A Stochastic Differential Equation Model for Spectrum Utilization and Its Verification Using Spectrum Measurement

Romesh Saigal, Yang Li
Dept. of Industrial and
Operations Engineering
University of Michigan
Ann Arbor, MI 48109
{rsaigal, youngli}@umich.edu

Mingyan Liu
Dept. of Electrical Engineering
and Computer Science
University of Michigan
Ann Arbor, MI 48109-2122
mingyan@eecs.umich.edu

Dawei Chen
Dept. of Computer Science
Hong Kong University of
Science and Technology
Hong Kong, China
{dwchen, qianzh}@cse.ust.hk

***Abstract*—Dynamic spectrum access has been a subject of extensive study in recent years. The increasing volume of literature calls for better understanding of the characteristics of current spectrum utilization as well as better tools for analysis. Toward this goal, a number of measurement studies have been conducted recently. They revealed previously little known features of spectrum utilization. However, to date a good spectrum utilization model (or more precisely perhaps, a good primary user model) has not emerged. There remains a gap between what we have learned from measurement studies and theoretical studies on dynamic spectrum access – the latter typically rely on overly simplified channel models, e.g., the two-state Markov (or Gilbert-Elliot) model. In this paper we present a stochastic differential equation based wireless spectrum utilization model that accurately captures the dynamic changes in the channel condition induced by the primary users' activities. Salient features of this model include: (1) it is based on fundamental physics of scattering and multi-path fading of the primary users' signals; (2) it is verified using real spectrum measurement data; (3) it is in closed form, and can thus be used in theoretical studies on optimal spectrum access and sharing; and (4) a desirable by-product of this model is a straightforward procedure to generate synthetic spectrum data that can subsequently be used to create a realistic spectrum environment simulating the presence of primary users' activities, within which spectrum sensing and access schemes may be evaluated.**

I. Introduction

With ever-increasing demand for wireless spectrum, it is quickly becoming a scarce resource. This makes the concept of dynamic spectrum access (as opposed to static allocation) highly desirable. It is also made feasible by advances in software defined radio (SDR) [1] and cognitive radio (CR) [2]. The basic idea is for wireless devices to quickly detect spectrum availability and exploit instantaneous opportunities for its own transmission needs.

This concept is particularly relevant for the licensed bands with one or more primary, licensed users. With spectrum agility, secondary unlicensed users may be allowed to access licensed bands provided that their activities do not cause excessive disruption to the primary users, or that they only operate during inactive periods of the latter. A good example is the licensed TV bands, to which the FCC has recently opened access.

Accordingly, there have been extensive studies on both technical and policy issues related to dynamic (and open) spectrum access. With this comes the need for better understanding of current spectrum utilization, especially in the presence of licensed primary users. Existing studies often rely on overly simplified spectrum models (e.g., the commonly used two-state Markov model, also known as the Gilber-Elliot model) that may not accurately capture the activities of primary users.

This motivated a series of spectrum measurement studies published recently, see e.g., [3], [4], [5], [6], [7], [8]. These studies have all revealed important features of wireless spectrum utilization and its dynamic changes over time. Despite these studies and the various statistics obtained therein, there remains a gap between what we have learned from measurement studies and how to use it in designing and evaluating the performance of various spectrum access schemes, especially in the presence of primary users. In particular, these measurement studies have not in general led to tools that can generate realistic spectrum utilization as a time process in simulations to evaluate spectrum sensing and access algorithms. In [9] a sequence of probability distributions of spectrum availability were derived using measurement data. However, these distributions captures the average behavior of spectrum rather than describing spectrum activity as a stochastic process as we aim to achieve in this paper.

The goal of this paper is to develop a comprehensive spectrum utilization model that can be used for analysis and for generating synthetic spectrum utilization process over time. To that end, we present a stochastic differential equation (SDE) based spectrum utilization model that accurately captures the dynamic changes in spectrum utilization due to primary users' activities. Specifically, our model (expressed as an SDE of the total measured power in a frequency band as a function of time) is shown to be a Radial Ornstein–Uhlenbeck process [10], specified by three parameters that represent the average power level in the frequency band, the volatility/drift of the

process (how much the mean changes over time), and the rate at which signal phases change over time, respectively.

We show how these parameters may be estimated if training data is available, and discuss how to obtain their values otherwise. We also verify the accuracy of this model using spectrum measurement data. Because this model takes on a relatively simple closed form, it can be directly used in analytical and optimization studies of opportunistic spectrum access. It also opens up new directions and opportunities as this model is continuous in time and can be used to describe non-stationary random process; these features are quite different from common assumptions.

Last but not least, we show how to use this model to straightforwardly generate *synthetic* spectrum utilization data in the form of a random process. In this sense our stochastic differential equation model is analogous to mobility models [11] that can be used to generate a mobile environment for the study and in particular the simulation of ad hoc networks.

To summarize, our model provides a way of generating a wireless spectrum environment, in the form of realistic background noise due to primary users' activities, in which spectrum sensing and access algorithms may be studied. In particular, it can be used to generate realistic spectrum environment in a TV band, which is highly desirable in a simulated or emulated environment. To the best of our knowledge, this is the first model of its kind in the literature.

The main contributions of this paper are as follows:

1) We present a stochastic differential equation model based on the physics of scattering and multi-path fading to capture the received energy (by a secondary user) process in the presence of a primary user's spectrum utilization.
2) We verify the accuracy of the model using spectrum measurement data and provide guideline on how the parameter values may be selected in the case of generating synthetic data.
3) This model leads to a straightforward procedure of generating realistic synthetic spectrum data that can subsequently be used to evaluate the performance of various spectrum access and allocation schemes.

The remainder of the paper is organized as follows. Section II introduces the stochastic differential equation model and how it is derived. Section III describes how the parameters are estimated from real spectrum data, while Section IV verifies the resulting model. Section V presents a way of synthesizing spectrum utilization data, and Section VI concludes the paper.

II. A Stochastic Differential Equation Model

The spectrum utilization model presented here uses stochastic differential equations (SDE) to model dynamic scattering and multipath fading channels, in particular, Rayleigh-distributed stationary channels. This is a technique developed and used in a number of studies, see e.g., [12], [13], [14], [15]. Specifically, our model is derived from a dynamic wireless channel model developed in [16] using similar techniques. Underlying this model is the assumption of either a single transmitter or many non-dominant transmitters stationary in space and in time. The model describes the complex signal received by a stationary receiver (thus with zero Doppler's effect) [1]. Building upon this work, our contribution lies in (1) extracting the received energy as a random process expressed as an SDE, and (2) developing a method to estimate the unknown parameters of the model.

