

Modeling and Generation of Space-Time Correlated Signals for Sensor Network Fields

Davide Zordan[§], Giorgio Quer^{§,*}, Michele Zorzi[§] and Michele Rossi[§]

[§]DEI, University of Padova, via Gradenigo 6/B – 35131, Padova, Italy

*University of California San Diego – La Jolla, CA 92093, USA.

Abstract—In the past few years, a large number of networking protocols for data gathering through aggregation, compression and recovery in Wireless Sensor Networks (WSNs) have utilized the spatio-temporal statistics of real world signals in order to achieve good performance in terms of energy savings and improved signal reconstruction accuracy. However, very little has been said in terms of suitable spatio-temporal models of the signals of interest. These models are very useful to prove the effectiveness of the proposed data gathering solutions as they can be used in the design of accurate simulation tools for WSNs. In addition, they can also be considered as reference models to prove theoretical results for data gathering algorithms. In this paper, we address this gap by devising a mathematical model for real world signals that are correlated in space and time. We thus describe a method to reproduce synthetic signals with tunable correlation characteristics and we verify, through analysis and comparison against large data sets from real world testbeds, that our model is accurate in reproducing the signal statistics of interest.

I. INTRODUCTION AND RELATED WORK

The temporal and spatial correlations are key statistical features of real signals, which are effectively exploited by many networking applications in the Wireless Sensor Network (WSN) domain. The temporal correlation captures the time evolution of the signal, making it possible to find appropriate sampling intervals for its accurate reconstruction. During these intervals, sensor nodes may go into a low power state, thus saving energy. The spatial correlation can instead be exploited in the deployment phase of, e.g., WSNs for environmental monitoring, to obtain suitable sensor densities as well as good deployment strategies [1]. Moreover, these features can be directly exploited in the design of networking protocols and signal compression techniques that make use of signal statistics. As an example, the authors of [2] design a distributed and collaborative Medium Access Control (MAC) protocol for WSNs that utilizes the spatial correlation of the monitored signal and exploits the fact that a sensor node can act as a representative node for other sensors in its neighborhood. [3] seeks to minimize the energy consumption of WSNs through the use of suitable spatio-temporal sampling rates. The objective of this work is to adapt the sleeping and spatial sampling behavior of the sensor nodes (readings from closely located sensors are almost equivalent when the signal is correlated in space) as a function of the signal statistics. The aim is to reduce the number of sensors that sample the signal per unit of time, while still allowing its accurate

reconstruction. The problem of designing a data gathering tree over a WSN is addressed, e.g., in [4], where the authors exploit the spatial correlation of the signal to design a proper gathering tree. [5] presents a system for data handling in WSNs, which takes into account the spatio-temporal statistics of the signal, incorporating long-term storage, multi-resolution data access and spatio-temporal pattern mining.

In addition, recent techniques for in-network aggregation and distributed data compression in WSNs use theoretical tools such as Compressive Sensing (CS) [6], [7]. In CS, convex optimization techniques exploit the sparsity of the data to achieve distributed compression. As shown in [8], [9], the spatio-temporal statistics of the signal can be used to design an ad hoc sparsification basis that allows the effective use of CS for reconstruction through random sampling.

Although many approaches in the WSN literature utilized the spatio-temporal characteristics of real world signals, very little attention has been paid to the definition of simple yet accurate models, including lightweight, fast and accurate tools for the reproduction of signals with the desired spatio-temporal statistics. We believe that these models are instead very much needed to prove performance limits of data gathering and distributed signal processing solutions, as well as to carry out their performance evaluation systematically.

In this paper, we address this gap by developing a framework to statistically characterize real world signals in space and time. This framework allows the accurate reproduction of the spatio-temporal behavior of such signals, obtaining synthetic models that can be effectively generated and subsequently used for protocol design and testing. These models are thus extensively validated against real world data, gathered from our indoor testbed at DEI [10], as well as climate data from [11]. The signal generation tool so obtained can be tuned to generate spatially and temporally correlated signals, where temporal and spatial correlations can be independently set. We remark that the usage of this tool is not limited to the field of WSN protocol optimization. As an example, the shadowing that affects radio transmissions can also be generated through our model. Note that in most previous papers only the spatial correlation has been kept into account, see [12].

Previous work to generate spatially correlated signals for WSNs [13] accounted for the spatial correlation of real world signals through variogram functions, by however neglecting the temporal correlation. Similar approaches can also be found in the field of geostatistics [14]. However, the studies within this field are mostly centered around finding optimal predictors and interpolators for spatio-temporal varying signals, e.g., Kriged Kalman filtering [15], rather than giving simple, fast

The work presented in this paper has been performed within the MOSAICS project, “MONitoring Sensor and Actuator networks through Integrated Compressive Sensing and data gathering”, a research project funded by the University of Padova under grant no. CPDA094077.

and accurate models for their generation. A further and very recent application is in cooperative cognitive radio sensing [16].

In this paper, we describe a general approach that can be used in the aforementioned applications. In detail, our main contributions are the following:

- we describe a method to generate synthetic signals with tunable correlation characteristics; in this way it is possible to extract the statistical characteristics from real datasets and use them to generate statistically similar signals;
- we verify with standard analytical techniques and through comparison against real and large data sets that these characteristics are preserved;
- we provide a useful simulation tool that can be applied in all the above fields [17].

The rest of the paper is organized as follows. In Section II we discuss the general correlation models for real world signals and introduce their properties. The model for space-time correlated signals is presented in detail in Section III and is validated against real world measurements in Section IV. Section V concludes our work.

