We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

3,900
Open access books available

116,000
International authors and editors

120M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
1. Introduction

The present Chapter describes a methodical approach to the design of a monitoring process for environmental applications. The case study and experimental results concern the remote control of exposure levels to electromagnetic pollution. The main contribution focuses on the design of a configurable network of wireless and smart web-sensors and the development of a two-levels data fusion procedure. The aim of an environmental monitoring process is to provide qualified information on the investigated phenomenon, so to minimize possible errors or hazardous consequences for the exposed population. That objective requires not only an appropriate choice of the measurement system, but care has to be paid about the maintenance of instrumentation in order to assure a suitable metrological state. On the contrary measured data may be inconsistent and meaningless so to provide an erroneous knowledge about the monitored process, often that occurrence is cause of underestimated risks. Therefore the operating state of the measurement system has to be guaranteed with the passing of time, in order to get reliable data. So the matter requires to evaluate the measurement system performances, (Neely et al., 2000). Moreover data processing stage plays a crucial role when decisions have to be taken with reference to warning limits. In fact several laws and regulations have been issued to limit the exposure levels to environmental parameters responsible for pollution, in order to keep under control the status of our habitat and so to guarantee an appropriate quality of life. But such decisions are the result of a comparison between the fixed limits and the measured data, consequently the uncertainty contribution due to measurement process may be reason of wrong decisions if its effect is not taken into consideration. In that context, the considered application field needs methodical monitoring processes being able to characterize and manage in real time warning and risky situations so to reduce possible harms for the exposed population. In the urban centers today it is possible to characterize several electromagnetic pollution sources to different working frequencies, as in example the lines of power supply or the antennas of the radio-television service, (Ahlbom et al., 1998; IEC 50166-2, 1995). Medical studies would seem to point out a relationship between the continual exposure to high electromagnetic field levels and the onset of some typologies of cancer. The World Health Organization suggests prudence and the observance of the quality
standard levels. Therefore, environmental data and available information have to be timely processed by means of suitable procedures in order to guarantee specific requirements of accuracy and reliability, so to minimize errors and risks. Furthermore sensor data may be erroneous, inaccurate or incomplete because of malfunctions, anomalies, delays, unsuitable maintenance or limited range. In addition, typically, environmental monitoring requires a wide network of sensors displaced on a large area, such sensors have inevitably different metrological characteristics. So a huge amount of data has to be processed in short time according to information on measurement uncertainty, topographical data, operating state of the sensors and further knowledge. At this aim the authors propose a sensor data fusion procedure being able to merge such information in order to optimize the processing so to get information of greater quality. The developments of the research in the environmental monitoring field show several open problems to be treated. For example, distributed systems have to be dynamically configurable according to changing conditions of the surrounding environment, or to environmental and topographical information, (Lin & Gerla, 1997). The actual sensor networks do not provide effective solutions and the used architecture is not able to configure itself, (Bertocco et al., 2002; Hou et al., 2004; Lee & Song, 2006; Mahfuz & Ahmed, 2005; Saripalli et al., 2006; Tsujita et al., 2005). Moreover sampling plans have to be flexible according to the desired accuracy and specifications. Also population density distribution has to be considered, in fact in presence of alarm occurrences, it is possible to characterize the zones which need more attention and have higher intervention priority as the most populated ones. The present Chapter focuses attention on such matters in order to propose possible useful tools for environmental monitoring applications. In the first part of the Chapter the design of a configurable network of wireless and smart web-sensors is described. Since electromagnetic field is characterized by spatial and temporal variations, a distributed network has been designed. As a matter of fact, weather conditions, the presence of metallic objects and antennas of radio-mobile service may affect the trend of the electromagnetic field. Therefore sensors, displaced along a wide urban area, measure the electromagnetic field levels in high and low frequency ranges according to suitable sampling plans. Then a remote processing Server allows to fuse the available information so to evaluate the overcoming of the law limits and the occurrence of alert states. By General Packet Radio Service (GPRS) communication the network exchanges data and commands with a remote client, then data are stored on WEB Pages. Each measurement unit consists of an isotropic smart sensor, a GPS module and a GPRS modem. In order to optimize the configuration of the sensor network, an algorithm has been developed. In this way the network is configurable according to the needs, and the area is divided in local zones. The size of the partition depends on the accuracy and the resolution required for the monitoring, so environmental and topographical information is used to optimize the network configuration and the monitoring map. If the specifications and accuracy requirements change, it is possible to configure dynamically the sensor network and the sampling plans by a new partition of the area. The single sensor has the task to monitor a specific zone and, according to the monitoring map, sends data to an ASP Web Page. A Server acquires the measured electromagnetic field levels for each zone, with information on the state of the single sensor like its measurement uncertainty, its reliability and its operating state; data are protected by a password system. The second part of the Chapter describes a sensor data fusion procedure developed by the authors in order to qualify the data processing stage. A two-levels approach allows the progressive fusion of
the huge amount of data so to get first local and then global information about the exposure to electromagnetic field in the monitored area. In fact, each sensor allows to collect data in a specific zone, such information provides only a small ‘image’ of the whole area concerning a restricted spatial portion correlated with the near zones. The data fusion procedure merges those partial ‘images’ so to provide a view of the electromagnetic field behavior in the area. In this way the limitation of the single sensor is overcome and a more accurate knowledge on the field exposure is got. By data fusion, the initial large amount of information may be used to get a more accurate, consistent and meaningful result. Errors are so reduced, as much as possible, using the same redundancy of data. The main problem is to characterize optimal fusion rules being able to merge complementary information so to minimize the global probability of fault or error. At this purpose a Statistical Model estimates the sensor reliability curves and its operating state. In this way measured data, information on the sensor network reliability and measurement uncertainty are processed for a more efficient interpretation of the data and a more accurate representation of the monitored area than that one provided from the single sensor; the useful information is so maximized. Furthermore it is interesting to notice that typically sensors are in a continuous working state so they may go to faulty occurrences. In such circumstance, data fusion represents an effective tool to get accurate and efficient information from faulty measurements. The proposed procedure fuses data and correlated information so to get an improved accuracy. The Server runs the procedure in order to verify the compliance of the exposure levels to electromagnetic field with the limits established from law. A fuzzy algorithm allows to perform a first level of data fusion. Data of the single zone are processed by a decision-making algorithm in order to take local decisions concerning warning or alarm occurrences. The erroneous decision probabilities are estimated according to the statistical distribution of process and the measurement uncertainty, in this way the initial data amount is reduced and local information is got. A second level of data fusion allows to get a global information about the exposure to electromagnetic field in the whole area, in this way local decisions and environmental information are fused in order to get a more accurate image about the pollution status of the area. A report shows the global situation. Consequently, the zones which require more attention are characterized, and suitable corrective interventions can be planned. Quality indexes provide information on reliability and consistency of the results. The originality of the present contribution is due to the design of a configurable network of smart sensors and to a two-levels data fusion procedure compliant with the quality assurance requirements, (ISO 9001, 2000). Knowledge on measurement uncertainty and sensor reliability is merged in order to optimize the available information and obtain a better estimation about the monitored process by using the same data redundancy.

