Chapter 12
Framework for Analyzing Environmental Indicators Measurements Acquired by Wireless Distributed Sensory Network – Air Pollution Showcase

Barak Fishbain

Abstract Distributed continuous in situ monitoring of the environment is an essential component in assessing environmental indicators as their changes take place on different spatial and temporal scales. Recent sensory and communication technological developments have led to the emergence of Wireless Distributed Environmental Sensing Networks (WDESNs) that consist mainly of Micro-Sensory-Units (MSUs). The set of skills and expertise that are called for developing these WDESNs are from different fields and research disciplines, where each discipline has its own jargon. To allow cross-disciplinary discussion, a comprehensive, yet simple framework for discussing data acquired by WDESNs is presented. The terminology presented here allows for describing complex multi-modal environmental sensory networks and the integration of the observed environmental indicators' into a holistic understanding of the environment. The usage of the presented framework is demonstrated in describing a recently reported data acquired by air-pollution WDESN.

Keywords Enviromatics • Environmental informatics • Distributed sensing • Sensory networks • Environmental sensing • Environmental data analysis • Environmental data fusion

12.1 Introduction

Continuous in situ monitoring of the environment is an essential component in assessing environmental indicators. Air, water and land characteristics and quality, and their change over time is fundamental to most environmental applications. Quantifying changes in environmental indicators over time is a most challenging

B. Fishbain (✉)
Technion Enviromatics Lab, Department of Environmental, Water and Agricultural Engineering, Faculty of Civil & Environmental Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel
e-mail: fishbain@technion.ac.il

R.H. Armon, O. Hänninen (eds.), *Environmental Indicators*,
DOI 10.1007/978-94-017-9499-2_12

task as their changes take place on different spatial and temporal scales. Local system characteristics and behavior affect overall system, while system level parameters govern the local scale processes.

Recent sensory and communication technological developments have led to the emergence of Wireless Distributed Environmental Sensing Networks (WDESNs) that consist of Micro-Sensory-Units (MSUs), mainly in air and atmospheric assessments (Mead et al. 2013; Chen 2008; Dutta et al. 2009; Williams et al. 2013; Etzion et al. 2013; Broday et al. 2013) and aquatic systems (Kroll and King 2007; Hall et al. 2007; Pickard et al. 2011; Research and Development, National Homeland Security Research Center 2005; Kramer 2009; Storey et al. 2011). These have greatly enhanced monitoring of complex infrastructure and the natural environment, allowing the study of fundamental processes in the environment, as well as hazard warnings, such as flood (Hart and Martinez 2006; Martinez and Hart 2004; Zhou and Roure 2007) and pollution alerts (Hochbaum and Fishbain 2011), in a new way. In conjunction new hardware designs and algorithmic approaches to support these networks have emerged (Ramanathan et al. 2006; Budde et al. 2012; Denzer 2005; Kranz et al. 2010; Bakken et al. 2001). The focus of most of these studies is an efficient operation of the network and failure tolerance and therefore they do not provide new insights on the monitored environments. Further, many problems studied in the field, have their roots in the desire to emulate cognitive capabilities of a human (e.g., Ramanathan et al. 2006; Chen 2008; Dutta et al. 2009). With the rapid advances in hardware technology, this anthropocentric and somewhat limited paradigm may no longer be the only source for inspiration. The variety and availability of sensors have made the accessible data much greater in quantity than the data that can be gathered and interpreted by a human being. However, considering the sheer amount of data and its diversity, building such multi-sensors systems becomes a great challenge.

Field deployments of low-cost air quality MSU-based WDESN have been recently reported (Mead et al. 2013; Williams et al. 2013). Williams et al. (2013) showed the use of metal-oxide micro-sensors for measuring ambient O_3 levels and Mead et al. (2013) used electrochemical probes for measuring CO, NO and NO_2. These studies do not present a network configuration for air pollution measurements. Therefore, their ability to capture the spatial pollutant variability has not been shown.

The ability and potential of WDESN to capture spatial and temporal variation in field campaigns has been recently demonstrated (Broday et al. 2013; Shashank et al. 2013). These studies regard only one indicator (pollutant) in their analysis. The integration of different modalities possesses the ability to exploit the sensors' cross-selectivity and to compensate each modality's inherent deficiencies by utilizing the strengths of the other modalities. This is still a relatively new discipline, and little effort has so far been devoted to the implementation of fully operational devices for environmental applications as seen in other fields such as computer vision and robotics. Such spatial and temporal analysis of several indicators on a well-defined environment, through WDESN, has been recently reported by Moltchanov et al. (2015).

This new emerging field, enviromatics (environmental informatics), calls for a new set of algorithms and methodologies originating from several distinct disciplines – chemistry, environmental sciences, optimization, multi-dimensional signal processing, data fusion and data communication. To facilitate any discussion between scholars from these different research fields, there is a need to build a comprehensive, yet simple, framework. Here such framework is presented. Section 12.2 presents the framework, its components and its lexis foundations. Section 12.3 demonstrates the usage of the presented lexis in the context of air-pollution distributed sensory network by describing the results reported by Moltchanov et al. (2015). Section 12.4 concludes this chapter.

12.2 Framework

12.2.1 Lexis

This section introduces the lexis that facilitates the Environmental Multimodal Distributed Sensory Network framework. *An environmental episode* is a disruption in the environmental conditions. To this end, an episode can be a major environmental happening, such as high or irregular gamma radiation levels (Hochbaum and Fishbain 2011), volcanic eruptions (Werner-Allen et al. 2006) or mass fish death (CNN Wire Staff 2011). It can also be a typical happening such as traffic rush hours. Such environmental episode consists of recorded *events*. Events are a set of *incidents*, which are recorded by a single or a set of sensors, and are found to be related to an environmental episode. Events in those cases are, for example, the deviations from the regular background radiation; the seismic-acoustic recordings and the oxygen levels in the aquatic system. Each incident represents one or more measured environmental indicators (Fig. 12.1). In the context of air-pollution, this can be ozone (O_3), nitrogen dioxide (NO_2) or particular matter (PM). This formulation ensures, by definition, the following:

Corollary 1

- Each environmental episode has a well-defined spatial and temporal window.
- An environmental episode manifests itself in the sensors' acquired data.