II. CHARACTERIZATION OF SIGNALS FROM SENSOR NETWORK FIELDS

In this section we want to capture the relevant statistics from real environmental signals, in order to exploit this information for the generation of new samples with similar features. In the following, we present the details of our mathematical model for space-time correlated signals, introducing our notation, reported in Tab. I, as well as our basic assumptions. We focus on time varying two-dimensional (2D) fields of real-valued measurements; with x and y we indicate the space coordinates, whereas $\mathcal{D} = [-x_D, x_D] \times [-y_D, y_D]$ is the space domain. We consider that time is slotted where the slot time has a fixed duration $\Delta T > 0$. Thus, the time index is $t = i\Delta T$ with $i = 0, 1, 2, 3, \dots$, and the corresponding time domain is denoted by \mathcal{T} . A point in space $\mathbf{p} \in \mathcal{D}$ is indicated by the pair $\mathbf{p} = (x, y)$. With $z(\mathbf{p}, t) : \mathcal{D} \times \mathcal{T} \rightarrow \mathbb{R}$ we indicate the multidimensional random process that represents the space-time correlated random field (which is the objective of our analysis). When we fix a specific point with coordinates \mathbf{p}_o in space and t_o in time, $z(\mathbf{p}_o, t_o)$ represents a random variable (r.v.) describing the value of this field at the specific point considered, with mean $\mu_z(\mathbf{p}_o, t_o)$ and variance $\sigma_z^2(\mathbf{p}_o, t_o)$. We assume that z is a stationary random process (weak-sense stationarity) in both space and time, so that $\mu_z(\mathbf{p}_o, t_o) = \mu_z$ and $\sigma_z^2(\mathbf{p}_o, t_o) = \sigma_z^2, \forall \mathbf{p} \in \mathcal{D}$ and $\forall t \in \mathcal{T}$. Moreover, we assume that the correlation function of the considered signal is separable in the temporal correlation and the spatial correlation:

$$\rho(\mathbf{p}_1, t_1, \mathbf{p}_2, t_2) = \rho_S(\mathbf{p}_1, \mathbf{p}_2)\rho_T(t_1, t_2). \quad (1)$$

This is a general assumption used in the meteorology and geology fields, e.g., see [18], [19]. In detail, we consider the following models to study separately the temporal and spatial characteristics of the real signals.

| Symbol | Meaning |
|------------------------------|--|
| \mathcal{D} | Spatial domain |
| \mathcal{T} | Time domain |
| \mathcal{F} | Frequency domain |
| \mathbf{p} | Point in space $\mathbf{p} = (x, y) \in \mathcal{D}$ |
| $\boldsymbol{\omega}$ | Point in frequency $\boldsymbol{\omega} = (u, v) \in \mathcal{F}$ |
| $z(\mathbf{p}, t)$ | Space-time correlated field |
| $\rho_T(\Delta T)$ | Temporal correlation coefficient of z |
| d | Euclidean distance between two points in \mathcal{D} |
| $\rho_S(d)$ | Spatial correlation function of z between two points \mathbf{p}_1 and \mathbf{p}_2 with distance $d = \mathbf{p}_1 - \mathbf{p}_2 $ |
| $R_S(\boldsymbol{\omega})$ | 2D Fourier transform of $\rho_S(\mathbf{p})$ |
| $w(\mathbf{p}, t)$ | i.i.d. random Gaussian field at time t |
| $W(\boldsymbol{\omega}, t)$ | 2D Fourier transform of $w(\mathbf{p}, t)$ |
| $\varepsilon(\mathbf{p}, t)$ | i.i.d. random Gaussian noise field |

TABLE I
NOTATION USED IN THE ANALYSIS.

A. Spatial correlation

$\rho_S(\mathbf{p}_1, \mathbf{p}_2)$ is the spatial correlation function between any two points $\mathbf{p}_1, \mathbf{p}_2 \in \mathcal{D}$. Formally, $\forall t \in \mathcal{T}$:

$$\rho_S(\mathbf{p}_1, \mathbf{p}_2) = \frac{\text{cov}(z(\mathbf{p}_1, t), z(\mathbf{p}_2, t))}{\sigma_z(\mathbf{p}_1, t)\sigma_z(\mathbf{p}_2, t)}, \quad (2)$$

where $\text{cov}(\cdot)$ is the covariance function. The weak-sense stationarity assumption made to define our multidimensional process implies also the correlation stationarity, i.e., for the spatial correlation, if the Euclidean distance $d_{1,2} = |\mathbf{p}_1 - \mathbf{p}_2|$ is equal to $d_{3,4} = |\mathbf{p}_3 - \mathbf{p}_4|$, then $\rho_S(\mathbf{p}_1, \mathbf{p}_2) = \rho_S(\mathbf{p}_3, \mathbf{p}_4)$. For this reason, the spatial correlation function can be defined as a function of the distance between two points, that is a scalar $d \in [0, d_M]$, where $d_M = \sqrt{(2x_D)^2 + (2y_D)^2}$, i.e., the maximum distance between two points in \mathcal{D} . In the following, with an abuse of notation we can write $\rho_S(\mathbf{p}_1, \mathbf{p}_2) = \rho_S(d_{1,2})$, without loss of generality.

