

A Maximum Entropy Framework for Statistical Modeling of Underwater Acoustic Communication Channels

François-Xavier Socheleau^{*‡}, Christophe Laot^{*}, Jean-Michel Passerieux[‡]

^{*}Institut TELECOM; TELECOM Bretagne, UMR CNRS 3192 Lab-STICC, Université européenne de Bretagne,
Email: {fx.socheleau, christophe.laot}@telecom-bretagne.eu

[‡]Thales Underwater Systems, Email: jean-michel.passerieux@fr.thalesgroup.com

Abstract—Based on a method of inductive inference known as the principle of maximum entropy, a time-varying underwater acoustic channel model is derived. The resulting model is proved to be *consistent* so that it only relies on the available knowledge of the environment to model. While requiring only a few parameters (e.g. channel average power and Doppler spread), it is shown through fading statistics and bit error rates measurements that accurate channel impulse responses can be obtained for communication applications. The Matlab code of the proposed model is available at <http://perso.telecom-bretagne.eu/fxsocheleau/software>.

Index Terms—Underwater acoustic communications, channel modeling, scattering function, entropy

I. INTRODUCTION

TIME-VARYING channel models tailored to underwater acoustic communications (UAC) are strongly needed since research in that domain mainly relies on sea-trials which are expensive and time-consuming. While deterministic physics-driven models [1] offer a fairly accurate modeling of channels average behavior, statistical modeling appears as the most suitable manner to take into account the fast time fluctuations of the channel response due to random phenomena such as scattering or random motions of the transmitter and the receiver. In UAC, where few real data are available, intuitive assumptions are often preponderant to fulfill the lack of knowledge on the channel properties. Unlike radio channels, there is no consensus on the statistical characterization of both the probability distribution and the power spectral density of UAC fading processes [2]. Such ad-hoc model construction may result in inconsistencies and is therefore non-consensual.

In a context where efforts are emerging to provide standardized channel models [3], the originality of this contribution is to provide a *consistent* time-varying channel model with an additional constraint of low-dimensionality or *conciseness* and this within an *open* theoretical framework.

Consistent modeling is defined as the requirement that independent modelers with the same state of channel knowledge must obtain identical models. More generally, the consistency desideratum is a key requirement in statistical inference and can be formulated as follows [4]: “If a conclusion can be reasoned out in more than one way, then every possible way must lead to the same result.” Consistency prevents modelers from invoking unjustified or irrational assumptions.

A *concise* model is defined as a model driven by few parameters which corresponds to cases where little is known about the channel or where its behavior can be expressed through few variables. Concise modeling is mainly driven by the will to provide a channel model that does not require difficult-to-obtain parameters. Our approach relies on a Bayesian viewpoint: rather than trying to model the complete and accurate physical reality of the UAC channel (which remains elusive), it is more pragmatic to only model the available information on this reality while leaving the unknown information in an unconstrained space. The model is not meant to represent the physical reality of the channel but rather a state of knowledge of this reality. A crucial task then consists in determining how much information is needed to characterize the environment to model with enough accuracy.

In [5], Debbah *et al.* address the question of consistent channel modeling in light of Bayesian probability theory and the principle of maximum entropy introduced in [6] by Jaynes¹. Classically, the maximum entropy problem consists in deriving a probability density function (pdf) from a finite set of expectations, where these expectations represent the available knowledge of the process to model. Entropy translates information into probability assignment. Moreover, maximizing entropy is shown to be the only way to provide a consistent model that is maximally noncommittal with respect to unavailable information [7]. Choosing the model with the greatest entropy therefore avoids the arbitrary introduction of information that is not known.

With the consistency and conciseness requirements in mind, we show in this paper that a time-varying UAC channel model can be found in light of inductive inference and the principle of maximum entropy. More precisely, an analytical model is obtained for both the probability distribution and the power spectral density of UAC fading processes without making any a priori assumption relying on intuition. A typical example addressed in this paper is the modeling of time-varying UAC channels when only some of their moments such as the average power or the Doppler spread are known. It is shown that

¹The principle of maximum entropy was introduced by Jaynes as an extension of the principle of insufficient reason of J. Bernoulli and Laplace. The latter states that if there is no reason to believe that out of a set of possible mutually exclusive events, no one event is more likely to occur than any other, then one should assume that all events are equally probable.

entropy maximization creates a consistent model for us out of this little information available. Moreover, this inductive approach provides an open framework that makes the model incremental so that a prior model can be updated to a posterior model when new information becomes available.

The paper is organized as follows. Section II reviews some basic characterizations of time-varying channels. After recalling some results of [5] to model the channel pdf, the principle of maximum entropy is applied to the derivation of power spectral density of UAC fading processes in section III. In section IV, we show some outputs provided by the resulting channel model and discuss the relevance of it when faced to real data recorded at sea. Finally, conclusions are given in section V.