In this model the signal detected at a receiver is viewed as a collection of a large number of reflected waves, and thus exhibits a multipath propagation phenomenon. This makes the received signal's phase random and hard to predict, and can possibly lead to large fluctuation in the received power. Assuming that the received signal on each path is random, the model developed in [16] is based on a continuous time description of the scattered electric field received at a stationary receiver with multipath reception along N paths, expressed as

$$\epsilon_t^{(N)} = \sum_{k=1}^{N} a_k \exp[i\varphi^{(k)}(t)] \ , \tag{1}$$

where a_k is the amplitude of the received signal along path k and i is the square root of -1. The phase factors $\exp[i\varphi^{(k)}(t)]$ are independent and uniformly distributed on a unit circle in the complex plane and for each t. In addition, it is assumed that the phase $\varphi^{(k)}(t)$ satisfies the following SDE:

$$d\varphi^{(k)}(t) = B^{\frac{1}{2}} dW^{(k)}(t) \ ,$$

where $\varphi^{(k)}(0)$ are uniformly distributed on $[0, 2\pi)$, $W^{(k)}$ are independent Weiner processes, and B is a constant that represents the rate of change in the phase of the received signal. By integrating the above SDE it is readily seen that $\text{VAR}(\varphi^{(k)}(t) - \varphi^{(k)}(0)) = Bt$.

Using stochastic calculus, it was established in [16] that the amplitude process is given by

$$\Psi(t) = I(t) + iQ(t) \ ,$$

where $I(t)$ and $Q(t)$ are the in phase and quadrature components of the incoming waves received at time t, and can be represented by the following two SDEs:

$$dI(t) = -\frac{1}{2} BI(t)dt + \frac{\sqrt{2}}{2}\sigma B^{\frac{1}{2}} dW^{(I)}(t) \ ; \tag{2}$$

$$dQ(t) = -\frac{1}{2} BQ(t)dt + \frac{\sqrt{2}}{2}\sigma B^{\frac{1}{2}} dW^{(Q)}(t) \ , \tag{3}$$

with $I(0) = 0$, $Q(0) = 0$, and $W^{(I)}(t)$ and $W^{(Q)}(t)$ are two independent standard Wiener processes.

The parameter B makes the above two SDEs *mean-reverting*, i.e., the process, in equilibrium, approaches the mean [17]. Such processes are also sometimes referred to as Ornstein-Uhlenbeck processes [17].

The parameter σ^2 represents the stationary magnitude of the average scattering power taken over an asymptotically large

[1] We discuss in Section VI how to potentially account for mobile users and the arrival and departure of users.

number of propagation paths (see [16]), and is shown to be the asymptotic (in t) variance of ϵ_t and satisfies:

$$\sigma^2 = \sum_{k=1}^{\infty} a_k^2$$

which is assumed to be finite, assuming no single path dominates: $\lim_{N\to\infty} \frac{a_j^2}{\sum_{k=1}^{N} a_k^2} = 0$.

The above summarizes what was developed in [16]. We now proceed to derive the power process received at the receiving antenna at time t (all proofs are given in the appendix). This is given by

$$R(t) = \sqrt{I^2(t) + Q^2(t)}. \tag{4}$$

Assuming processes (2) and (3), and using standard arguments from stochastic calculus [17], we have the following lemma:

Lemma 1: Let $\bar{I}(t) = e^{\frac{1}{2}Bt} I(t)$ and $\bar{Q}(t) = e^{\frac{1}{2}Bt} Q(t)$. Then

$$d\bar{I}(t) = \frac{\sqrt{2}}{2}\sigma B^{\frac{1}{2}} e^{\frac{1}{2}Bt} dW^{(I)}(t) \tag{5}$$

$$d\bar{Q}(t) = \frac{\sqrt{2}}{2}\sigma B^{\frac{1}{2}} e^{\frac{1}{2}Bt} dW^{(Q)}(t). \tag{6}$$

Furthermore, both $\bar{I}(t)$ and $\bar{Q}(t)$ are normally distributed, with mean 0 and variance $\frac{\sigma^2}{2}(e^{Bt} - 1)$.

An immediate consequence of the above lemma is that

$$\bar{R}(t) = \sqrt{\bar{I}(t)^2 + \bar{Q}(t)^2} \tag{7}$$

has a Rayleigh distribution with parameter $\sqrt{\frac{\sigma^2}{2}(e^{Bt} - 1)}$.

The main theorem is given as follows:

Theorem 1: The power process $R(t)$ given in Eqn 4 satisfies the following SDE:

$$dR(t) = -\frac{BR(t)}{2}dt + \frac{1}{4}\frac{B\sigma^2}{R(t)}dt + \frac{1}{\sqrt{2}}\sigma B^{\frac{1}{2}} dW(t) \tag{8}$$

with $R(0) = r_0$, r_0 being some constant, and W a standard Wiener process.

The power process we propose to use in this paper is the above with an added drift term:

$$dR(t) = Adt - \frac{BR(t)}{2}dt + \frac{1}{4}\frac{B\sigma^2}{R(t)}dt + \frac{1}{\sqrt{2}}\sigma B^{\frac{1}{2}} dW(t)\ . \tag{9}$$

The reason for adding this drift term (which steers the mean of the process in a certain direction over time) is primarily due to the observation that the received energy process is in general highly non-stationary; the cyclic nature of most wireless services (peak vs. off-peak hours) is a clear example of this non-stationarity. Therefore in general one would expect the parameters associated with the above model (if it were to fit well the actual data) to vary over time. The appropriate time constant may be determined via testing over measurement data (the time scale we used in subsequent sections is one hour; more is discussed in Section IV). Because of this non-stationarity an additional drift term is expected to make the model more accurate as we verify in Section IV.

To summarize, our model given in (9) consists of four terms: the first a drift term, the second the "mean-reverting" term (or the O-U process term), the third a "radial" term, and the fourth a volatility term. Using the terminology of [10], we will call this process a *Radial Ornstein-Uhlenbeck process with a drift*.