Moreover, there are other issues to be considered when studying a real signal. Since the signal is sensed only at specific locations, we can not evaluate the process $z(\mathbf{p}, t)$ at every point $\mathbf{p} \in \mathcal{D}$, but we can observe the process only at the points where the signal is sensed, i.e., $\mathbf{p}_i \in \mathcal{D}$. This translates to the fact that also $\rho_S(d)$ may not be defined $\forall d \in [0, d_M]$. Thus, in the literature there exist many models to capture the spatial correlations for real signals [20]. In this paper we consider the following two models:

Power Exponential (PE) model: the spatial correlation $\rho_S(d)$ is modeled with the function:

$$\tilde{\rho}_{S_{PE}}(d) = \exp\{-(d/\zeta)^\nu\}, \quad \text{for } 0 < \nu \leq 2; \quad (3)$$

Rational Quadratic (RQ) model: the spatial correlation $\rho_S(d)$ is modeled with the function:

$$\tilde{\rho}_{S_{RQ}}(d) = \frac{1}{1 + (d/\zeta)^\nu S_\nu}, \quad \text{for } 0 < \nu \leq 2; \quad (4)$$

Both models above depend on the parameters ζ and ν , that are the correlation length and the order of the function, respectively, while $S_\nu = 20^{1/\nu} - 1$ is a scaling factor. The two algorithms set ζ and ν to best fit the correlation of the real signal considered, that is defined only for a finite set of distances d_j , with $j = 1, \dots, J$. The best fit is obtained minimizing the Root Mean Square Error (RMSE), i.e.,

$$\xi_F = \sqrt{\frac{1}{J} \sum_{j=1}^J (\tilde{\rho}_{S_M}(d_j) - \rho_S(d_j))^2}, \quad (5)$$

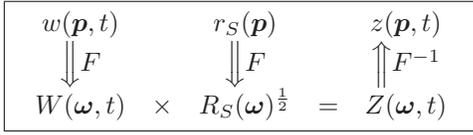


Fig. 1. Diagram for the 2D filtering procedure.

where $F = PE$ for the PE model and $F = RQ$ for the RQ model.

B. Temporal correlation

$\rho_T(t_1, t_2)$ is the spatial correlation function between any two time samples $t_1, t_2 \in \mathcal{T}$. Formally, $\forall \mathbf{p} \in \mathcal{D}$:

$$\rho_T(t_1, t_2) = \frac{\text{cov}(z(\mathbf{p}, t_1), z(\mathbf{p}, t_2))}{\sigma_z(\mathbf{p}, t_1)\sigma_z(\mathbf{p}, t_2)}. \quad (6)$$

Similar to the case of the spatial correlation, also in this case, as a consequence of the weak-sense stationarity assumption, the stationarity in the temporal correlation holds, i.e., we can write with an abuse of notation $\rho_T(t_1, t_2) = \rho_T(t_2 - t_1) = \rho_T(i\Delta T)$, where $i\Delta T = t_2 - t_1$. Since we aim at generating a synthetic signal with an autoregressive model, as will be detailed in the next section, we only consider the one step time correlation, i.e., we are only interested in calculating the coefficient $\rho_T(\Delta T)$.

III. MODEL FOR SPACE-TIME CORRELATED SIGNALS

Our objective is to devise a suitable and tractable model for the generation of the colored signal $z(\mathbf{p}, t)$. Specifically, we want to obtain a dynamic model, evolving at discrete time instants and thus allowing the generation of $z(\mathbf{p}, i\Delta T)$ for $i = 0, 1, 2, 3, \dots$. This model should retain, as much as possible, the correlation characteristics of the original signal that we sample from the sensor network field. Next, we present a suitable method for the generation of such signal, whereas its accuracy is demonstrated later on in Section IV by comparing the generated signals against real data measurements.

The correlated signal $z(\mathbf{p}, t)$ is attained through the following procedure:

- S1 At time $t_o = 0$, we generate an i.i.d. random Gaussian field $w(\mathbf{p}, t_o) : \mathcal{D} \times \mathcal{T} \rightarrow \mathbb{R}$, which for any specific location is a Gaussian r.v. with zero mean and unit variance, $\mathcal{N}(0, 1)$. $w(\mathbf{p}, t_o)$ is a stationary process (strict sense) by construction.
- S2 $z(\mathbf{p}, t_o)$ is obtained by coloring $w(\mathbf{p}, t_o)$ through a 2D filtering procedure. In detail, we first obtain $W(\boldsymbol{\omega}, t_o)$ from $w(\mathbf{p}, t_o)$ using the 2D Fourier transform $F[\cdot]$, i.e., $W(\boldsymbol{\omega}, t_o) = F[w(\mathbf{p}, t_o)]$, where $\boldsymbol{\omega} = (u, v) \in \mathcal{F}$ and \mathcal{F} is the frequency domain. Given the reference point $\mathbf{p}_o = (0, 0)$, for any point $\mathbf{p} \in \mathcal{D}$ we enforce the correlation $\rho_S(|\mathbf{p} - \mathbf{p}_o|)$ and compute its Fourier transform, $R_S(\boldsymbol{\omega}) = F[\rho_S(|\mathbf{p} - \mathbf{p}_o|)]$.
- S3 Thus, we multiply $W(\boldsymbol{\omega}, t_o)$ by $R_S(\boldsymbol{\omega})^{1/2}$ and compute the inverse Fourier transform of the result so obtained to attain the wanted colored random field $z(\mathbf{p}, t_o)$. Note that z is still Gaussian and stationary. This procedure is sketched in the diagram of Fig. 1. This (S1–S3) is a known method for coloring a random Gaussian field, whose proof as well as further mathematical details can

be found in [21]. The field z so obtained is correlated in space and its spatial correlation can be controlled through a *valid* correlation function $\rho_S(|\mathbf{p} - \mathbf{p}_o|)$ [20].