Additionally, to contribute to the effort of channel model standardization, the Matlab code of the proposed model is available at <http://perso.telecom-bretagne.eu/fxsocheleau/software>.

II. LINEAR TIME-VARYING CHANNEL

A propagation channel can be usually modeled as a linear random time-varying system defined by its impulse response $h(\tau, t)$ or equivalently by its (delay-Doppler) spreading function $S(\tau, \nu)$ so that the input $x(t)$ and the output $y(t)$ of this system satisfy

$$y(t) = \int_{-\infty}^{\infty} h(\tau, t) x(t - \tau) d\tau \quad (1)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S(\tau, \nu) x(t - \tau) e^{2i\pi\nu t} d\tau d\nu. \quad (2)$$

$S(\tau, \nu)$ characterizes the attenuation and scatterer reflectivity associated with paths of delay τ and Doppler ν . It verifies $S(\tau, \nu) = \int_{-\infty}^{\infty} h(\tau, t) e^{-2i\pi\nu t} dt$.

In a band-limited baseband equivalent discrete-time setting, the channel input-output relation satisfies

$$y[k] = \sum_{l=0}^{L-1} h_l[k] x[k-l] \quad (3)$$

where $x[k]$, $y[k]$, and $h_l[k]$ are the sampled versions of $x(t)$, $y(t)$, and $h(\tau, t)$, with sampling frequency larger than the system bandwidth. L denotes the total number of channel taps.

Generally, the statistical characterization of the channel is preferred to a deterministic description since it provides a more concise channel representation. In this case, the impulse response is modeled as a multi-variate random process $H(k) = [h_0[k], \dots, h_{L-1}[k]]$. The only assumption we made with regard to the statistical properties of this process, is that it is wide-sense stationary. This means that the first and the second moments of the channel do not vary with respect to time.

III. MAXIMUM ENTROPY MODELING

A. Gaussian channel model

The vast majority of models assume that the channel impulse response is a Gaussian process. The assumption usually relies on the physical argument that for a fixed delay the

received signal is the sum of the contribution of a large number of scattered reflections. Instead of being assumed, Gaussianity can be formally justified in light of inductive inference and the principle of maximum entropy [5].

For instance, in the case where the only known information about the UAC channel is that its total energy is finite and equal to σ^2 , the channel probability density function $P(H)$ can be obtained by maximizing the entropy

$$- \int \log P(H) dP(H) \quad (4)$$

under the constraint that

$$\int dP(H) = 1 \text{ and } \int \sum_{l=0}^{L-1} |h_l|^2 dP(H) = \sigma^2. \quad (5)$$

Note that the index k is voluntarily omitted when there are no references to time-related properties. This maximization problem can be solved using the method of Lagrange multipliers

$$\begin{aligned} \mathcal{L}(P) = & - \int \log P(H) dP(H) + \alpha \left(1 - \int dP(H) \right) + \\ & \lambda \left(\sigma^2 - \int \sum_{l=0}^{L-1} |h_l|^2 dP(H) \right) \end{aligned} \quad (6)$$

and setting the functional derivative of $\mathcal{L}(P)$ to 0

$$\frac{\delta \mathcal{L}(P)}{\delta P} = -\log P(H) - 1 - \alpha - \lambda \sum_{l=0}^{L-1} |h_l|^2 = 0. \quad (7)$$

Solving the system of equations (5) and (7) then leads to i.i.d and Gaussian distributed channel taps [5], i.e.,

$$P(H) = \frac{1}{(\pi\sigma^2/L)^L} \prod_{l=0}^{L-1} e^{-\frac{L|h_l|^2}{\sigma^2}}. \quad (8)$$

As opposed to classical channel model derivation, Gaussianity as well as taps independence were not prior assumptions but a consequence of the only constraint of finite average power.

To refine our model, deterministic ray-tracing models such as [8] can prove to be helpful. For instance, such physics-driven models can provide the channel power-delay profile for a given transmission set-up (see Figure 1). This means that for each tap l , we can get its average power $\mathbb{E}[|h_l|^2]$. Using [9], this average power can even be split-up as the sum of the power of the scatterers and a specular component such that $\mathbb{E}[|h_l|^2] = \sigma_l^2 + |s_l|^2$ where $s_l = \mathbb{E}[h_l]$ is the expression of the specular component (that can be null). In this case, the channel pdf that maximizes the entropy becomes

$$P(H) = \prod_{l=0}^{L-1} \frac{1}{(\pi\sigma_l^2)} e^{-\frac{|h_l - s_l|^2}{\sigma_l^2}}. \quad (9)$$

This result indicates that when the channel knowledge is limited to its first and second order statistics, the less arbitrary model to implement, according to the principle of maximum entropy, is the Gaussian model. A Gaussian pdf induces Rice or Rayleigh fading depending on whether s_l is null or not (note that $|s_l|^2/\sigma_l^2$ is usually known as the Rice factor). Subjecting the model to higher order constraints than second

order expectations would result to a different pdf. As shown in section IV-B, knowledge of the second order statistics is usually sufficient to induce accurate models.