Three unknown parameters uniquely define this model: A, B, and σ. A will be referred to as the drift constant; it is an indication of how much the mean of the received power changes over time. B will be referred to as the phase constant; it corresponds to the rate at which the received signal phase changes. σ^2 will be referred to as the power constant; it is the sum of signal magnitudes received over multiple paths. In the next two sections we show how these three parameters can be estimated using spectrum measurement data for training. We also give interpretation and guidelines on how these parameters may be chosen for specific channels when such training is not available.

III. Parameter Estimation

In order to estimate the three unknowns A, B and σ from real measurement data, we first rearrange terms in (9) to obtain the following:

$$\frac{dW(t)}{\sqrt{dt}} = \frac{\sqrt{2}}{\sigma B^{\frac{1}{2}}}\left\{\frac{dR(t)}{\sqrt{dt}} - A\sqrt{dt} + \frac{BR(t)}{2}\sqrt{dt} - \frac{B\sigma^2}{4R(t)}\sqrt{dt}\right\}. \tag{10}$$

Note that the left hand side of the above equation is now a zero-mean, unit-variance normally distributed random variable. The idea behind our parameter estimation process is to use real measurement data to generate data points corresponding to the right hand side of Eqn (10), and then match the first three (or more) sample moments to that of the 0-mean unit-variance normal distribution, thereby solving three unknowns.

Specifically, for a given frequency band our measurements are in the form of a time series of energy readings, denoted as $\hat{R}(t_i)$, $i = 0, 2, \cdots, N$. From these measurements we can now obtain successive differences between these readings, denoted as $d\hat{R}(t_i) = \hat{R}(t_i) - \hat{R}(t_{i-1})$, $i = 1, 2, \cdots, N$. We can also obtain the differences in sampling times, denoted as $dt_i = t_i - t_{i-1}$, $i = 1, 2, \cdots, N$. For our measurement data, sampling times are evenly spaced. Therefore in our case dt_i is treated as a constant.

Following this, the original measurement data may be viewed as a collection of N triples $(\hat{R}(t_i), d\hat{R}(t_i), dt_i)$, $i = 1, 2, \cdots, N$. Each such triple will now be referred to as a *sample* within the context of estimation and testing. From this collection of samples, we now select a random subset $\mathcal{N}_{est}$ of size N_{est} for estimation. We plug in each selected sample into the RHS of Eqn (10) and obtain the following data point $\hat{w}_i$ for $i \in \mathcal{N}_{est}$:

$$\hat{w}_i = \frac{\sqrt{2}}{\sigma B^{\frac{1}{2}}}\left\{\frac{d\hat{R}(t_i)}{\sqrt{dt_i}} - A\sqrt{dt_i} + \frac{B\hat{R}(t_i)}{2}\sqrt{dt_i} - \frac{B\sigma^2}{4\hat{R}(t_i)}\sqrt{dt_i}\right\}. \tag{11}$$

This gives us a total of N_{est} data points $\{\hat{w}_i, i \in \mathcal{N}_{est}\}$, each a function of A, B and σ. These three unknown parameters can now be estimated by matching (1) the sample mean of the data set $\{\hat{w}_i, i \in \mathcal{N}_{est}\}$ to 0; (2) its sample variance to 1; and (3) its sample 3rd moment to 0, and so on. That is, the parameters are estimated by matching the first three sample moments to the first three moments of the 0-mean unit-variance normal distribution. One can also try to match even higher moments.

In our experiments, we try to match the first four moments, and obtain the estimates by solving the following minimization problem:

$$\min_{A,B,\sigma} \quad (m_1 - 0)^2 + (m_2^{1/2} - 1)^2 + (m_3^{1/3} - 0)^2 + (m_4^{1/4} - 3^{1/4})^2$$

where m_i denotes the i-th sample moment of the data set $\{\hat{w}_i, i \in \mathcal{N}_{est}\}$, and 0, 1, 0, and 3 are the first four moments of a standard Normal distribution (0-mean, unit-variance).

Once these parameters are estimated, we will use the remaining $N - N_{est}$ samples for testing and model verification. This is done in a very similar way as in estimation. Specifically, the testing samples are also plugged into the RHS of Eqn (10). However, this time the computation is done with the estimated values of A, B and σ. This gives us $N - N_{est}$ data points, also commonly referred to as the *residual of the test data*. The model verification test lies in checking whether the residual follows the standard normal distribution.

In the next section we present results on the above estimation and testing procedure.

IV. Model Verification

A. Data sets used

The model verification uses spectrum data from our measurement study reported in [8], which was done over a period of multiple days continuously, and simultaneously at multiple locations. The resolution of the measurements is such that one energy reading (in μV) is produced for each band of width 200KHz, from 20MHz to 3GHz, and for roughly every 75 seconds of sweep time over this range. This means that the sweep time of each band is roughly 5ms. So for a given frequency band, the energy process is essentially sampled once every 75s, which results in a discrete-time process with a fairly large gap between successive samples compared to the 5ms of each sample. Consequently this data set does not directly reveal whether our model works with samples taken at a higher frequency. However, since the model is continuous in time, it may be validated at arbitrary time scales using the same methods. As we will show it fits very well at this sampling frequency. More validation can be conducted as data becomes available.

From this collection of data, we present results from primarily the following 12 sets as listed in Table I, with the exception of Tables IV and V. More experiments were conducted, but not all can be presented here due to space limit; the results shown here are fairly representative. These correspond to three frequency bands, starting at 518, 738, and 1842MHz, respectively, and each of width 200KHz. The primary user/service in these bands are TV (listed as TV1 and TV2 for distinction) and GSM downlink. We show two time periods: 10-11am and 3-4pm. Half the data sets are from one location, the other half from a second location. The two locations are about a few miles apart.

Data Set	Location	Time	MHz	Primary user
1	1	10-11am	518	TV1
2	1	10-11am	738	TV2
3	1	10-11am	1842	GSM
4	1	3-4pm	518	TV1
5	1	3-4pm	738	TV2
6	1	3-4pm	1842	GSM
7	2	10-11am	518	TV1
8	2	10-11am	738	TV2
9	2	10-11am	1842	GSM
10	2	3-4pm	518	TV1
11	2	3-4pm	738	TV2
12	2	3-4pm	1842	GSM

TABLE I
DATA SETS FOR MODEL VERIFICATION

Each data set consists of the same one-hour time series of measured power level on *six* consecutive days, Monday through Saturday. On each day, the one-hour period contains roughly 55 energy readings, so the entire data sets contains around 330 data points. For each data set, we randomly select 60% as the estimation/training data set and the remaining 40% the testing data set[2]. As detailed in the previous section, the training data set is used to estimate the parameters and the testing data set to verify the model by checking the normality of the residual w_i obtained by plugging testing samples into the RHS of Eqn (10) with the estimated parameters. In selecting the right amount of data for estimation one needs to strike a balance in the following sense: insufficient training leads to estimation error while insufficient testing leads to unreliable comparison. Our experiments suggest that the 60-40 split results in a good balance.

