IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 26, 2019, accepted August 15, 2019, date of publication September 4, 2019, date of current version October 4, 2019.

*Digital Object Identifier 10.1109/ACCESS.2019.2939498*

# Rain Attenuation Along Terrestrial Millimeter Wave Links: A New Prediction Method Based on Supervised Machine Learning

**SPIROS N. LIVIERATOS[1] AND PANAYOTIS G. COTTIS[2]**

[1]Department of Electrical and Electronic Engineering, School of Pedagogical and Technological Education, 14121 Athens, Greece
[2]School of Electrical and Computer Engineering, National Technical University of Athens, 15780 Athens, Greece

Corresponding author: Spiros N. Livieratos (slivieratos@aspete.gr)

This work was supported by the Special Account for Research of ASPETE through the funding program "Strengthening ASPETE's Research."

**ABSTRACT** During the current decade, wireless data traffic has been increasing very rapidly, a trend which is expected to accelerate over the next decade driven by the widespread use of video streaming and the rise of the Internet-of-Things (IoT). In this framework, cellular technology is rapidly moving towards its 5$^{th}$ generation (5G) that will employ millimeter wave (mmWave) frequencies in the attempt to exploit more spectrum and offer multi-Gigabit-per-second (Gbps) data rates to mobile devices. Various propagation phenomena affect adversely mmWave communications, rain fading being the most severe one. The existing ITU-R prediction model for rain induced attenuation over terrestrial line-of-sight (LOS) links does not perform accurately on a global level. This weakness constitutes the main motivation to formulate enhanced models which, by employing appropriate attributes, apply more satisfactorily to specific locations or climatic zones. ITU-R databank includes experimental data of real LOS links operating in various locations that can be used to facilitate supervised machine learning (SML) to formulate methods towards accurate prediction of rain attenuation. Based on a set of past examples or instances, SML aims at exploring/identifying the relationship between a set of descriptive features (inputs) and a target feature (output). After been appropriately trained with past data, SML can be used to make predictions about new instances. This paper proposes a new prediction method which uncovers the latent dependence of rain attenuation on predictors such as path length, operation frequency, wave polarization, rain rate distribution, etc. ensuring high prediction accuracy without necessitating complex mathematical expressions.

**INDEX TERMS** Gaussian processes, machine learning, millimeter wave communications, rain fading, regression, supervised learning, wireless networks.

## I. INTRODUCTION

Learning means finding patterns from previous experience in the attempt to deal with unknown situations. Learning comes as the result of repeatedly observing meaningful indicators that affect the problem each time in hand. When computers (machines) are involved, the repeated observations come in the form of data whereas the solution to a new problem may be perceived/obtained as the output of an algorithm. Machine learning (ML) aims at automating the process of extracting knowledge from experience in order to make a prediction concerning an unknown situation. ML emerged as a sub-discipline of artificial intelligence and has been applied in areas such as computer perception, communication and reasoning [1]. ML constitutes an alternative for data-driven decision making or prediction and has become one of the most powerful artificial intelligence tools [2].

ML is usually classified as supervised or unsupervised machine learning [3]. In supervised machine learning (SML), the goal is to learn (determine) how a set of inputs is related to a set of outputs, given a labeled (known) set of input-output pairs. In unsupervised learning, sometimes called knowledge discovery, where only inputs exist, the goal is to extract interesting patterns governing a set of input data. For two reasons, problems addressed employing unsupervised learning are not well-defined. First, it is not a priori known what kinds of patterns to search for. Second, unlike SML - where the prediction of the output given the inputs can be compared to observed values-, there is no error metric to use. Another type of ML, known as reinforcement learning, is also in use,

The associate editor coordinating the review of this manuscript and approving it for publication was Sajjad A. Madani.

aiming at learning how to act or behave forced by appropriate reward or punishment signals.

To address the increasing, at a rate of over 50% per year per subscriber, demand in wireless data traffic, wireless networks are migrating to the 5th generation (5G) standard which will use millimeter wave (mmWave) frequencies ranging from 30GHz to 300GHz, as they offer channel bandwidths more than ten times wider than the bandwidth offered by 4G long-term evolution (LTE) networks [4]. Using service-driven 5G networks, the operators aim at flexibly and efficiently providing services such as enhanced mobile broadband, ultra-reliable and low-latency communications and massive machine type communications [5]. 5G networks should also support backward compatibility with 4G-LTE and Wi-Fi. Since, compared to the microwave frequencies currently used by LTE, the wavelengths at mmWave frequencies are shorter by an order of magnitude ranging from 1mm to 10mm, atmospheric phenomena such as precipitation and diffraction cause stronger attenuation/fading. Hence, the impact of atmospheric phenomena in the design of new mmWave communication systems becomes critical, necessitating accurate prediction. Over the past years, measurements and prediction models concerning a plethora of propagation scenarios regarding terrestrial line-of-sight (LOS) and satellite links have been proposed by many companies and research groups [6]–[8]. The propagation problems related to mmWave communications, most importantly attenuation due to rain, affect the physical layer and, subsequently, the medium access control layer and higher layers; hence, their expected severe impact on 5G wireless networks necessitates proper handling.

Due to its stochastic behavior with regard to duration, location and occurrence frequency, rainfall is a complex meteorological phenomenon. Since, for any location on Earth, the statistical distribution of rain attenuation is obtained from local data concerning the rain rate distribution, the accuracy of rain rate measurements drastically affects the estimation accuracy of rain induced attenuation. The existing ITU-R prediction model for rain induced attenuation over LOS links does not perform accurately on a global level. This constitutes the main motivation to formulate alternative models which, by employing appropriate attributes, apply better to specific locations or climatic zones. In any case, it is not an easy task to employ a complex prediction model for rain attenuation that applies to any location and climatic region. However, the availability of a plethora of statistically stable measurements concerning rain rate and rain induced attenuation over real microwave links operating in various locations encourages the application of SML based regression methods expecting to achieve more accurate predictions. By appropriately training the relevant algorithms, a prediction method may be formulated which:

i. uncovers the latent dependence of rain attenuation on factors such as path length, operation frequency, wave polarization, rain rate distribution, etc., without having to employ complex mathematical expressions
ii. ensures high prediction accuracy

Concluding, SML can enhance the procedure of uncovering the latent relationship between rain rate and rain attenuation which cannot be captured by classic statistical/analytic methods [5]. By improving the estimation of the spatiotemporal behavior of rain attenuation, SML gives rise to a novel procedure towards calculating the rain fade margins necessary for optimizing the deployment and operation of wireless networks operating above 5GHz, especially above 10GHz.

The rest of the paper is organized as follows. Section II presents the rain attenuation prediction models currently in use. Section III presents the basic mathematical background referring to (i) SML, (ii) Gaussian processes (GPs) and (iii) regression techniques that constitute the necessary steps towards developing the proposed prediction method. In Section IV, the proposed method is analyzed. Also, its prediction accuracy is validated by performing performance comparison with relevant prediction models, taking into account real data for rain rate and rain attenuation extracted from the experimental databank of ITU-R. Finally, Section V concludes the paper and presents fields for further study and application of the proposed method.

## II. EXISTING PREDICTION MODELS FOR RAIN ATTENUATION

In LOS terrestrial links or earth-space links operating above 5GHz, especially above 10GHz, the occurrence of rain along the transmission path constitutes the most important factor degrading system performance. The rain attenuation along a terrestrial path is determined by multiplying the specific attenuation $\gamma_R$ (dB/km) with the effective propagation path length $d_{eff}$ (km). $\gamma_R$, which is the main parameter characterizing rain attenuation on a local basis, depends on the operation frequency, the wave polarization, and the geographical coordinates [9], [10]. The ITU-R Recommendation P.838–3 [11] establishes the procedure relating $\gamma_R$ to the local rain intensity, particularly to the parameter $R_{p_{exc}}$(mm/h), which is determined as the rain rate level exceeded for $p_{exc}$ time percentage, that is

$$p_{exc} = Prob\left(Rain\ Rate > R_{p_{exc}}\right) \tag{1}$$

normalized on a per year basis. Based on $R_{p_{exc}}$, $\gamma_R$ is determined using the power law relationship

$$\gamma_R = k(R_{p_{exc}})^{\alpha} \tag{2}$$

where $k$ and $a$ depend on the frequency and polarization of the electromagnetic wave and on the link elevation angle. Tables tabulating $(k, a)$ pairs for many locations on Earth are provided in [11]. Moreover, actual values of $k$ and $a$ can be obtained via interpolation using a logarithmic scale for $k$ and a linear scale for $a$. Then, the rain attenuation $A$(dB) which is exceeded for $p_{exc}$ time percentage on a per year basis is calculated from

$$A = \gamma_R d_{eff} = \gamma_R L r \tag{3}$$

where $d_{eff}$ is determined by multiplying the actual radio link length $L$ (km) with the path reduction factor $r$ evaluated for a $p_{exc}$ time percentage. The purpose of introducing the path reduction factor is to replace the actual path length with a hypothetical length equivalently affected by uniform point rainfall. $r$ is addressed by the various existing rain models for rain attenuation prediction as a parameter intended to render the prediction of rain attenuation more accurate, if it is properly determined.

