Live E! Sensor Network: Correlations in Time and Space

Elyes Ben Hamida[1], Pierre Borgnat[2], Hiroshi Esaki[3], Patrice Abry[2], Eric Fleury[4]

[1]Université de Lyon, INSA Lyon, INRIA/ARES, France

[2]Université de Lyon, ENS Lyon, CNRS, Laboratoire de Physique, UMR 5672, France

[3]The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan

[4]Université de Lyon, ENS Lyon, INRIA/ARES, France

elyes.ben-hamida@insa-lyon.fr, pierre.borgnat@ens-lyon.fr, hiroshi@wide.ad.jp, patrice.abry@ens-lyon.fr, eric.fleury@inria.fr

Résumé – Le projet *Live E!* est un consortium de recherche qui réunit des établissements universitaires et industriels japonais et qui a pour objectif le développement d'une plate-forme expérimentale afin de collecter et de partager des informations en relation avec l'environnement. Cette plate-forme consiste en un large nombre de capteurs déployés à travers plusieurs pays et mesurant plusieurs phénomènes physiques, tels que la température, l'humidité, la pression, *etc.* Dans ce papier, nous présentons une analyse préliminaire des mesures collectées. Nous proposons une méthode de décomposition des séries temporelles grâce à laquelle nous étudions les structures de corrélation temporelle et spatiale des mesures observées par les capteurs.

Abstract – The *Live E!* project is a Japanese research consortium among industry and academia to develop a platform collecting and sharing digital information related with the earth. This platform consists of a large number of spatially distributed weather stations measuring different environmental quantities such as temperature, humidity, pressure, etc. In this paper, we conduct the first analysis on this huge data set. To that end, we perform a synchronous-average based decomposition of the collected time series, and we explore the structures of correlations both in time and space on the observed data.

1 Motivation and context

Recent natural disasters, such as heat island effect or hurricane, have highlighted the need for concerted research efforts to better understand the causes and consequences of such natural phenomena whose impacts on social life and business activity is significant. In this context, the *live E!* project [1], developed by the WIDE consortium and the IPv6 Promotion Council Japan, deploys a global infrastructure aiming at collecting and sharing environmental information.

Live E! provides a multi-domain sensor networking platform, which is composed of several weather stations deployed across several countries of Asia, including Japan, Thailand, *etc.* Each weather station is operated in autonomous and distributed manner and embeds several sensors measuring standard environmental quantities, such as temperature, pressure, humidity, wind speed, *etc.* As depicted in Fig. 1, each weather station is managed by an operational-unit such as an university or a company which has sensors for their own activities. An operational-unit has a server, which collects data from weather stations basically in *real-time* via the Internet. Those servers collaboratively exchange the data among them using the *Live E!* standard protocol over the Internet, providing multi-attribute sensor search and data retrieval interface for users. In February 2009, the *Live E!* platform has 106 weather stations deployed across 13 countries and 11 servers across 9 organizations. Readers are referred to [1, 2] for further information on the *live E!* platform.

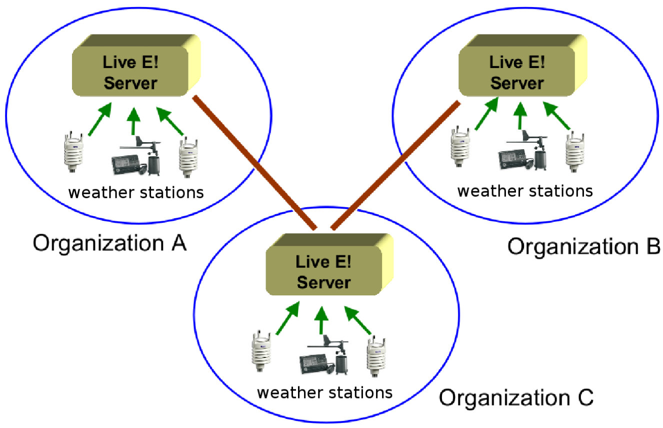


Fig. 1: *Live E! network architecture.*

In this paper, we provide a first analysis of the space and time correlations of the collected *live E!* measurements. We propose an original synchronous average-based model to decompose each time series in order to separate dependencies stemming from expected cyclic trends,