B. Verification using Q-Q plots

To check whether the residual follows a standard normal distribution, and in particular to check how far it is from the standard normal distribution, we will use the Quantile-Quantile(Q-Q) plot, a commonly used graphical statistical tool, see e.g, [18], [19], [20]. The quantiles are points taken at regular intervals from the cumulative distribution function (CDF) of a random variable. The p-quantile for a random variable X is the value x such that Prob$(X < x) = p$. A Q-Q plot shows the quantiles of the first data set against the quantiles of the second data set, and is therefore an intuitive (as well as visual) and efficient way to determine if two data sets follow a common distribution. For two random data sets S_1 and S_2, the Q-Q plot is generated by first sorting each set in increasing order, and then sequentially placing points on the

[2]Randomly selecting a set for estimation is a standard procedure in statistical analysis; the resulting estimation represents the data better in case the underlying process is non-stationary.

plot. The i-th point is placed at coordinates (s_i^1, s_i^2), where s_i^1 and s_i^2 are the values of the i-th data point in the sorted sets S_1 and S_2, respectively.

Typically a 45-degree reference line is plotted on the Q-Q plot: if the two data sets come from the same distribution with exactly the same parameters, then their quantiles will match and the points should fall approximately along this reference line.

However, if two data sets come from the same type of distribution but with different parameters, then on the Q-Q plot the points will still fall approximately along a line, but this line is not the reference line. To interpret the difference between the parameters through the Q-Q plot, we need to mention the concept of a *location-scale family*, which is a family of probability distributions parameterized by a *location parameter* μ and a *scale parameter* $\sigma \geq 0$. If X is a random variable whose probability distribution belongs to such a family, then $Y = \mu + \sigma X$ is also in the family, and every distribution in the family is of this form. If two data sets come from the same location-scale family, but with two different location parameters, then the points on the Q-Q plot will fall approximately along a line with a slope close to 1 and intercept (on the Y-axis) close to the difference between the location parameters. If two data sets come from the same location-scale family but with different scale parameters, then the points will fall approximately along a line with slope close to the ratio between the scale parameters and intercept close to 0.

Normal distribution is one example of the location-scale family, where the mean is the location parameter and the standard deviation is the scale parameter. In order to check whether a data set follows a standard normal distribution, we will make the first data set the theoretical quantiles of the standard normal distribution and the second data set the residual of the test data on the Q-Q plot. If the points fall along the reference line, then it is strong evidence that the residual follows the standard normal distribution. If the points fall along a line, but not the reference line, then this suggests that the residual follows normal distribution but is not exactly standard.

Figure 1 shows the above normality test results for the three frequencies at Location 1 during the hour of 10-11am. The dashed line represents the best linear fit of the points (the residual) – the closer the points are to the dashed line, the more normally distributed the residual is. The solid line represents the 45-degree reference line. It can be seen that in all three cases the points have very good linear fit, indicating strong normality of the residuals. In addition, all three dashed lines are very close to the reference line, indicating the residuals follow close-to-standard normal distributions. In the case of 518MHz, the residual has a mean very close to 0 (i.e., the dashed line approximately crosses the origin – point (0, 0) – on the plot), but the variance is slightly above 1 (compare to the 45-degree line). The same observation applies to 738MHz. In the case of 1842MHz, the variance is approximately unit, but the mean is slightly below zero.

The same test is performed over the hour of 3-4pm, as well as at the second location. These are shown in Figures 2, 3, and 4, respectively. Similar results were obtained over other time periods. Our general conclusion is the model fits extremely well the measured data [3].

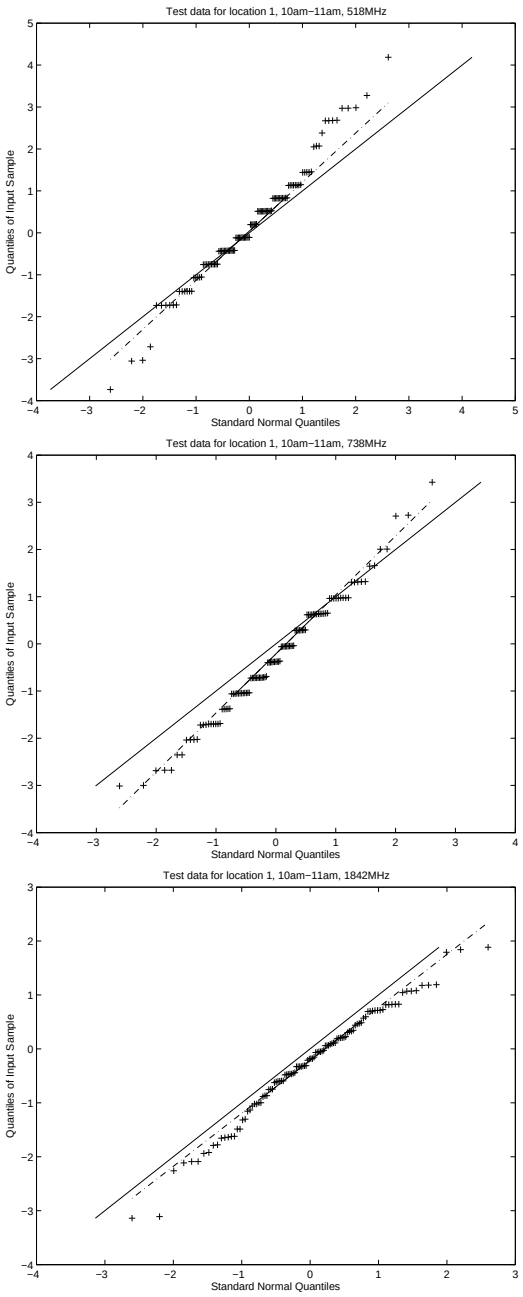

Fig. 1. Q-Q plot, location 1, 10-11am: 518MHz (top), 738MHz (middle), and 1842MHz (bottom)

C. Estimated parameter values

We next examine the values of the estimated parameters, in an attempt to provide useful observations and guidelines in

[3]Note that the fit cannot be perfect, i.e., the matching of the first 3-4 moments cannot be done precisely, so there will always be small discrepancies when dealing with real data.