Events are extracted from streams of *incidents* and only then are cross-referenced. The key justification behind this dichotomy is that different set of algorithms are called for extracting events from a stream of incidents, and for relating these events. The separation of these tasks also allows for distributed processing, as the event extraction can be held on the sensing platform itself. The

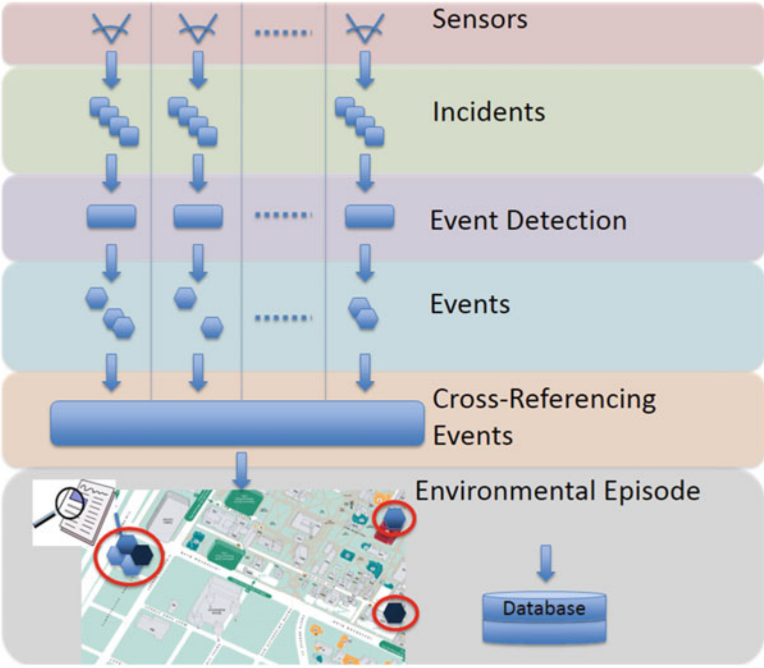

Fig. 12.1 Multimodal distributed sensory networks dataflow and terminology

set of extracted events is processed by a set of analytics tools for individual data streams as well as multiple data streams of various types. Both kinds of analytics are needed for unveiling the information embedded within each data stream/source as well as the information that is spread across various data sources. This formulation facilitates an efficient and reliable processing of exceptionally large environmental data sets (of size in the tens of thousands or millions entries).

12.2.2 Event Detection in Streams of Incidents

The small size and low power-consumption of MSU allow for mobile measurements. Placing these units on vehicles enables the coverage of a wider area with a lower number of units, while keeping the spatial and temporal resolution high enough. Few recent studies showed the possibility and advantages of such use, and the relatively easy adaptation of such MSU's to function as mobile sensing units, with the addition of GPS (Al Ali et al. 2010; Devarakonda et al. 2013; Levy et al. 2012). Thus, WDESN may consist of stationary and mobile nodes. Therefore the discussion is a dual-facet one, where incident streams generated from stationary and mobile WDESN nodes should be addressed differently. The following sections address event detection in streams received from stationary and mobile nodes.

12.2.2.1 Event Detection in Streams of Stationary WDESN Nodes

The common practice for extracting events from a set of incidents is finding a deviation from a pattern that represents the normal or routine set of incidents. The patterns may be computed from the indicators' time series themselves (e.g., Hall et al. 2007; Pickard et al. 2011), or from any representation of the data. Typical data representation techniques are Principal Component Analysis (PCA) (e.g., Huang et al. 2010), spectral exploration (e.g., Eder et al. 2012; Koscielny-Bunde et al. 1998; Whitcher and Jensen 2000) or De-trended Fluctuation Analysis – FDA (e.g., Koscielny-Bunde et al. 1998; Talkner and Weber 2000). All aforementioned methods require that the data samples are acquired in a uniform fashion, i.e., with the same time and spatial intervals between samples. However, this requirement, typically, does not hold for the dataset in hand. To tackle this hurdle, data interpolation methods that interpolate the data in a regular grid fashion from the non-uniform samples are sought. These methods include, for example, the Discrete Sampling Theorem (Yaroslavsky et al. 2009) and, in the context of air pollution, the Inverse Distance Weighted (IDW) (Moore et al. 2008) and Kriging (Sarigiannis and Saisana 2008) interpolations.

Regardless of the representation method, an incident can be regarded as an event if the deviation from the pattern is above a certain threshold. The threshold can be set manually (e.g., Pickard, et al. 2011) or inferred in an automatic fashion based on the acquired data and trends observed (Arad et al. 2013). This approach, however, is vulnerable to faulty measurements that present a large deviation from the pattern.

Classification approaches such as Support Vector Machine (SVM) (e.g., Oliker and Ostfeld 2012), Minimum Volume Ellipsoid (MVE) (Becker and Gather 1999) or the Supervised Normalized Cut (SNC) (Yang et al. 2013) present a more robust event detection schemes as they allow for the integration of several environmental indicators into one recorded incident. SVM aims at separating events from the background stream of incidents. The SVM method consists of two stages – At the first stage, incidents that are known to be an event and incident that are known to be non-event (i.e., routine) are applied on an environmental indicators space. Then, based on the pre-classified data, a classifier is constructed on this space. At the second stage each new incident is compared against the classifier and tagged accordingly. SVM requires that normal and abnormal incident patterns are known before-hand. While routine patterns can be extracted from continuous streams of incidents (e.g., Kroll and King 2007; Hall et al. 2007; Pickard et al. 2011), considering all possible abnormal indicators' conditions and values is not a tractable task.

MVE solves this problem by using only the non-event incidents in the training phase (Becker and Gather 1999). The MVE, in its training phase, constructs a minimum volume ellipsoid on the indicators space, which includes all non-event incidents. Then any measurement that falls outside this ellipsoid is considered to be an event.