2. The sensor network design

2.1 The configurable architecture

In the present paragraph is described the design of the monitoring network used. The aim is to verify the compliance of the electromagnetic field levels with the limits established by laws, and to get global information about the field behaviour on the whole monitored area. Today networking and communication sectors have achieved remarkable developments in real-time applications, but the increasing complexity of systems and networks is cause of further problems concerning the management and maintenance of the used instrumentation. So not always the available network architecture is suitable for obtaining a
consistent and significant view of the observed phenomenon. As a result in order to
guarantee reliability and accuracy, it is required an accurate choice of instrumentation, but
also suitable procedures estimating its metrological state and measurement uncertainty. In
this view, the authors propose the project of an original configurable network of wireless
and smart web-sensors. Main features of the contribution regard the possibility to configure
dynamically the network according to the needs and the topographical information. So the
sampling plans, the partition of the area and the sensors layout are updated if requirements
and specifications change. By a methodical approach, a distributed network has been
projected so to monitor a wide urban centre (Reggio Calabria city, in the south of Italy). The
network and the smart web-sensors have been projected according to the guidelines of the
IEEE 1451 Standard, (IEEE 1451, 2001). The whole monitoring process has been optimized by
innovative procedures being able to manage the network and maintain the sensors, so to
check the operating state of the instrumentation and estimate the next calibration interval.
The choice of a distributed architecture is due to the geographical extent. The single
measurements units are displaced according to the monitoring map along the area. The size
of the area to be monitored requires a complex and wide measurement network, so in order
to reduce the computational burden of each sensing unit and to manage easily the network,
an algorithm performs a geographical partitioning of the area. In this way the region is
divided in several small local zones. The algorithm allows to make an efficient partition, in
order to guarantee a suitable size for each zone according to the topographical information
on the area and the population density distribution. The information collected in the single
zone provide a meaningful view of the electromagnetic field trend, and the correlation
among different measurements in the area is assured. Such constraint is an important
requirement for the data fusion stage. The number \( N \) of the zones depends on the sampling
specifications, and zones size is not necessarily equal. According to the desired accuracy and
resolution of monitoring, the algorithm designs a specific partition of the area and so a new
network configuration. A major/minor severity level of the sampling plan is therefore cause
of a major/minor partition. The user can modify and configure the partition by setting some
parameters, like the distribution of population density and the topographical data. The map
of the area is acquired, and a Cartesian axes system is selected. According to the dimensions
of the area, the number of sensors and the available resources, the user chooses the initial
width of the grid to be applied. A first partition is so made on the area map, and a
preliminary sampling plan is shown for each local zone. At first, the algorithm does a simple
subdivision of the area which is partitioned by zones with equal size. A high number of
zones would allow a more accurate monitoring, but it is also cause of high costs. So the
matter requires a compromise between the desired resolution, the available resources and
the tolerable costs. The best solution is therefore to characterize the zones which need a
higher attention level, in fact it is possible to single out the “sensible zones” where the
partition grid has to be thickened. In details, the cost constraints limit the maximum number
of possible zones, so an opportune criterion has to be considered. Surely zones with greater
population density require a more accurate monitoring. In fact in presence of an alarm
occurrence or of high exposure levels to electromagnetic field, the risk for the exposed
population is more high considering the possible impact on a greater number of people.
Therefore the algorithm reduces the size of partition in the zones characterized by a high
population density or by the presence of sensible targets like schools, orphanages, hospitals
or known electromagnetic pollution sources. As a result, such zones are monitored with
more care. So the user provides the function of the population density $pd(x,y)$ of the area, or specifies the site of the sensible targets or otherwise a generic zone of interest. Then automatically, the algorithm thickens the partition grid of the sensible zones, such zones are divided in further four sub-zones with equal size (see Fig. 1).

![Partition of the monitored area](image)

Fig. 1. (a) Partition tool. (b) Partition of the monitored area

According to topographical information about the area and specifications, the best configuration of the sensor network is got. The final partition allows to characterize the sampling map for each zone, while specific sampling plans are projected for the sub-zones in order to perform a more detailed monitoring with greater accuracy and resolution. Each sensor has to monitor a specific local zone with reference to a particular sampling plan. The sampling time, the refresh time of the monitoring map, the thickness of the grid, the number of samples and their locations depend on the desired confidence and severity level, on the tolerable costs, on the available resources and on the tolerable reliability for the monitoring.