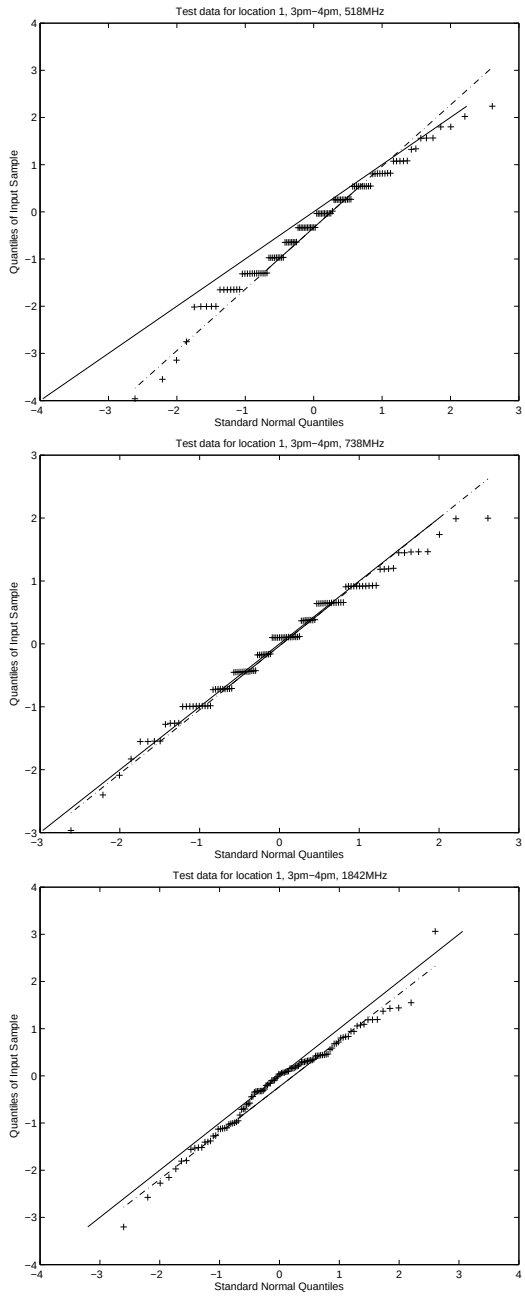

Fig. 2. Q-Q plot, location 1, 3-4pm: 518MHz (top), 738MHz (middle), and 1842MHz (bottom)

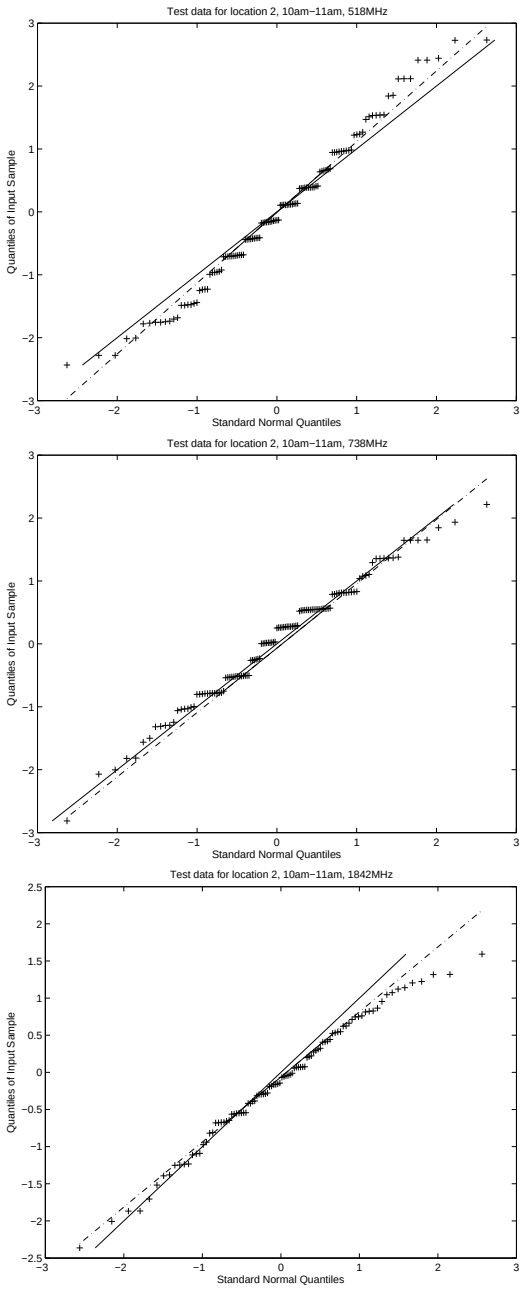

Fig. 3. Q-Q plot, location 2, 10-11am: 518MHz (top), 738MHz (middle), and 1842MHz (bottom)

selecting appropriate values when real data is not available for training. These values can then be used in synthesizing one's own spectrum utilization data, described in the next section.

We start by listing the sample statistics (mean received power (in μV) and power range) of the data sets used in our model verification in Table II. In Table III we show the estimated parameter values for each data set.

There are a few observations to make by inspecting the results listed in Table III. Firstly, the drift constant A is quite small for the two TV bands, and much smaller than that for the GSM band in general. This suggests that the TV bands are much more stable, to be expected. GSM signals (downlink in this case) are driven by user demand so it fluctuates more. Secondly, this drift constant for the GSM band varies from location to location, and from time to time, while staying relatively stable for the TV bands. The phase parameter B is generally very small in value, and does not appear to have a clear trend across time/location [4]. The power parameter σ

[4]Note that B is negative for data sets 9 and 12. As B is non-negative by construction, this means that in this case the model did not perform as expected. This is likely a numerical consequence of B being very small in absolute quantities in both cases.