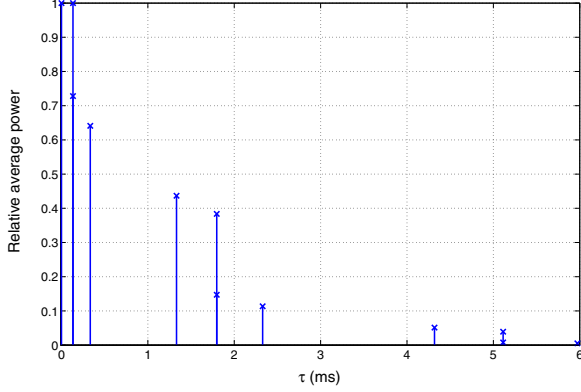


Fig. 1. Example of a shallow-water power-delay profile provided by ray-tracing. The transmitter and the receiver are at 20 m and 5 m above the sea-floor respectively, and separated by 1000 m in a 25 m deep water.

B. Power spectral density derivation of UAC fading processes

Eq. (8) and (9) indicate that without explicit knowledge of correlation between channel taps, these taps have to be considered as independent. This result justifies the use of the wide-sense stationary uncorrelated scattering (WSSUS) paradigm [10]. For WSSUS channels, the channel correlation is reduced from four to two dimensions so that scatterers with different delay or different Doppler are uncorrelated, i.e.,

$$\mathbb{E}[S(\tau, \nu) S^*(\tau', \nu')] = C(\tau, \nu) \delta(\tau - \tau') \delta(\nu - \nu') \quad (10)$$

where $C(\tau, \nu)$ is the so-called scattering function defined as

$$\begin{aligned} C(\tau, \nu) &= \mathbb{E}[|S(\tau, \nu)|^2] \\ &= \int_{-\infty}^{\infty} \mathbb{E}[h(\tau, t) h^*(\tau, t + \Delta t)] e^{-2i\pi\nu\Delta t} d\Delta t. \end{aligned} \quad (11)$$

In discrete-time, the scattering function becomes

$$C_l(\nu) = \sum_{u=-\infty}^{+\infty} \mathbb{E}[h_l[k] h_l^*[k+u]] e^{-2i\pi\nu u}. \quad (12)$$

For l fixed to $l = l_0$, $C_{l_0}(\nu)$ corresponds to the power spectral density or Doppler power spectrum of the random process $h_{l_0}[k]$. Note that $C_l(\nu)$ is discrete in delay but continuous in Doppler.

From Eq. (8) and (9), it can also be deduced that the joint entropy of the multivariate random process H can be expressed as the sum of the entropy of each process h_l . Thus, maximizing the overall entropy of H is equivalent to maximizing the entropy of the different h_l independently. Therefore, the scattering function that maximizes the entropy is obtained by independent modeling of the Doppler spectrum of each channel tap. For this reason and for the sake of readability the index l is voluntarily omitted throughout this subsection.

The dynamic of a stationary Gaussian process is fully characterized by its second order statistics or “color” that is expressed under the stationary assumption by its power spectral density or Doppler power spectrum $C(\nu)$. Thanks to Kolmogorov, the entropy rate $\bar{h}$ of a discrete-time Gaussian processes $h[k]$ can be expressed as a function of this power spectral density so that [11]

$$\bar{h} = \frac{1}{2} \log(2\pi e) + \frac{1}{2} \int_{-1/2}^{1/2} \log C(\nu) d\nu. \quad (13)$$

The process $h[k]$ is implicitly assumed to be sampled at a period that verifies Shannon’s theorem such that $C(\nu)$ has a bounded normalized support $[\nu_{\min}, \nu_{\max}]$ with $-1/2 \leq \nu_{\min} \leq \nu \leq \nu_{\max} \leq 1/2$.

To be consistent with the principle of maximum entropy as articulated previously, our objective is to find, for each tap, the $C(\nu)$ that maximizes the entropy rate $\bar{h}$ subject to some constraints that translates our state of knowledge of the channel time fluctuations. In general, this knowledge can be expressed for each tap as a set of spectral moments μ_p of known functions $m_p(\nu)$, $p \in \mathcal{P}$ where $\mathcal{P}$ denotes the set of constraint orders, i.e.,

$$\mu_p = \int_{\nu_{\min}}^{\nu_{\max}} m_p(\nu) C(\nu) d\nu. \quad (14)$$

In maximum entropy spectral estimation [12], it is classically assumed that at least N equally-spaced points of the autocorrelation function are known. In this case, μ_p corresponds to the autocorrelation function $\mu_p = \mathbb{E}[h[k] h^*[k+p]]$ that can also be expressed as trigonometric spectral moments such that $m_p(\nu) = e^{2i\pi p\nu}$ and $\mathcal{P} = \{0, 1, \dots, N-1\}$. Such representation of the channel may not be adapted to concise modeling since it requires the knowledge of N correlation points. A shortcut toward concise modeling could be to choose N small but this may not be a good option since it may not characterize the channel with enough accuracy. An alternative approach is to express our state of knowledge of the channel time fluctuations in terms of geometrical moments such that $m_p(\nu) = \nu^p$. Such moments have the main advantage of being easy to physically interpret so that their order of magnitude is usually known for a given environment. Typically, μ_0 is the total average power of $h[k]$, μ_1 is the barycenter of $C(\nu)$ (that can be seen as an indicator of spectral symmetry) and μ_2 expresses the (root-mean-square) Doppler spread D_s of $h[k]$ such that $D_s = \sqrt{\mu_2/\mu_0 - (\mu_1/\mu_0)^2}$.

Therefore, from Eq. (13) and (14), the proposed scattering function model is obtained by solving, for each tap, the following optimization problem

$$\begin{aligned} \max_{C(\nu)} \quad & \int_{\nu_{\min}}^{\nu_{\max}} \log C(\nu) d\nu, \\ \text{subject to} \quad & \int_{\nu_{\min}}^{\nu_{\max}} \nu^p C(\nu) d\nu = \mu_p, \forall p \in \mathcal{P}. \end{aligned} \quad (15)$$

As articulated in the previous subsection, the solution is provided thanks to the Lagrange multipliers

$$\mathcal{L}(C) = \int \log C(\nu) d\nu - \sum_{p \in \mathcal{P}} \lambda_p \left(\mu_p - \int \nu^p C(\nu) d\nu \right). \quad (16)$$

By differentiating $L(C)$ with respect to C , we get

$$C(\nu) = \frac{1}{\sum_{p \in \mathcal{P}} \lambda_p \nu^p}. \quad (17)$$

The multipliers λ_p are then found by solving the set of non-linear equations

$$\int_{\nu_{\min}}^{\nu_{\max}} \frac{\nu^q}{\sum_{p \in \mathcal{P}} \lambda_p \nu^p} d\nu = \mu_q, \quad \forall q \in \mathcal{P}. \quad (18)$$

Depending on the set $\mathcal{P}$, an analytical solution of such system of equations is in general not straightforward to obtain. Therefore, to facilitate the computation of the multipliers λ_p , a numerical approach such as the one presented in appendix A is advocated.

Eq. (15) highlights the flexibility of the proposed model. The set of constraints or parameters μ_p to feed into the model is not predefined so that it can be updated each time a new information on the environment to model becomes available.

IV. ILLUSTRATIONS

To illustrate the interest of the proposed maximum entropy framework, a simulation example of a shallow-water channel is first presented in subsection IV-A. The accuracy and the validity of our inductive approach is then discussed in subsection IV-B where fading statistics as well as bit error rates are assessed and compared with real data recorded at sea.

A. Example of a shallow-water channel impulse response

Thanks to the proposed inductive approach and from the only knowledge of some moments of the channel to model, it is possible to generate channel impulse response realizations that can be useful to assess the performance of UAC systems. Random channel realizations can be generated from the scattering function using sum-of-sinusoids methods or by Gaussian processes filtering [13]. Figure 2 shows an example of the output of a maximum entropy channel simulator. The power delay-profile used to generate the impulse response is the one of Figure 1 provided by ray-tracing. As we focus on concise modeling, the channel time fluctuations are here only constrained by the spectral moment μ_2 that is homogeneous to the square of the Doppler spread (for $p = 1$ and $\forall p > 2$, the moments μ_p are assumed to be unknown). The maximum Doppler frequency is set to 5 Hz and the power spectral density of each tap is obtained such that the rms Doppler spread increases linearly, from 0.5 to 2 Hz, with the tap delay (i.e. $\nu_{\max} = -\nu_{\min} = \frac{1}{2}$ and μ_2 ranges from $\left(\frac{0.5}{2 \times 5}\right)^2$ to $\left(\frac{2}{2 \times 5}\right)^2$). The Rice factor ($|s_l|^2 / \sigma_l^2$) is exponentially decreasing with the tap delay [14]. Note that this channel corresponds to the example of the Matlab toolbox available at <http://perso.telecom-bretagne.eu/fxsocheleau/software2>.

²For each channel tap to model, this toolbox also includes the Thorp filtering as presented in [15] (each path of an acoustic channel acts as a lowpass filter, introducing its own dispersion).

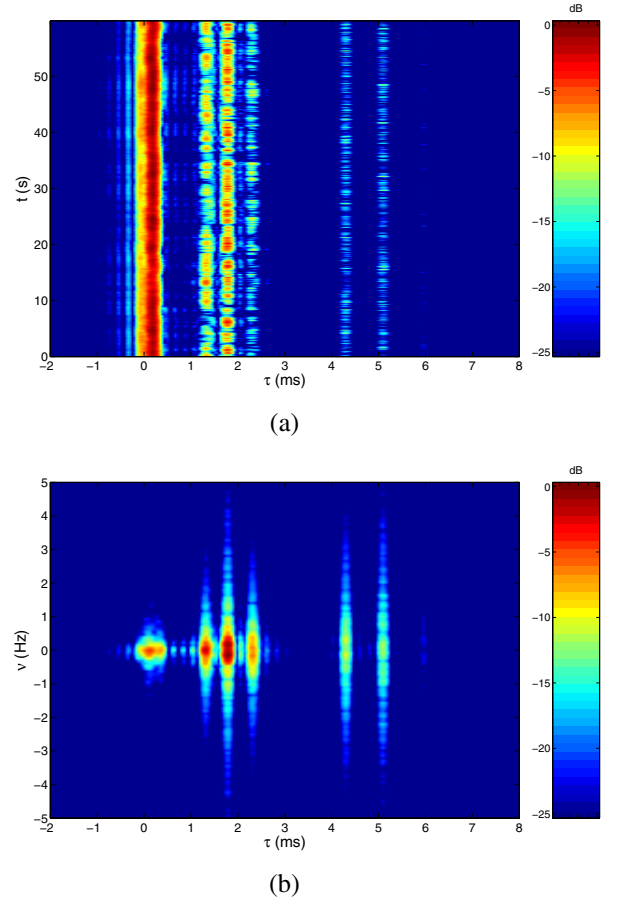


Fig. 2. Example of the maximum entropy channel simulator applied to UAC. (a) Envelope of the band-limited channel impulse response, (b) Channel scattering function without the specular components s_l . The transmission bandwidth is set to 5 kHz and the carrier frequency to 20 kHz.

B. Model versus real UAC channel

As emphasized in the introduction, the model is not meant to represent the physical reality of the channel but rather a state of knowledge of this reality. The “big question” is then: how much information does it need to characterize the environment to model with enough accuracy? The answer depends on what we intend to do with the model. In UAC, channel simulator are mainly employed

- 1) to help the design of communication systems and
- 2) to measure their performance in terms of bit error rate (BER).

Therefore, to assess the accuracy of the simulator, we suggest

- 1) to compare some fading statistics of a real channel recorded at sea and its equivalent maximum entropy model and
- 2) to compare the BER of a given communication system when it is faced to this real channel and when it is faced to the model.

Ideally, we would like to characterize the channel with the minimum of parameters which corresponds to the requirement of conciseness presented in the introduction. Typically, we would like to limit the parameters of the model to the statistics of order two or less with $p \leq 2$. Such statistics are easy to

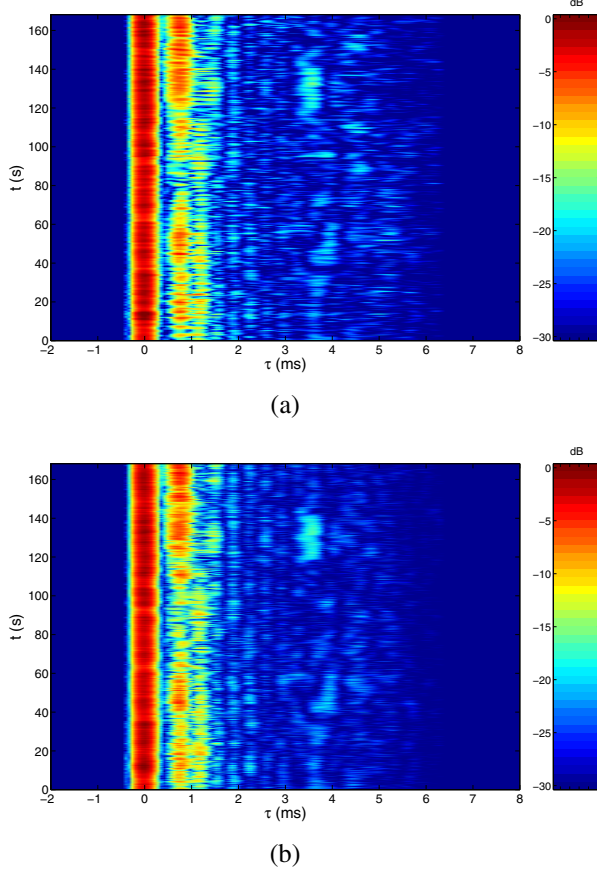


Fig. 3. Comparison between a real underwater acoustic channel probed in the Atlantic ocean and a realization of its equivalent maximum entropy model. (a) Real probed channel, (b) Maximum entropy channel. The estimated rms Doppler spread of each tap ranges from 1.1 to 4.8 Hz and the measured maximum Doppler frequency is 6.2 Hz.

physically interpret so that their order of magnitude is usually known for a given environment.

Figure 3-(a) shows an example of a real underwater acoustic channel probed in the Atlantic ocean at 17.5 kHz and over a bandwidth of 3 kHz. Figure 3-(b) displays a realization of the maximum entropy simulator that tries to mimic this probed channel. The simulator is constrained, for each tap, with the values of μ_0 and μ_2 measured on the real data recorded at sea. μ_0 is estimated by calculating the average power of each tap and μ_2 is estimated using the algorithm presented in [16] (for $p = 1$ and $\forall p > 2$, the moments μ_p are assumed to be unknown). The power of the specular component s_l of each tap is obtained thanks to the empirical mode decomposition as detailed in [14]. The maximum Doppler frequency $\nu_{\max}$ is measured using the method presented in [17].

The two channel impulse responses look visually similar. It then remains to verify that they have the same impact on the design and the performance of communications systems.

During the design phase of communication systems, the selection of error-correcting codes and interleavers is primarily driven by burst error statistics. These statistics are commonly given by the level crossing rate (LCR) and the average fade duration (AFD) that provide a useful means of characterizing the severity of the fading over time [13]. The LCR is defined as

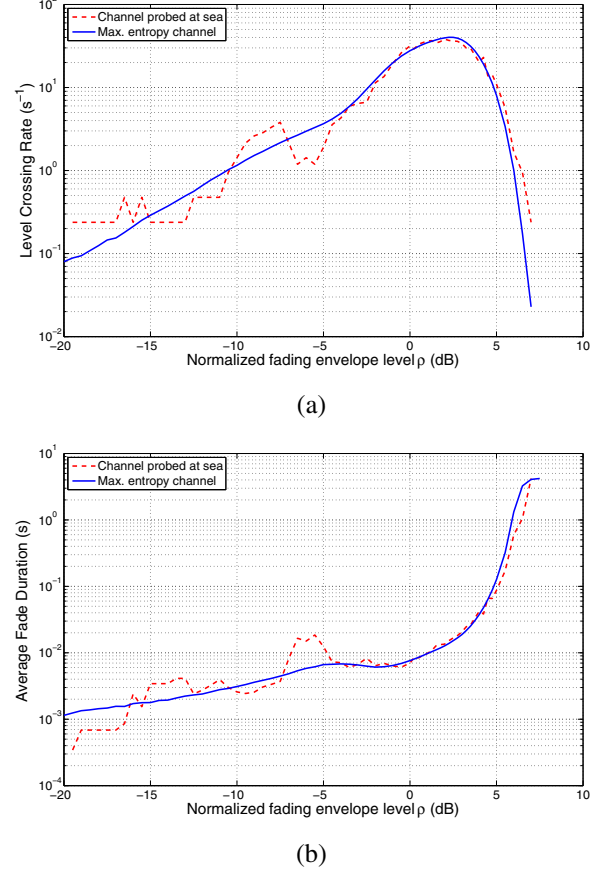


Fig. 4. LCR and AFD estimation on a channel recorded over 160 s in the Atlantic Ocean (the results of the maximum entropy model are averaged over 250 realizations). (a) LCR, (b) AFD.

the rate at which the envelope $r(k) = |\sum_l h_l[k - \lfloor lF_t/F_\tau \rfloor]|$ crosses a specified level ρ in the positive slope. $\lfloor \cdot \rfloor$ denotes the flooring operator, F_t and F_τ denote the channel sampling frequency in the time and delay domain respectively. In the discrete-time setting, this can be expressed in terms of probability as

$$\text{LCR} = P[r(k) \geq \rho, r(k-1) < \rho] \cdot F_t, \quad (19)$$

The AFD is defined as the average time that the fading envelope remains below a specified level after crossing that level in a downward direction

$$\text{AFD} = \frac{P[r(k) < \rho]}{\text{LCR}}. \quad (20)$$

Note that both the LCR and the AFD are strongly dependent upon the power spectral density of the fading process.

The results of the measured LCR and AFD on both channels are shown in Figure 4. It can be seen that there is a good match between the statistics of the two channels, especially for ρ between -5 and 5 dB. Outside this range, the LCR and AFD difference is imputable to the measurements on the original impulse response recorded at sea that provide estimates of the fading statistics with a relatively large variance. This impulse response indeed corresponds to a single realization of the probed channel over a finite time window of 160 seconds

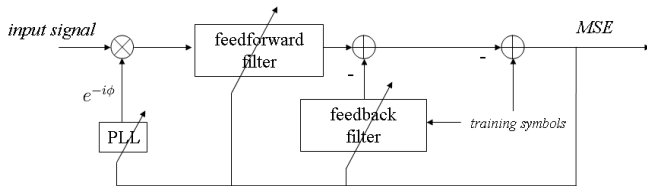


Fig. 5. Adaptive data-aided intersymbol interference canceler.

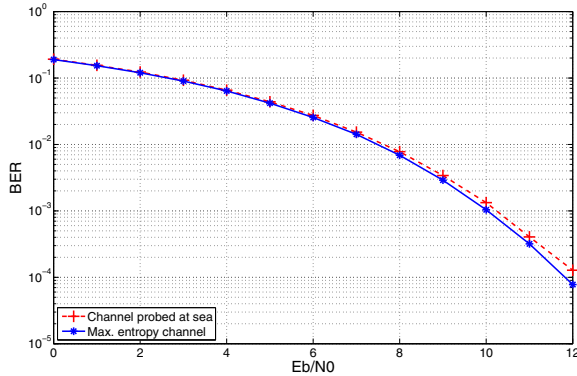


Fig. 6. BER comparison of the QPSK IC receiver when faced to the probed channel and the maximum entropy channel.

whereas the statistics of the maximum entropy model are averaged over 250 channel realizations.

To further illustrate the relevance of our model, we have also measured the performance of a QPSK communication link for several signal-to-noise ratios. As shown in Figure 5, the receiver is made of an adaptive data-aided intersymbol interference canceler (IC) [18] with joint phase tracking. Filters' coefficients are updated by a data-aided least-mean square algorithm (LMS-DA) and phase tracking is performed by a second order PLL. As the objective here is not to tune the receiver to obtain the best possible performance but to exhibit some possibilities offered by the channel simulator, the IC is data-aided over the entire input signal duration and takes into account the causal as well as the anti-causal interference. The IC is initialized by centering the strongest peak of the power delay profile in the feedforward filter at the start of the equalization process.

The performance of the QPSK IC receiver is plotted on Figure 6. It can be seen that the BER of this receiver is almost the same with both channels, the real one and the one provided by the simulator. A slight difference can be observed for low BER. This, once again, corresponds to an artifact due to the time limitation of the channels to 160 seconds which results in nonnegligible BER uncertainty.

As a conclusion, similar fading statistics and bit error rates in both configurations indicate that second order moments convey enough information to make the model accurate.

V. CONCLUSION

Throughout this paper, we have shown that the maximum entropy principle proves to be a relevant framework for a concise and consistent modeling of underwater acoustic communication channels. An exhaustive knowledge of the propagation phenomena is not required to obtain a channel model. Fading statistics and bit error rate measurements indicate that limiting this knowledge to the channel average power and the Doppler spread leads to accurate models. Moreover, the framework developed in this paper relies on the information theory formalism which communication engineers are familiar with. Finally, since the proposed model meets the consistency, conciseness and openness requirements, it is a good candidate for UAC channel standardization.

VI. ACKNOWLEDGMENT

The authors are grateful to the GESMA (Groupes d'Etudes Sous-Marines de l'Atlantique) for providing the experimental data used in this paper.

REFERENCES

- [1] P.C. Etter, *Underwater Acoustic Modeling and Simulation*, 3rd ed., Spon Press, 2003.
- [2] M.Stojanovic and J.Preisig, "Underwater Acoustic Communication Channels: Propagation Models and Statistical Characterization," *IEEE Commun. Mag.*, pp. 84–89, Jan. 2009.
- [3] R. Otnes, T. Jensenrud, J.E. Voldhaug, and C. Soldberg, "A Roadmap to Ubiquitous Underwater Acoustic Communications and Networking," in *Proc. Underwater Acoustic Measurement: Technologies and Results*, Jun. 2009.
- [4] E. T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge Univ. Press, 2003.
- [5] M. Debbah and R. R. Muller, "MIMO Channel Modeling and the Principle of Maximum Entropy," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1667–1690, 2005.
- [6] E.T. Jaynes, "Information Theory and Statistical Mechanics," *Physical review*, vol. 106, no. 4, pp. 620–630, 1957.
- [7] J. Shore and R. Johnson, "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy," *IEEE Trans. Inf. Theory*, vol. 26, no. 1, pp. 26–37, 1980.
- [8] M. Porter, "Bellhop gaussian beam/finite element beam code," 2007, <http://oalib.hlsresearch.com/Rays/index.html>.
- [9] X. Cristol, "NARCISSUS-2005: A Global Model of Fading Channel for Application to Acoustic Communication in Marine Environment," in *Proc. IEEE Oceans Conf.*, Jun. 2005.
- [10] P. A. Bello, "Characterization of randomly time-variant linear channels," *IEEE Trans. Commun. Systems*, vol. 11, no. 4, pp. 360–393, 1963.
- [11] T. Cover and J. Thomas, *Elements of Information Theory*, Wiley, 1991.
- [12] J. P. Burg, *Maximum Entropy Spectral Analysis*, Ph.D. thesis, Department of Geophysics, Stanford University, 1975.
- [13] M. Patzold, *Mobile fading channels*, Wiley, 2002.
- [14] F.-X. Socheleau, J.-M. Passerieux, and C. Laot, "Characterisation of Time-Varying Underwater Acoustic Communication Channel with Application to Channel Capacity," in *Proc. Underwater Acoustic Measurement: Technologies and Results*, Jun. 2009.
- [15] M. Stojanovic, "Underwater Acoustic Communications: Design Considerations on the Physical Layer," in *Proc. IEEE Conf. on Wireless On demand Network Systems and Services*, Jan. 2008.
- [16] P. Bello, "Some Techniques for the Instantaneous Real-Time Measurement of Multipath and Doppler Spread," *IEEE Trans. Commun. Tech.*, vol. 13, no. 3, pp. 285–292, 1965.
- [17] J. Holtzman and A. Sampath, "Adaptive Averaging Methodology for Handoffs in Cellular Systems," *IEEE Trans. Veh. Technol.*, vol. 44, no. 1, 1995.
- [18] M. S. Muller and J. Salz, "A unified theory of data-aided equalization," *Bell Syst. Techn. Journal*, vol. 60, no. 11, 1981.
- [19] A. M. Djafari, "A Matlab Program to Calculate the Maximum Entropy Distributions," http://djafari.free.fr/pdf/me91_1.pdf.

APPENDIX A

NUMERICAL SOLUTION OF EQUATION (18)

To obtain the lagrange multipliers λ_p of Eq. (18), a simple gradient based method as the one presented in [19] is suggested. The vector $\mathbf{\Lambda} = [\lambda_p]_{p \in \mathcal{P}}$ is found by solving the system

$$G_q(\mathbf{\Lambda}) = \int_{\nu_{\min}}^{\nu_{\max}} \frac{\nu^q}{\sum_{p \in \mathcal{P}} \lambda_p \nu^p} d\nu = \mu_q, \quad q \in \mathcal{P}. \quad (21)$$

This equation is solved iteratively by first developing the $G_q(\mathbf{\Lambda})$ in Taylor's series around an initial vector $\mathbf{\Lambda}^0$.

$$G_q(\mathbf{\Lambda}) \approx G_q(\mathbf{\Lambda}^0) + (\mathbf{\Lambda} - \mathbf{\Lambda}^0)^T \text{grad}(G_q(\mathbf{\Lambda}))_{(\mathbf{\Lambda}=\mathbf{\Lambda}^0)}. \quad (22)$$

If ϵ and v are defined as

$$\epsilon = \mathbf{\Lambda} - \mathbf{\Lambda}^0, \quad (23)$$

$$v = [\mu_q - G_q(\mathbf{\Lambda}^0)]_{q \in \mathcal{P}}^T \quad (24)$$

and

$$\Gamma = (\gamma_{qn}) = \left(\frac{\partial G_q(\mathbf{\Lambda})}{\partial \lambda_n} \right)_{(\mathbf{\Lambda}=\mathbf{\Lambda}^0)} \quad (25)$$

then,

$$\Gamma \epsilon = v. \quad (26)$$

Note that the matrix Γ is symmetric so that

$$\gamma_{qn} = \gamma_{nq} = - \int_{\nu_{\min}}^{\nu_{\max}} \frac{\nu^{q+n}}{\left(\sum_{p \in \mathcal{P}} \lambda_p \nu^p \right)^2} d\nu \quad (27)$$

At each iteration, ϵ is computed and a new $\mathbf{\Lambda}^0$ that verifies $\mathbf{\Lambda}^0 = \mathbf{\Lambda} - \epsilon$ is set. The iterations continue until ϵ is small enough.