- S4 In order to enforce a temporal correlation as well, we adopt an autoregressive filter as follows:

$$w(\mathbf{p}, t_o + \Delta T) = \rho_T(\Delta T)w(\mathbf{p}, t_o) + \sqrt{1 - \rho_T(\Delta T)^2}\varepsilon(\mathbf{p}, t_o + \Delta T), \quad (7)$$

where $\varepsilon(\mathbf{p}, t_o + \Delta T)$ is an i.i.d. random Gaussian noise, $\mathcal{N}(0, 1)$. $\rho_T(\Delta T)$ is a temporal correlation coefficient that we enforce in the model. Thus, the procedure S1–S4 is iterated for all future time steps to calculate $w(\mathbf{p}, t_o + i\Delta T)$ from $w(\mathbf{p}, t_o + (i-1)\Delta T)$, for $i = 1, 2, 3, \dots$. Note that $\rho_T(\Delta T)$ is obtained from actual field measurements, as further discussed in Section IV. Also, note that $w(\mathbf{p}, t)$ is still i.i.d. in the space domain, whereas thanks to (7), it is time-correlated with correlation coefficient $\rho(\Delta T)$. As we demonstrate shortly, $w(\mathbf{p}, t)$ is again a stationary and Gaussian random process which maintains the same mean and variance of the original process $w(\mathbf{p}, t_o)$.

In what follows, we show that the autoregressive model that we superimpose to the spatially correlated signal preserves the statistical properties of the random field $w(\mathbf{p}, t)$, i.e., the autoregressive filtering is stable and the signal that it generates is time-stationary. To this end, we need to show that both mean and variance of $w(\mathbf{p}, t)$ are preserved using (7), and we do it using standard statistical techniques for continuous time processes, e.g., see [22].

Conservation of the mean of $w(\mathbf{p}, t)$. For the mean, we have that:

$$\mu_w(\mathbf{p}, t + \Delta T) = \mathbb{E}[w(\mathbf{p}, t + \Delta T)] = 0, \quad (8)$$

that is obtained using (7) and the linearity of the expectation. It is easy to verify (8) for $t = t_o$, then the result is proven inductively for all time steps. Finally, note that this result holds $\forall \mathbf{p} \in \mathcal{D}$.

Conservation of the variance of $w(\mathbf{p}, t)$. For the variance, we have:

$$\begin{aligned} \sigma_w^2(\mathbf{p}, t + \Delta T) &= \\ &= \mathbb{E}[(w(\mathbf{p}, t + \Delta T) - \mu_w(\mathbf{p}, t + \Delta T))^2] = 1, \end{aligned} \quad (9)$$

where we use the definition of autoregressive filter in (7) and a mathematical reasoning similar to the one used in (8). Again, the result in (9) holds for $t = t_o$, so we have that $\sigma_w^2(\mathbf{p}, t_o + \Delta T) = 1$ by construction, and the result $\forall t \in \mathcal{T}$ follows inductively.

Conservation of the temporal correlation of $z(\mathbf{p}, t)$. In what follows, we prove that the random process $z(\mathbf{p}, t)$ that we obtain through steps S1–S4 is also correlated in time and that its correlation coefficient is $\rho(\Delta T)$ at all time instants and for all points $\mathbf{p} \in \mathcal{D}$. Without loss of generality (the result also holds in the continuous case), we consider our spatial signal as being sampled from a rectangular sensor grid of $N \times M$ points evenly spaced in \mathcal{D} . Owing to this assumption, let the space point $\mathbf{p} = (x, y)$ be defined with $x \in \{1, 2, \dots, N\}$ and $y \in \{1, 2, \dots, M\}$. Also, let us define $r_S(\mathbf{p})$ as:

$$r_S(\mathbf{p}) = F^{-1} \left[F[\rho_S(|\mathbf{p} - \mathbf{p}_o|)]^{\frac{1}{2}} \right] = F^{-1} \left[R_S(\boldsymbol{\omega})^{\frac{1}{2}} \right]. \quad (10)$$

$$\begin{aligned}
& \text{cov}(z(\mathbf{p}, t), z(\mathbf{p}, t + \Delta T)) \stackrel{(a)}{=} \mathbb{E}[z(\mathbf{p}, t)z(\mathbf{p}, t + \Delta T)] \\
& \stackrel{(b)}{=} \mathbb{E} \left[\sum_{i=1}^N \sum_{j=1}^M w(i, j, t) r_S(x - i, y - j) \sum_{k=1}^N \sum_{q=1}^M w(k, q, t + \Delta T) r_S(x - k, y - q) \right] \\
& \stackrel{(c)}{=} \mathbb{E} \left[\sum_{i=1}^N \sum_{j=1}^M w(i, j, t) r_S(x - i, y - j) \sum_{k=1}^N \sum_{q=1}^M \left(\rho_T(\Delta T) w(k, q, t) + \sqrt{1 - \rho_T(\Delta T)^2} \varepsilon(k, q, t + \Delta T) \right) r_S(x - k, y - q) \right] \\
& \stackrel{(d)}{=} \sum_{i,k=1}^N \sum_{j,q=1}^M \rho_T(\Delta T) \mathbb{E} \left[w(i, j, t) w(k, q, t) \right] + \sqrt{1 - \rho_T(\Delta T)^2} \mathbb{E} \left[w(i, j, t) \varepsilon(k, q, t + \Delta T) \right] r_S(x - i, y - j) r_S(x - k, y - q) \\
& \stackrel{(e)}{=} \rho_T(\Delta T) \sigma_w^2 \sum_{i=1}^N \sum_{j=1}^M r_S(x - i, y - j)^2
\end{aligned} \tag{13}$$