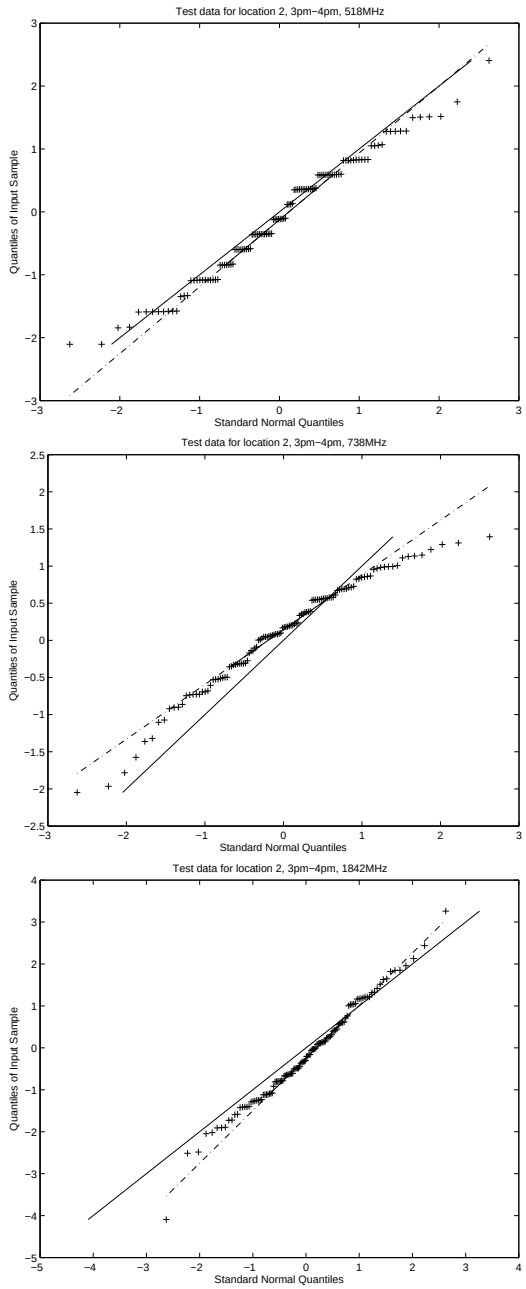

Fig. 4. Q-Q plot, location 2, 3-4pm: 518MHz (top), 738MHz (middle), and 1842MHz (bottom)

for the TV bands appears very consistent over time for the same frequency but differ at different locations. The GSM band clearly has much higher power than the TV bands, and at different locations and times shows quite different characteristics.

We further estimated the parameter values for TV1 and GSM bands at location 1 for each hour between 10am and 4pm. These are listed in Tables IV and V. In general we see that for both bands all three parameters stay fairly constant over time at the same location (with the exception of the first hour for the GSM band), while having very different A and σ

Data Set (Loc, time)	MHz	Mean(μ)	Range
1 (1, 10-11am)	518	0.125	0.093-0.166
2 (1, 10-11am)	738	0.121	0.095-0.148
3 (1, 10-11am)	1842	795	537-1202
4 (1, 3-4pm)	518	0.125	0.100-0.309
5 (1, 3-4pm)	738	0.123	0.095-0.145
6 (1, 3-4pm)	1842	876	616-1445
7 (2, 10-11am)	518	0.374	0.309-0.447
8 (2, 10-11am)	738	0.963	0.759-1.202
9 (2, 10-11am)	1842	3501	1584-5623
10 (2, 3-4pm)	518	0.382	0.309-0.479
11 (2, 3-4pm)	738	0.949	0.646-1.230
12 (2, 3-4pm)	1842	3856	2344-5623

TABLE II
SAMPLE DATA STATISTICS: MEAN RECEIVED POWER (μV) AND RANGE

Data Set	μ	A	B	σ	σ/μ
1	0.125	-0.341	0.067	1.58	12.63
2	0.121	-0.395	0.018	3.23	26.70
3	795	85.500	0.436	730.06	0.92
4	0.125	-0.161	0.053	1.23	9.85
5	0.123	-0.507	0.134	1.36	11.07
6	876	20.395	0.050	184.36	0.21
7	0.374	-1.359	0.049	6.55	17.52
8	0.963	-3.741	0.046	17.80	18.49
9	3501	-2.967	-0.001	3585.51	1.02
10	0.382	-1.420	0.022	10.13	26.52
11	0.949	-2.727	0.042	15.86	16.71
12	3856	-16.873	-0.007	1563.12	0.41

TABLE III
ESTIMATED PARAMETER VALUES

values in each band. The quantities shown here may be used as a reference in the SDE model to simulate similar channels.

Time	A	B	σ
10am-11am	-0.341	0.067	1.579
11am-12pm	-0.441	0.095	1.507
12pm-1pm	-0.455	0.099	1.499
1pm-2pm	-0.327	0.113	1.196
2pm-3pm	-0.346	0.015	3.364
3pm-4pm	-0.161	0.053	1.232

TABLE IV
PARAMETER VALUES OVER TIME AT LOCATION 1, 518MHZ

Time	A	B	σ
10am-11am	85.50	0.436	730.06
11am-12pm	20.03	0.052	136.20
12pm-1pm	14.64	0.037	122.29
1pm-2pm	21.64	0.052	167.24
2pm-3pm	20.35	0.051	159.77
3pm-4pm	20.39	0.050	184.36

TABLE V
PARAMETER VALUES OVER TIME AT LOCATION 1, 1842MHZ

V. SYNTHETIC SPECTRUM DATA GENERATION

An important reason for developing such a model is to provide a way to generate synthetic channel data (sample paths of energy levels in a channel) that are statistically close to

variations observed in a real channel, so that one can easily generate a realistic "spectrum environment" in which to test and evaluate various algorithms and protocols. Below we show such synthetic data can be easily generated under our SDE model.

Taking Eqn (9) and integrating over a small interval ϵ, we get

$$\begin{aligned} & R(t+\epsilon) - R(t) \\ = \;& A(t+\epsilon-t) - B\int_t^{t+\epsilon} R(\tau)d\tau \\ & + \frac{1}{4}B\sigma^2 \int_t^{t+\epsilon} \frac{1}{R(\tau)}d\tau + \sigma\frac{1}{\sqrt{2}}B^{\frac{1}{2}}(W(t+\epsilon)-W(t)) \\ \approx \;& A\epsilon - BR(t)\epsilon + \frac{1}{4}B\sigma^2\frac{1}{R(t)}\epsilon \\ & + \sigma\frac{1}{\sqrt{2}}B^{\frac{1}{2}}(W(t+\epsilon)-W(t)) \end{aligned} \tag{12}$$

where the approximation holds when ϵ is sufficiently small.