The SNC method forms the event detection task as graph-cuts optimization problem. In this formulation an undirected graph, *G(V,E)*, is constructed, where

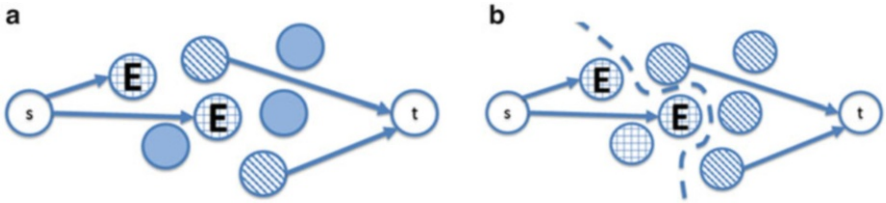

Fig. 12.2 SNC procedure

V is the set of incidents and E is the set of edges connecting adjacent incidents. The set of incidents contains both tagged (either as event or as non-event) and untagged incidents. A node that corresponds to an event-incident, is referred to as an event-node. Each edge $[i,j] \in E$ carries a similarity weight, w_{ij}, that represents the similarity between incidents i and j. The method is more generalized than SVM and MVE as adjacency (i.e., what characteristic qualifies two incident to be considered adjacent) and the similarity measures are not dictated by the SNC formulation and can be set by the characteristics of the specific problem in hand. Once the graph is constructed two additional nodes, s and t, are added to the graph. An arc with infinite weight is drawn from s to all event-nodes and from all non-event nodes to t. The construction of the graph is illustrated in Fig. 12.2a, where event-nodes are marked with an "E", non-event nodes are presented with diagonal markings and untagged nodes are solid.

The classification is then held by solving the s-t minimum-cut problem (Ford and Fulkerson 1956) on the graph. The minimum s-t cut problem partitions the nodes in the graph into two partitions, S and T, such that $s \in S$ and $t \in T$ and the sum of similarity weights on edges going between S and T (i.e., the cut) is minimized. After the partition, all incidents that correspond to nodes in S, are considered as an event and all incidents that are associated with nodes in T are tagged as non-event. The classification process is illustrated in Fig. 12.2b. All untagged nodes are now tagged based on the cut.

12.2.2.2 Event Detection in Streams of Mobile WDESN Nodes

Streams received from portable sensors are obtained from the same sensor in different locations at different times. Thus, mobile sensors' measurement are time-space variant. While networks that consist of mobile sensory platforms (Al Ali et al. 2010; Devarakonda et al. 2013; Levy et al. 2012) have been suggested, little effort has been invested in fusing the readings (i.e., incidents) into one holistic spatio-temporal patterns. Pattern recognition in mobile sensory networks has been suggested through the *concentrated alert* (CA) *problem*. The CA problem aim at optimizing two goals: One goal is to identify a small region; another goal is to have a large number of sensors indicating high pollution levels. These two goals are potentially conflicting – focusing on a large number of sensors reporting high levels of pollution within an area is likely to result in the entire region; on the other hand focusing on concentration alone would result in a single block of the area containing the highest number of sensors reporting high concentrations, thus disregarding information provided by other detectors in neighboring areas. The goal then is to identify, at every period of time, a region that relatively to its size has high concentration of sensors reporting high pollution levels. One way to achieve this is by minimizing a ratio function of the size divided by the number of detected events in the region. Another, is to optimize a weighted combination of the goals. As was shown in (Hochbaum and Fishbain 2011) the latter problem is efficiently solvable in polynomial time.

One iteration of the CA problem finds one region with the highest concentration of sensors reporting highest levels of pollution, hence a binary decision, which is based on solving a minimum-cut problem on an associated graph (Hochbaum and Fishbain 2011). After each iteration the region that was tagged as risky with its corresponding sensors are removed and the process is repeated so equi-pollution contour lines are produced. The final output is the patterns and deviations from these patterns are events.

Another graph-theory based method, firstly presented here, for generating patterns from distributed sensory network is based on Markov Random Fields (MRF) (Hochbaum 2001). The MRF optimization problem aims at finding a value such that two functions are minimized: a deviation cost function that depends on the distance between an observed value and a modified one; and a penalty function that grows with the distance between values of related (adjacent) pairs, i.e., *a separation function*. Specifically, within the suggested framework, the goal is to modify the indicators' values so that an objective function, consisting of one term due to the deviation of the events' values from the measurements, and a second term that penalizes differences in assigned values to adjacent spatial sensors, is minimized. This is illustrated in Fig. 12.3.

Adjacency is not mandated by the method and can be derived from the characteristics of the problem in hand. For example incidents can be considered to be adjacent if they are less than a predefined distance from each other. Once adjacency is determined, one has to assign the deviation and separation weights. This implies that for computing the separation cost function one should extract the affinity function for all adjacent incidents pairs.

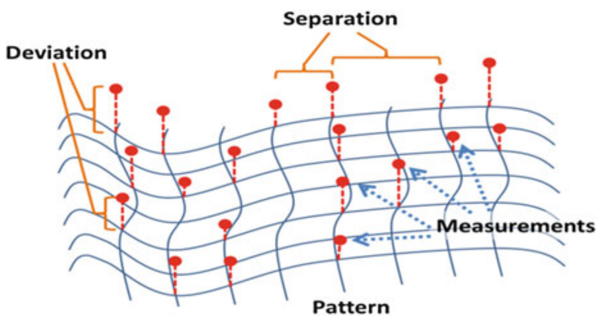

Fig. 12.3 Markov Random Field

For *G()* and *F()* the deviation and separation cost functions respectively, g_j – events value in node *j*, the problem is to find a pattern *X* that consists of a set of $x_i \in X$ that minimizes the following objective function:

$$min_X \left\{ \sum_{i \in V} G_j(g_i, x_i) + \sum_{(i,j) \in A} F_{ij}(x_i - x_j) \right\}$$

When the separation function is linear and the deviation function is linear, quadratic or piecewise linear convex, the MRF problem can be solved efficiently (in strongly polynomial time) through network flow algorithm (Hochbaum 2001).