$$\begin{aligned}
\sigma_z(\mathbf{p}, t) \sigma_z(\mathbf{p}, t + \Delta T) & \stackrel{(a)}{=} \mathbb{E}[z(\mathbf{p}, t)^2] = \mathbb{E} \left[\sum_{i=1}^N \sum_{j=1}^M w(i, j, t) r_S(x - i, y - j) \sum_{k=1}^N \sum_{q=1}^M w(k, q, t) r_S(x - k, y - q) \right] \\
& = \sum_{i=1}^N \sum_{j=1}^M \mathbb{E} [w(i, j, t)^2] r_S(x - i, y - j)^2 = \sigma_w^2 \sum_{i=1}^N \sum_{j=1}^M r_S(x - i, y - j)^2
\end{aligned} \tag{14}$$

From this definition and the diagram of Fig. 1, we see that $z(\mathbf{p}, t)$ is obtained in the domain \mathcal{D} as the spatial convolution between $w(\mathbf{p}, t) = w(x, y, t)$ and $r_S(\mathbf{p}) = r_S(x, y)$, i.e.,

$$z(\mathbf{p}, t) = \sum_{i=1}^N \sum_{j=1}^M w(i, j, t) r_S(x - i, y - j), \tag{11}$$

where $i \in \{1, 2, \dots, N\}$ and $j \in \{1, 2, \dots, M\}$. We now compute the mean of process $z(\mathbf{p}, t)$, that is $\mu_z(\mathbf{p}, t) = \mathbb{E}[z(\mathbf{p}, t)]$:

$$\begin{aligned}
\mu_z(\mathbf{p}, t) & = \mathbb{E} \left[\sum_{i=1}^N \sum_{j=1}^M w(i, j, t) r_S(x - i, y - j) \right] \\
& = \sum_{i=1}^N \sum_{j=1}^M \mathbb{E} [w(i, j, t)] r_S(x - i, y - j) = 0,
\end{aligned} \tag{12}$$

where the result follows from the linearity of the expectation, the fact that $r_S(\mathbf{p})$ is a deterministic function (as it is directly derived from the spatial correlation function, which is a known quantity) and the result in (8).

We are now ready to calculate the numerator of (6), which is given by (13): in this equation, equality (a) follows from (12), whereas (b) follows from (11) and (c) follows from (7). Also, (d) follows as the temporal correlation coefficient $\rho_T(\Delta T)$ and $r_S(\mathbf{p})$ are a constant and a deterministic function, respectively. For (e), note that $\mathbb{E}[w(i, j, t) \varepsilon(k, q, t + \Delta T)] = 0$, as process ε is independent of w by construction and its mean is zero. Moreover, when indices $i \neq k$ or $j \neq q$, we have $\mathbb{E}[w(i, j, t) w(k, q, t)] = \mathbb{E}[w(i, j, t)] \mathbb{E}[w(k, q, t)] = 0$, since $w(\mathbf{p}, t)$ is i.i.d. in the space domain and $\mathbb{E}[w(\mathbf{p}, t)] = 0$, $\forall \mathbf{p} \in \mathcal{D}$. Thus, for the only non zero case in which $i = k$ and $j = q$, we have $\mathbb{E}[w(i, j, t) w(k, q, t)] = \mathbb{E}[w(i, j, t)^2] = \sigma_w^2$.

The denominator of (6) is computed in (14). From (11), it follows that the variance of $z(\mathbf{p}, t)$ does not depend on t , since $w(\mathbf{p}, t)$ is stationary and $r_S(\mathbf{p})$ is a deterministic mapping function that only depends on the spatial correlation. Hence,

equality (a) follows as $\sigma_{z(\mathbf{p}, t)} = \sigma_{z(\mathbf{p}, t + \Delta T)}$ and $\mu_z(\mathbf{p}, t) = 0$, see (12). The remaining equalities follow using the same reasonings as those done above for Eq. (13).

Taking the ratio of (13) and (14), we have proved the conservation of the temporal correlation, $\forall t \in \mathcal{T}$ and $\forall \mathbf{p} \in \mathcal{D}$.

Discussion. The method presented in S1–S4 above allows one to obtain a stationary signal which is correlated in space and time according to an arbitrary spatial correlation function $\rho_S(\mathbf{p})$ and a temporal correlation coefficient $\rho_T(\Delta T)$. These parameters are tunable and can be fit to any real data set.

From (12) we see that the random process so obtained has zero mean; however, we observe that any non-zero mean can be enforced and all the results obtained here still hold.

Finally, from (14) and recalling that $\sigma_w^2 = 1$, we see that the variance of $z(\mathbf{p}, t)$ equals $\sigma_z^2(\mathbf{p}, t) = \sum_{i=1}^N \sum_{j=1}^M r_S(x - i, y - j)^2$. Thus, the variance follows directly from the spatial correlation properties of the signal. However, we can tune the amplitude of this variance through the multiplication of $z(\mathbf{p}, t)$ by a constant. Also in this case the properties discussed above remain unchanged.