Usually a homogeneous sample is obtained by placing the monitoring points uniformly along the zone, but not always topographical data are compatible with this choice, so an accurate plan is necessary to assure a suitable reliability of collected information. The configurability features of the network allow so to adapt the sampling plans and the network configuration to environmental and topographical data. In this way the single zone is characterized by a sample of electromagnetic field levels measured in specific points, the representativeness of the sample is guaranteed by an experimental design. According to the local monitoring plan, each sensor moves along the associated zone and collects environmental data in the fixed points of interest. In this way the mobile architecture allows the measurement system to monitor a wide area with a reduced number of sensors. Network configuration, monitoring maps, number and size of the zones can be updated dynamically if the sampling specifications change, so a new partition is performed and new sensors may be added. The network security is guaranteed by a password access, thus only the authorized administrator can get access and manage the whole network and its configuration.

### 2.2 The wireless and smart web-sensors

If the electromagnetic field levels overcome the warning limits fixed by law, timely corrective actions are required in order to reduce possible risks for the population health. In
these circumstances only a real time monitoring process of the area allows to characterize such alarm events. So the matter demands the remote control of the used measurement instrumentation and a prompt data processing. The designed distributed network is composed of several wireless and smart web-sensors projected according to the guidelines of the IEEE 1451 Standard, (De Capua & Morello et al., 2004b; De Capua & Morello et al., 2005c). Each measurement system is an isotropic sensor sensitive to the electromagnetic field, it is equipped with a GPRS modem for wireless communication, and a GPS module for its localization. Three transducers, displaced orthogonally along the three axes of a Cartesian system by a plexiglas support, constitute the sensing unit. The single output is a $\mu$V–mV voltage proportional to the electromagnetic field applied perpendicularly to the chip surface. A suitable circuit has been projected in order to condition and amplify the voltage signals by an embedded auto-scale setting with noise compensation (see Fig. 2).

![Image of Electromagnetic Field Sensor and Signal Conditioning Circuit](image)

Fig. 2. (a) Electromagnetic Field Sensor. (b) Signal conditioning circuit

The system is able to configure automatically itself so to select the smaller available scale, in this way the smart sensor optimizes the accuracy and the resolution of measurement, reducing the measurement uncertainty of collected data. A 16 bits A/D converter has been used to digitalize the three signals in order to make possible the communication between the several functional blocks of the sensor. A microcontroller architecture manages the exchange of data between the transducers, the GPRS modem, the GPS module and the internal memories. A LCD display shows the measured level of electromagnetic pollution. A flash memory storages metrological information about the sensor concerning its operating status, measurement uncertainty, calibration and reliability curves, next calibration interval, probability of faulty functioning, the actual sensor reliability, and the probabilities of erroneous decision relating to the decision making process. Such information is updated at the successive sensor calibration because of the inevitable decline of performances with time affecting any measurement system. It represents an important and essential knowledge for the data processing stage, in fact the data fusion procedure uses such information in order to optimize data and credibility of the results for a more accurate, complete and fault tolerant computing (for more details see Paragraph 3). Measured data are stored in a Secure Digital (SD) Memory. Then each sensor sends data (electromagnetic field levels, GPS position and
A Sensor Data Fusion Procedure for Environmental Monitoring Applications
by a Configurable Network of Smart Web-Sensors

metrological information) to its own Web Page in Active Server Pages (ASP) format. In this way the single Web Page contains information about the pollution status of the associated zone, the sensor history and its operating state. Such data are accessible by password, only authorized users can read data and information for processing and statistical purposes. The Server acquires and processes data by the sensor data fusion procedure in order to test the electromagnetic field compliance with the exposure limits, (see Fig. 3).

Fig. 3. Control panel of data acquisition and processing

Merely a free WEB Page is accessible to general public, it contains final reports and computing results on the pollution status of the monitored area. Wireless communication has been made possible by a programmable GPRS modem with M2M technology. Queries are performed by AT commands. In this way the administrator of network can get direct access to the single sensor by a PIN number. He may ask for specific measures or exchange information about the desired sampling specifications (sampling time and monitoring map).

