into account information ranging from validated technical data to economic and social factors. The technical information on the status of the environment needs however to be pre-processed before it can be readily used in order to integrate in the most efficient, valid, and user-friendly manner the array of different data types characterising environmental systems. In the case of air pollution control and abatement this information includes data on the causes of pollution, on atmospheric transport processes, and on the effect pollution has on the affected ecosystem. The user requirements for this information include:

- dynamic character of the input data regarding the evaluation of the actual state of pollution;
- validated detailed spatial distribution of pollution during its generation and transport;

- timeliness in providing the necessary input to the decision-makers;
- accuracy in the amplitude and spatial-temporal profile of pollution and its effects on humans and the ecosystem;
- objectivity in the assessment of the sources, distribution and impacts of pollution (especially when environmental treaties and relevant legislation at the national or Community level are implemented)

Earth observation can help reduce the errors inherent to dynamic atmospheric transport models by calibrating and validating the transport models using high resolution satellite information and the appropriate information processing algorithms (Sifakis and Deschamps 1992). Furthermore, EO-derived information being a direct depiction of reality, it can be used for the verification of the degree of implementation of international environmental treaties (e.g. LTRAP) and Community directives by signatory parties and member states. Through the development of a knowledge-base based on EO-derived data and the appropriate inference engine the necessary input to the users community can be provided timely and economically with reduced uncertainty.

Inference System for Decision Support

The integrated air quality assessment and dynamic urban system modelling methodology delineated above can be linked with a multi-criteria-based Decision-Support System (DSS) in order to couple in a cost-effective way urban policy measures with the implementation of international environmental treaties and EC legislation in member states. The information deficiency that is demonstrated by the current state of the art in environmental observation technologies may be filled through the use of knowledge inference systems such as neural networks and case-based reasoning (Schocken and Ariav 1994; Keller 1995). With regard to urban air pollution assessment, neural networks can be used to operationally couple the calculation of aerosol optical thickness with forecasting of extreme pollution incidents in local and regional applications. Inference methods may support strategic environmental decision-making as well. In the case of the optimisation of air pollution monitoring networks at the urban and regional scales, inference engines may support the analysis of the optimal solution robustness. Network optimisation is better treated as a multiple objective optimisation problem. Issues such as network efficacy in signalling the alert necessary to reduce the exposure of vulnerable population to extreme pollution levels, or its efficiency in monitoring the effects of pollution on sensitive natural or cultural monuments need to be counterbalanced with cost and societal acceptance. The overall network performance, however, depends on its robustness to changing environmental pressure due to evolving land use and technology deployment, its reliability and interoperability.

The starting point of all multiple-criteria decision aid methods is the identification of the criteria for problem evaluation. In the case of integrated air quality assessment an array of different environmental factors has been identified as fundamental for the comprehensiveness of the assessment:

- effect on human health
- toxicological damage to ecosystems
- effect on climate change
- threat to cultural heritage
- disruption of the socio-economic system

At least one indicator is needed for each impact category. The valuation procedure defines the combination of results of each impact category regarding the set of questions posed in the multi-criteria assessment definition. In normative analysis the multi-criteria decision analysis problem can be interpreted as one of deriving a decision rule according to the decision-makers' preferences, and applying this rule to determine the most desirable set of decision alternatives. This approach has two fundamental assumptions. First, that the preference structure of the decision-makers can be represented by a real-valued function of attributes; and second, that the decision-makers are able to reveal their preference among more or less uncertain prospects with an adequate discriminatory power (Wang 1994). Whatever the set of "optimal" solutions revealed by the preference prioritisation scheme employed in the multi-criteria decision analysis algorithm employed, each solution has to be tested for robustness to changing environmental pressure. Case-based reasoning may reveal the sensitivity of the solutions to changing specific features of the socio-economic system dynamics underlying the pressure exercised on the environment.

In the application of ICAROS® for the computational assessment of air quality at the urban and regional scale a 6x6x1 back-propagation neural network was built to approximate the functional relationship between the aerosol optical thickness and its concentration at each pixel of the processed satellite image. The structure of the ANN is shown in figure 4.