12.2.3 Cross Referencing Events

In order to understand the nature of the entire environmental episode based on all extracted events, one has to find a set of events that co-align with each other and best explain the observed environmental episode. To this end, the common methods are the Maximum Likelihood Estimation (MLE) (e.g., Álvarez, et al. 2005), Artificial Neural Networks (ANN) (Reich, et al. 1999; Gardner and Dorling 1998) and the Dempster-Shafer Theory (D-S) (e.g., Wang, et al. 2006; Yadav, et al. 2013).

The MLE approach assumes an underlying model and aims at finding the model's parameters that are most probable, given the observations. For example, Álvarez et al. present a Markov chain model to study air pollution, where daily maximum ozone measurements in Mexico City are assumed to follow a Markov chain of order $K > 0$. The parameter *K* is then inferred using MLE (Álvarez et al. 2005). The necessity of a model is a drawback of these types of methods as describing environmental and atmospheric phenomena by an accurate model is an extremely difficult task.

In its most general formulation ANN consists of a system of simple interconnected neurons, or nodes (as illustrated in Fig. 12.4), which represent a nonlinear mapping between an input vector (events in this case) and an output vector (episodes). The nodes are connected by weights and output signals which are a function of the inputs to the node modified by a simple nonlinear transfer, or activation, function. It is the superposition of many simple nonlinear transfer functions that enables the multilayer perceptron to approximate extremely non-linear functions. The output of a node is scaled by the connecting weight and fed forward to be an input to the nodes in the next layer of the network. This implies a feed-forward neural network, where each input, I_i, goes through several such functions. Each *i*th transfer function in layer L,f_i^L, can be of a different form. Due to its easily computed derivative a commonly used transfer function is the logistic function, $y = (1 - e^{-x})^{-1}$, but any other function would work. As both the input

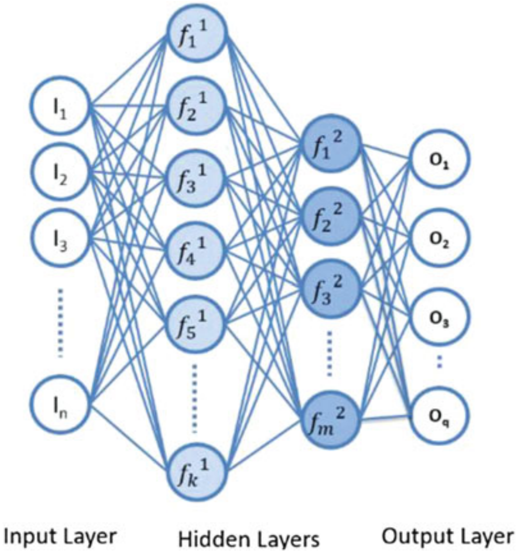

Fig. 12.4 Artificial neural network for associating events with episodes

events $\{I_1, I_2, I_3, I_4, \cdots, I_n\}$ and observed episodes $\{O_1, O_2, O_3, \cdots, O_q\}$are known, the goal is to find the various functions' parameters that explain best the linkage between the events of episodes. The resulted network and its parameters explain the role of various events in the observed episodes. Thus the goal is not to detect the happening of an episode, but to infer the events that contribute to its happening.

In its core the D-S methodology is a generalization of the Bayesian theory of subjective probability. For illustration of the problem, let us consider an episode, O, and a set of related events, Ω. All subsets of events, $\{\omega\} \subseteq \Omega$, are assigned with a probability [0,1] that represents their post-priori probability to actually happen given observation O, such that $\sum_{\omega \subseteq \Omega} P(\omega) = 1$. Two values are then calculated for each subset, ω_i:

- **Belief Value** – the sum of probabilities of all subsets that include ω_i, i.e., $Bel(\omega_i) = \sum_{\omega_j \supseteq \omega_i} P(\omega_j)$.
- **Plausibility value** – the sum of probabilities of all subsets that intersect with, ω_i, thus $Pl(\omega_i) = \sum_{\omega_j \cap \omega_i \neq \phi} P(\omega_j)$.

$Bel(\omega_i)$ represents the evidence we have for ω_i directly. So $P(\omega_i)$ cannot be less than this value. $Pl(\omega_i)$ represents the maximum share of the evidence, if, for all sets that intersect with ω_i, the part that intersects is actually valid. Thus, $Pl(\omega_i)$ is the maximum possible value of $P(\omega_i)$. The difference, $\left(Pl(\omega_j) - Bel(\omega_j) \geq 0\right)$, called the *belief interval*, represents how certain a belief, ω_i is. Thus, how accurate is the initial assumption about its post-prior conditional probability given O.

12.3 Neighborhood Scale Air-Quality Environmental Indicators

Next we illustrate the use of the suggested framework within the context of air-quality monitoring by describing the results of (Moltchanov et al. 2015) in the above terminology. In their study Moltchanov et al. (2015) report ambient measurements of gaseous air pollutants by a WDESN of MSUs that have been deployed in three urban sites, about 150 m apart. This study of Moltchaqnov et al. is the first time a network's capability to capture spatiotemporal variations is demonstrated at a sub-neighborhood spatial resolution, which suits the requirement for highly spatiotemporal resolved measurements at the breathing-height when assessing exposure to urban air pollution.

12.3.1 Study Design

The study region is the city of Haifa, located in the north of Israel with a population of ~265,000. The city is located at the northern part of the Israeli Mediterranean Sea shore, ranging from sea level to the Carmel mountain ridge at ~500 m above sea level (ASL). The Haifa bay area is a busy metropolitan, combining both densely populated residential areas and a large industrial complex which includes a primary sea-port, petrochemical industries and oil refineries. The entire region is burdened with traffic, including heavy vehicles, cargo trains and a large number of diesel buses, as Haifa functions as a transportation center for the entire northern region of Israel.