3. The sensor data fusion approach

3.1 Sensor reliability

Any measurement result is characterized by the estimated value of the measurand and the associated measurement uncertainty. The latter depends strongly on the metrological state of the used measurement system. Measurement uncertainty is a parameter characterizing the dispersion of the quantity values being attributed to the measurand by a measurement process, (IEC-ISO GUM, 1995; ISO/TS 17450-2, 2002). Typically the uncertainty of a generic measurement system changes inevitably with time. Therefore in order to guarantee reliable measures, the operating state of the used equipment and measurement systems has to be checked periodically by calibration, (ANSI/NCSL Z540-1, 1994; ISO/IEC 17025, 1999; UNI EN ISO 10012-1, 1993; UNI EN ISO 10012-2, 2001). In this view in the present paragraph the authors propose an original procedure being able to estimate the reliability and the performances of the sensors of the network, such an approach allows to...
qualify and maximize the measured data. Clearly data with higher quality and accuracy are indispensable for a reliable processing. For that reason the developed procedure provides information about the credibility of the measured data and the actual reliability level of each sensor. Such knowledge is then used in order to characterize the integrity of the whole network and of the computing results. Thus if the reliability of the network or of the single sensor decreases below a fixed tolerable level, the sensor/sensors have to be calibrated. A proper maintenance of the used measurement instrumentation is an important prerequisite in order to guarantee the credibility of sensor data fusion. In details, by a Statistical Model the procedure estimates optimal calibration intervals for the single sensor. The purpose is to reduce the probability of an out-of-tolerance state during the maintenance time. So according to a desired reliability target, the next calibration time is estimated, assuring a suitable operating state for the sensor. About that matter, numerous models have been proposed by several authors, (Castrup & Johnson, 1994; Nunzi et al., 2004; Wyatt & Castrup, 1991). The present original approach bases oneself on evaluating the impact of the uncertainty associated with the calibration process on the estimation of the maintenance interval, (De Capua & Morello et al., 2005b). An erroneous analysis of the calibration results could be cause of unsuitable calibration intervals. Consequently it may be the reason for possible out-of-tolerance states during the maintenance time, so unreliable data would be processed. The basic problem concerns the decision, made during a calibration process, about the real operating state of a sensor. The used approach allows to characterize the decision reliability by the erroneous decision probabilities about the real state of the tested sensor. The estimated decision risk is subsequently used in order to optimize the reliability function of the sensor concerning its in-tolerance state. A sensor is in a tolerance operating state if its metrological characteristics are compliant with fixed tolerance limits in order to guarantee a tolerable measurement uncertainty. According to information on the last calibration process of the sensor, the Model estimates the erroneous decision probability $\Pi_\beta$ concerning the occurrence to have taken a wrong decision about its in-tolerance state. In other words $\Pi_\beta$ represents the probability to have supposed erroneously in the last calibration that the sensor was in an in-tolerance state, and so in accordance with the tolerance limits, whereas in truth it was in an out-of-tolerance state, and so not compliant with the tolerance limits. If $\delta$ is the tolerance limit of the tested parameter; $x_m$ is the random variable representing the tested parameter; $x$ is the random variable of the calibration process, its distribution around the expected parameter value represents the randomness of the measurement result, in other words its standard deviation is the standard uncertainty due to calibration process. We can assert that the tested parameter or the considered metrological characteristic is compliant with the tolerance limit, and consequently the sensor is in a tolerance state, if $x_m \in [0, \delta]$. So the erroneous decision probability associated with the sensor calibration is estimated by the equation:

$$\Pi_\beta = \int_{x_m - \delta}^{x_m} F(x) \cdot f(x) \, dx$$

(1)

Where $F(x)$ is the Cumulative Distribution Function (CDF) associated with the variable $x$, while $f(x)$ is the Probability Density Function (PDF) of the variable $x_m$. The next step requires the definition of the sensor reliability curve $R(t)$, it can be estimated by the following expression based on the Weibull Model:
That function allows to evaluate the in-tolerance probability of the sensor in a precise moment $t$ according to a desired quality target. In order to optimize the estimation of the reliability function, the parameter $R_o$ has been set equal to $1 - \Pi_\beta$ in this way the erroneous decision probability weighs the information about the reliability level concerning the operating state of the sensor. In fact high probability values of erroneous decision are cause of low reliability levels for the sensor, because of a high probability to have supposed erroneously in the last calibration that the sensor was in an in-tolerance state. The parameters $\lambda$ and $\beta$ in the equation (2) are evaluated by a maximum likelihood estimation, in order to characterize the best fit function according to past information on previous calibrations of the considered sensor. In other words, the two parameters are settled in order to guarantee the best fit with the data history of the previous calibrations. In fact the reliability function can be sampled by historical time series of previous calibration data, by means of the ratio between the number of in-tolerance operating states and the total number of calibrations in a definite time $t$. In the Fig. 4 (a) the trend of reliability function is shown for different values of the two parameters. The Fig. 4 (b) shows the reliability curve estimated for a sensor of the network starting from data of six previous calibrations.

$$R(t) = R_o * e^{-(\lambda t)^{\beta}}$$

(2)

Fig. 4. (a) Weibull Model. (b) Sensor reliability curve

The reliability function allows to get information about the temporal evolution of the sensor performances. In this way it is possible to know how its measurement reliability decreases over time. So a low reliability level is sign of a sensor which may be faulty because of out-of-tolerance occurrences; differently a high reliability level is sign of a sensor with an appropriate in-tolerance state, and therefore the measured data are reliable. The reliability curve of each sensor is stored in its own internal flash memory and subsequently in the relative Web Page. In addition the procedure allows to estimate the next calibration interval $t^*$ by fixing a desired maximum tolerable reliability $R^*$ for the single sensor:

$$t^* = \frac{1}{\lambda} \sqrt{-\ln \left( \frac{R^*}{R_o} \right)}$$

(3)

In this way during the maintenance interval $t^*$, the sensor reliability $R(t)$ will not decrease below the target level $R^*$, so to guarantee a suitable in-tolerance state. The described procedure has therefore two benefits: first it provides a suitable maintenance interval for the
single sensor in order to assure an appropriate operating state and therefore reliable data for the following fusion stage; finally the reliability function permits to evaluate the actual reliability of each sensor $R_i(t')$ in a precise moment $t'$, such information is used during the processing stage in order to assure a fault tolerant fusion of data. In this way each zone is characterized by a value $R_i(t')$ which is representative of the data reliability collected in the zone and consequently it gives a measure of the credibility of the fusion results.