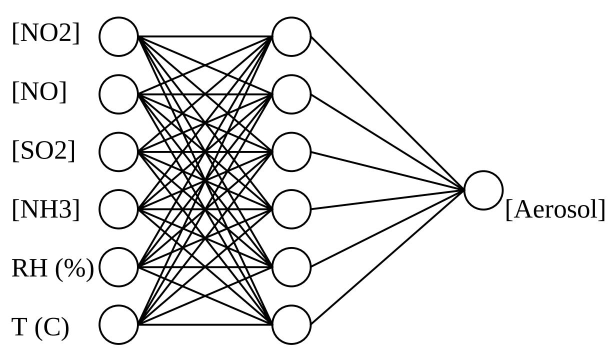

Fig. 4: 6x6x1 ANN inferring the functional relationship between the atmospheric aerosol concentration and primary pollutants concentrations per satellite image pixel.

The algorithm employed here is the following:

a) Starting from the gaseous pollutants concentrations measured by the existing monitoring network and from the temperature and relative humidity of the air, calculate the amount of aerosol formed by the chemical transformation of primary pollutants. These values are based on point measurements of gaseous pollutants at the ground-based monitoring network.

b) Through appropriate satellite image processing, calculate the amount of optical thickness of the secondary atmospheric aerosol for each pixel of the satellite image. The spatial resolution herein may reach as low as 600 m (to ensure low data uncertainty).

c) Statistically correlate the calculated optical thickness values with the amount of aerosol per image pixel calculated by the chemical transformation model. Linear multivariate regression has given very good results (R^2=99.98%).

d) Use the calculated values of aerosol optical thickness as input to calculate the spatial distribution of aerosol from the statistical model. These values are now continuous in space.

e) Apply the 6x6x1 neural network to fit the functional relationship between the total amount of aerosol calculated in step *d* and the concentrations of individual gaseous pollutants contributing to atmospheric aerosol formation, namely NO_x, SO_2, NH_3, as well as fine particulate matter. Relative humidity and temperature of air at the respective pixel are auxiliary input variables. The neural network has allowed us to calculate the inverse transfer function of the more complex chemical transformation model used in step *a*. The resulting error is reduced to less than 7% after 45 epochs of training. Its evolution is showed in figure 5.

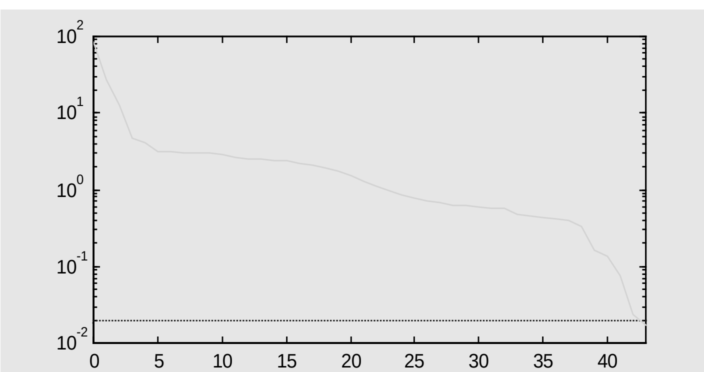

Fig. 5: Summed square error evolution as a function of epochs in the 6x6x1 ANN

Method Implementation and Results

The output of the algorithm described above is a continuous in space distribution of several major primary air pollutants such as PM10 (fine particulate matter with a diameter lower than 10 microns), NO_x, and SO_2. The algorithms calculating the aerosol optical thickness from the visible and near-infrared channels of the sensors on board SPOT and Landsat satellites have a standard error of the order of 5%. The statistical model correlating the output of these algorithms with the aerosol formed by the measured concentrations of gaseous pollutants has an error of the order of 0.02%. Finally, the error of the neural network fitting the inverse function relating the aerosol concentration with the individual gaseous pollutant concentrations is on the order of 5-7%. The overall error of the algorithm becomes approximately 12%.