Specifically, the study was performed at the Neve Sha'anan neighborhood – a residential neighborhood located on a relatively leveled region of the Carmel Ridge, about 200 m ASL. The neighborhood is roughly divided by a major road (Trumpeldor Ave.), which also serves as the main commercial center for the residents of this area (Fig. 12.5).

In the study described in (Moltchanov et al. 2015), six units were deployed at three different locations some 100–150 m apart – sites A, B and C in Fig. 12.5. The campaign took place for 71 days in the summer of 2013 (between 16/06/2013 and 26/08/2013), a season that is characterized by meteorological stability and persistence. Site A is a balcony on small residential street. The units were placed 3 m high above ground level (AGL) and 5 m off the centerline of the road, which experience sparse traffic density. Site B is a second story balcony. The first story of the building houses a pizzeria with an exhaust at the rear side of the building. The units were placed 4 m AGL and 7 m off an adjacent busy street. The building is a corner building of the Ziv junction, which is a busy local (neighborhood-scale) commercial zone (see Fig. 12.5b and c). Site C is located on the Trumpeldor Blvd., the main street of the neighborhood, ~80 m NE from a busy bus stop. The units were placed on a roof of a kindergarten courtyard, 3 m AGL and 4 m off road.

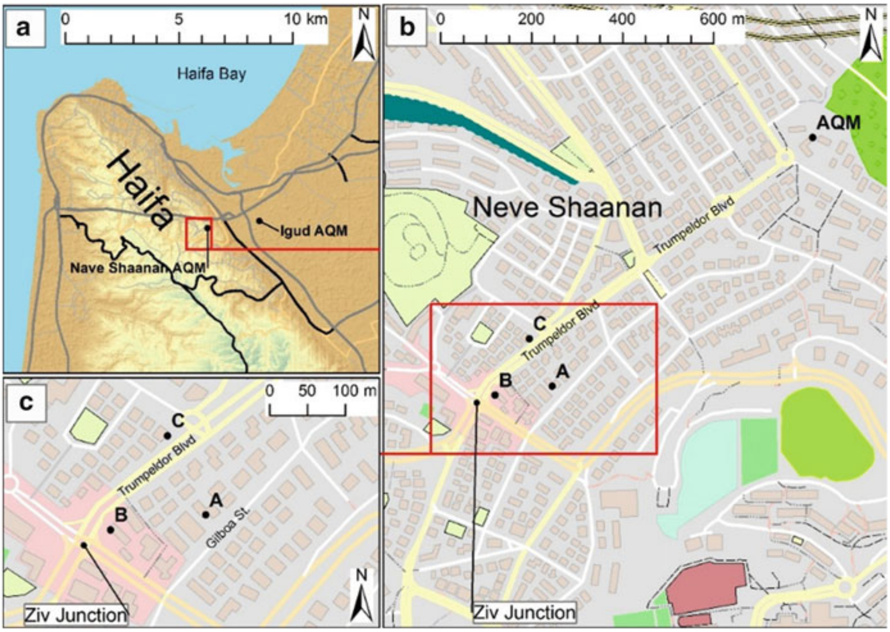

Fig. 12.5 Map of the study region showing the Haifa bay area and the Carmel Ridge (**a**), Nave Shaanan neighborhood (**b**) and the Ziv Junction area in greater detail (**c**)

Wind data were obtained from the Nave Shaanan AQM station (marked AQM in Fig. 12.5b), located 750 m down the road from site C.

12.3.2 *Incident Streams*

The incident streams are produced by CanarIT^{TM} air quality MSUs (Airbase Systems LTD). Each unit generates streams of ambient levels of the following indicators – O_3, NO_2, total VOC (tVOC), Total Suspended Particles (TSP), noise, temperature (Temp) and relative humidity (RH). In this analysis we focus on NO_2 and O_3 measurements only. Each apparatus gauges the indicators' level over a short time period (Meas.T) and reports the average level, computed over Meas.T, at a given Frequency (Freq). Thus, Meas.T is shorter than the time between reports (1/Freq). The technical specifications of the various sensors are detailed in Table 12.1. Measurements are transmitted to a cloud-based storage, by an on-board GSM embedded chip.

Throughout the campaign, one unit was fixed at each site. The three remaining units were rotated twice between the sites, on days 28 and 51 from the beginning of the campaign, so that each of these three units operated at each of the three sites. The rotation allowed for comparison within and between the sites, thus enabling to

Table 12.1 Sensors technical specifications

Stream	Man.	Model	Meas.T [s]	Freq. [Hz]	Dyn.R [ppb]	Res. [ppb]
O_3	Aeroqual	SM50	1	60	0–150	1
NO2	AppliedSensors	iAQ-100	2	20	0–2,000 ppb	5
tVOC[1]	AppliedSensors	iAQ-100	2	20	0–2,000 ppm[2]	
TSP			0.5	20		
Noise			0.25	20		
Temp.			0.5	20		
RH			0.5	20		

Man. manufacturer, *Meas.T* acquisition time, *Freq.* reports frequency, *Dyn.R* sensors' dynamic range, *Res.* the sensors' resolution

Table 12.2 Location of units during each of the study periods

Site	Fixed units	Period I – days 1–28	Period II – days 28–50	Period III – days 50–71
		Jun. 16 – Jul. 14	Jul. 14 – Aug. 5	Aug. 5 – Aug. 26
A	418	422	413	424
B	420	424	422	413
C	414	413	424	422

compare the obtained streams across the different locations and the sensor nodes. Table 12.2 details the sensors' location throughout the field campaign. Because of the persistent meteorological conditions over the region in the summer, 21–28 days' time periods are assumed sufficient to characterize the pollution levels at each of the microenvironments.

As detailed in (Moltchanov et al. 2015), in general, high correlations were found among incidents of collocated WDESN nodes, suggesting consistency across different MSUs' measurements. In contrast, low correlations are depicted between concentrations of both NO_2 and O_3 measured by MSUs at different locations, indicating that the measured concentrations echoed some local conditions and responded to the specific microenvironment where the nodes were placed.