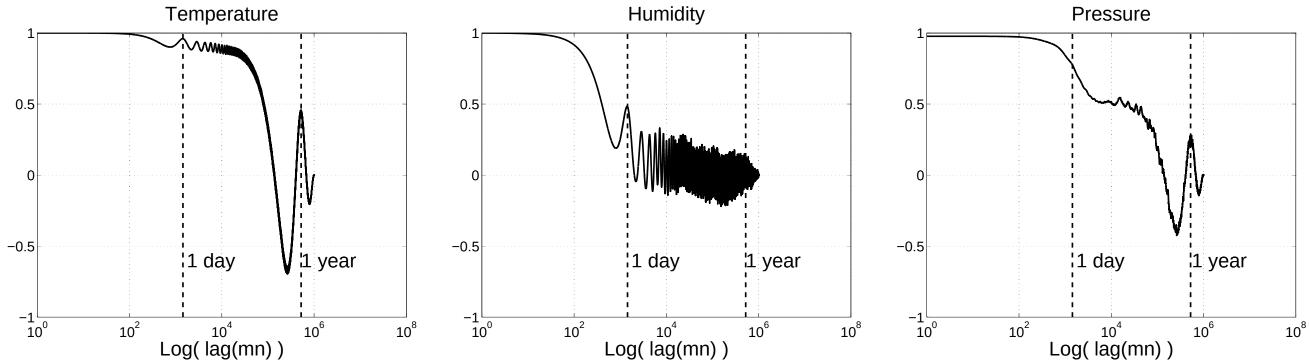


FIG. 2: *Autocorrelation functions (ACF) on Temperature, Humidity and Pressure.*

such as seasonal (year) and daily effects from dependencies actually observed on residual fluctuations. Then, we investigate the auto- and cross- correlations of both the measurements observed at a given weather station and between measurements collected at adjacent stations.

The remainder of this paper is organized as follows. Section 2 describes the data set. In Section 3, we propose an original synchronous average-based decomposition of time series. Then, we investigate, in Section 4, the space and time correlations both for the entire data and the fluctuations. Finally, we conclude in Section 5 and we give some prospects of future work in the area.

2 Dataset

In the present contribution, we consider the dataset collected at *kurashiki* city, located west of Okayama, Japan, from November 3, 2006 to October 10, 2008 (for almost two years). It consists of 25 weather stations, configured to measure information every $60s$. Here, we concentrate our analysis on the *Temperature*, *Humidity* and *Pressure* measurements.

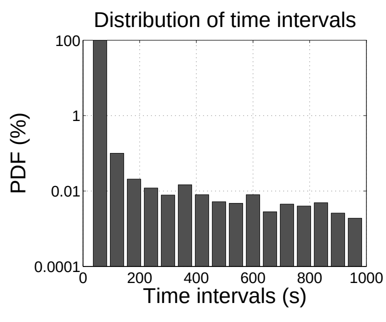


FIG. 3: *PDF of time intervals.*

A preliminary analysis indicates ***i)*** that weather station clocks are not perfectly synchronized and suffer sometimes from a low level jitter, i.e., in the extreme case the sampling period slightly fluctuates around the nominal $T_S = 60s$ (order of magnitude $\approx 13\%$) ***ii)*** and that there are missing data, cf. Fig. 3 displaying the histograms of delays between actual successive measurements (order of magnitude of missing data is 1%). Fig. 3 shows that fortunately long rows of missing data are extremely rare. To ease correlation structure assessment, data are preprocessed by standard (linear) interpolation techniques aiming at both replacing missing data and ensuring a regular and synchronous time sampling.

3 Trends, fluctuations and their autocorrelations in time

From pre-processed data, autocorrelation functions (ACF) (obtained from standard estimation procedures) of the *Temperature*, *Humidity* and *Pressure* measurement are depicted in Fig. 2, for a particular station (*ajino junior high school* weather station) chosen as representative of the dataset. For *Temperature*, *Humidity* and *Pressure*, ACF shows clear periodicities corresponding to $D = 1$ day and to 1 year.

Inspired by analysis of cyclostationary process, a model is supposed for each time series, $X(t)$, that consists in the addition of 3 modes:

$$X(t) = S(t) + \sum_{n=1}^{N_d} [A_n(d) \times G_J(t - nD)] + R(t)$$

where N_d is the total number of days, $S(t)$ represents the trend over the year, $A_n(d) \times G_J(t)$ is the daily trend and $R(t)$ is a short term fluctuations (or residual).

The yearly mode is defined as being an smoothed averaged over one day of the data, so that it is not affected by the daily variations (represented in Fig. 4 (b)). Then, the daily trend is estimated by taking a cyclic mean over a period of day, after taking out the yearly mode:

$$G_J(t) = << X(t) - S(t) >>_d$$

where $<< \, . \, >>_d$ represents the median value over one day. The amplitude of a specific day is defined as:

$$A_n(d) = [\max(X(t)) - \min(X(t))]_{t \in d}$$

Thanks to this additive model, each time series, $X(t)$, can be decomposed into three components: the trend over