### 3.2 The two-levels data fusion procedure

After the monitoring process each zone is characterized by a set of electromagnetic field levels collected according to the sampling plans. But information on the sensor reliability and on measurement uncertainty allows in addition to improve the available knowledge minimizing errors during the computation. Now data from different sensors have to be merged in order to optimize the information about possible alarm occurrences regarding pollution status of the monitored area. Sensor data fusion is the best solution to manage data and correlated knowledge from different sources which may be faulty. Moreover the sensors are distributed on a wide area to monitor a particularly complex phenomenon with temporal and spatial evolutions; besides a wireless sensor network is more complicate than a wired network because of communication errors, delay and topographical constraints. So each sensor has only a restricted view of the monitored phenomenon limited to the own local zone. Data fusion solves the limitation of the single source, and it is indispensable when a wide distributed network has to be used in order to get a more accurate view of the process behaviour in the whole area, (Linn et al., 1991; Wu et al., 2002). It allows a more efficient interpretation of data reducing the initial amount with less uncertainty and error than that obtained from a single source. The complexity of the matter needs not only to analyze and characterize these features, but also a careful analysis and integration of correlated information are necessary. The proposed approach consists in a two-levels data fusion, (De Capua & Morello et al., 2005a; De Capua & Morello et al., 2007). The procedure is able to improve the accuracy of data and to detect possible faulty states of the sensors. First for each zone, a fuzzy algorithm provides local decisions regarding alert occurrences due to the overcoming of the law limit, (De Capua & Morello et al., 2004a). In this way the initial amount of data is reduced by a preliminary fusion, and information on the measurement uncertainty is used to qualify the decision making process reducing errors and risks. Subsequently a final fusion of data and of correlated knowledge provides global information about the behaviour of electromagnetic field in the area so to characterize the general pollution status. In detail, after the partition of the area, each sensor must acquire a sample of $N_i$ measures along the own local zone. Starting from the collected data and information on sensors performances, the Server processes electromagnetic field levels in order to single out possible hazardous events in the monitored area. For each measured value, the conformity with the exposure limit has to be evaluated. In fact in presence of warning situations, corrective interventions must be undertaken so to safeguard the population health. The problem requires suitable decision making rules in order to take reliable decisions. The matter concerns the choice of the more plausible decision about the conformity of measured data with the exposure limit. The uncertainty, which affects the measures, could be cause of possible wrong decisions, so a measured value may seem erroneously over the fixed limit, (Carbone et al., 2002; Castrup, 1995; Catelani et al., 1998; Zingales, 1996). Consequently the comparison between measured value and exposure limit
cannot be performed by a simple mathematical comparison. Qualified procedures are necessary to take account of costs and risks associated to erroneous decisions. It is possible to use the same data redundancy in order to improve the results accuracy, and correlated information as environmental and topographical knowledge may qualify the computation. The fuzzy algorithm is able to perform a decision making process on the acquired data. Each zone is characterized by a sample of \( N_i \) measures, which are representative of the electromagnetic field behavior according to the desired severity level of the monitoring. The procedure is performed in the single zone, to determine if an acquired measurement value is conformance or non-conformance with the specification limit. So starting from the statistical distribution correlated with the monitored process and the metrological characteristics of the measurement system, the algorithm takes a decision about the overcoming occurrence of the exposure limit for the single measured datum in the zone. The possible alternatives of decision are two: \( A_1 \) indicates the conformance occurrence, that is the measured level is below the limit; while \( A_2 \) indicates the non-conformance occurrence, in other words the level is beyond the specification limit. The information stored in the flash memory of the sensor, regarding its metrological operating status, is used in order to estimate the erroneous decision probabilities associated with the measurement process. So data are fused with information regarding the statistical distribution of process and the measurement uncertainty. Starting from the \( N_i \) measured values, the Statistical Model described in the previous Paragraph allows now to calculate the probability \( P_\alpha \). It represents the probability to have supposed erroneously the measured value being above the exposure limit, because of the associated measurement uncertainty, when the measurand is really in conformity with the limit:

\[
P_\alpha = \int_\delta^{\infty} \left[ 1 - F_x \left( \delta - x \right) \right] \cdot f_x \left( x \right) dx
\]  

(4)

Similarly to the equation (1) \( \delta \) is the exposure limit; \( x_m \) is the random variable associated with the measured quantity; and \( x \) is the random variable associated with the measurement process, so its standard deviation represents the standard measurement uncertainty. With the same symbolism it is possible to define the erroneous decision probability \( P_\beta \). Differently it represents the probability to have supposed erroneously the measured value being below the exposure limit, because of the associated measurement uncertainty, when the measurand is really in non-conformity with the limit:

\[
P_\beta = \int_{-\infty}^{\delta} F_x \left( \delta - x \right) \cdot f_x \left( x \right) dx
\]  

(5)

The reader has to notice that the equations (1) and (5) are equal, but the meaning of the expression is different. Because in the first equation the probability concerns the calibration process and so the decision regards the tolerance state of sensor. The second equation estimates the probability concerning the measurement process of the electromagnetic field levels and so the respective decision refers to the overcoming of the exposure limit. In the same way, the variables \( \delta, x_m \) and \( x \) have a different connotation in the two equations being associated with the calibration and the measurement process respectively. The Fig. 5 shows, in example, the trend of the function \( P_\alpha \left( u, \delta \right) \), where \( \sigma_0 \) represents the standard deviation of the measured data. The function characterizes the relation between the erroneous decision probability and the measurement uncertainty \( u \) and the exposure limit \( \delta \).
Increasing values of $u$ are cause of more and more high values of the erroneous decision probability; while a restrictive limit $\delta$ lead to a higher error probability. In order to explain the developed fuzzy decision making algorithm, we must start from the guidelines of the Standard EN ISO 14253-1, (ISO EN 14253-1, 1998). According to the Standard decisional rule it is possible to single out three zones when the measured value is put in comparison with the exposure limit: the conformance and non-conformance zones and the uncertainty range. So, if $U$ is the expanded uncertainty, the measured value belongs to the conformance zone if it falls in the left interval $[0, \delta - U]$; the right interval $[\delta + U, +\infty]$ represents the non-conformance zone; while the middle interval $[\delta - U, \delta + U]$ is the uncertainty range. The Fig. 6 shows the fuzzy decisional criterion developed according to the guidelines of the ISO Standard rule.