12.3.3 Events Detection in Streams of Incidents

Diurnal patterns of NO_2 and O_3 concentrations among collocated MSUs were highly correlated. NO_2 and O_3 daily patterns are presented in Fig. 12.6. Figures (a) and (c) depict weekday's pattern of NO_2 and O_3 respectively. Figures (b) and (d) present the corresponding weekend's patterns. Both NO_2 and O_3 reveal distinct patterns in each site. The reason for this is twofold – first each microenvironment is governed by different conditions. Second, the sensors themselves present inherent biases.

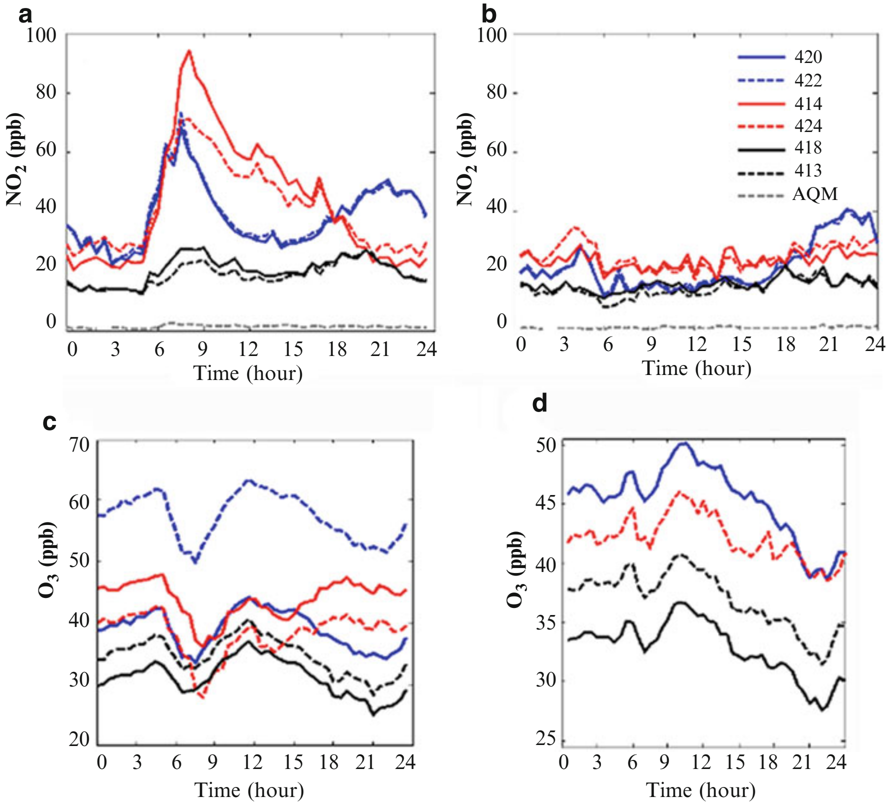

Fig. 12.6 Daily patterns of NO2, (**a**) and (**b**), and O3, (**c**) and (**d**), concentrations (30 min. averages) measured by collocated nodes in locations A (*black lines*), B (*blue lines*) and C (*red lines*) during Period II. Plates (**a**) and (**c**) are weekdays (Sunday–Thursday) patterns, plates (**b**) and (**d**) weekends (Saturdays) patterns. *Dashed grey lines* – simultaneous AQM monitoring data (Source Moltchanov et al. 2015)

This is especially true for O_3 sensors (see Fig. 12.6c and d). Regardless of the reason for these differences, the phenomenon itself emphasizes the advantage of extracting events from each stream of incidents separately.

In the study of Moltchanov et al. (2015), the goal was to assess the MSUs' capability to capture real-life phenomena. Therefore, let us consider incidents acquired from a specific stream as events if they present values that are higher or lower (by a standard deviation) than the specific stream's average. This should correspond to higher and lower values of NO2 and O3 respectively during morning and evening traffic peaks (Levy et al. 2012; Nazaroff and Alvarez-Cohen 2001; Zalel et al. 2008). On weekday, events were recorded, for example, between 7 and 8 am. On weekends, on the other hand, the daily peaks diminish (Fig. 12.6b, d) and therefore no events are recorded.

Site A is located in a small residential street with little traffic and is characterized by lower NO_2 concentrations than the two other sites (See sensor # 413 and 418 in Fig. 12.6a and b). Nevertheless, events are recorded at site A at morning (7:00–7:30) and at evening (~19:30) rush hour peaks. Site B is located at a busy junction in a commercial area. Indeed, the diurnal NO_2 patterns on weekdays reveal a dual-peak daily pattern (Fig. 12.6a), where events are indeed recorded around these peaks. On Saturdays the concentrations remain low and constant until ~19:00, when traffic resumes (Fig. 12.6b). However no events are recorded on weekends. The signature of traffic is revealed also in the weekday ozone concentration patterns that show a dramatic decrease from 05:30 am and reach a minimum at about 07:00–07:30, in parallel to a peak in NO_2 concentrations. Typically, O_3 concentrations increase afterwards and reach peak values at around noon (Fig. 12.6c). Likewise, events extracted from NO_2 and O_3 streams at site C indicate traffic related origin. Peak NO_2 concentrations and minimum O_3 concentrations occurred between 08:00 and 09:00 am.

12.3.4 Cross Correlating Events

Once the set of events are extracted one can cross-correlate them and infer the observed episode. Based on the recorded events, the morning episode at sites A and B starts on average at 7:00 am. Thus, the morning episode's spatial boundaries contain both sites. The episode at site C starts somewhat later than that, forming a different episode. It is worthwhile noting that the suggested reasoning, given in (Moltchanov et al. 2015), is that while the episodes in A and B are due to traffic behavior, the episode at site C is due to children drop off at the kindergarten, located at C. The association of events recorded in sites B and C with the same morning episode is supported by the low mean wind speed (Fig. 12.7a) and the high traffic volume at the Ziv junction, which result in high concentrations at both B and C.