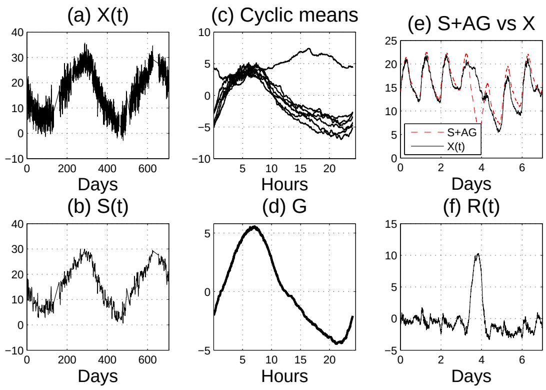


Fig. 4: *Decomposition of Temperature in modes: **a)** original time series; **b)** trend over the year; **c)** cyclic superposition of data over the day (for a given week); **d)** estimated cyclic daily trend; **e)** sum of the first 2 modes compared to $X(t)$; **f)** residual $R(t)$.*

the year $S(t)$, the daily trend $A_n(d) \times G_J(t)$ (day and night related movements) and the residual $R(t)$ (short term fluctuations).

Fig. 4 illustrates the output of this decomposition procedure for the *ajino_junior_high_school* weather station. After this decomposition, the ACF of the residual $R(t)$ is computed (shown in Fig. 5). For the Temperature T and Humidity H measurements, some periodicity over the day can still be observed because of the structure of the errors. As seen on Fig. 4 (e) and (f), the most important departures are when a day does not follow closely the usual cyclic pattern. This creates shapes of one day, hence the remaining periodicity.

4 Correlations amongst sensors and stations

To estimate correlations at a given weather station, amongst sensors, we compute the cross-correlations (XCF) between the collected environmental quantities (*Temperature*, *Humidity* and *Pressure*). The correlation coefficients are averaged over all weather stations and the results are summarized in Tab.1.

Tab. 1: Cross-correlation (XCF) coefficients between environmental quantities

	T/H	**T/P**	**H/P**
$R(t)$	-0.46 ± 0.08	-0.24 ± 0.13	-0.09 ± 0.10

Results show ***i)*** an inverse correlation between temperature and humidity ($\approx -0.46 \pm 0.08$ for $R(t)$); ***ii)*** a low negative correlation between temperature and pressure ($\approx -0.24 \pm 0.13$ for $R(t)$); ***iii)*** and finally no correlation between pressure and humidity ($\approx -0.09 \pm 0.1$ for $R(t)$). These results indicate that some environmental quantities are significantly inter-correlated even in their short time fluctuations (or residuals, i.e., after removal of the year and day trends). Such correlations can be exploited to forecast missing measurements or to perform optimized and efficient data aggregation.

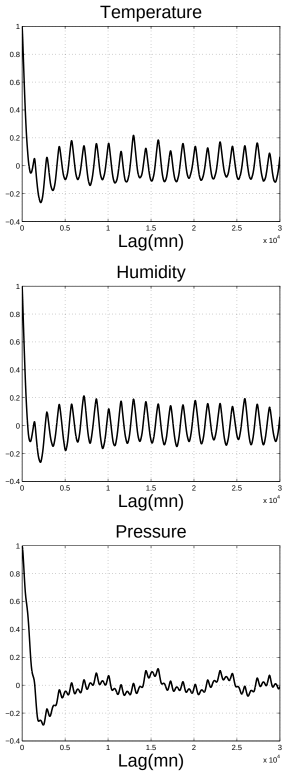
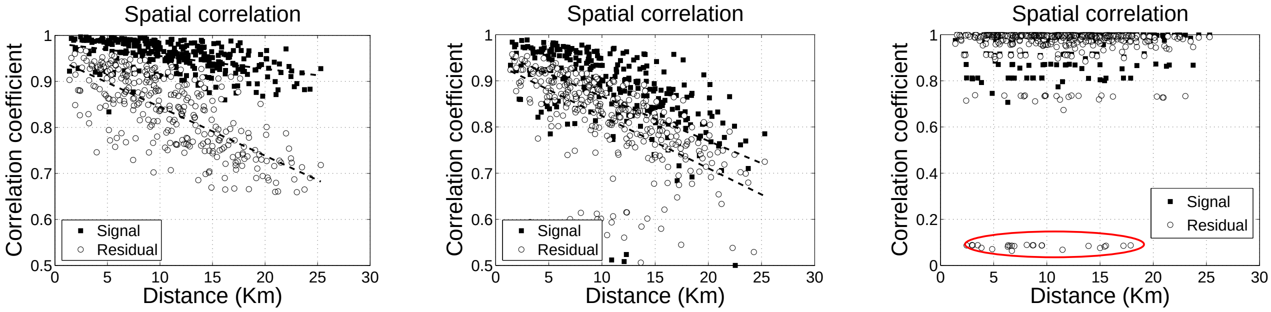