Various prediction models for rain attenuation are employed in the design of terrestrial or satellite links based either on cumbersome statistical regression - in case sufficient local experimental data is available- or on analytical models- in case only local rain rate measurements are available. However, it is neither straightforward nor sometimes feasible to apply a unique prediction model for rain attenuation employing a complex algebraic statistical expression that fits to any location or climatic region. Four of the most frequently employed rain attenuation models are following.

### A. ITU-R P. 530-16 MODEL

Based on the previous approach for the calculation of the specific attenuation, the ITU-R Recommendation P.530-16 [12] determines the path attenuation exceeded for 0.01% of the time. $R_{0.01}$, which is the rain rate exceeded for 0.01% of the time in a year, is employed in the numerical calculations. If this information is not locally available, an estimate can be obtained using the information given in ITU-R Recommendation P.837-7 [13]. Employing an empirical formula, the results obtained are scaled to percentages of time that range from 1% to 0.001%. This method is proposed for locations over the world where the respective national authority for telecommunications recommends rain attenuation be considered for any operating frequency from 5GHz to 100GHz with path lengths up to 60 km. The relevant calculations can be found in the Appendix.

### B. SILVA-MELO MODEL

This model uses the numerical coefficients that are derived for effective rain rate and equivalent rain cell diameter that were obtained by multiple non-linear regressions, using the measured data available in the ITU-R databank. Details of the model are fully reported in [14] whereas the relevant calculations can be found in the Appendix.

### C. MOUPFOUMA MODEL

This model uses only the parameter $R_{0.01}$(mm/h) which, in the area of interest, represents the rain rate value exceeded for 0.01% per year. This model does not need rain rate numerical values for all time percentages. The detailed approach can be found in [15] and the relevant calculations can be found in the Appendix.

### D. LIN MODEL

The methodology proposed by Lin in [16] employs the path reduction factor to estimate rain attenuation statistics on terrestrial links. The method accounts for partially correlated rain rate variations along the propagation path length. The relevant calculations can be found in the Appendix.

## III. SUPERVISED MACHINE LEARNING / REGRESSION AIDED BY GAUSSIAN PROCESSES

SML can be employed in solving either regression problems or classification problems. In contrast to classification problems where discrete classes of outputs are sought, regression problems deal with the prediction of continuous quantities. A training set includes the inputs $x_1^{in}, x_2^{in}, \ldots, x_d^{in}$ and the output $y$ incorporating $n$ observations. The inputs and outputs are alternatively called predictors/features and targets, respectively. An $n \times d$ matrix $\boldsymbol{X} = \left[\boldsymbol{x_1^{in}} \boldsymbol{x_2^{in}} \ldots \boldsymbol{x_d^{in}}\right]$, where $\boldsymbol{x_1^{in}}, \boldsymbol{x_2^{in}}, \ldots, \boldsymbol{x_d^{in}}$ are $n \times 1$ column vectors, is introduced to denote the $n$ observations of the inputs and an $n \times 1$ column vector $\boldsymbol{y}$ is used to denote the $n$ observations of the output. The known inputs and outputs that constitute the training dataset are organized as a single $n \times (d+1)$ matrix $\boldsymbol{D} = [\boldsymbol{X}\ \boldsymbol{y}]$ [17]. In addition to the training set, a test set exists which is the dataset including the test observations known for certain pairs of inputs and outputs. The test set will be used to check the accuracy of the prediction method after the application of the SML prediction algorithm. By properly processing the training data, the main objective of SML is to make inference about the relationship between the inputs and the output, i.e. about the conditional probability distribution of the output given the inputs. To replicate/predict the output observations $\boldsymbol{y}$ without knowing the exact multivariable function $\boldsymbol{y} = f(\boldsymbol{X})$, the optimal approach is to infer a conditional probability distribution, $p(f|\boldsymbol{X}, \boldsymbol{y})$, over possible regression functions $f$ given the training data $\boldsymbol{X}$ (inputs) and $\boldsymbol{y}$ (output). Next, $p(f|\boldsymbol{X}, \boldsymbol{y})$ will be used to determine the estimated output following the Bayesian prediction property for the test data [18]

$$p\left(\boldsymbol{y}_*|\boldsymbol{X}_*, \boldsymbol{X}, \boldsymbol{y}\right) = \int p\left(\boldsymbol{y}_*|f, \boldsymbol{X}_*\right) p\left(f|\boldsymbol{X}, \boldsymbol{y}\right) df \qquad (4)$$

where $\boldsymbol{X}_*$ is the test input matrix and $\boldsymbol{y}_*$ is the test output vector. The values of the test output will be compared with the predicted ones, denoted as $\boldsymbol{f}_*$, which will be defined later in the paper. $\boldsymbol{f}_*$ will be calculated taking as input arguments the test inputs $\boldsymbol{X}_*$. There is no need to determine the statistical distribution of the inputs. Moreover, the prediction method will be built based on the training dataset $\boldsymbol{D}$. Applying the proposed prediction method, the latent relationship between the training inputs $\boldsymbol{X}$ and the training output$\boldsymbol{y}$ will be captured.

Under the assumption that the $n$ output observations are normally distributed, the proposed SML based prediction method will be built employing GPs. As GP regression (GPR) models are non-parametric kernel-based probabilistic models, the proposed method constitutes a GPR approach, provided that the covariance between any pair of the previous output observations is calculated from the kernel function adopted. A GP defines a prior distribution over possible regression functions which can be converted into a posterior

distribution over possible regression functions if sufficient experimental data is available [18]. A new output can be predicted from any new set of inputs by combining the GP prior distribution with a Gaussian likelihood function for each of the observations. The posterior distribution that comes up is also Gaussian with mean and covariance that can be easily computed by properly processing the observation data. Although it seems difficult to represent a distribution over a function, it turns out that it suffices to define a distribution over a finite set of observations of the function at the points, say $\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n}$, where $\boldsymbol{x_i}$ is a $1 \times d$ row vector representing the inputs of the $i^{th}$ observation or equivalently the $i^{th}$ row of $\boldsymbol{X}$ matrix. A GP assumes that the probability distribution $p\left(f\left(\boldsymbol{x_1}\right), \ldots, f\left(\boldsymbol{x_n}\right)\right)$ is jointly Gaussian, with mean value

$$m\left(\boldsymbol{x_i}\right) = E\left[f\left(\boldsymbol{x_i}\right)\right] \tag{5}$$

and covariance given by

$$\begin{aligned} K_{ij} &= cov\left(f\left(\boldsymbol{x_i}\right), f\left(\boldsymbol{x_j}\right)\right) = k\left(\boldsymbol{x_i}, \boldsymbol{x_j}\right) \\ &= E\left[\left(f\left(\boldsymbol{x_i}\right) - m\left(\boldsymbol{x_i}\right)\right)\left(f\left(\boldsymbol{x_j}\right) - m\left(\boldsymbol{x_j}\right)\right)\right] \end{aligned} \tag{6}$$

where $k(.,.)$ is a positive definite kernel function. The key idea of a kernel function is that if the inputs $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ are almost the same, the respective outputs will also be almost the same. Let the prior distribution over possible regression functions $f$ be a GP with parameters expressed in (5) and (6) [18]. For any finite set of inputs, the conditional distribution over the regression function $f$ is a joint Gaussian distribution

$$p\left(f|\boldsymbol{X}\right) = N\left(\boldsymbol{f}|\boldsymbol{\mu}, \boldsymbol{K}\right) \tag{7}$$

where

$$N\left(f|\boldsymbol{\mu}, \boldsymbol{K}\right) \triangleq \frac{1}{2\pi^{n/2}|\boldsymbol{K}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{K}^{-1}(\mathbf{x}-\boldsymbol{\mu})} \tag{8}$$

$$\boldsymbol{\mu} = \left(m\left(\boldsymbol{x_1}\right), \ldots, m\left(\boldsymbol{x_n}\right)\right) \tag{9}$$

and the elements of $\boldsymbol{K}$ are given from (6).