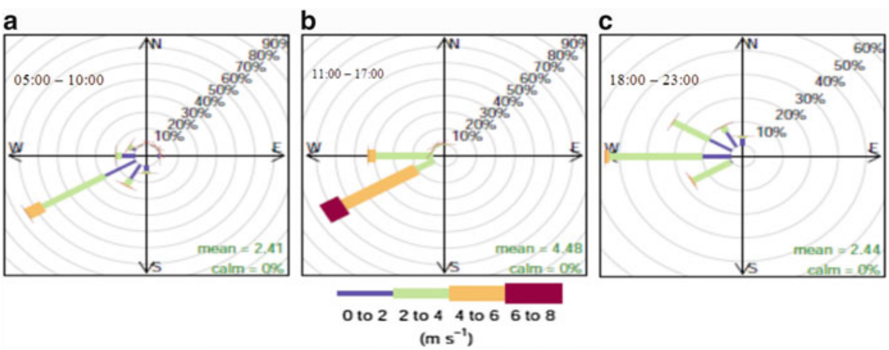

Fig. 12.7 Wind rose plots based on data measured at the AQM station on period I. (**a**) Morning, (**b**) noon and afternoon, and (**c**) evening (Source Moltchanov et al. 2015)

In the afternoon, however, reduced traffic volume and higher wind speeds (Fig. 12.7b) result in significantly lower concentrations at site B whereas site C maintains its relatively higher concentrations. Thus, the evening episode is bounded to site C only.

This detailed discussion demonstrates the capability of the suggested framework to capture both spatial and temporal boundaries of an environmental episode – traffic pollution in this case. The episodes correspond to intra-urban pollutant "hot-spots", which in return can be treated. The aggregation of the events into set of episodes allows for investigating the episode separately, while incorporating all supporting relevant data in the process and eliminating all non-relevant data. Hence, the use of the suggested framework has a huge potential for providing a comprehensive, yet simple framework for capturing environmental indicators' impact and their dynamic spatial and temporal variability. As such it presents a great potential to become a common tool in examining environmental indicators acquired through WDESN.

Conclusion

In this chapter a framework and its lexis foundations for analyzing measurements acquired by Wireless Environmental Distributed Sensory Network (WDESN) is presented. A set of sensors acquire observations of environmental indicators. These are formed by the sensing platform as streams of incidents. Incidents that deviate from the normal pattern are regarded as events. The event extraction can be held on the sensing platform itself, which facilitates distributed analysis of the acquired data. Then the events are cross-referenced for inferring environmental episodes. This formulation ensures, by definition, that an environmental episode is well defined in time and space and manifests itself in the recorded incidents by the sensors. Algorithms and methods that constitute the framework are also described.

A practical example of using this framework for describing the results of (Moltchanov et al. 2015) is given. In this example, it is shown how the framework set the spatial and temporal boundaries of an environmental episode and how it captures its dynamic behavior. This shows the great potential of such framework for any future environmental research that is based on WDESN.

References

Al Ali A, Zualkernan I, Aloul F (2010) A mobile GPRS-sensors array for air pollution monitoring. IEEE Sens J 10(10):1666–1671

Álvarez L, Fernández-Bremauntz A, Rodrigues E, Tzintzun G (2005) Maximum a posteriori estimation of the daily ozone peaks in Mexico City. J Agric Biol Environ Stat 10(3):276–290

Arad J, Housh M, Perelman L, Ostfeld A (2013) A dynamic thresholds scheme for contaminant event detection in water distribution systems. Water Res 47(5):1899–1908

Bakken G, Kauffmanl G, Jurs P, Albert K, Stitzel S (2001) Pattern recognition analysis of optical sensor array data to detect nitroaromatic compound vapors. Sens Actuators B 79:1–10

Becker C, Gather U (1999) The masking breakdown point of multivariate outlier identification rules. J Am Stat Assoc 94(447):947–955

Broday D, Fishbain B, Etzion Y, Levy I (2013) Evaluation of low-cost PM sensors, intended for use in a dense monitoring grid. Portland Oregon, s.n.

Budde M, Berning M, Busse M, Miyaki T, Beigl M (2012) The TECO envboard: a mobile sensor platform for accurate urban sensing — and more. In: Ninth International Conference on Networked Sensing Systems (INSS), Antwerp, Belgium

Chen C (2008) Transform methods for remote sensing environmental monitoring. IEEE Signal Processing Society, Las-Vegas, pp 5165–5168

CNN Wire Staff (2011) Lack of oxygen likely cause of fish kill in California. Cable News Network (CNN), Los Angeles

Denzer R (2005) Generic integration of environmental decision support systems – state-of-the-art. Environ Model Softw 20(10):1217–1223

Devarakonda S, Sevusu P, Liu H, Liu R, Iftode L, Nath B (2013) Real-time air quality monitoring through mobile sensing in metropolitan areas. In: Proceedings of the 2nd ACM SIGKDD international workshop on urban computing. ACM

Dutta P, Aoki P, Kumar N, Mainwaring A, Myers C, Willett W, Woodruff A (2009) Common sense: participatory urban sensing using a network of handheld air quality monitors. In: 7th ACM Conference on Embedded Networked Sensor Systems (SenSys '09). ACM, Berkeley, CA

Eder B, Appel W, Pierce T (2012) On the use of principal component and spectral density analyses to evaluate the community multiscale air quality (CMAQ) model. In: Air pollution modeling and its application XXI. Springer, Dordrecht, pp 439–445

Etzion Y, Broday D, Fishbain B (2013) Analysis of image color and effective bandwidth as a tool for assessing air pollution at urban spatiotemporal scale. SPIE, San Francisco

Ford LR, Fulkerson DR (1956) Maximal flow through a network. Can J Math 8:399–404

Gardner M, Dorling S (1998) Artificial neural networks (the multilayer perceptron) – a review of applications in the atmospheric sciences. Atmos Environ 32(14–15):2627–2636