In the model above we do not keep into account the mean and variance characteristics of the signal. From our previous discussion, we see that these can be promptly accounted for by our model in case the application is also influenced by them.

IV. RESULTS

In this section we present the real datasets that we have used to validate our model. We want to show that the signal model of Section III can effectively capture the correlation characteristics of real signals. To this end, we first compute the correlation characteristics of the real data. These are thus used to tune the spatial correlation models of Section II and obtain the temporal correlation coefficient, which are subsequently used with the procedure of Section III to obtain the synthetic signals with the desired correlation properties. These are

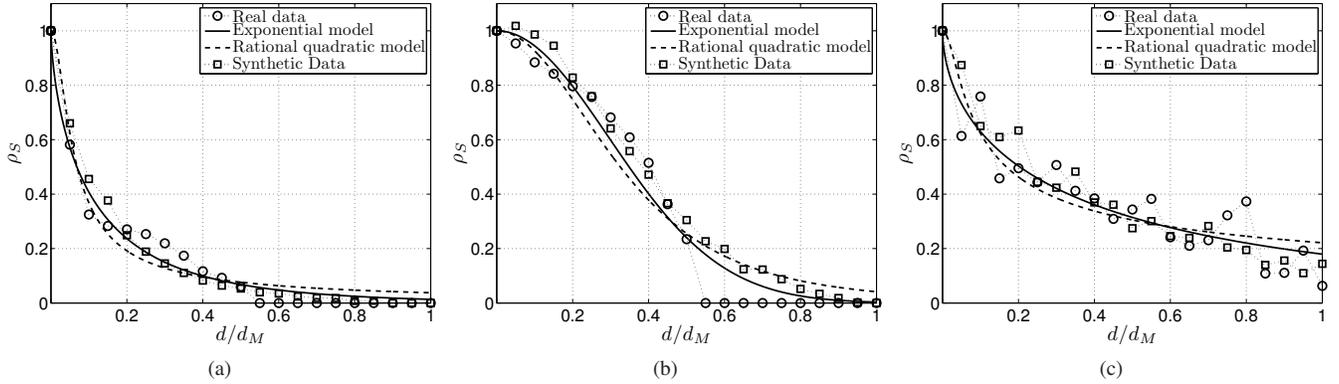


Fig. 2. Spatial correlation for the real data, the two correlation fitting models (PE and RQ), and the correlation of the synthetic data generated according to the best fitting model, for the following signals: (a) global precipitation, CP; (b) global temperature, CT; and (c) indoor luminosity, DL₁.

finally examined to check how well they can reproduce the correlation characteristics of the real data sets.

A. Real Signals analyzed

We hereby consider two types of real signals, both available online. The former is a dataset of global climate data from the Center for Climatic Research (CCR), Department of Geography, University of Delaware, available at [11]. This dataset is obtained interpolating the observations from about 10^4 climatic stations, and sampling uniformly the sensed field into a total of 85794 different points. The maximum distance between any two points is $d_M \simeq 2 \cdot 10^4$ km, while the time step is $\Delta T = 1$ month. Within the large set of data available in the website, we selected eight representative signals, namely, Air Temperature (CT), Total Precipitation (CP), EvapoTranspiration (CE), the Difference among Precipitation and EvapoTranspiration (CD), Snow Cover (CC), Snow Melt (CM), Soil Moisture (CS) and Moisture Indices (CI). For more information about these signals, see [11].

For the latter dataset, we considered a number of indoor signals gathered from the environmental monitoring WSN testbed deployed on the ground floor of the Department of Information Engineering (DEI), University of Padova, Italy [10]. The data from this dataset were sampled from $N = 68$ IEEE 802.15.4-compliant TelosB wireless nodes deployed according to an irregular topology, with maximum distance among any two sensors of $d_M \simeq 36$ m. The time step in this case is $\Delta T = 360$ s. These sensors can measure five different signals: temperature, humidity, luminosity in two different ranges (DL₁ : 320 – 730 and DL₂ : 320 – 1100 nm), and their battery voltage. For the performance analysis we considered temperature (DT) and humidity (DH), which have high spatial and temporal correlation, and luminosity in the two ranges (DL₁ and DL₂).

For each signal in both CCR and DEI databases, we considered about 500 time samples, we calculated the spatial correlation function using (2) and the one step temporal correlation coefficient using (6). Thus, we fit the spatial correlation of each signal through the PE model of (3) and the RQ model of (4). In the columns of Tab. II we show: 1. all the signals considered, 2-7. the parameters inferred and the values of the corresponding RMSE for the two correlation models (PE and

| Signal | Spatial Correlation | | | | | | Temporal Corr. $\rho_T(\Delta T)$ |
|-----------------|---------------------|-------|--------------|----------|-------|--------------|-----------------------------------|
| | PE Model | | | RQ Model | | | |
| | ζ | ν | ξ_{PE} | ζ | ν | ξ_{RQ} | |
| CD | 0,12 | 0,65 | 0,042 | 0,91 | 0,47 | 0,058 | 0,85 |
| CP | 0,11 | 0,68 | 0,034 | 0,71 | 0,52 | 0,050 | 0,90 |
| CI | 0,25 | 0,96 | 0,046 | 1,05 | 0,71 | 0,073 | 0,88 |
| CM | 0,06 | 0,53 | 0,040 | 0,49 | 0,51 | 0,034 | 0,27 |
| CE | 0,24 | 1,00 | 0,057 | 0,87 | 0,80 | 0,085 | 0,85 |
| CT | 0,42 | 2,00 | 0,073 | 0,94 | 2,00 | 0,103 | 0,98 |
| CC | 0,31 | 0,96 | 0,121 | 1,62 | 0,60 | 0,117 | 0,99 |
| CS | 0,05 | 0,62 | 0,022 | 0,33 | 0,60 | 0,019 | 0,94 |
| DT | 272 | 0,18 | 0,119 | 12916 | 0,13 | 0,146 | 0,99 |
| DH | 11640 | 0,31 | 0,008 | 9273 | 0,17 | 0,024 | 0,99 |
| DL ₁ | 0,38 | 0,56 | 0,081 | 25,46 | 0,23 | 0,095 | 0,98 |
| DL ₂ | 0,43 | 0,58 | 0,081 | 28,47 | 0,23 | 0,089 | 0,99 |