Assuming we start from some initial condition $R(t_o)$ at time t_o, we can generate a sequence of data $R(t_o + k\epsilon)$ at times $t_o + k\epsilon$ for $k = 1, 2, \cdots$ with time resolution (or time step) of ϵ as follows. Note that $W(t+\epsilon)-W(t)$ is normally distributed with zero mean and variance ϵ.

1) We begin by generating a random (independent across successive generations) sample from the 0-mean ϵ-variance normal distribution. Denote the sample by $dW = W(t_o+\epsilon) - W(t_o)$.
2) Take this sample value into Eqn (12), replacing the corresponding part in the last term on the RHS.
3) Now we can compute completely the RHS; this gives the difference between $R(t_o+\epsilon)$ and $R(t_o)$, hence we have generated a value for $R(t_o+\epsilon)$.
4) Now repeat the above procedure indefinitely to produce a time series of desired length.

The end result of this procedure is a sequence of synthesized $R(t)$ values, representing a particular realization of the channel.

VI. Conclusion and Future work

In this paper we presented a stochastic differential equation based spectrum utilization model that accurately captures the dynamic changes in the channel condition induced by primary users' activities, and verified it using real spectrum measurement data. It is in closed form, and therefore can be used in theoretical studies on optimal spectrum access and sharing. Furthermore, we illustrated a straightforward procedure based on this model to generate synthetic spectrum data.

As we have mentioned in introducing the basic assumptions underlying this model, the model is meant for scenarios with stationary transmitters (or primary users) both in time and in space. This means that the model assumes the primary users do not change either their location or their quantity over the modeling time period. Indeed this model works less satisfactorily in cases where there are variations in the population of transmitters, e.g., the GSM uplink channel. In such a case the primary users (cellphone users) arrive and depart over time, causing the population to fluctuate over time. Part of our on-going work is to construct and add a birth-death-like process to the current model to capture this arrival and departure process in the population.

References

[1] J. Kennedy and M. Sullivan, "Direction Finding and "Smart Antennas" Using Software Radio Architechtures," *IEEE Communications Magazine*, pp. 62–68, May 1995.

[2] S. Haykin, "Cognitive Radio: Brain-Empowered Wireless Communications," *IEEE Journal on Selected Areas of Communications (JSAC)*, vol. 23, no. 2, pp. 201–220, February 2005.

[3] M. A. McHenry, P. A. Tenhula, D. McCloskey, D. A. Roberson, and C. S. Hood, "Chicago spectrum occupancy measurements & analysis and a long-term studies proposal," in *The first international workshop on Technology and policy for accessing spectrum*. 2006, ACM Press New York, NY, USA.

[4] M. A. McHenry, "NSF spectrum occupancy measurements project summary," in *Shared Spectrum Company Report*, August 2005.

[5] M. H. Islam, C. L. Koh, S. W. Oh, X. Qing, Y. Y. Lai, C. Wang, Y.-C. Liang, B. E. Toh, F. Chin, G. L. Tan, and W. Toh, "Spectrum Survey in Singapore: Occupancy Measurements and Analyses," *Cognitive Radio Oriented Wireless Networks and Communications, 2008. CrownCom 2008. 3rd International Conference on*, pp. 1–7, May 2008.

[6] R. I. C. Chiang, G. B. Rowe, and K. W. Sowerby, "A Quantitative Analysis of Spectral Occupancy Measurements for Cognitive Radio," *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th*, pp. 3016–3020, April 2007.

[7] M. Wellens, J. Wu, and P. Mahonen, "Evaluation of Spectrum Occupancy in Indoor and Outdoor Scenario in the Context of Cognitive Radio," *Cognitive Radio Oriented Wireless Networks and Communications, 2007. CrownCom 2007. 2nd International Conference on*, pp. 420–427, Aug. 2007.

[8] D. Chen, S. Ying, Q. Zhang, M. Liu, and S. Li, "Mining spectrum usage data: a large-scale spectrum measurement study," in *ACM International Conference on Mobile Computing and Networking (MobiCom)*, Beijing, China, September 2009.

[9] P. F. Marshall, "Closed-form analysis of spectrum characteristics for cognitive radio performance analysis," in *Proc. 3rd IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, Chicago, IL, October 2008, pp. 1–12.

[10] A. N. Borodin and P. Salminen, *Handbook of Brownian Motion - Facts and formulae*, Birkhäuser, Berlin, 2000.

[11] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research," in *Wireless Communication and Mobile Computing (WCMC): Special issue on Mobile Ad Hoc Networking: Research, Trends and Applications*, 2002.

[12] T. R. Fields and R. J. A. Tough, "Stochastic dynamics of the scattering amplitude generating K-distributed noise," *Journal of Mathematical Physics*, vol. 44, no. 11, pp. 5212–5223, November 2003.

[13] T. R. Fields and R. J. A. Tough, "Diffusion processes in elctromagnetic scattering amplitude generating K-distributed noise," *Journal of Mathematical Physics*, vol. 44, no. 11, pp. 5212–5223, November 2003.

[14] T. R. Fields and R. J. A. Tough, "Dynamical models of weak scattering," *Journal of Mathematical Physics*, vol. 46, pp. 13302–13320, 2005.

[15] C. D. Charalambous and N. Menemenlis, "A state-space approach in modeling multipath fading channels via stochastic differential equations," in *Proc. IEEE International Conference on Communications (ICC)*, June 2001, vol. 7, pp. 2251–2255.

[16] T. Fang and T. R. Fields and S. Haykin, "Stochastic differential equation theory applied to wireless channels," *IEEE Transactions on Communications*, vol. 55, pp. 1478–1489, 2007.

[17] B. Oksendal, *Stochastic Differential Equations - An Introduction With Applications*, Springer, Berlin, 1998.

[18] G. Blom, *Statistical estimates and transformed beta variables*, John Wiley and Sons, New York, 1958.

[19] J. Chambers, W. Cleveland, B. Kleiner, and P. Tukey, *Graphical methods for data analysis*, Wadsworth, 1983.

[20] W. Cleveland, *The Elements of Graphing Data*, Hobart Press, 1994.