The results of the neural network are fed into a multi-objective optimisation algorithm which takes into account the exposure of population to high levels of pollutants concentrations (exceeding WHO and EC threshold guidelines), protection of relatively more vulnerable nature (including agricultural crops) and cultural monuments, and cost. Using Monte-Carlo simulation to vary various factors of environmental pressure (traffic flows, position of main industrial activities, use of free space for residential purposes, variation in land use classes, etc.) scenarios of land use evolution are obtained. Using a historical time series of satellite images and---through the above algorithm---the spatial distribution of pollution in the area of interest, the optimal solution set is obtained. Through case-based reasoning the link between the robustness of the optimal solutions and the variation of critical land use classes is identified. Inference rules are derived by normalising the outcome of alternative scenarios for air pollution distribution in the fuzzy space of model parameters. Using the relative semantic distance of different alternatives of land use evolution and their relationship with the pressure on air quality, the responses of the optimisation algorithm with regard to changes in land use are hierarchised. The optimal solution is then the set of monitoring network configurations that presents robustness in the same uncertainty class.

An example of a solution suggested by this algorithm from its application in the case of the optimisation of the air quality monitoring network in the larger area of Brescia in Northern Italy is given in figure 6. Here, the dots represent the location of the currently existing monitoring stations, the small triangles the sites of major industrial activity in the zone, and the large triangles the suggested locations of the "optimised" network. The numbers given to the large triangles represent the order of importance (i.e. of information gain) of each of the monitoring stations suggested by the algorithm. The “optimised” network configuration departs somewhat from the one of the original network. The location of stations 2 and 4 approximately coincides with the site of existing stations, while station 3 could substitute for two existing stations without any information loss. Stations 1, 5 and 6, finally, are set towards the south-west of the city of Brescia in areas where the prevalent wind field transfers pollution produced in the city of Brescia and the industries found between Milan and Brescia.

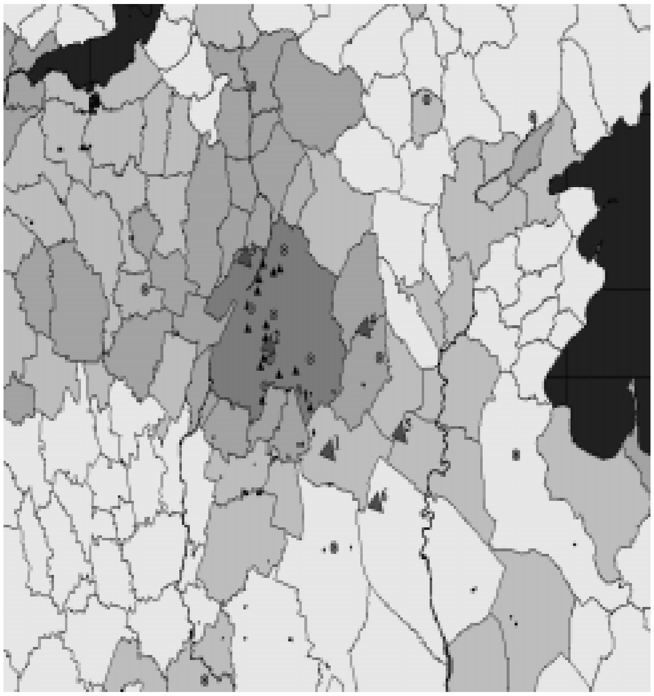

Fig. 6: Optimal spatial configuration and typology of the air quality monitoring stations network in the larger area of Brescia as calculated by the JRC inference algorithm.

The new configuration represents an enhancement of the information obtained regarding significantly poor air quality by three orders of magnitude compared to the existing monitoring network.

Conclusions

The need for advanced information systems supporting integrated assessment of urban environmental quality and the respective decision-making processes is widely recognised nowadays. Research at the Joint Research Centre of the European Commission in compliance with its mandate to support Commission policies and regional development in the EU member states through technology and know-how transfer supports the development of such information technology tools. Environmental data fusion using object-oriented technology and neural networks as inference engines are the main information technology approaches employed. The methodology outlined here fills the gaps in spatial and temporal information exhibited by current air quality assessment methods. The inference techniques employed (ANN and case-based reasoning) account for the complexity characterising environmental systems at large scale. Monte Carlo tests at the end of the information processing chain ensure the robustness of the algorithm against periodic and random change in the environmental system. The functional integration of Earth observation systems for strategic and operational environmental management introduces an objective source of synoptic information on the state of the environment.