TABLE II
CORRELATION OF REAL SIGNALS

RQ), and 8. the value of the temporal correlation $\rho_T(\Delta T)$ estimated from the real data. Boldface text is used in the table to indicate the best fitting model for each signal.

In order to calculate the spatial correlation (shown in Fig. 2 and discussed shortly), we had to consider each pair of points in the dataset. Since this number for the CCR dataset is very large, we picked at random only a subset of the total number of pairs (10^5 pairs, randomly selected, were considered for the results in this paper), while for the DEI dataset we considered all possible pairs. For each pair of points p_i and p_j , we calculated the distance $d_{i,j}$ and the spatial correlation $\rho_S(d_{i,j})$, using (2). For representation purposes, we considered the maximum distance d_M and we divided it into 20 intervals. For each interval and for each pair of points whose distance falls within such interval, we calculated the average spatial correlation. In this way, we obtained one spatial correlation value for each interval (associating it with the center of the interval), similarly to the procedure adopted in [13] for the variogram calculation.

B. Model Validation

In Fig. 2 we show on the y -axis the spatial correlation $\rho_S(d)$ for three selected signals *vs* the normalized distance $d/d_M \in [0, 1]$. For comparison, in this figure we plot the empirical correlation obtained from the real data, the auto-correlation function obtained from the PE ($\tilde{\rho}_{S_{PE}}(d)$) and RQ ($\tilde{\rho}_{S_{RQ}}(d)$) models, as well as the spatial correlation obtained

| Signal | Model | ξ_F | ξ_S | $\Delta\rho_T$ |
|-----------------|-------|---------|---------|----------------|
| CD | PE | 0,042 | 0,021 | 3,92E-3 |
| CP | PE | 0,034 | 0,021 | 2,71E-3 |
| CI | PE | 0,046 | 0,030 | 1,18E-2 |
| CM | RQ | 0,034 | 0,035 | 3,82E-3 |
| CE | PE | 0,057 | 0,040 | 2,60E-3 |
| CT | PE | 0,073 | 0,035 | 2,16E-3 |
| CC | RQ | 0,117 | 0,069 | 2,14E-6 |
| CS | RQ | 0,019 | 0,031 | 2,60E-4 |
| DT | PE | 0,119 | 0,560 | 1,39E-4 |
| DH | PE | 0,008 | 0,542 | 1,16E-5 |
| DL ₁ | PE | 0,081 | 0,040 | 6,62E-4 |
| DT ₂ | PE | 0,081 | 0,072 | 2,24E-4 |

TABLE III
CORRELATION OF SYNTHETIC SIGNALS

from the synthetic signal. Note that the synthetic signals are generated through the procedure of Section III, using the best fitting model among PE and RQ, in terms of ξ_F . As shown in this figure, the chosen correlation models both nicely fit empirical correlation values. Moreover, the spatial correlation obtained from the synthetic signal is also very close to the empirical one.

In Figs. 2-(a) and 2-(b) we respectively show ρ_S for the Total Precipitation (CP) and the Air Temperature (CT) signals from the CCR dataset. In Fig. 2-(c) we plot ρ_S for the Luminosity signal (DL₁) from the DEI dataset. From these figures we see that our model can very nicely fit the spatial characteristics of the real signals for both PE and RQ models. Furthermore, with the proposed method we are able to generate a synthetic signal that also follows with good accuracy the real signal correlation. Tab. III shows the results of the fitting of the synthetic signal for all the considered datasets. In particular, in the columns of this table we represent: 1. the considered signal, 2. the best fitting model among PE and RQ, 3. the corresponding RMSE ξ_F for the best model, 4. ξ_S , the RMSE between the chosen fitting model and the spatial correlation of the synthetic signal, 5. the relative error $\Delta\rho_T$ among the temporal correlation $\rho_T(\Delta T)$ and $\tilde{\rho}_T(\Delta T)$, of the real and the synthetic signal, respectively. $\Delta\rho_T$ is calculated as follows:

$$\Delta\rho_T = \frac{|\rho_T(\Delta T) - \tilde{\rho}_T(\Delta T)|}{\rho_T(\Delta T)}. \quad (15)$$

From Tab. III, we observe that for all signals: (a) $\Delta\rho_T$ is sufficiently small, so we represent with high accuracy the temporal correlation, (b) both fitting models PE and RQ accurately reproduce the spatial correlation of the real signals. Furthermore, (c) the synthetic data very nicely follows the real spatial characteristics for all the CCR signals and for the DEI luminosity signals DL₁ and DL₂, (d) we introduce some error while representing the spatial characteristics for the DEI signals DT and DH. This is due to the fact that these two signals have a correlation length ζ , see Tab. II, that is very large compared to the size of the network d_M . In this case, we are not able to reproduce with high accuracy the actual correlation of the signal due to border effects.