APPENDIX

A. Proof of Lemma 1

Let $f(x,t) = e^{\frac{1}{2}Bt}x$. Then $f_t = \frac{\partial f}{\partial t} = \frac{1}{2}Be^{\frac{1}{2}Bt}x$, $f_x = \frac{\partial f}{\partial x} = e^{\frac{1}{2}Bt}$ and $f_{xx} = \frac{\partial^2 f}{\partial^2 x} = 0$. Using Ito's formula (Theorem 4.2.3 [17]; this is provided below for the convenience of the reader), and replacing x with $I(t)$, we have

$$\begin{aligned} d\bar{I}(t) &= f_t dt + f_x dI(t) + \frac{1}{2}f_{xx}(dI(t))^2 \\ &= \frac{1}{2}Be^{\frac{1}{2}Bt}I(t)dt + e^{\frac{1}{2}Bt}(-\frac{1}{2}BI(t)dt \\ &\quad + \frac{\sqrt{2}}{2}\sigma B^{\frac{1}{2}}dW^{(I)}(t)) \\ &= \frac{\sqrt{2}}{2}\sigma B^{\frac{1}{2}}e^{\frac{1}{2}Bt}dW^{(I)}(t) \end{aligned}$$

where we have substituted $dI(t)$ with Eqn (2). The second, Eqn (6), can be obtained by the same argument. Integrating, say the equation for $\bar{I}(t)$, we see that

$$\bar{I}(t) = \bar{I}(0) + \frac{\sqrt{2}}{2}B^{\frac{1}{2}}\sigma \int_0^t e^{\frac{1}{2}Bs}dW^{(I)}(s) \ .$$

Using the facts that the mean of the Ito's integral $\int_0^t e^{\frac{1}{2}Bs}dW^{(I)}(s)$ is zero and its variance $\int_0^t e^{Bs}ds$ we get the result.

B. Proof of Theorem 1

Consider $\bar{R}(t)$ as in Eqn (7) and note that $\bar{R}(t) = e^{\frac{1}{2}Bt}R(t)$. Now, consider the function $f(x,y) = \sqrt{x^2+y^2}$, for which we have the following first and second order partial derivatives: $f_x = \frac{x}{\sqrt{x^2+y^2}}$, $f_y = \frac{y}{\sqrt{x^2+y^2}}$, $f_{xx} = \frac{1}{\sqrt{x^2+y^2}} - \frac{x^2}{(\sqrt{x^2+y^2})^3}$, $f_{yy} = \frac{1}{\sqrt{x^2+y^2}} - \frac{y^2}{(\sqrt{x^2+y^2})^3}$ and $f_{xy} = -\frac{xy}{(\sqrt{x^2+y^2})^3}$.

Substituting the above into Ito's formula when differentiating (7), and using standard results on Wiener processes: $dW^{(I)}(t)^2 = dt$, $dW^{(Q)}(t)^2 = dt$ and $dW^{(I)}(t)dW^{(Q)}(t) = 0$, we get

$$d\bar{R}(t) = \frac{\bar{I}(t)d\bar{I}(t)}{\bar{R}(t)} + \frac{\bar{Q}(t)d\bar{Q}(t)}{\bar{R}(t)} + \frac{1}{4}\sigma^2 Be^{Bt}\frac{1}{\bar{R}(t)}dt. \quad (13)$$

Consider the first two terms in the above expression. It is seen that

$$\begin{aligned} &\frac{\bar{I}(t)d\bar{I}(t)}{\bar{R}(t)} + \frac{\bar{Q}(t)d\bar{Q}(t)}{\bar{R}(t)} \\ &= \frac{\sqrt{2}}{2}\sigma B^{\frac{1}{2}}e^{\frac{1}{2}Bt}[\frac{I(t)dW^{(I)}(t) + Q(t)dW^{(Q)}(t)}{R(t)}] \ . \end{aligned}$$

We have that

$$\begin{aligned} &\frac{I(t)dW^{(I)}(t) + Q(t)dW^{(Q)}(t)}{R(t)} \\ &= \frac{1}{R(t)}[I(t), Q(t)] \begin{bmatrix} dW^{(I)}(t) \\ dW^{(Q)}(t) \end{bmatrix} \end{aligned}$$

and by the definition of $R(t)$

$$\frac{1}{R(t)^2}[I(t), Q(t)] \begin{bmatrix} I(t) \\ Q(t) \end{bmatrix} = 1 \ .$$

Therefore, using Theorem 8.4.2 of [17] (also provided below), we conclude that $\int_0^t \frac{I(s)dW^{(I)}(s) + Q(s)dW^{(Q)}(s)}{R(s)}ds$ has the same law as a Wiener process, denoted as $W(t)$, independent of $W^{(I)}$ and $W^{(Q)}$. This means that we can write

$$\frac{\bar{I}(t)d\bar{I}(t)}{\bar{R}(t)} + \frac{\bar{Q}(t)d\bar{Q}(t)}{\bar{R}(t)} = \frac{\sqrt{2}}{2}\sigma B^{\frac{1}{2}}e^{\frac{1}{2}Bt}dW(t) \ .$$

Substituting the above into (13) we obtain:

$$d\bar{R}(t) = \frac{\sqrt{2}}{2}\sigma B^{\frac{1}{2}}e^{\frac{1}{2}Bt}dW(t) + \frac{1}{4}\sigma^2 Be^{Bt}\frac{1}{\bar{R}(t)}dt. \quad (14)$$

Since $R(t) = e^{-\frac{1}{2}Bt}\bar{R}(t)$, we have

$$dR(t) = -\frac{1}{2}Be^{-\frac{1}{2}Bt}\bar{R}(t)dt + e^{-\frac{1}{2}Bt}d\bar{R}(t) \ . \quad (15)$$

Replacing $d\bar{R}(t)$ with (14) in the above equation gives the desired result.

C. Ito's formula

Let $X(t)$ be an Ito process given by

$$dX(t) = udt + vdW(t)$$

Let $f(t,x) \in C^2([0,\infty) \times \mathbf{R})$ (i.e. f is twice continuously differentiable on $[0,\infty) \times \mathbf{R}$. Then

$$Y(t) = f(t, X(t))$$

is again an $Itô$ process, and

$$dY(t) = f_t dt + f_x dX(t) + \frac{1}{2}f_{xx}(dX(t))^2$$

where $(dX(t))^2 = (udt + vdW(t))^2$ is computed using the following standard results of Wiener processes:

$$dt \cdot dt = dt \cdot dW(t) = dW(t) \cdot dt = 0, \quad dW(t) \cdot dW(t) = dt.$$

D. Theorem 8.4.2 of [17]

An Ito process

$$dY(t) = vdW(t); \qquad Y_0 = 0$$

coincides with the n-dimensional Brownian motion *iff*

$$vv^T = I_n$$

where I_n is the n-dimensional identity matrix.