Hall J, Zaffiro A, Marx R, Kefauver P, Krishnan R, Haught R, Herrmann J (2007) On-line water quality parameters as indicators of distribution system contamination. J Am Water Works Assoc 99(1):66–77

Hart J, Martinez K (2006) Environmental sensor networks: a revolution in the earth system science? Earth Sci Rev 78(3–4):177–191

Hochbaum D (2001) An efficient algorithm for image segmentation, Markov Random Fields and related problems. J ACM 48(2):686–701

Hochbaum D, Fishbain B (2011) Nuclear threat detection with mobile distributed sensor networks. Ann Oper Res 187(1):45–63

Huang J, Choi H, Hopke P, Holsen T (2010) Ambient mercury sources in Rochester, NY: results from principle components analysis (PCA) of mercury monitoring network data. Environ Sci Technol 44(22):8441–8445

Koscielny-Bunde E, Bunde A, Havlin S, Roman H, Goldreich Y, Schellnhuber H (1998) Indication of a universal persistence law governing atmospheric variability. Phys Rev Lett 81 (3):729–732

Kramer K (2009) Continuous monitoring of waters by biological early warning systems. In: Rapid chemical and biological techniques for water monitoring. Wiley, West Sussex, pp 197–219

Kranz M, Holleis P, Schmidt A (2010) Embedded interaction: interacting with the Internet of things. Internet Comput IEEE 14(2):46–53

Kroll D, King K (2007) Operational and laboratory verification testing of a heuristic on-line water monitoring system for security. Int J High Speed Electron Syst 17(4):631–642

Levy I, Mihele C, Lu G, Narayan J, Hilker N, Brook J (2012) Elucidating multipollutant exposure across a complex metropolitan area by systematic deployment of a mobile laboratory. Atmos Chem Phys Discuss 12:31585–31627

Martinez K, Hart J (2004) Environmental sensor networks. Computer 37(8):50–56

Mead M, Popoola O, Stewart G, Landshoff P, Calleja M, Hayes M, Baldovi JJ, Hodgson TF, McLeod MW, Dicks J, Lewis A, Cohen J, Baron R, Saffell JR, Jones R (2013) The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. Atmos Environ 70:186–203

Moltchanov S, Levy I, Etzion Y, Lerner U, Broday DM, Fishbain B (2015) On the feasibility of measuring urban air pollution by wireless distributed sensor networks. Sci Tot Environ 502:537–547

Moore K, Neugebauer R, Lurmann F, Hall J, Brajer V, Alcorn S, Tager I (2008) Ambient ozone concentrations cause increased hospitalizations for asthma in children: an 18-year study in Southern California. Environ Health Perspect 116(8):1063–1070

Nazaroff W, Alvarez-Cohen L (2001) Environmental engineering science. Wiley, New York

Oliker N, Ostfeld A (2012) A weighted support vector machine classifier for contamination event detection in water distribution systems. Adelaide, s.n.

Pickard B, Haas A, Allgeier S (2011) Optimizing operational reliability of the Cincinnati contamination warning system. J Am Water Work Assoc 103(1):60–68

Ramanathan N, Balzano L, Estrin D, Hansen M, Harmon T, Jay J, Kaiser B, Sukhatme G (2006) Designing wireless sensor networks as a shared resource for sustainable development. In: IEEE International Conference on Information and Communication Technologies and Development (ICTD 2006). IEEE Society on Social Implications of Technology (SSIT), Berkeley, pp 256–265

Reich S, Gomez D, Dawidowski L (1999) Artificial neural network for the identification of unknown air pollution sources. Atmos Environ 33(18):3045–3052

Research and Development, National Homeland Security Research Center (2005) Technologies and techniques for early warning systems to monitor and evaluate drinking water quality: a state-of-the-art review. US Environmental Protection Agency (USEPA), Cincinnati

Sarigiannis D, Saisana M (2008) Multi-objective optimization of air quality monitoring. Environ Monit Assess 136:87–99

Shashank J, Levy I, Fishbain B, Broday D, (2013) Application of distributed sensor networks for estimating exposures to air pollution in urban areas. Basel, s.n.

Storey M, Gaag B v d, Burns B (2011) Advances in on-line drinking water quality monitoring and early warning systems. Water Res 45(2):741–747

Talkner P, Weber R (2000) Power spectrum and detrended fluctuation analysis: application to daily temperatures. Phys Rev Ser E 62:150–160

Wang Y, Yang J, Xu D (2006) Environmental impact assessment using the evidential reasoning approach. Eur J Oper Res 174(3):1885–1913

Werner-Allen G, Lorincz K, Welsh M, Marcillo O, Johnson J, Ruiz M, Lees J (2006) Deploying a wireless sensor network on an active Volcano. IEEE Internet Compu 10(2):18–25

Whitcher B, Jensen M (2000) Wavelet estimation of a local long memory parameter. Explor Geophys 31:94–103

Williams D, Henshaw G, Bart M, Laing G, Wagner J, Naisbitt S, Salmond J (2013) Validation of low-cost ozone measurement instruments suitable for use in an air-quality monitoring network. Meas Sci Technol 24(6):5803–5814

Yadav J, Kharat V, Deshpande A (2013) Evidence theory and fuzzy relational calculus in estimation of health effects due to air pollution. J Intell Syst 22(1):9–23

Yang Y, Fishbain B, Hochbaum D, Norman E, Swanberg E (2013) The supervised normalized cut method for detecting, classifying, and identifying special nuclear materials. INFORMS J Comput 26(1):45–58

Yaroslavsky L, Shabat G, Ideses I, Salomon B, Fishbain B (2009) Non-uniform signal sampling, signal and image recovery from sparse data. J Opt Soc Am A 26(3):566–575

Zalel A, Yuval, Broday D (2008) Revealing source signatures in ambient BTEX concentrations. Environ Pollut 156(2):553–562

Zhou J, Roure D (2007) FloodNet: coupling adaptive sampling with energy aware routing in a flood warning system. J Comput Sci Technol 22(1):121–130