V. CONCLUSIONS

In this paper we have presented a model for the statistical characterization of real world signals that are correlated in space and time. Our model allows the efficient generation of

synthetic signals with the desired correlation properties, where spatial and temporal correlations can be independently set and fit to those of the real signals of interest. The accuracy of the proposed model has been verified through comparison against real data sets from large sensor testbeds. Future extensions of our work include the application of our statistical model to other types of signals, such as sensor data from smart grids, e.g., to model the distributed space and time dependent energy production and consumption process.

REFERENCES

- [1] B. Wang, K. C. Chua, V. Srinivasan, and W. Wan, "Information Coverage in Randomly Deployed Wireless Sensor Network," *IEEE Transactions on Wireless Communications*, vol. 6, no. 8, pp. 2994–3003, Aug. 2007.
- [2] M. C. Vuran and I. F. Akyildiz, "Spatial Correlation-Based Collaborative Medium Access Control in Wireless Sensor Networks," *IEEE/ACM Trans. on Networking*, vol. 14, no. 2, pp. 316–329, Apr. 2006.
- [3] S. Bandyopadhyay, Q. Tian, and E. J. Coyle, "Spatio-temporal sampling rates and energy efficiency in wireless sensor networks," *IEEE/ACM Trans. on Networking*, vol. 13, no. 6, pp. 1339–1352, Dec. 2005.
- [4] Y. Yu, B. Krishnamachari, and V. K. Prasanna, "Data Gathering with Tunable Compression in Sensor Networks," *IEEE Trans. on Parallel and Distributed Systems*, vol. 19, no. 2, pp. 276–287, Feb. 2008.
- [5] D. Ganesan, D. Estrin, and J. Heidemann, "DIMENSIONS: Why do we need a new data handling architecture for sensor networks?" in *ACM Workshop on Hot Topics in Networks*, Princeton, NJ, US, Oct. 2002, pp. 143–148.
- [6] D. Donoho, "Compressed sensing," *IEEE Trans. on Information Theory*, vol. 52, no. 4, pp. 4036–4048, Apr. 2006.
- [7] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [8] R. Masiero, G. Quer, D. Munaretto, M. Rossi, J. Widmer, and M. Zorzi, "Data Acquisition through Joint Compressive Sensing and Principal Component Analysis," in *IEEE Globecom 2009*, Honolulu, Hawaii, US, Nov.-Dec. 2009.
- [9] G. Quer, D. Zordan, R. Masiero, M. Zorzi, and M. Rossi, "WSN-Control: Signal Reconstruction through Compressive Sensing in Wireless Sensor Networks," in *IEEE SenseApp Workshop*, Denver, CO, US, Dec. 2010, pp. 937–944.
- [10] R. Crepaldi, S. Friso, A. F. Harris III, M. Mastrogiovanni, C. Petrioli, M. Rossi, A. Zanella, and M. Zorzi, "The Design, Deployment, and Analysis of SignetLab: A Sensor Network Testbed and Interactive Management Tool," in *IEEE Tridentcom*, Orlando, FL, US, May 2007.
- [11] C. J. Willmott and K. Matsuura, "Global Climate Resource Pages," Aug. 2009. [Online]. Available: <http://climate.geog.udel.edu/~climate/>
- [12] D. Catrein and R. Mathar, "Gaussian Random Fields as a Model for Spatially Correlated Log-Normal Fading," in *ATNAC*, Adelaide, South Australia, Dec. 2008.
- [13] A. Jindal and K. Psounis, "Modeling Spatially Correlated Data in Sensor Networks," *ACM Transactions on Sensor Networks*, vol. 2, no. 4, pp. 466–499, Nov. 2006.
- [14] N. Cressie, *Statistics for Spatial Data*. John Wiley, 1993.
- [15] C. K. Wikle and N. Cressie, "A dimension-reduced approach to space-time Kalman filtering," *Biometrika*, vol. 86, no. 4, pp. 815–829, 1999.
- [16] S.-J. Kim, E. Dall'Anese, and G. B. Giannakis, "Cooperative Spectrum Sensing for Cognitive Radios Using Kriged Kalman Filtering," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 1, pp. 24–36, Feb. 2011.
- [17] D. Zordan, G. Quer, and M. Rossi, "MATLAB Implementation of the Space-Time correlated signals generator," May 2010. [Online]. Available: <http://www.dei.unipd.it/~rossi/software.html>
- [18] P. Abrahamsen, "A review of Gaussian random fields and correlation functions," Norwegian Computing Center, Oslo, Tech. Rep., 1997.
- [19] P. Whittle, "On Stationary processes in the plane," *Biometrika*, vol. 41, no. 3-4, pp. 434–449, 1954. [Online]. Available: <http://biomet.oxfordjournals.org/content/41/3-4/434.short>
- [20] B. Matérn, "Spatial variation. Stochastic models and their application to some problems in forest surveys and other sampling investigations," *Meddelanden fran Statens Skogsforskningsinstitut*, vol. 49, no. 5, 1960.
- [21] A. M. Yaglom, *An Introduction to the Theory of Stationary Random Functions*. Dover Phoenix Editions, 1962.
- [22] N. Benvenuto and G. Cherubini, *Algorithms for Communications Systems and their Applications*. Wiley, 2002.