Fig. 5. Trend of $P_{\alpha}(u, \delta)$

Fig. 6. Fuzzy decisional rule
Now in order to understand the criterion, it is necessary to analyze the previous ISO Standard. The rule of the ISO Standard is a simple decisional criterion: if the measured value belongs to the conformance or non-conformance zone then the measurand has to be considered respectively compliant or not compliant with the limit. Several gaps make such decisional rule not optimal. In the first instance, it is possible to point out that no decisional criterion is suggested for the uncertainty range, so if the measured value belongs to the interval \([δ-U, δ+U]\), the ISO criterion does not provide any reliable decision. In the second instance, the decisional rule is lacking in information about the reliability or consistency of decision, so no knowledge is given about the decision goodness. It is possible to interpret such an approach as a binary one, so if the measured value belongs to the conformance zone, we could assign a credibility level equal to 1 to the conformity alternative \(A_1\), while the non-conformity alternative \(A_2\) has a null credibility level; in fact according to the ISO Standard in this case the ‘conformity’ decision has to be taken. By a first analysis that choice is not always the most reliable, because the comparison does not take into account the measurement uncertainty influence. So also the ‘non-conformity’ alternative could be plausible with a credibility level being not null. In this view, a fuzzy approach may improve the decisional rule when data affected by uncertainty are available, (Sousa & Kaymak, 2001; Triantaphyllou & Chi-Tun, 1996). The proposed fuzzy algorithm improves the ISO Standard rule by fusing information about the measurement uncertainty and the erroneous decision probabilities. In fact the previous probability \(P_β\) allows to estimate the probability to have supposed erroneously the measured value in conformity with the exposure limit; since this probability is not necessarily null, it is possible to assert by a fuzzy approach that also the non-conformity alternative may be plausible. As a result if the measured value belongs to the conformance zone we can assign a credibility level equal to 1-\(P_β\) to the alternative \(A_1\), while the credibility level of the alternative \(A_2\) is not exactly equal to zero but equal to \(P_β\), see Fig. 6. In this way the decisional criterion weighs up the occurrence of a possible wrong decision due to the measurement uncertainty. With the same reasoning, we can observe that the probability \(P_α\) to have supposed erroneously the measured value in non-conformity with the limit, is not necessarily null. So if the measured value belongs to the non-conformance zone, it is reasonable to assign a credibility level equal to 1-\(P_α\) to the alternative \(A_2\), and \(P_α\) to the alternative \(A_1\). The Fig. 6 shows the membership functions of the decisional criterion. For each decision alternative along the x-axis is represented the measured value, whereas the y-axis refers to the credibility assignment. In this way for each alternative, the function assigns a credibility level according to the measured value. The probabilities \(P_α\) and \(P_β\) weigh the decisional criterion in order to guarantee the decision consistency. Therefore the credibility level depends on the erroneous decision probabilities and so on the measurement uncertainty. A second decisional criterion takes in account risks and costs associated with the available alternatives, so to improve the discernment properties of the algorithm. Practical visual tools provide information about the confidence level of decision, generating a graphical representation of the global reliability of the two possible alternatives \(A_1\) and \(A_2\). Starting from the credibility levels, two fuzzy triangular sets show the global satisfaction of the decisional rules for the two alternatives (see Fig. 7). Each alternative is then described by a triangular set, which shows the alternative reliability. The single triangle is got by fixing the x-axis of the vertex equal to the previous credibility levels in Fig. 6 and its y-axis equal to 1. In this way the credibility levels of the membership functions are characterized by the most high level of reliability. By a fuzzy approach also the
Fig. 7. Global reliability of the alternatives near credibility levels are possible with a lower reliability degree, and it is visible by the decreasing trend of the triangle sides. A rapid visual analysis of the Fig. 7 allows to single out the best reliable alternative, in other words the alternative characterized by highest credibility levels. In the example reported in Fig. 7, the best alternative is $A_2$, because its triangle is on the right side with highest credibility levels. So the measured value has to be assumed non-compliant with the exposure limit. A quality index provides information on the decision reliability or decision consistency $DC$:

$$DC = \left( \frac{(p_{m})_{\text{max}} - (p_{m})_{\text{min}}}{2} \right) \times (1 - \left( (p_{n})_{\text{max}} - (p_{m})_{\text{max}} \right)) \times 100$$ (6)

where:

i. $(p_{m})_{\text{max}}$ is the vertex abscissa of the triangle associated with the chosen alternative;

ii. $(p_{m})_{\text{min}}$ is the vertex abscissa of the triangle associated with the rejected alternative;

iii. $(p_{n})_{\text{max}}$ is the abscissa of the right base value of the triangle associated with the chosen alternative.

The proposed fuzzy decision making algorithm allows to perform a first level of data fusion. So starting from the data collected in the single zone, each measured value is put in comparison with the exposure limit. The most reliable decision is taken about the conformity or non-conformity with the limit. The available information stored in the flash memory of each sensor is so used in order to minimize risks of possible decisional errors. The further knowledge on the reliability of sensor and the next calibration interval represents an useful information in order to evaluate the actual performances of the sensor and its operating status. The same computing is executed in each zone of the investigated area. This allows the data reduction in order to manage the computational load. In this way the i-th local zone is merely characterized by $N_i$ decision results. In order to evaluate the pollution status of the single zone, the frequency of non-conformities $f_i$ is estimated:

$$f_i = \frac{N_{io}}{N_i}$$ (7)

where $N_{io}$ is the number of values which overcome the exposure limit, and $N_i$ is the total size of the sample. In Fig. 8 the warning report of the local zones is shown. The report depicts the
pollution status in the area according to the made partition. The graduate color scale shows the different levels of pollution severity. The green color characterizes the zones with electromagnetic field levels in conformity with the exposure limit. The yellow color underlines the zones where warning situations are happened. The red color indicates alarm/alert events which may be cause of possible risks for the exposed population. The analysis of the pollution status in the area is an useful tool in order to single out the zones in which warning occurrences have happened. In this way it is possible to plan local corrective interventions, so to safeguard the involved population. Such information is not yet functional to characterize the risk state of the whole area, so an additional level of fusion is required in order to merge further information on topographical data and sensor reliability.