In general, when dealing with realistic situations

$$y = f\left(\boldsymbol{x}\right) + \varepsilon \tag{10}$$

as there are a lot of external random factors that may affect the observations by adding noise $\varepsilon$. Assuming additive independent identically distributed Gaussian noise $\varepsilon$ with variance $\sigma_n^2$, the covariance of the noisy observations is given from [19]

$$\begin{aligned} cov\left(y_p, y_q\right) &= k\left(\boldsymbol{x_p}, \boldsymbol{x_q}\right) + \sigma_n^2 \delta_{pq} \quad \text{or} \\ cov\left(\boldsymbol{y}\right) &= \boldsymbol{K}\left(\boldsymbol{X}, \boldsymbol{X}\right) + \sigma_n^2 \boldsymbol{I} = \boldsymbol{K_y} \end{aligned} \tag{11}$$

where $\delta_{pq}$ is the Kronecker delta, which is one, if $p = q$, and zero, otherwise. The joint distribution of the observed values $\boldsymbol{y}$ of the output and the predicted (function) values $\boldsymbol{f_*}$ at the test points under the prior distribution is

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{f_*} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \boldsymbol{K}\left(\boldsymbol{X}, \boldsymbol{X}\right) + \sigma_n^2 \boldsymbol{I} & \boldsymbol{K}\left(\boldsymbol{X_*}, \boldsymbol{X}\right) \\ \boldsymbol{K}\left(\boldsymbol{X}, \boldsymbol{X_*}\right) & \boldsymbol{K}\left(\boldsymbol{X_*}, \boldsymbol{X_*}\right) \end{bmatrix}\right) \tag{12}$$

If there are $n$ training observations and $n_*$ test observations, then $\boldsymbol{K}\left(\boldsymbol{X}, \boldsymbol{X_*}\right)$ is the $n \times n_*$ matrix of the covariances evaluated at all pairs of training and test observations. Similarly, $\boldsymbol{K}\left(\boldsymbol{X}, \boldsymbol{X}\right)$ is the $n \times n$ matrix of the covariances evaluated at all pairs of training observations, $\boldsymbol{K}\left(\boldsymbol{X_*}, \boldsymbol{X_*}\right)$ is the $n_* \times n_*$ matrix of the covariances evaluated at all pairs of test observations and $\boldsymbol{K}\left(\boldsymbol{X_*}, \boldsymbol{X}\right)$ is the $n_* \times n$ matrix of the covariances evaluated at all pairs of test and training observations. The conditional posterior distribution required for the predictions is given from

$$p\left(\boldsymbol{f_*}|\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{X_*}\right) \sim N\left(\overline{f_*}, cov\left(\boldsymbol{f_*}\right)\right) \tag{13}$$

where

$$\begin{aligned} \overline{f_*} &\triangleq E\left[\boldsymbol{f_*}|\boldsymbol{X}, \boldsymbol{y}, \boldsymbol{X_*}\right] \\ &= \boldsymbol{K}\left(\boldsymbol{X_*}, \boldsymbol{X}\right)\left[\boldsymbol{K}\left(\boldsymbol{X}, \boldsymbol{X}\right) + \sigma_n^2 \boldsymbol{I}\right]^{-1} \boldsymbol{y} = \boldsymbol{K}\left(\boldsymbol{X_*}, \boldsymbol{X}\right) \boldsymbol{K_y^{-1}} \boldsymbol{y} \end{aligned} \tag{14}$$

and

$$\begin{aligned} cov\left(\boldsymbol{f_*}\right) &= \boldsymbol{K}\left(\boldsymbol{X_*}, \boldsymbol{X_*}\right) - \boldsymbol{K}\left(\boldsymbol{X_*}, \boldsymbol{X}\right)\left[\boldsymbol{K}\left(\boldsymbol{X}, \boldsymbol{X}\right) + \sigma_n^2 \boldsymbol{I}\right]^{-1} \\ \boldsymbol{K}\left(\boldsymbol{X}, \boldsymbol{X_*}\right) &= \boldsymbol{K}\left(\boldsymbol{X_*}, \boldsymbol{X_*}\right) - \boldsymbol{K}\left(\boldsymbol{X_*}, \boldsymbol{X}\right) \boldsymbol{K_y^{-1}} \boldsymbol{K}\left(\boldsymbol{X}, \boldsymbol{X_*}\right) \end{aligned} \tag{15}$$

To simplify the notation, the compact notations $\boldsymbol{K} = \boldsymbol{K}\left(\boldsymbol{X}, \boldsymbol{X}\right)$ and $\boldsymbol{K_*} = \boldsymbol{K}\left(\boldsymbol{X}, \boldsymbol{X_*}\right)$ are introduced. In case there is only one test observation $\boldsymbol{x_*}$, $\boldsymbol{k}(\boldsymbol{x_*})= \boldsymbol{k_*}$ is set to denote the vector of covariances between the test observation and the $n$ training observations. Using the above compact notations, (14) and (15) reduce to

$$\overline{f_*} = \boldsymbol{k_*^T}(\boldsymbol{K} + \sigma_n^2 \boldsymbol{I})^{-1} \boldsymbol{y} = \boldsymbol{k_*^T} \boldsymbol{K_y^{-1}} \boldsymbol{y} \tag{16}$$

$$\begin{aligned} var\left(\boldsymbol{f_*}\right) &= \boldsymbol{k}\left(\boldsymbol{x_*}, \boldsymbol{x_*}\right) - \boldsymbol{k_*^T}(\boldsymbol{K} + \sigma_n^2 \boldsymbol{I})^{-1} \boldsymbol{k_*} \\ &= \boldsymbol{k}\left(\boldsymbol{x_*}, \boldsymbol{x_*}\right) - \boldsymbol{k_*^T} \boldsymbol{K_y^{-1}} \boldsymbol{k_*} \end{aligned} \tag{17}$$

The function values $f_*$-corresponding to test inputs $\boldsymbol{X_*}$-can be properly sampled from the joint posterior distribution given from (13), by evaluating the mean and covariance matrix using (16) and (17). The predictive performance of GPs depends on the suitability of the type of kernel selected to represent the covariance function. A stationary kernel $k(\boldsymbol{x}, \boldsymbol{x}')$ is a function of the difference $\boldsymbol{x} - \boldsymbol{x}'$, where $\boldsymbol{x}$ and $\boldsymbol{x}'$ are $1 \times d$ vector input observations. Moreover, if the kernel is a function of $|\boldsymbol{x} - \boldsymbol{x}'|$ is called isotropic. Indicative isotropic kernels combined with GPs are the following:

Squared exponential (SE) class

$$k\left(\boldsymbol{x}, \boldsymbol{x}'\right) = \exp\left(-\frac{|\boldsymbol{x} - \boldsymbol{x}'|^2}{2l^2}\right) \tag{18}$$

Exponential (EXP) class

$$k\left(\boldsymbol{x}, \boldsymbol{x}'\right) = \exp\left(-\frac{|\boldsymbol{x} - \boldsymbol{x}'|}{l}\right) \tag{19}$$

Rational Quadratic (RQ) class

$$k\left(\boldsymbol{x}, \boldsymbol{x}'\right) = \exp\left(-\alpha\, ln\left(1 + \frac{|\boldsymbol{x} - \boldsymbol{x}'|^2}{2\alpha l^2}\right)\right) \tag{20}$$

Matern class

$$k\left(\boldsymbol{x}, \boldsymbol{x}'\right) = \frac{1}{2^{v-1}\Gamma\left(v\right)}\left(\frac{\sqrt{2v}}{l}\left|\boldsymbol{x} - \boldsymbol{x}'\right|\right)^v K_v\left(\frac{\sqrt{2v}}{l}\left|\boldsymbol{x} - \boldsymbol{x}'\right|\right) \tag{21}$$

The parameter $l$ defines the characteristic length-scale. For the Matern class in particular, the parameters $v$ and$l$ are positive, whereas $K_v$ is the modified Bessel function [20]. The Matern kernel function becomes simple when $v$ is half-integer, that is $v = z + 1/2$, where $z$ is a non-negative integer. The exponential class comes as a special case of the Matern class by setting $v = 1/2$. The rational quadratic class with $\alpha > 0$ and $l > 0$ can be seen as a scale mixture (an infinite sum) of squared exponential kernel functions with different characteristic length-scales.