Fig. 5: *Autocorrelation functions (ACF) on the residual of Temperature, Humidity and Pressure time series (ajino_junior_high_school weather station).*

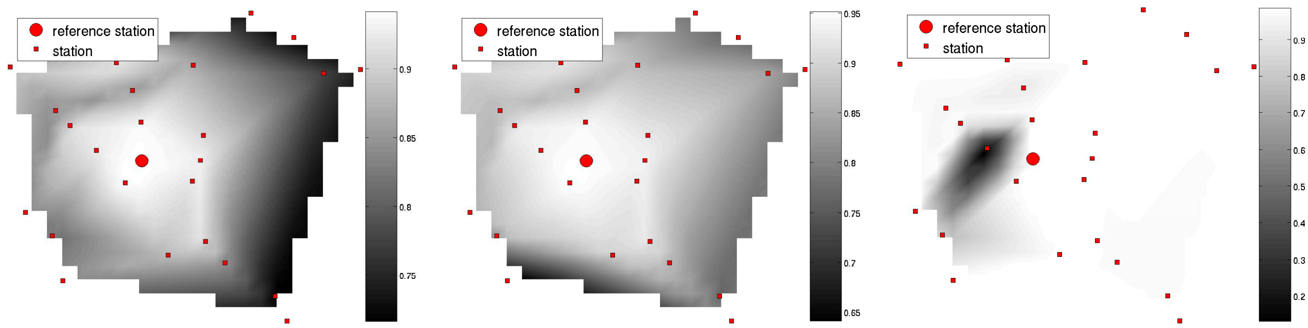
Regarding the correlation amongst adjacent weather stations, we evaluate the correlation coefficients between each pairs of weather stations, both for the entire data $X(t)$ and the fluctuations $R(t)$. The coefficients of T, H and P are compared in Fig. 6 as a function of the distance between stations (calculated from the station Latitude and Longitude available information). Moreover, spatial correlation maps are depicted in Fig. 7 for both T, H and P. These maps measure the degree of correlation of $R(t)$ between a reference weather station (located in the middle of the deployment area) and all other stations.

For T and H, while correlation coefficients of the complete data only very slightly decrease with distance for the entire data (a signature of the correlation of the daily cycles), for the residuals they decrease of about 30% in a range of 25km distances (quasi linearly). This explain the color gradients, obtained in Fig.7-(a) and Fig.7-(b), which vary from white (*i.e.,* high correlation) to black (*i.e.,* lower correlation).



(a) Temperature (b) Humidity (c) Pressure

FIG. 6: *Spatial correlation as a function of the distance.*



(a) Temperature (b) Humidity (c) Pressure

FIG. 7: *Spatial correlation maps corresponding to $R(t)$.*

For P, the situation is different as, both for entire data and residuals, correlation coefficients do not decrease with distance and remain very close to 1, hence showing that even short term residual fluctuations remain highly correlated over distances. This can be interpreted as a confirmation of the expected fact that Pressure is mostly controlled by altitude and that all station in the dataset are located at comparable altitudes. However, Fig. 6-(c) shows that for a number of data points, residuals are quasi not correlated to other stations. The weather station involved in these low coefficients is the *nishi_junior_high_school* station (located inside the shaded area of Fig.7-(c)). As all weather stations are located at similar altitude, this low correlation is thus a consequence of a failure or a lack of calibration for the *Pressure* sensor of that particular weather station.

5 Conclusions and developments

In this paper, we presented a preliminary analysis of the time and space auto- and cross- correlations of the *Live E!* measurements collected at the *kurashiki* city. We proposed an original synchronous average-based model to decompose each time series in order to separate dependencies stemming from expected cyclic trends, such as seasonal (year) and daily effects from dependencies actually observed on residual fluctuations. Results show that some environmental quantities (*Temperature* and *Humidity*) are significantly inter-correlated even in their short time fluctuations and present a high spatial correlation between adjacent weather stations, which is not the case of the *Pressure* measurements.

This preliminary work suggests several directions for future research. First, the careful analysis of such correlation structures is envisaged as a necessary preliminary phase enabling the detection of sensor fault or anomalous meteorological event. Moreover, the knowledge of the covariance structures as well as the temporal and spatial correlation is likely to ease the development of new efficient data aggregation, dissemination and forecasting algorithms.

References

[1] Homepage of the Live E! Project: http://www.live-e.org/, 2009.

[2] S. Matsuura, H. Ishizuka, H. Ochiai, S. Doi, S. Ishida, M. Nakayama, H. Esaki, and H. Sunahara., Live e! project: Establishment of infrastructure sharing environmental information, in *Applications and the Internet Workshops, 2007. SAINT Workshops 2007. International Symposium on*, Jan. 2007, pp. pp.67–67.