Fig. 8. Pollution status of the zones

The second level of data fusion has the task to fuse the obtained knowledge in order to get a view of the global pollution status of the area, so to plan general actions on the area characterizing the zones with higher priority of intervention. So information concerning the results of the decision making process in the single zone, the sensor reliability and topographical data are fused. The final result has the objective to improve the available knowledge turning the attention no more to the single zone but to the whole area. In this way the local results of each zone are weighed by the information concerning the reliability of the respective sensor. It is reasonable to observe that in presence of an alarm status, the zones, which have a higher priority of intervention, are those characterized by a higher reliability degree and a higher population density. So the zones with a higher priority of intervention may be singled out by means of the relative population density $\sigma_i$:

$$\sigma_i = \sigma_{t \sigma} / \sigma_t$$

(8)

where $\sigma_i$ is the population density of the i-th zone, whereas $\sigma_t$ is the total population density of the whole area. So high levels of $\sigma_i$ point out zones with high intervention priority, because, in presence of alarm situations, the impact on the population would be more disastrous. By the equation (2), it is possible to estimate the reliability $R_i(t')$ of each sensor in a specific instant $t'$. This parameter represents a practical information about the consistency of the data fusion in the first level processing. So it can be used in order to get an indication...
about the possibility of false alarm in the i-th zone due to a faulty functioning of the sensor. Now we can define a Loss Function $L_i$ for the single zone by the following expression:

$$L_i = 100^*q_i^*R_i^2(\sigma_{ri})^2$$  \hspace{1cm} (9)

This parameter is a percentage measure of the cost of non-quality, it shows how the non-quality of the monitored process is high, and so how the process is in an out-control situation. The Fig. 9 shows the trend of the Loss Function over the relative population density $\sigma_{ri}$ and the reliability of sensor $R_i$ weighed by the frequency of non-conformities $f_i$. The function represents an effective parameter evaluating the risk state of the i-th zone. A graduated color scale shows the risk level for the zone, where the red color characterizes the zones with higher risk.

Fig. 9. Risk state of the i-th zone

The aim of that parameter is to characterize the risk of the zones in the whole area. So, assuming $f_i$ to be constant, zones with high population density are characterized by high loss of quality. While the zones, which have a high probability of false alarm occurrences, have less weight in the definition of intervention plans. Finally a global index $F$ provides information about the average frequency of the alarm occurrences in the $N$ zones of the area:

$$F = \frac{\sum_{i=1}^{N} f_i * R_i}{N}$$  \hspace{1cm} (10)

By a weighed sum of the non-conformities frequencies $f_i$, which are weighed by the respective sensor reliability $R_i$, the index provides a measure of the global alarm occurrence in the area. The final reports and computing results are stored in free-accessible Web Pages, so ordinary users can get knowledge about the environmental pollution of the area.

In conclusion, the first level of data fusion is functional to get restricted information about the local zone and the conformity of the electromagnetic field with the exposure limit fixed by laws. So local plans of intervention can be designed. A report shows the pollution status of the single zone. The knowledge of the measurement uncertainty allows to qualify the processing stage, in order to take reliable decisions during the comparison between
measured values and exposure limit. The used fuzzy algorithm permits to minimize the possible occurrence of errors. The second level of fusion has the task to fuse the previous information with topographical data and performances of sensor. The zones are put in comparison so that an outline of the global electromagnetic field behavior is got. The purpose is to design a global plan of intervention. In this way the zones with higher priority of intervention are singled out, and information about the risk for population health is obtained. Indexes and parameters provide knowledge about the reliability of the computing results and the risk state of each zone. Correlated information is used in order to guarantee a fault tolerant data fusion, so that zones with lower reliability levels, or in other words zones with a greater possibility of false alarm occurrences have less weight in the computing.

4. Experimental results

The sensor data fusion procedure and the network of wireless and smart web-sensors have been tested in order to verify the consistency of the developed models. Experimental results have been obtained by a monitoring process carried out in a wide urban centre (Reggio Calabria city, in the south of Italy). Starting from the map of the area, the partition algorithm has subdivided the area in several local zones. According to the available resources, the topographical data and the desired accuracy for the monitoring, 30 local zones have been singled out. The initial partition grid has been thickened in two specific zones, because of the presence of sensible targets (schools and hospitals). So 8 new sensible zones have been got with a smaller size, they have required a more accurate monitoring, (see Fig. 10).

Fig. 10. Partition of the monitored area

In these zones the monitoring maps have been designed with more attention. The sampling plans for the remaining zones have been realized by an experimental design in order to optimize the choice of the specifications. In details, information on the severity level of monitoring, the desired accuracy and the population density distribution have been used in order to get for each zone the single monitoring map, the sampling frequency and the sample size. In this way the single zone has been characterized by a representative sample of electromagnetic field levels. The monitoring of the several zones has been executed in different temporal windows. In fact because of the limited number of the available
prototypes of the smart sensors, the same sensors have been used for monitoring different zones, though the sensor network design would expect the presence of a specific sensor for each zone. However that may be, that choice does not compromise the consistency of the monitoring. The Fig. 11 (a) shows the moment of data acquisition. The Server acquires for each zone the set of the measured values and information about the operating state of the sensor. In Fig. 11 (b) the data trend of a specific zone is shown.