The ICAROS® computational environment is designed to address regional and urban scale problems regarding air quality. Future research will focus on the effect of scale variation in the computational assessment of environmental condition produced by this methodology. Telematic links between ICAROS® platforms should enable us to provide early warnings of conditions leading towards extreme natural and human-made pressure on the environment at an inter-regional scale.

This advanced data fusion computational environment is coupled with dynamically updated and calibrated knowledge based on an urban model allowing the simulation of the non-linear interactions between economic activities, environmental pressure, energy and resource flows, and human mobility (and transport) in cities. Currently, this set of urban policy support tools is being validated through a series of pilot projects in different local settings, namely, the conurbation of Brescia (Italy), Cayenne and Kourou (French Guiana). Future applications of this comprehensive environmental information processing and policy- and decision-support methodology include peripheral cities of the European North (the Euro-Arctic regions) and ex-Eastern European regions. Integrated environmental knowledge assessment and management is of primary importance for the harmonisation of the environmental legislation in ex-Eastern European countries with the current and future legal framework in the European Union.

References

CEOS 1992. The Relevance of Satellite Missions to the Study of the Global Environment, British National Space Centre, London, UK.

Fedra, K. 1990. A Computer-based Approach to Environmental Impact Assessment. In A.G. Colombo and G. Premazzi (eds.) Proceedings of the Workshop on Indicators and Indices for Environmental Impact Assessment and Risk Analysis, 11-40. EUR 13060 EN, JRC Ispra, Italy.

Günther, O., Radermacher, F.J., Riekert, W.-F. 1995. Environmental Monitoring: Models, Methods, and Systems. In N.M. Avouris and B. Page (eds.) Environmental Informatics, 13-38. Kluwer Academic Publishers, the Netherlands.

Keller, H.B. 1995. Neural Nets in Environmental Applications. In N.M. Avouris and B. Page (eds.) Environmental Informatics, 127-145. Kluwer Academic Publishers, the Netherlands.

Lipfert, F.W. 1994. Air Pollution and Community Health: A Critical Review and Data Sourcebook. Van Nostrand Reinhold, New York.

Pope, A.C., Thun, M.J., Namboodir, M.M., Dockery, D.W., Evans, J.S., Spezer, F.E, Heath, C.W.Jr. 1995. Particulate Air Pollution as a Predictor of Mortality in a Prospective Study for US Adults. *Am. J. Respir. Crit. Cara Med.*, 151: 669-674.

Sarigiannis, D.A. 1997, Complex System Modelling for Decision Support. In ISIS Annual Report 1996, EUR 17321 EN, EC, Brussels-Luxembourg.
Sarigiannis, D.A., Assimakopoulos, D., Bonetti, A., Huynh, F., Lointier, M., Schäfer, K., Sifakis, N.I., Soulakelis, N., Tombrou, M. 1998. ICAROS – Integrated Computational Assessment via Remote Observation System. In Proceedings Annual Conference of UK Remote Sensing Society, 72-79. Chatham Maritime, Kent, UK.
Sifakis, N.I., Deschamps, P.Y. 1992. Mapping of Air Pollution using SPOT Satellite Data. Phot. Eng. and Rem. Sens., LVIII: 1433-1437.
Sifakis, N.I. 1995. Remote sensing of Veils of Atmospheric Pollution and Degradation of the Environment in the Region of Athens", Photo-interprétation: images aériennes et spatiales, vol. 33, 1995/4, 220-225 & 233-236.
Sifakis, N.I., Soulakellis, N. 1996. Air Pollution Observations and Mapping from Earth-observation Satellites: Potential and Limitations. Proceedings Int. Conference on Protection and Restoration of the Environment, 370-378. Chania, Greece.
Schocken, S., Ariav, G. 1994. Neural Networks for Decision Support: Problems and Opportunities, Decision Support Systems, 11: 393-414, Elsevier Science B.V.
Wang, J. 1994. Artificial Neural Networks Versus Natural Neural Networks—A Connectionist Paradigm for Preference Assessment, Decision Support Systems, 11: 415-429, Elsevier Science B.V.