To estimate the parameters of the selected kernel function, an empirical Bayes approach is employed which will allow to use much faster continuous optimization methods. In particular, the marginal likelihood as expressed in (4) will be maximized. The relevant mathematical analysis, which is quite cumbersome, can be found in [18]. The quantity to be maximized is

$$logp\left(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{\vartheta}\right) = -\frac{1}{2}\boldsymbol{y^T K_y^{-1} y} - \frac{1}{2}\log\left|\boldsymbol{K_y}\right| - \frac{n}{2}\log(2\pi) \quad (22)$$

where $\boldsymbol{\vartheta}$ represents the vector of kernel parameters, which are denoted as $l$ , $\alpha$ and $v$ in the previously mentioned kernel classes. The kernel parameters are also called hyper-parameters where $\boldsymbol{K_y}$ has been defined in (11). To obtain the hyper-parameters as a result of maximization of the marginal likelihood, the partial derivatives of the marginal likelihood with respect to the hyper-parameters are taken

$$\begin{aligned}\frac{\partial}{\partial\vartheta_j} logp\left(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{\vartheta}\right) &= \frac{1}{2}\boldsymbol{y^T K_y^{-1}}\frac{\partial \boldsymbol{K_y}}{\partial\vartheta_j}\boldsymbol{K_y^{-1} y} - \frac{1}{2}tr\left(\boldsymbol{K_y^{-1}}\frac{\boldsymbol{\partial K_y}}{\boldsymbol{\partial\vartheta_j}}\right)\\ &= \frac{1}{2}tr\left(\left(\boldsymbol{\alpha\alpha^T} - \boldsymbol{K_y^{-1}}\right)\frac{\partial \boldsymbol{K_y}}{\partial\vartheta_j}\right) \quad (23)\end{aligned}$$

where

$$\boldsymbol{\alpha} = \boldsymbol{K^{-1} y} \quad (24)$$

and the trace function of a $n \times n$ square matrix M is defined as

$$tr\left(\boldsymbol{M}\right) = \sum_{i=1}^{n} M_{ii} \quad (25)$$

The form of $\frac{\partial \boldsymbol{K_y}}{\partial\vartheta_j}$ depends on the kernel class. It is also mathematically related to the parameter with respect to which the partial derivative is taken.

Finally, to avoid overfitting, i.e. avoid the danger to fit exactly the training data and fail to reliably predict future observations, cross-validation is employed. The k-fold cross-validation splits the data into k disjoint, equally sized subsets. Validation is done on a single subset, which has the role of the so called test set and training is done using the union of the remaining k − 1 subsets. The former has the role of the so called training set. The entire procedure is repeated k times, each time with a different subset for validation/testing. Thus, a large proportion of the data can be used for training whereas all cases appear as validation cases. Typical values for k are in the range 3 to 10. In this work, k = 5.

Following the procedure mentioned above, the numerical implementation of GPRs is shown in the Algorithm below [19]:

1. inputs: $\boldsymbol{X}$ (predictors), $\boldsymbol{y}$ (target), $\boldsymbol{k}$ (covariance function), $\boldsymbol{\sigma_n^2}$ (noise level), $\boldsymbol{x_*}$ (test input), $\boldsymbol{y_*}$ (test output)
2. $\boldsymbol{L} = Cholesky(\boldsymbol{K_y})$ : decomposition of $\boldsymbol{K_y}$ following the Cholesky method where $\boldsymbol{K_y} = \boldsymbol{LL^T}$
3. $\boldsymbol{\alpha} = \boldsymbol{L^T}\left(\boldsymbol{L}\backslash\boldsymbol{y}\right)$ : this calculation is $\boldsymbol{\alpha} = \boldsymbol{K_y^{-1} y} = \boldsymbol{L^{-T} L^{-1} y}$
4. $\bar{f}_* = \boldsymbol{k_*^T a}$ : calculation of the predictive mean, as in (16)
5. $\boldsymbol{V} = \boldsymbol{Lk_*} = \boldsymbol{L^{-1} k_*}$ : calculation of the predictive variance, as in (17)
6. $var\left(\boldsymbol{f_*}\right) = k\left(\boldsymbol{x_*}, \boldsymbol{x_*}\right) - \boldsymbol{V^T V}$ : calculation of the predictive variance, as in (17)
7. $logp\left(\boldsymbol{y}|\boldsymbol{X}\right) = -\frac{1}{2}\boldsymbol{y^T a} - \sum_i logL_{ii} - \frac{n}{2}\log(2\pi)$ : calculation of the log marginal likelihoo
8. return: $\bar{f}_*$(mean), $var\left(\boldsymbol{f_*}\right)$ (variance), $logp\left(\boldsymbol{y}|\boldsymbol{X}\right)$ (log marginal likelihood);

If there are more than one set of inputs for which the output must be calculated, steps 4-6 should be repeated. To assess the deviation between the actual output value $y_*$ and the predicted output value $y_{predict}$ prediction, a loss function, denoted as $Loss(.)$, should be employed. To evaluate the accuracy of such a prediction, the expected loss

$$\begin{aligned}&E\left[Loss\left(\boldsymbol{y_* y_{predict}}\right)|\boldsymbol{x_*}\right]\\ &\qquad = \int Loss\left(\boldsymbol{y_*}, \boldsymbol{y_{predict}}\right)p\left(\boldsymbol{y_*}|\boldsymbol{x_*}, \boldsymbol{D}\right)dy_* \quad (26)\end{aligned}$$

should be minimized [19]. Hence, the optimal prediction is

$$\boldsymbol{y_{optimal}}|\boldsymbol{x_*} = arg_{y_{predict}}^{min} E\left[Loss\left(\boldsymbol{y_*}, \boldsymbol{y_{predict}}\right)|\boldsymbol{x_*}\right] \quad (27)$$

In general, the value of $\boldsymbol{y_{optimal}}$ that minimizes the expected loss function $|\boldsymbol{y_{predict}} - \boldsymbol{y_*}|$ is the median of $p\left(\boldsymbol{y_*}|\boldsymbol{x_*}, \boldsymbol{D}\right)$, whereas for the squared loss function $(\boldsymbol{y_{predict}} - \boldsymbol{y_*})^2$ it is the mean of the distribution. When the predictive distribution is Gaussian, the mean and median values coincide.

As a general remark, it should be noted that failing to build an analytical prediction method due to the high number of predictors and to their complex relation with the output, which is rain induced attenuation in the present work, an SML based regression method is employed. The prediction algorithm is built taking into account the training dataset and is tested with the test dataset. If the prediction accuracy is satisfactory the proposed method can be used for predicting the attenuation due to rain over working links or links to be deployed.

## IV. THE PROPOSED PREDICTION METHOD. NUMERICAL RESULTS AND VERIFICATION

ITU-R has proposed a model for the prediction of rain attenuation on terrestrial radio links. However, the ITU-R model does not perform satisfactorily in all climatic zones [21]. In general, it is not possible to apply a prediction model to any location or any climatic region. As mentioned in Section I, this weakness constitutes the main motivation to formulate an alternative prediction method employing SML based regression techniques which, by adopting appropriate attributes, apply better to specific locations or climatic zones.

In general, the inputs required by most prediction models for rain attenuation over terrestrial links are the rainfall rate exceeded for a specific time percentage, the propagation path length, the operation frequency, and the wave polarization. This paper proposes a new method for the prediction of rain attenuation over terrestrial microwave links. It is an SML based regression algorithm employing specific GP compatible kernel functions. The available training dataset is extracted from the ITU-R databank and contains 89 experimental links located in the following, alphabetically ordered, countries: Brazil (14 links), China (2 links), Congo (1 link), Czech (5 links), Germany (4 links), France (6 links), Italy (6 links), Japan (15 links), Malaysia (1 link), Netherlands (1 link), Norway (1 link), Poland (2 links), Sweden (3 links), Russia (3 links), the United Kingdom (22 links), the United States of America (3 links). Apparently, the employed geographical footprint is not limited to specific regions. It spans over various rain climatic zones, namely C, E, F, G, H, K, L, N, M and P, and tends to have a global character. From this perspective the generalization of the proposed method has been addressed and facilitated. However, the dataset employed may become more general and possibly more accurate as long as more locations are included spanning various climatic zones and the operational characteristics of the adopted links take a wider range of values. Additional information is provided within the experimental data concerning the specific relation between specific factors, such as local rain rate, path length, operation frequency and wave polarization, with the rain attenuation for specific exceedance probability levels, namely 0.001%, 0.002%, 0.003%, 0.006%, 0.01%, 0.02%, 0.03%, 0.06% and 0.1%. Hereafter, these levels will be called reference exceedance probability levels (REPLs). Moreover, it has to be underlined that 28 experimental links of the employed databank, located in Germany, Japan, Malaysia, Netherlands, Sweden and the United Kingdom, operate at a frequency higher than 30GHz and below 300GHz, a zone which is considered to be the pure mmWave frequency band.

The preliminary version of the proposed prediction method comes from formulating an SML regression algorithm for each one rain attenuation REPL. The 12 inputs (predictors) of this preliminary version are the path length ($PL$), the operation frequency ($FREQ$), the wave polarization ($POL$) as well as the 9 rain rate levels that correspond to the 9 REPLs just mentioned. These 9 rain rate levels are denoted by the vector $\boldsymbol{RR_{exc}}$ $(all\ p\%\ levels)$. To train the algorithm of the preliminary version of the proposed method, the output (target feature) is the rain attenuation level exceeded for the corresponding REPL. This real rain attenuation REPL is known as part of the experimental data. Symbolically, the mathematical formulation of the regression function is given from

$$RA_{exc}\,(p\%) = f_{GPR}(PL, FREQ, POL, \boldsymbol{RR_{exc}}\,(allp\%levels)) \tag{28}$$

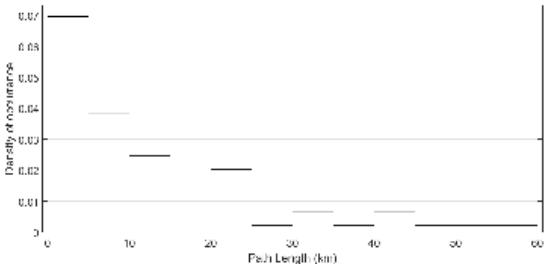


**FIGURE 1.** Distribution of path lengths.

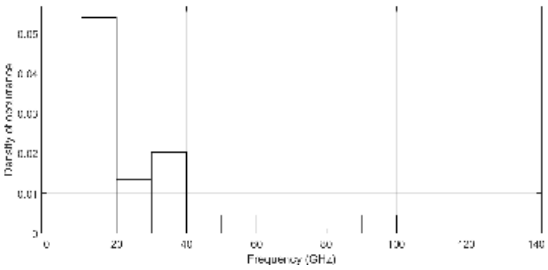


**FIGURE 2.** Distribution of operating frequencies.

where the output (target feature) $RA_{exc}\,(p\%)$ is the rain attenuation in dB exceeded for $p\%$ on a per year basis and the inputs (predictors) are:

| | |
|---|---|
| $PL$: | path length that ranges from 0.5km to 58km. A histogram of its occurrence density is given in Fig. 1 |
| $FREQ$: | operating frequency that ranges from 7GHz to 137GHz. A histogram of its occurrence density is given in Fig. 2 |
| $POL$: | electromagnetic wave polarization that ranges from 0 degrees to 90 degrees. A histogram of its occurrence density is given in Fig. 3 |
| $\boldsymbol{RR_{exc}}$ $(all\ p\%\ levels)$: | rain rate in mm/h for the 9 available REPLs which jointly represent the corresponding local rain rate distribution. The ranges of the rain rates corresponding to the 9 REPLs are given in Table I. As the dataset employed covers a wide range of locations, path lengths, operation frequencies, wave polarization angles, and rain rates, the proposed method is expected to have a general applicability. |

To investigate which kernel function performs better, the training of the algorithm is performed employing the four kernel functions given by (18) - (21) under a 5-fold cross-validation scheme. At this point, a brief discussion on cross-validation should be done. Cross-validation estimates the performance of an algorithm when employing a new

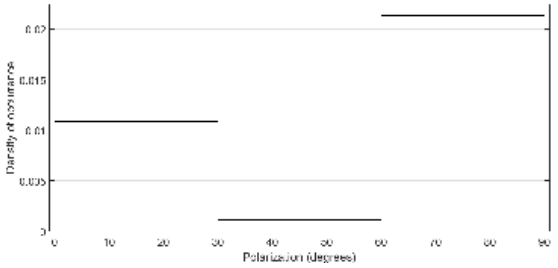


**FIGURE 3.** Distribution of polarization angles.

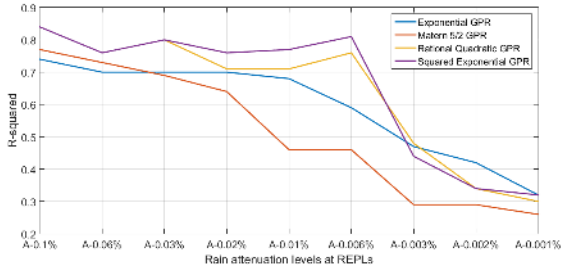


**FIGURE 4.** Regression effectiveness expressed in the form of $R^2$ employing the four GPs referred to in Sec.III, and taking into account 12 predictors, namely the path length, operation frequency, wave polarization, and the 9 REPLs of rain rate.

**TABLE 1.** Range of rain rate for the 9 available REPLs.

| Rain Rate REPLs | Min Rain Rate (mm/h) | Max Rain Rate (mm/h) |
|---|---|---|
| $\boldsymbol{R_{0.001\%}}$ | 47 | 230 |
| $\boldsymbol{R_{0.002\%}}$ | 39 | 179 |
| $\boldsymbol{R_{0.003\%}}$ | 32 | 167.2 |
| $\boldsymbol{R_{0.006\%}}$ | 23 | 150.7 |
| $\boldsymbol{R_{0.01\%}}$ | 18 | 133.5 |
| $\boldsymbol{R_{0.02\%}}$ | 13 | 118.5 |
| $\boldsymbol{R_{0.03\%}}$ | 10 | 105.8 |
| $\boldsymbol{R_{0.06\%}}$ | 6 | 84 |
| $\boldsymbol{R_{0.1\%}}$ | 4.5 | 67.8 |

dataset compared to the training dataset. It is used to select the model which minimizes the prediction error. Furthermore, cross-validation protects against overfitting. First, a number of folds is assumed to partition the dataset. This partitioning is made in a stratified way, that is, both the training and the test datasets have roughly the same class proportions as in the entire dataset and are formed as representative as possible. If a 5-fold cross-validation is selected, as it is done in this work, then (i) the entire dataset is partitioned into 5 disjoint sets or folds, (ii) for each fold a model is trained using the out-of-fold observations and its performance is assessed using in-fold data and (iii) the average test error over all folds is calculated. This method gives a good estimate of the predictive accuracy of the final algorithm trained employing the entire dataset. Although it requires multiple fits, it makes efficient use of the entire dataset.

After the algorithm training, the rain attenuation level is predicted applying (28). As expected, there is a difference between the predicted and the actual values of the outputs, which is known as the residual (error) of the prediction. The effectiveness of the regression is expressed employing the $R^2$ metric that ranges from 0 to 1. $R^2$ is a statistical measure called coefficient of determination and provides a measure of how well the observed outputs are replicated by the algorithm. $R^2$ essentially expresses the proportion of the variance in the dependent response that is predictable from the independent predictors. Indicatively, the interpretation of $R^2 = 0.85$ is that 85% of the variance in the response can be attributed to the predictors employed whereas the remaining 15% can be attributed to unknown, lurking predictors or inherent variability. In Fig. 4, $R^2$ is plotted for the set of REPLs given in Table I. The horizontal axis represents the levels of rain attenuation exceeded for the time percentages correspondingly dictated by the 9 REPLs. As readily observed from Fig. 4, the regression effectiveness decreases as the exceedance time level of the rain attenuation decreases. Such a low prediction accuracy does not allow to use the preliminary version by itself without enhancement.

Next, the training process of the proposed SML regression algorithm is enhanced by introducing as an additional predictor the rain attenuation estimate$RA_Est_{exc}^{ITU}\ (p\%)$ as calculated from the ITU-R model [12]. Hereafter, this estimate will be called ITU-R estimate. The kernel functions employed and the cross-validation scheme are kept the same. Hence, in the enhanced version of the proposed method, 12 + 1 predictors are involved in the regression function, which is symbolically written as

$$RA_{exc}\ (p\%) = f_{GPR}(PL, FREQ, POL, \boldsymbol{RR_{exc}}\ (all\ p\%\ levels\ ), RA\_Est_{exc}^{ITU}\ (p\%)) \quad (29)$$

The $R^2$ results of the enhanced SML based prediction method are plotted in Fig. 5.

Two important conclusions can be deduced from Fig. 5. First, the incorporation of the ITU-R estimate as an additional input (predictor) has enhanced the training process leading to a significant improvement in the regression accuracy. This improvement is verified by the significantly higher $R^2$ values observed in Fig. 5 compared to the respective ones observed in Fig. 4. Second, the Rational Quadratic GPR performs better than the other GPRs for all the REPLs considered.

It is important to examine how the enhanced version of the proposed method performs in comparison to the established prediction models for rain attenuation presented in Section II. In this framework, numerical calculations were performed using rain rate data extracted from the same ITU-R databank. To become comparable, the respective results have been normalized adopting the test variable $\rho_V$ proposed by ITU-R in Recommendation P.311-13 [22]. According to P.311-13, for each time percentage examined and each radio link of the ITU-R databank considered, say the $i^{th}$link, the ratio of

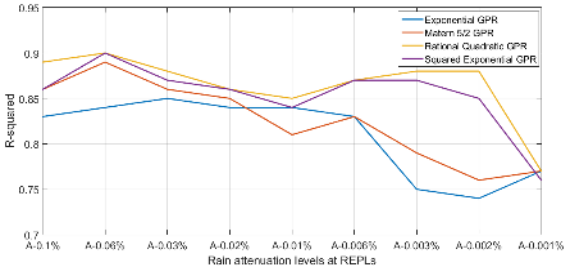


**FIGURE 5.** Same as in Fig. 4, but employing the ITU-R estimate as an additional predictor.

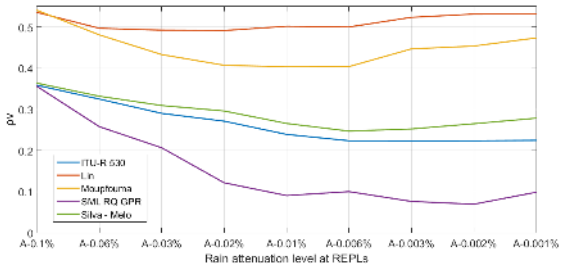


**FIGURE 6.** Regression effectiveness expressed in the form of the test variable $\rho_V$ employing various rain attenuation prediction models and the enhanced version of the proposed SML prediction method. For this comparison, the Rational Quadratic kernel function is employed.

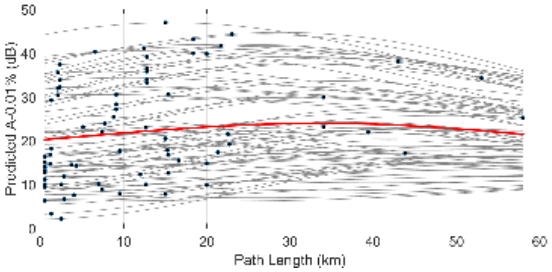


**FIGURE 7.** ICE plots of the predicted $A_{0.01}$ with respect to the path length as a predictor.

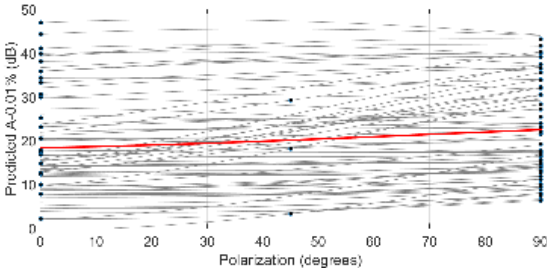


**FIGURE 8.** ICE plots of the predicted $A_{0.01}$ with respect to the wave polarization as a predictor.

the predicted rain attenuation, $A_p(dB)$, to the measured rain attenuation, $A_m(dB)$, is calculated from

$$S_i = \frac{A_{p,i}}{A_{m,i}} \tag{30}$$

Next, the variable $V_i$ is calculated from

$$V_i = \begin{cases} (\frac{A_{m,i}}{10})^{0.2} lnS_i, & for\ A_{m,i} < 10dB \\ lnS_i, & for\ A_{m,i} \geq 10dB \end{cases} \tag{31}$$

Then, the mean $\mu_V$ and the standard deviation $\sigma_V$ of the $V_i$ values for each time percentage are calculated. Finally, the test variable is determined as the rms (root mean square) value

$$\rho_V = \sqrt{\mu_V^2 + \sigma_V^2} \tag{32}$$

In comparing the various prediction methods, it should be noted that the lower the test variable is the better the prediction method. The numerical results obtained following the above comparison procedure are depicted in Fig. 6.

As readily observed from Fig. 6, the proposed SML based prediction method performs significantly better than the four prediction models under comparison. The proposed method is much more accurate than those of ITU-R [12] and Silva-Melo [14], which are considered to perform more effectively than the others.

At this point, it is worthwhile to investigate the individual significance of the 13 predictors employed in the proposed method, expecting that the importance of the ITU-R estimate will be proven high, since it is determined independently of the method under consideration. In this course, the relationship between a predictor and the predicted responses is visualized. Typically, this visualization process has to be repeated for all the $(12 + 1)$ predictors employed. The visualization is based on the notion of the individual conditional expectation (ICE) plot [23]. In the ICE plots there are:

- Black circles that represent the predicted responses for the predictor each time considered, which for the problem in hand might be the path length, the wave polarization, the operation frequency, any of the 9 REPLs, or the ITU-R estimate.
- Gray lines that visualize the response of the algorithm for each experiment allowing the predictor examined, e.g. the path length, to take values over its entire range.
- A red line, called partial dependence plot (PDP), which visualizes the average relationship between the selected predictor and predicted responses. In other words, PDP is the average of the gray lines.

The ICE plots highlight the variation in the fitted values over the range of the predictor examined each time. As a criterion, a high variance of the gray lines or alternatively slow variations of the PDP indicate a low dependence of the predicted responses on the predictor examined. For the proposed sensitivity assessment, the ICE plot regarding the prediction of $A_{0.01}$ is indicatively considered. $A_{0.01}$ refers to 0.01% exceeded time percentage for rain attenuation (in dB) which is the most frequently encountered REPL in rain attenuation prediction models.

Next, ICE plots are used to examine the relationship between the predicted response $A_{0.01}$, which has been indicatively chosen for the sensitivity assessment, with the various

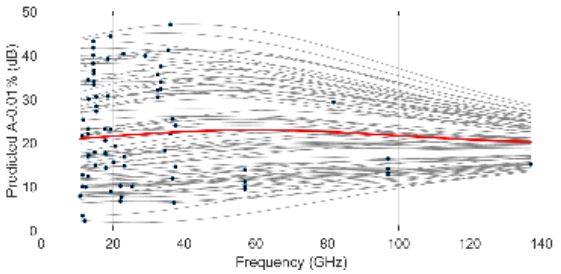


**FIGURE 9.** ICE plots of the predicted $A_{0.01}$ with respect to the operation frequency as a predictor.

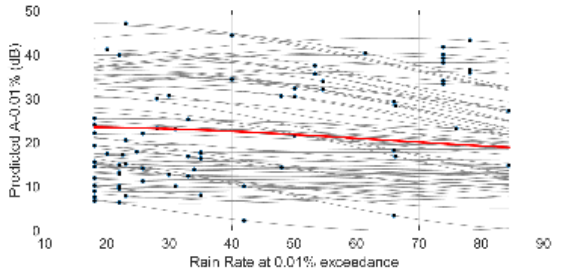


**FIGURE 10.** ICE plots of the predicted $A_{0.01}$ with respect to the $R_{0.01}$ level as a predictor.

predictors examined. Specifically, in Figs. 7 to 10, ICE plots are presented concerning, respectively, the predictors (i) path length, (ii) wave polarization, (iii) operation frequency and (iv) REPL $R_{0.01}$, which was indicatively chosen to represent the set of the 9 REPLs considered. From these plots, it is readily observed that the variance of the gray lines is high and the PDP lines look almost straight. Hence, the ICE plots indicate that the above four predictors do not affect significantly the predicted response $A_{0.01}$. On the other hand, from Fig. 11 presenting the ICE plot concerning the relationship between the same indicative predicted response $A_{0.01}$ with the ITU-R estimate which was employed as an additional predictor, it is readily observed that the variance of the gray lines is much smaller than their variance observed in Figs. 7 to 10 whereas, at the same time, the PDP slope exhibits drastic fluctuations extending to the entire range of $A_{0.01}$ values taken into account. Hence, it is verified that the ITU-R estimate -which is used as an additional predictor in addition to the 12 predictors used by the preliminary version of the proposed method- constitutes the predictor that dramatically improved the prediction accuracy of the proposed SML based method. It should also be noted that the ITU-R estimate, which is used as an additional predictor, is not drawn directly from the experimental data, but it is a macroscopic estimate obtained employing another prediction model, namely the ITU-R model [12], which employed the path length, the wave polarization, the operation frequency and the rain rate $RR_{0.01}$. The prediction of rain attenuation for REPLs other than 0.01% is based on scaling methods described in [12].

From the engineering point of view, it is of high importance to investigate the exceedance probability of rain attenuation,

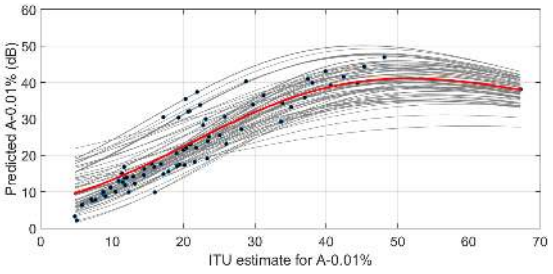


**FIGURE 11.** ICE plots of the predicted $A_{0.01}$ with respect to the ITU-R estimate as a predictor.

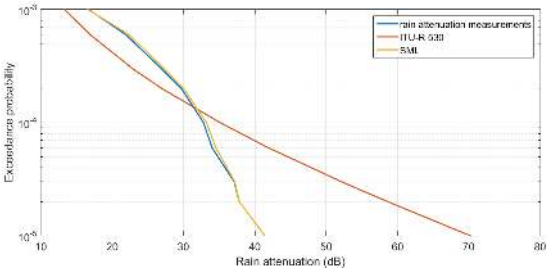


**FIGURE 12.** Exceedance probability of rain attenuation in Yotsua, Japan.

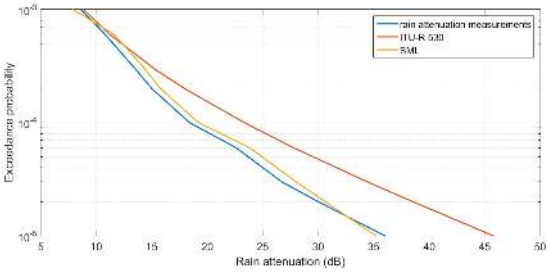


**FIGURE 13.** Exceedance probability of rain attenuation in Paranapiacaba, Brazil.

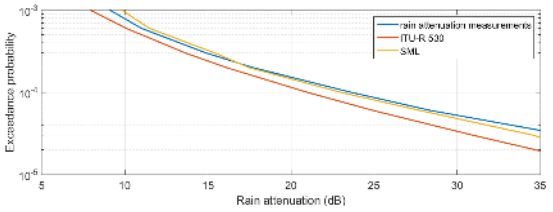


**FIGURE 14.** Exceedance probability of rain attenuation in Novara, Italy.

i.e. *Prob*(*Rain attenuation* > *level*), or, equivalently, the complementary cumulative distribution function (ccdf) of rain attenuation for various links covering various locations, path lengths, operation frequencies and rain conditions. Particularly, Figs. 12-18 depict the exceedance probability along experimental links located in Yotsua (Japan), Paranapiacaba (Brazil), Novara (Italy), Stockholm (Sweden), Chibolton (UK) and Tokyo (Japan), the operation characteristics of which are provided in Table II. In all these Figs, the ccdf obtained employing the proposed SML based method fits better to the actual measurements than the ccdf obtained

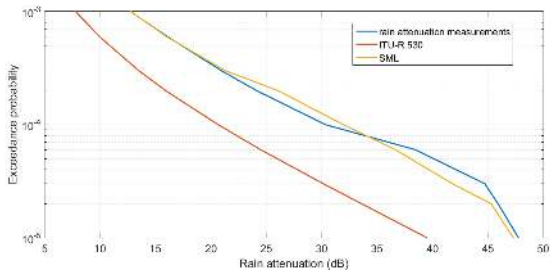


**FIGURE 15.** Exceedance probability of rain attenuation in Stockholm, Sweden.

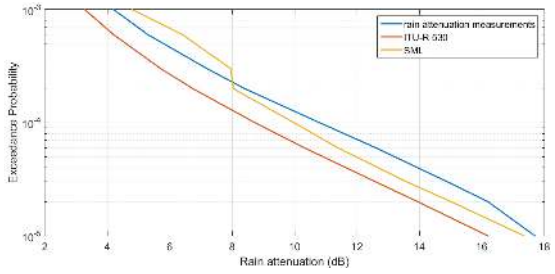


**FIGURE 16.** Exceedance probability of rain attenuation in Chibolton, UK (57GHz).

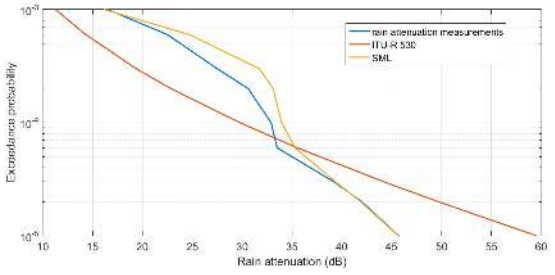


**FIGURE 17.** Exceedance probability of rain attenuation in Tokyo, Japan.

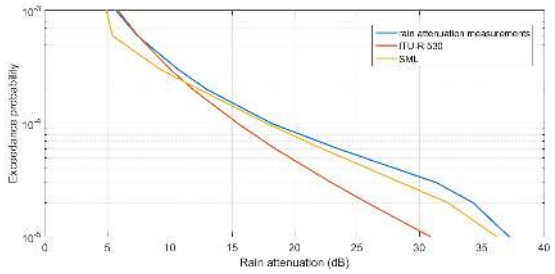


**FIGURE 18.** Exceedance probability of rain attenuation in Chibolton, UK (137GHz).

employing the ITU-R model. This happens for a wide variety of locations and a wide range of path length and operation frequency values, including the mmWave band as well.

The proposed method is founded on concrete mathematical models of stochastic processes. On the one hand, it offers the advantage of avoiding complex mathematical expressions. On the other hand, it possesses an empirical profile as it is tuned by regressing data available from the ITU-R databank. The reliability of the available data affects the accuracy of the proposed method. Inefficient experimental procedures, incorrectly tuned instruments, geographical limitations and, in general, not properly performed experiments affect adversely the accuracy of the proposed SML method. As deduced from the comparison with other prediction methods currently in use, it is confirmed that, if sufficient and reliable data is available, the proposed SML based prediction method can effectively capture the propagation impairment related to rain attenuation employing rain rate measurements. Compared to analytic models, SML based approaches can take into account more predictors, thus being capable of achieving significantly higher accuracy. Moreover, to enhance the efficiency of SML based prediction methods, it is of high importance to look for and employ additional predictors, not necessarily coming directly from the measurement field but calculated employing other prediction models. The insertion of such additional predictors to the training process can lead to a meta-training process in the sense that a more accurate prediction method is expected to be deployed. Accurate SML based prediction methods could also be formulated to represent the local rain fading mechanism concerning locations for which sufficient local data is available. The incorporation of new experimental data can improve the training of the algorithm which, subsequently, may result into a more accurate prediction.

**TABLE 2.** Operating characteristics of experimental links for figs 12-18.

| Experiment City | Frequency (GHz) | Path Length (km) | Rain Rate range (mm/h) |
|---|---|---|---|
| Yotsua | 14.6 | 12.8 | 28.2-117.4 |
| Paranapiacaba | 20.7 | 22.7 | 4.5-50 |
| Novara | 29.3 | 12.7 | 7.3-70 |
| Stockholm | 32.6 | 2.1 | 15.5-101.9 |
| Chibolton | 57 | 0.5 | 7-47 |
| Tokyo | 97 | 0.5 | 8-57 |
| Chibolton | 137 | 0.5 | 8-58 |

## V. CONCLUSION

In this paper, the concept of SML is combined with GPs for regression to formulate a new prediction method for rain attenuation. The proposed method has been validated taking into account experimental data from the ITU-R databank concerning LOS terrestrial links. The experimental data used for the training of the preliminary SML algorithm included 12 predictors which were the path length, the operation frequency, the wave polarization and 9 rain rate levels (REPLs). The prediction results obtained applying the preliminary version were not found more accurate than the corresponding ones obtained applying the well-established ITU-R model. The dataset used to develop the SML algorithm has a wide range of values for all the predictors and does not show geographical limitations, To enhance the prediction accuracy,

the rain attenuation estimate obtained applying the existing ITU-R model for each REPL was considered as an additional predictor along with the existing 12 predictors coming from the available experimental data. The numerical results obtained applying the enhanced version showed a significant prediction accuracy improvement over the prediction accuracy offered by the preliminary version. The performance comparison conducted to validate the enhanced version of the proposed SML based prediction method showed a significant superiority over other prediction models currently in use, reaching very high accuracy levels. The applicability of the proposed method is practically global. However, its adaptation to specific locations, climatic regions or frequency zones can be facilitated employing the proper training dataset that can convey the experimental information related to the specific problem each time in hand. New experiments can be properly planned so that additional propagation data can be collected and processed following the proposed methodology towards building prediction tools dealing with various propagation phenomena. Finally, if sufficient and both quantitatively and qualitatively homogeneous actual data can be included in the dataset for training and testing the ML algorithm, the proposed SML based prediction method can get a geographically and climatologically wider scope.

## APPENDIX

### A. ITU-R P. 530-16 MODEL

Determine the effective path length of the link, $d_{eff}$, by multiplying the actual path length $L$ (km) by the path reduction factor $r$, which can be expressed as:

$$r = \frac{1}{0.477L^{0.633}R_{0.01}^{0.073a}f^{0.123} - 10.579\,(1-exp\,(-0.024L))} \tag{A.1}$$

where $f$(GHz) is the frequency and $a$ is the exponent of the specific attenuation model. Maximum recommended $r$ is 2.5, so if the denominator of (A.1) is less than 0.4 use $r = 2.5$. $R_{0.01}$ is the rain rate exceeded for 0.01% of the time in a year. If this information is not locally available, an estimate can be obtained from the information given in Recommendation ITU-R P.837 [17].

The path attenuation exceeded for 0.01% of the yearly time is calculated as

$$A_{0.01} = \gamma_R d_{eff} = \gamma_R L r \tag{A.2}$$

The prediction of the full $P(A)$, for $0.001\% \leq P \leq 1\%$, is given by

$$A\,(P) = A_{0.01}C_1P^{-(C_2+C_3log_{10}P)} \tag{A.3}$$

with $C_1$, $C_2$, and $C_3$ being empirical coefficients depending on frequency $f$ [16].

### B. SILVA-MELO MODEL

In the model proposed by Silva-Mello et al. in [18], the effective path length $d_{eff}$ is calculated as:

$$d_{eff} = \frac{1}{1+\frac{L}{d_0}}L \tag{A.4}$$

where

$$d_0 = 119R(P)^{-0.244} \tag{A.5}$$

The prediction of the rain attenuation exceeded for $P$ percent of the time is achieved as:

$$A\,(P) = kR_{eff}^{a}d_{eff} \tag{A.6}$$

where $R_{eff}$, i.e., the effective rain rate, is:

$$R_{eff} = 1.763R(P)^{0.753+0.197/L} \tag{A.7}$$

### C. MOUPFOUMA MODEL

Similarly to the ITU-R model, the prediction method proposed by Moupfouma in [19] receives $R_{0.01}$ as the input to predict $A$ as:

$$A\,(P) = kR_{0.01}^{a}L_{eq}(P, L) \tag{A.8}$$

$L_{eq}$ in (A.8) is the equivalent path length calculated as:

$$L_{eq}\,(P, L) = Lexp\left(-\frac{R(P)}{1+\zeta\,(L)R(P)}\right) \tag{A.9}$$

where

$$\zeta\,(L) = \begin{cases} -100, & L \leq 7(km) \\ [\frac{44.2}{L}]^{0.78}, & L > 7(km) \end{cases} \tag{A.10}$$

### D. LIN MODEL

According to this model, the rainfall attenuation exceeded for a percentage $P$ of the yearly time can be calculated as:

$$A\,(P) = kR(P)^{a}Lr \tag{A.11}$$

$R(P)$ is the rain rate exceeded for the same percentage $P$ of the time. The factor $r$ takes the followingsimple expression:

$$r = \frac{1}{1+\frac{L}{L(R)}} \tag{A.12}$$

where

$$L\,(R) = \frac{2623}{R\,(P) - 6.2} \tag{A.13}$$

## ACKNOWLEDGMENT

The authors Spiros N. Livieratos and Panayotis G. Cottis acknowledge financial support for the dissemination of this work from the Special Account for Research of ASPETE through the funding program "Strengthening ASPETE's Research."

## REFERENCES

[1] M. Schuld and F. Petruccione, "Machine Learning," in *Supervised Learning With Quantum Computers*. Cham, Switzerland: Springer, 2018, p. 21.
[2] C. Jiang, H. Zhang, Y. Ren, Z. Han, K.-C. Chen, and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 98–105, Apr. 2017.
[3] K. P. Murphy, "Introduction," in *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: Massachusetts Institute of Technology, 2012, p. 2.
[4] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.
[5] M. G. Kibria, K. Nguyen, G. P. Villardi, O. Zhao, K. Ishizu, and F. Kojima, "Big data analytics, machine learning, and artificial intelligence in next-generation wireless networks," *IEEE Access*, vol. 6, pp. 32328–32338, 2018.
[6] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.
[7] G. R. MacCartney and T. S. Rappaport, "Rural macrocell path loss models for millimeter wave wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1663–1677, Jul. 2017.
[8] T. A. Thomas, M. Rybakowski, S. Sun, T. S. Rappaport, H. Nguyen, I. Z. Kovacs, and I. Rodriguez, "A prediction study of path loss models from 2-73.5 GHz in an urban-macro environment," in *Proc. IEEE 83rd Veh. Technol. Conf.*, Nanjing, China, May 2016, pp. 1–5.
[9] S. Shrestha and D.-Y. Choi, "Rain attenuation statistics over millimeter wave bands in South Korea," *J. Atmos. Solar-Terr. Phys.*, vols. 152–153, pp. 1–10, Jan. 2017.
[10] S. N. Livieratos, V. Katsambas, and J. Kanellopoulos, "A global method for the prediction of the slant path rain attenuation statistics," *J. Electromagn. Waves Appl.*, vol. 14, no. 5, pp. 713–724, Jan. 2000.
[11] *Specific Attenuation Model for Rain for Use in Prediction Methods*, document ITU-R Recommendation P.838-3, ITU, Geneva, Switzerland, 2005.
[12] *Propagation Data and Prediction Methods Required for the Design of Terrestrial Line-of-Sight Systems*, document ITU-R Recommendation P.530-16, ITU, Geneva, Switzerland, 2015.
[13] *Characteristics of Precipitation for Propagation Modelling*, document ITU-R Recommendation P.837-7, ITU, Geneva, Switzerland, 2017.
[14] L. A. R. Da. Silva Mello, M. S. Pontes, R. M. De Souza, and N. A. P. Garcia, "Prediction of rain attenuation in terrestrial links using full rainfall rate distribution," *Electron. Lett.*, vol. 43, no. 25, pp. 1442–1443, Dec. 2007.
[15] F. Moupfouma, "Electromagnetic waves attenuation due to rain: A prediction model for terrestrial or L.O.S SHF and EHF radio communication links," *J. Infr., Millim., Terahertz Waves*, vol. 30, no. 6, pp. 622–632, Jun. 2009.
[16] S. H. Lin, "11-GHz radio: National long-term rain statistics and empirical calculation of 11-GHz microwave rain attenuation," *Bell Syst. Tech. J.*, vol. 56, no. 9, pp. 1581–1604, Nov. 1977.
[17] C. E. Rasmussen and C. K. I. Williams, "Regression," in *Gaussian Processes for Machine Learning*, Cambridge, MA, USA: MIT Press, 2006, p. 7.
[18] K. P. Murphy, "Gaussian Processes," in *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: Massachusetts Institute of Technology, 2012, pp. 515–521.
[19] C. E. Rasmussen and C. K. I. Williams, "Regression," in *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006, pp. 16–22.
[20] M. Abramowitz and I. A. Stegun, "Bessel Functions," in *Handbook of Mathematical Functions*, New York, NY, USA: Dover, 1964, p. 358.
[21] F. Moupfouma and L. Martin, "Modelling of the rainfall rate cumulative distribution for the design of satellite and terrestrial communication systems," *Int. J. Satell. Commun.*, vol. 13, no. 2, pp. 105–115, Mar./Apr. 1995.
[22] *Acquisition, Presentation and Analysis of Data in Studies of Tropospheric Propagation*, document ITU-R Recommendation P.311-13, ITU, Geneva, Switzerland, 2009.
[23] A. Goldstein, A. Kapelner, J. Bleich, and E. Pitkin, "Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation," *J. Comput. Graph. Statist.*, vol. 24, no. 1, pp. 44–65, 2015.

**SPIROS N. LIVIERATOS** received the Diploma degree in electrical engineering and the Dr. Eng. degree from the National Technical University of Athens (NTUA), Greece, in 1992 and 1998, respectively, and the Executive MBA degree from the Athens University of Economics and Business (AUEB), Greece, in 2000. For more than 15 years, he worked in the telecommunication market at various managerial positions (SIEMENS, OTEGLOBE, and Greek Regulatory Authority for Telecommunications). In 2009, he joined the Department of Electrical and Electronic Engineering, School of Pedagogical and Technological Education, where he has been a Professor, since 2018. He has published more than 60 articles in international technical journals and conference proceedings. His current research interests include radio wave propagation in wireless and satellite networks and data analytics for communications and energy applications.

**PANAYOTIS G. COTTIS** received the Diploma degree in mechanical and electrical engineering and the Dr. Eng. degree from the National Technical University of Athens (NTUA), Greece, in 1979 and 1984, respectively, and the M.Sc. degree from The University of Manchester, Manchester, U.K., in 1980. In 1986, he joined the School of Electrical and Computer Engineering, NTUA, where he has been a Professor, since 1996. He has published more than 250 articles in international technical journals and conference proceedings. His current research interests include resource allocation in wireless and satellite networks, power-line communications, and smart grid applications.