Fig. 11. (a) Data acquisition. (b) Electromagnetic field trend in the i-th zone

By a first analysis of the figure, it is possible to observe that some values (highlighted by a red oval line) would seem to overcome the exposure limit depicted by a dash red line. After the data acquisition, the Server runs the sensor data fusion procedure in order to compute data and information, (see Fig. 12).

Fig. 12. (a) Processing stage. (b) Fuzzy triangular sets of the alternatives

In the Fig. 12, it is possible to see the several processing steps of the fuzzy decision making algorithm, with an example of fuzzy triangular sets for a specific measured value. In detail, considering the zone n°4 as an example, the measured electromagnetic field level $y=5.96$ V/m would seem to not overcome the exposure limit $\delta=6$ V/m, as it is shown in Fig. 11 (b). The estimated erroneous decision probabilities are $P_\alpha=0.0045$ and $P_\beta=0.0061$. By observing the triangular sets, the decision making procedure has computed that the measured value overcomes the exposure limit with a decision consistency $DC=1.5\%$. In the same way, the procedure has been performed in each zone for all measured values. So each zone has been
characterized by $N_i$ decision results. By calculating the frequency of non-conformities $f_i$ for the single zone, the following report concerning the pollution status of the area has been got, (see Fig. 13).

![Fig. 13. Pollution status of the zones](image)

By a simple analysis of the report it has been possible to detect the presence of three zones which require more attention. In fact the figure shows the presence of three zones with a different colouring; such zones are also highlighted in the Fig. 10 by a red contour. In the Table 1, the experimental results relating to the second level of data fusion are shown, in detail, as an example, the data of the previous three sensible zones are reported.

<table>
<thead>
<tr>
<th>Zone #</th>
<th>Frequency of non-conformities $f_i \in [0,1]$</th>
<th>Sensor Reliability $R_i \in [0,1]$</th>
<th>Risk State $L_i %$</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>7</td>
<td>0.2</td>
<td>0.9965</td>
<td>0.00089%</td>
</tr>
<tr>
<td>8</td>
<td>0.1</td>
<td>0.9961</td>
<td>0.00022%</td>
</tr>
<tr>
<td>9</td>
<td>0.34</td>
<td>0.996</td>
<td>0.0026%</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Table 1. Experimental results

These zones are characterized by frequencies of non-conformities $f_i$ above average, because of warning occurrences. The matter has required a more detailed monitoring in order to perform a more careful analysis. The investigation has made possible the identification of the sources of pollution in the zones. In fact the high exposure levels were caused mainly by antennas of radio-mobile service located in the vicinity. Only for such zones the monitoring campaign has therefore highlighted the necessity of local corrective actions in order to reduce the emission power of the sources, so to conform the electromagnetic field levels to the law’s exposure limit. The average frequency of alarm occurrences $F=0.03$ is rather low similarly to the risk state levels of the zones, so no global plan of intervention in the area is necessary but only local ones for the three examined zones. Finally, the estimated sensor reliability levels point out that the sensors are in an optimal operating state, therefore the probability of a faulty functioning and of false alarm is very low, assuring a suitable reliability for the results of the data fusion process.
5. Conclusion

In this Chapter an original approach to sensor data fusion for environmental monitoring applications has been proposed. Care has been paid for an innovative design of the distributed network of wireless and smart web-sensors with remote data processing. The architecture of the network is dynamically configurable by an algorithm according to the requirements and the accuracy desired for the monitoring. The sensors are able to measure the environmental electromagnetic field. A GPRS modem and a GPS module allow the single measuring unit to communicate remotely and to acquire information about its geographic location. In order to distribute the burden of measurement among several sensors, the monitored area (an urban centre) is divided in several local zones, where a set of sensors acquire a fixed number of data according to designed monitoring maps. Sampling time and location of the measurement points depend on topographical and environmental knowledge. The collected data are sent to ASP Web Pages accessible in reading only from identified users. The alone administrator has authorization to manage the whole network and to exchange information and commands with the sensors. Data and correlated information are remotely processed by an innovative data fusion procedure. The data fusion approach represents a suitable solution in order to manage a wide network and to process data from different sources. The developed procedure allows to minimize errors and faulty computing, by using information about the measurement uncertainty and the performances of the sensors. The data amount is processed and the same redundancy is used in order to increase the reliability and the accuracy of the results. A first level of data fusion provides information on alarm occurrences. So measured values are put in comparison with the exposure limit. The fuzzy decision making algorithm permits to qualify the comparison process minimizing possible occurrences of wrong decisions. A quality index values the consistency of the final decision alternative chosen. The estimation of the erroneous decision probabilities and the measurement uncertainty improve the computing results. A warning report shows the pollution status of the zones. Successively the results of the decisional process, information on sensor reliability and population density distribution are fused so to obtain a global view on the risk state in the whole area. The procedure is fault tolerant and permits the maximization of the useful information. In this way a more reliable result is got than that obtained from the single sensor, so greater efficacy and efficiency are achieved. The projected measurement process is a complete solution in support of monitoring, controlling and alarm reporting applications for environmental purposes.

6. References


Data fusion is a research area that is growing rapidly due to the fact that it provides means for combining pieces of information coming from different sources/sensors, resulting in ameliorated overall system performance (improved decision making, increased detection capabilities, diminished number of false alarms, improved reliability in various situations at hand) with respect to separate sensors/sources. Different data fusion methods have been developed in order to optimize the overall system output in a variety of applications for which data fusion might be useful: security (humanitarian, military), medical diagnosis, environmental monitoring, remote sensing, robotics, etc.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:
