



OPEN

Outdoor THz fading modeling by means of gaussian and gamma mixture distributions

Evangelos N. Pappasotiriou^{1✉}, Alexandros-Apostolos A. Boulogeorgos^{1,2} & Angeliki Alexiou¹

Terahertz (THz) band offers a vast amount of bandwidth and is envisioned to become a key enabler for a number of next generation wireless applications. In this direction, appropriate channel models, encapsulating the large and small-scale fading phenomena, need to be developed for both indoor and outdoor communications environments. The THz large-scale fading characteristics have been extensively investigated for both indoor and outdoor scenarios. The study of indoor THz small-scale fading has recently gained the momentum, while the small-scale fading of outdoor THz wireless channels has not yet been investigated. Motivated by this, this contribution introduces Gaussian mixture (GM) distribution as a suitable small-scale fading model for outdoor THz wireless links. In more detail, multiple outdoor THz wireless measurements recorded at different transceiver separation distance are fed to an expectation-maximization fitting algorithm, which returns the parameters of the GM probability density function. The fitting accuracy of the analytical GMs is evaluated in terms of the Kolmogorov-Smirnov, Kullback-Leibler (KL) and root-mean-square-error (RMSE) tests. The results reveal that as the number of mixtures increases the resulting analytical GMs perform a better fit to the empirical distributions. In addition, the KL and RMSE metrics indicate that the increase of mixtures beyond a particular number result to no significant improvement of the fitting accuracy. Finally, following the same approach as in the case of GM, we examine the suitability of mixture Gamma to capture the small-scale fading characteristics of the outdoor THz channels.

Terahertz (THz) wireless systems have been identified as a key enabler of the next generation networks era, since it can provide the required radio resources for a number of killer-applications, including wireless backhauling, mobile ad-hoc backhauling, as well as massive connectivity of bandwidth-hungry applications, like virtual and holographic reality¹⁻³, as well as enabling sensing^{4,5} and cm-level localization capabilities⁶. Novel wireless concepts, such as the internet of everything, connected and autonomous vehicles, and unmanned aerial vehicle, are also expected to benefit from the usage of the THz band². The first step towards designing and optimizing THz wireless systems is the development of indoor and outdoor channel models that can accurately capture the particularities of the propagation medium in this band. In particular, the THz wireless channel model can be seen as the joint contribution of the large and small scale fading⁷. The large scale fading can be expressed in terms of the deterministic pathloss and shadowing, while the fast channel amplitude fluctuations are described in terms of the stochastic small-scale fading⁷.

The large scale fading characteristics have been extensively investigated in both outdoor and indoor environments⁸⁻²³. In more detail, urban outdoor double directional channel measurements in the range of 141.1–148.5 GHz for distances over 100 m have been conducted⁸. These measurements by means of the many identified angles of arrival and angles of departure have verified the existence of line-of-sight (LoS) and non-line-of-sight (NLoS) THz multipath components. An outdoor measurement campaign in an urban microcell environment at 140 GHz recorded omni and directional LoS and NLoS links at a maximum distance of 117.4 m⁹. Based on the aforementioned measurements, omni and directional pathloss exponent models have been implemented, where the shadowing is expressed by means of a lognormal distribution. A ray-based deterministic tool has been employed to model the large scale pathloss of an urban outdoor scenario in the range of 90–200 GHz¹⁰. The pathloss has been modeled by the exponent model, where the shadowing due to vegetation has been modeled by means of a lognormal distribution. Based on 142 GHz multiple-input-multiple-output (MIMO) urban microcell propagation measurements, the channel spatial statistics of the number of spatial clusters and the cluster power distribution have been identified¹¹. A detailed spatial statistical MIMO channel generation

¹Department of Digital Systems, University of Piraeus Research Center, 18534 Piraeus, Greece. ²Department of Electrical & Computer Engineering, University of Western Macedonia, 50100 Kozani, Greece. ✉email: vangpapasot@unipi.gr

procedure has been introduced based on the empirical channel statistics. An extensive set of wireless LoS and NLoS measurements in the range of 145–146 GHz for distances between 1–100 m have been conducted¹¹. Building upon the measurements, the deterministic pathloss, shadowing, delay spread, angular spread and multipath component power distribution have been modeled. A vehicle to infrastructure channel has been developed for an urban scenario by means of ray tracing for the operational frequency of 110 GHz¹³. Accordingly, the channel statistics of pathloss, time-of-arrival and direction-of-arrival have been characterized. An initial review on the impact of the weather conditions to the deterministic attenuation of THz wireless links has been conducted¹⁴. In more detail, the channel impairments caused by the water vapor, dust particles, fog, clouds, and rain have been investigated. Meanwhile, deterministic THz polynomial pathloss models for the ranges of 100–450 GHz, 200–450 GHz and 275–400 GHz have been developed^{15,18,19}. In these models, the THz channel has been assumed to consist of a single deterministic LoS coefficient, which has been expressed as the sum of the free space and molecular absorption loss. Various LoS and NLoS indoor measurements are performed for wireless links operating at 28 GHz and 140 GHz^{16,17}. Therein, based on the received signal strength of the multipath components of the links, the millimeter wave (mmWave) and THz channels have been deterministically modeled as the logarithmic scale sum of the exponential pathloss and lognormal shadowing. A single path theoretical THz channel model for THz wireless communications within vegetation has been developed^{21,23}. In this model, the wireless channel consists of two coefficients, namely the deterministic pathloss and the lognormal shadowing.

The indoor THz small-scale fading channel modeling has recently gained a momentum^{7,16,17,20,24–31}. Specifically, for the case of wireless backhaul THz links, the small-scale fading has been theoretically modeled by means of the α - μ distribution^{24,26}. Then, the system performance has been quantified under different levels of transceiver hardware impairments, antennas misalignment and fading severity. Furthermore, the suitability of the α - μ distribution to describe the small-scale fading channel amplitude of indoor THz wireless channels has been experimentally validated in several studies^{7,30,31}. Experimental LoS and NLoS THz wireless measurements have been performed in an anechoic chamber²⁷. Based on this model, a stochastic indoor THz channel model has been developed, where the small-scale fading attenuation factor has been expressed in terms of a Rayleigh or Nakagami-m distribution under NLoS and as a Rice or Nakagami-m in LoS propagation conditions, respectively. A two dimensional stochastic geometric channel model has been developed for indoor THz wireless communications^{28,29}. Then, a parametric multipath Rice fading model has been derived. A measurement based indoor channel model for the range of 126–156 GHz for both LoS and NLoS conditions has been developed²⁰. The exponential pathloss and shadowing have been used to model the large scale fading, whereas the small-scale fading amplitude has been given by a novel distribution. Meanwhile, THz wireless measurements have been conducted within an anechoic chamber in the range of 240–300 GHz²⁵. Then, by exploiting the measurements and various fitting accuracy metrics, it has been concluded that the small-scale fading amplitude of the links can be accurately modeled by means of the Gamma and Gaussian mixture models. Also, the mixture Gamma (MG) has been employed in investigating the capacity of a wireless channel and expressions for the optimal and power rate adaptation, the channel inversion with fixed and truncated rate were derived. The expressions were verified by means of Monte-Carlo simulations³². Furthermore, the Gamma mixture has been used for analytical performance assessment of composite fading channels in terms of received signal-to-noise-ratio³³. In continuation of the previously mentioned work, the Gaussian mixture has been employed in the performance analysis of an energy detector. In more detail, analytical expressions for the performance parameters of average detection and area under the receiver probabilities were derived³⁴.

The aforementioned contributions underline the importance of not only the large-scale, but also the small-scale fading THz channel modeling. However, to the best of the authors knowledge, results on THz small-scale fading channel modeling in outdoor environments have not been published so far. Motivated by this, in this work, outdoor THz measurements performed in the campus area of Aalto university in Finland are exploited. In more detail, multiple LoS and NLoS links have been measured at different transceiver separation distances. For each link, multiple channel gain measurements were recorded, which have been used to perform fitting analysis of the empirical channel gain distribution amplitude to Gaussian Mixtures (GMs) analytical distributions. The evaluation of the suitability of GMs to describe the small-scale fading channel gain amplitude of outdoor THz wireless links is very useful. An appropriate GM is capable of describing complicated fading scenarios, where multiple peaks can occur in the fading amplitude of the empirical distribution^{25,35}. The GM is expressed as the sum of independent Gaussian distributions. Hence, it offers mathematical tractability, which is of great importance in analytical expressions evaluations. By taking this into account it should be noted that, the fluctuating-two-ray (FTR) model has also been employed in THz channel modeling^{36,37}. However, the FTR uses an infinite number of components to approximate the empirical distribution. As a consequence simpler distributions like the GM and Gamma mixture are preferred to accommodate the channel modeling and the analytical evaluation needs. Moreover, it should be noted that, in this work the suitability of GMs to model the small-scale fading amplitude of the outdoor THz links is more thoroughly investigated in comparison with MGs distributions. The reason for this is that the analytical expression of the GMs are more tractable in comparison with those of the MGs and have been employed in various performance evaluation works^{35,38,39}. Also, the suitability of MG to model the small-scale fading amplitude of short range indoor THz wireless links has been previously investigated²⁵. Moreover, the support of a GM is defined in the $(-\infty, \infty)$, which aids in achieving a good fit to the tails of the empirical distributions.

In this work the measurements of each link are preprocessed to obtain the channel gain of each of the recorded multipath components. Subsequently, in order to increase the number of the different channel realizations in each link, a method based on adding random phases to the path amplitudes will be employed. Then, by making use of the resulting channel realizations of each link, the empirical probability density function (PDF) and cumulative density function (CDF) are fitted to the analytical GMs. Also, MGs distributions are fitted to some indicative links and the fitting performance is compared to that of the GMs. Then, the parameters and weights

of each Gaussian and Gamma distribution of a GM and MG expression, respectively, are obtained by fitting it to the empirical channel gain distribution of the investigated link. This is accomplished by means of the expectation maximization (EM) algorithm^{25,35,38,40}. The accuracy of the fit of the analytical distributions to the corresponding empirical ones is quantified in terms of the Kolmogorov-Smirnov (KS), Kullback-Leibler (KL) and root-mean-square-error (RMSE) tests^{41–43}. However, the evaluation of the fitting accuracy of the analytical GMs and MGs to the empirical ones is performed only in terms of the KL and RMSE tests, because, for all the GMs and MGs of all the investigated links, the KS yields a good fit. As a result, the KS poses as a non strict fitting criterion. According to the KL and RMSE metrics for all the links, it is observed that, as the number of mixtures increases the resulting analytical GMs and MGs perform a better fit to the empirical distributions. On the other hand, as the number of mixtures decreases, the resulting analytical GMs and MGs perform worse in terms of fitting even for single peak empirical distributions. Furthermore, the KL and RMSE metrics indicate that the increase of mixtures above a particular threshold does not improve drastically the fitting accuracy performance of the analytical GMs and MGs to the empirical ones.

In order to further elucidate, the key contribution of this work lies in the approach that is followed to derive the empirical small-scale fading amplitude distribution of the investigated THz links. In more detail, the principle of transfer learning combined with the EM algorithm is employed for the measured data of an outdoor static THz propagation environment⁴⁴. These THz wireless link measurement data contain deterministic pathloss measurements and during each link measurement session there were no moving scatterers. Yet, in a realistic THz wireless signal propagation scenario moving scatterers may influence the channel characteristics. This can be adequately modeled by the methodology initially proposed by Molisch et al.⁴⁵. In this work this methodology is employed to populate the herein used link measurements datasets⁴⁴. Next, after observing the resulting empirical PDF of each measured THz link, we propose the GM distribution as a suitable target distribution. In order to identify the number of Gaussian distributions needed and their corresponding weights and parameters we follow a fitting methodology based on an interactive EM algorithm.

Results

Measurement setup and sites. Figure 1 illustrates the top-view of the outdoor premises of Aalto University in Finland, where the THz measurements are conducted. In more detail, each link is defined by a unique transmitter (Tx) and receiver (Rx) pair. Both the Tx and Rx are equipped with a single antenna. During each measurement session both the Tx and Rx were in fixed positions, while only the Tx-Rx pair of interest was active, i.e., no interference is induced by neighbor links. Figures 1a and b show that individually Rx₁ and Rx₂ are employed to perform the wireless THz measurements. The Txs marked with green dots denote a LoS link between the Tx and the Rx of interest, whereas the Txs marked with a yellow dot stand for a NLoS transceiver link. However, it should be noted that for the investigated outdoor THz measurements no paths were able to be received in the NLoS transmissions scenarios. The THz transmissions of all the investigated links are performed at the center radio frequency (RF) of 142 GHz with a total bandwidth of 4 GHz⁴⁴. The transmit power is set equal to 5 dBm and the transceivers antennas heights are 1.85 m. The Rx is equipped with a sectoral horn antenna with a gain of 19 dBi, whereas the Tx is equipped with an omni-directional antenna. Also, during the measurement of each Tx–Rx link, the Rx antenna is rotated with an angular step of 5° and no moving objects are present.

Fitting of the gaussian & gamma mixtures to the channel gain measurements. In this section, the fading channels are approximated using the GM distribution. Also, some indicative fitting results of modeling the fading channels by means of the MG distribution are presented. In more detail, Figs. 2, 3, 4 and 5 serve as an illustrative example of the fitting achieved by the analytical GMs and MGs expressions, which are obtained as the weighted sum of K Gaussian and K Gamma distributions respectively, to the empirical channel gain measurements of the investigated links. Table 1 quantifies the fitting achieved by the GMs to the empirical measurements of the links in terms of the KL and RMSE fitting accuracy metrics. The link, d , KL, \hat{R} and K columns stand for the TX–RX link index, the transceiver antennas separation distance, the achieved KL and RMSE metric

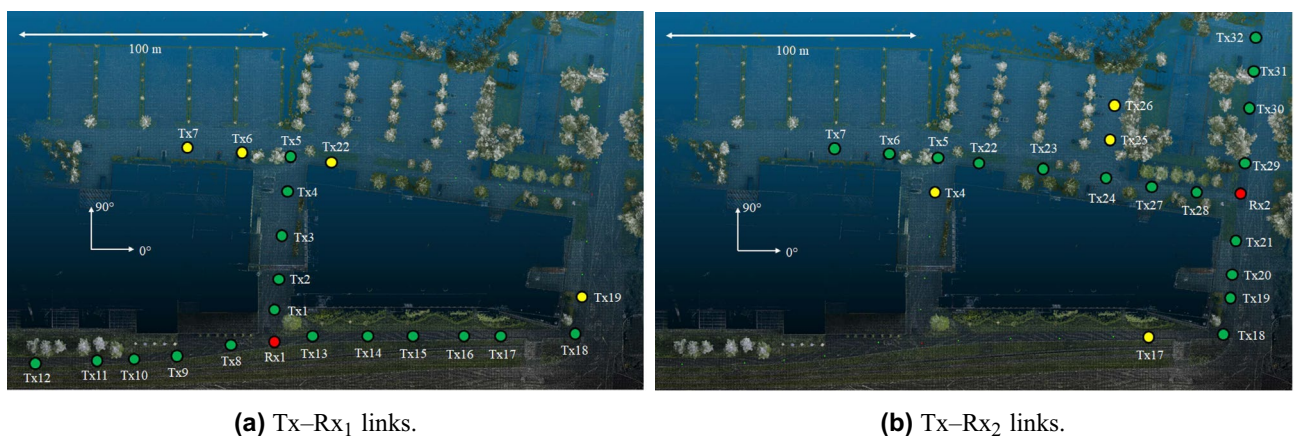


Figure 1. Top-view of the outdoor campus premises.

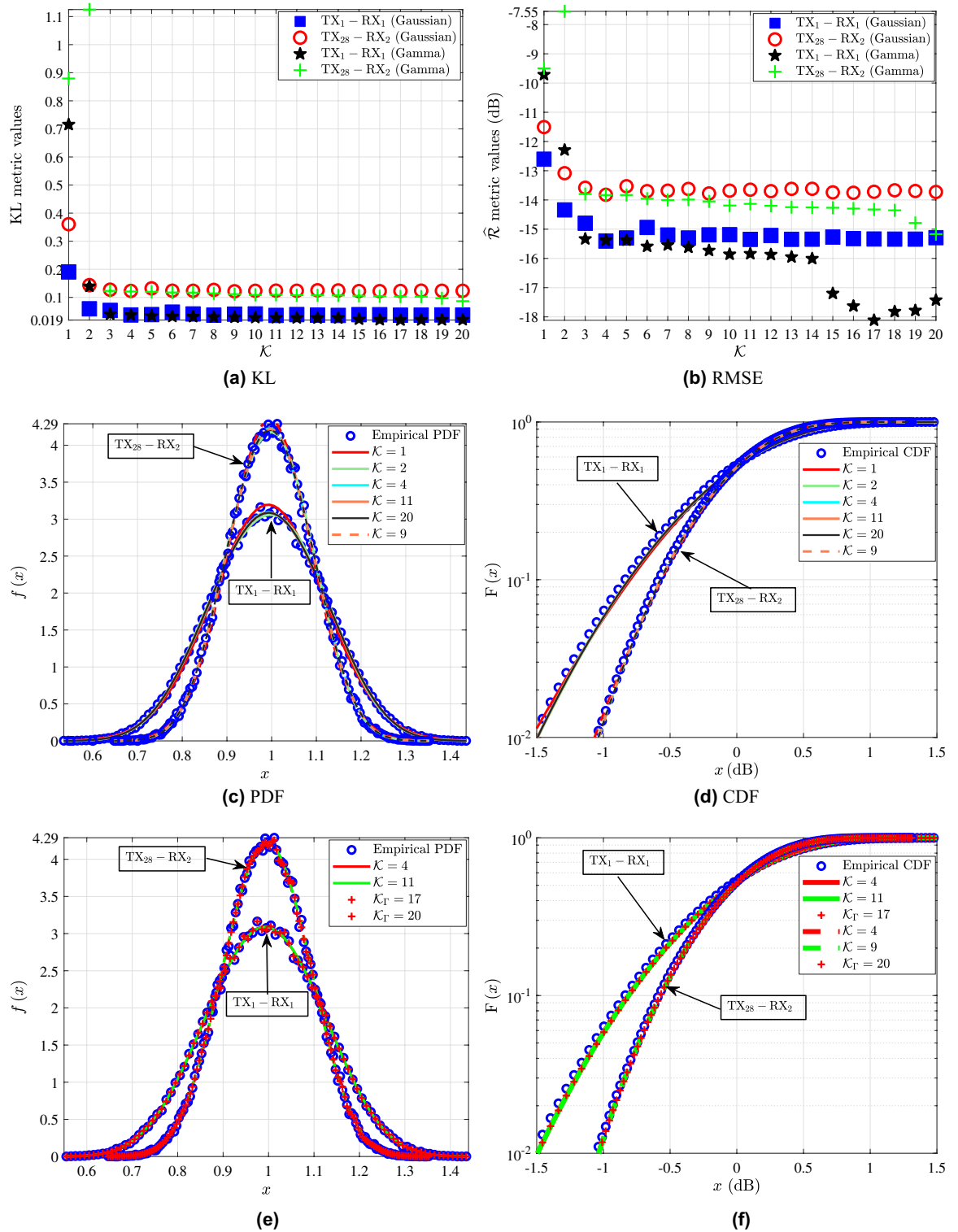


Figure 2. TX₁-RX₁ and TX₂₈-RX₂: (a) KL and (b) \hat{R} metrics for different values of K for the GMs and MGs. (c) Fitting of the PDF and (d) CDF analytical GM expressions to the empirical channel gain data. (e) Fitting of the PDF and (d) CDF of the best fitting analytical GM and MG expressions to the empirical channel gain measurements.

values and the corresponding K of the GM, respectively. The K GM components that yield the most accurate fit to the empirical channel gain measurements, are selected by using as a criterion the minimization of the KL metric. Meanwhile, the KS metric for $K \in [1, 20]$ for all of the presented links yields a good fit. Hence it is a non

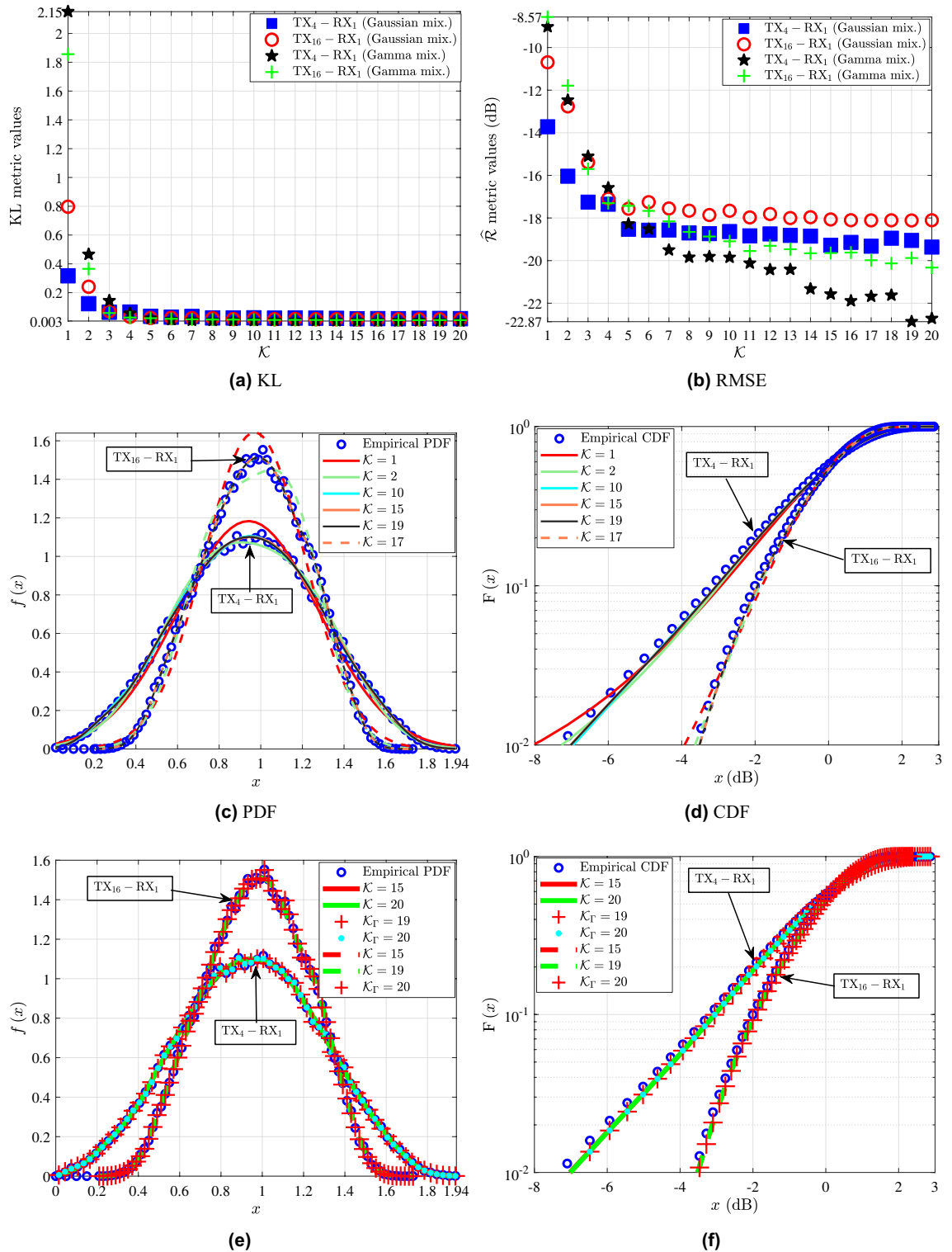


Figure 3. TX₄–RX₁ and TX₁₆–RX₁: (a) KL and (b) $\hat{\mathcal{R}}$ metrics for different values of K for the GMs and MGs. (c) Fitting of the PDF and (d) CDF analytical GM expressions to the empirical channel gain data. (e) Fitting of the PDF and (d) CDF of the best fitting analytical GM and MG expressions to the empirical channel gain measurements.

strict fitting criterion and cannot be employed to identify the K that corresponds to the GM with the best fit to the empirical measurements. Furthermore, the RMSE metric serves as the second best fitting criterion after the KL. Moreover, it should be noted that, the $K \in [1, 20]$ MGs passed the KS test for all of the examined links. As a

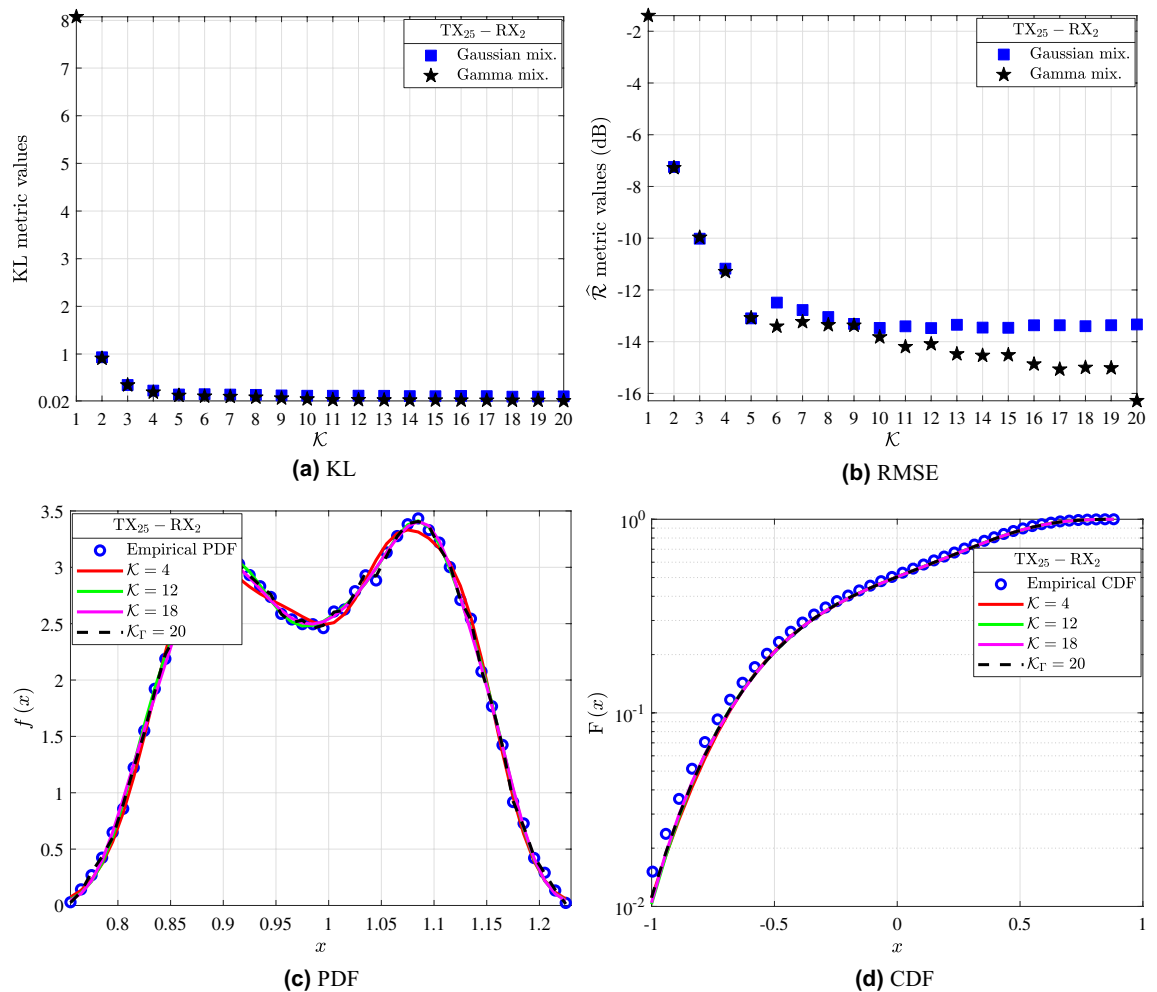


Figure 4. TX₂₅-RX₂: (a) KL and (b) \hat{R} metrics for different values of K for the GMs and MGs. (c) Fitting of the PDF and (d) CDF analytical GM and MG expressions to the empirical channel gain data.

result, the KS test cannot be employed to evaluate the fitting accuracy for the MG distributions. Meanwhile, as it can be observed from Figs. 2, 3, 4 and 5 and especially from 2(a)(b), 3(a)(b), 4(a)(b)s and 5(a)(b); the KL and RMSE tests are reliable fitting accuracy tests not only for the GMs but also for the MGs distributions. Note that, for the interested reader the parameters of the GMs and MGs extracted in this work; can be found on the following link: <https://github.com/T34gr/Gaussian-and-Gamma-mixture-distribution-parameters.git>.

Figure 2 illustrates the statistical characterization of the TX₁-RX₁ and TX₂₈-RX₂ links. In more detail, Fig. 2a shows the KL values of GMs and MGs with different K for both TX₁-RX₁ and TX₂₈-RX₂. As expected, for a given link, as K increases, the KL value of the GMs generally decreases. After achieving a minimum KL value, as K further increases, a short variation towards this value is observed. According to Table 1, for both of the links the maximum KL value is achieved for $K = 1$. Meanwhile, for $K = 4$ the first local minimum of KL is observed for the GMs of both TX₁-RX₁ and TX₂₈-RX₂, which is equal to 0.037 and 0.123, respectively. For the TX₁-RX₁ link, the global minimum value of KL is achieved for the GM with $K = 11$, which can be found in Table 1. On the other hand, for TX₂₈-RX₂ according to Table 1 the global minimum value of KL is achieved for the GM with $K = 9$. For the case of MG modeling, from Fig. 2a it is observed that for both TX₁-RX₁ and TX₂₈-RX₂ as K increases the KL is reduced. Also, for both the links $K = 1$ yields the worst fit, where KL is 0.715 and 0.879, respectively. Furthermore, for TX₁-RX₁ the KL results of the MGs tend to stabilize for $K \geq 15$ and the best fit is achieved for $K = 17$ with KL = 0.019. Also, it is observed that the KL results for both the MGs and GMs for the TX₁-RX₁ link are similar for $K \geq 3$. For the TX₂₈-RX₂ link the MG KL results stabilize for $K \geq 10$ and the best fit is accomplished for $K = 20$ with KL = 0.087. The KL values of the GMs in Table 1 and those of the MGs in Fig. 2a denote that for both links, the MG yields a better fit than the GM. However, as shown in Fig. 2e, both the examined mixture distributions achieve an accurate fit to the empirical channel gain measurements. Meanwhile, in Fig. 2b, the RMSE for different values of K of the GMs and MGs for both the TX₁-RX₁ and TX₂₈-RX₂ links is depicted. According to Table 1, for both of the aforementioned links the maximum RMSE value is achieved for $K = 1$. Meanwhile, for both the TX₁-RX₁ and TX₂₈-RX₂, the GM with $K = 4$ yields the minimum RMSE, which is reported in Table 1. Also, Fig. 2b shows that for both TX₁-RX₁ and TX₂₈-RX₂ the RMSE values of the MGs are lower compared to those of the GMs. In more detail, the MG with $K = 17$ yields the best fit to the

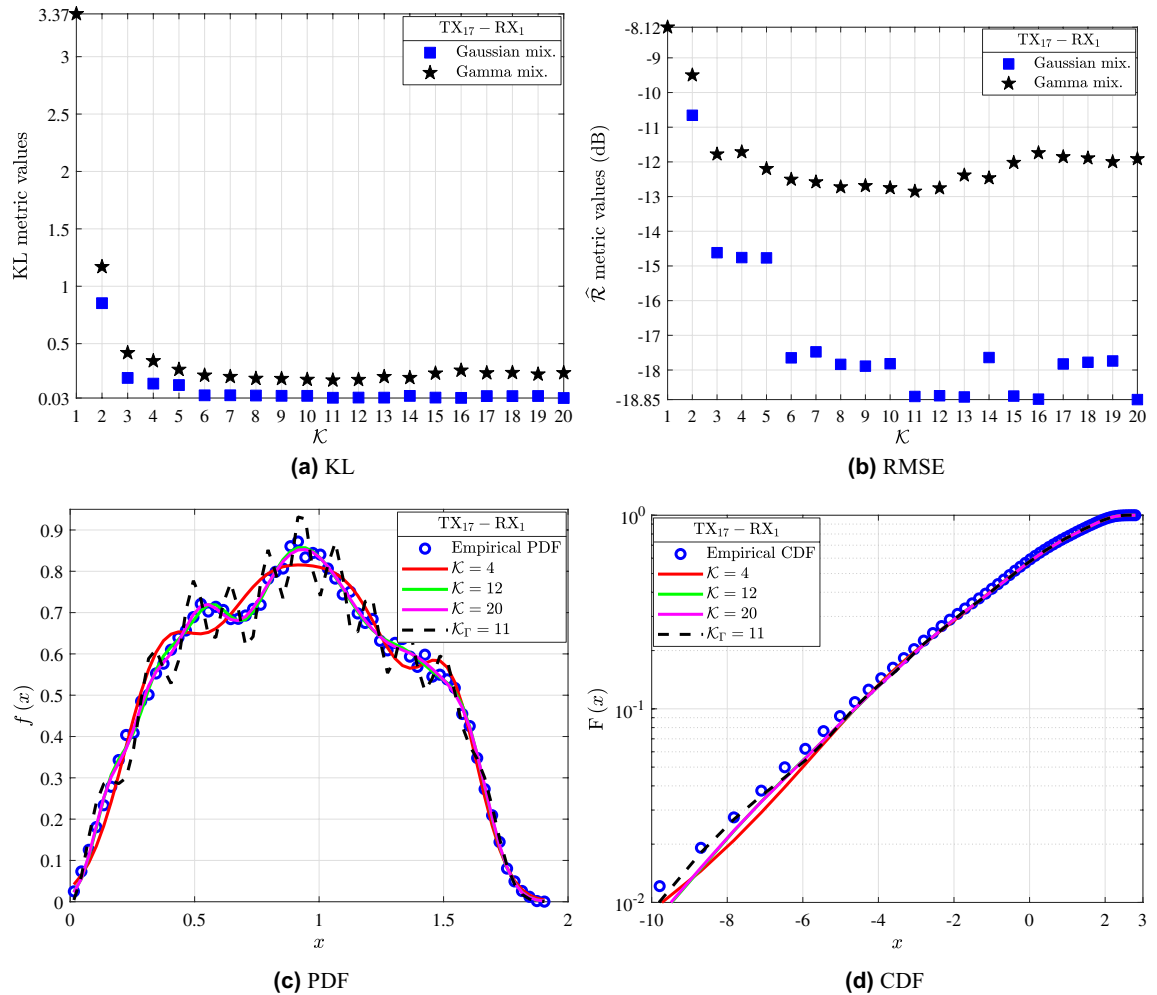


Figure 5. TX₁₇-RX₁: (a) KL and (b) \widehat{R} metrics for different values of K . (c) Fitting of the PDF and (d) CDF analytical GM and MG expressions to the empirical channel gain data.

empirical distribution of TX₁-RX₁ with $\widehat{R} = -18.12$ dB. For the TX₂₈-RX₂ the best GM fit is accomplished for $K = 20$ with $\widehat{R} = -15.17$ dB.

Figure 2c and d serve as an illustrative example of the fitting achieved by the analytical GM expressions with different K to the empirical channel gain PDFs and CDFs for the links TX₁-RX₁ and TX₂₈-RX₂, respectively. Specifically, the blue circles represent the empirical channel gain distributions of the investigated links, while the continuous and dashed lines stand for the fitted GMs of different K for the links TX₁-RX₁ and TX₂₈-RX₂, respectively. Note that, unless otherwise is stated, the continuous and dashed lines of the same color denote GMs with the same K . By taking into account the KL and RMSE values of Table 1 and by examining the fitting of the PDFs and CDFs of the GMs to the empirical channel gain distributions of Fig. 2c and d, it can be ascertained that the increase of K leads to analytical GM expressions that better fit the empirical ones. Fig. 2e and f illustrate the fitting achieved by the analytical PDFs and CDFs of the GMs and MGs with different K to the empirical channel gain measurements of TX₁-RX₁ and TX₂₈-RX₂ links. In these figures, the blue circles represent the empirical channel gain PDFs and CDFs of the links. The continuous and dashed red and green lines stand for the GMs with K equal to 4 and 11, which denote the best fitting GMs to the empirical distributions according to the RMSE and KL metrics of the TX₁-RX₁ and TX₂₈-RX₂ links, respectively. Moreover, the red crosses indicate the MGs that yield the best fit to the empirical distributions according to both the metrics. In more detail, the MG with K equal to $K_{\Gamma} = 17$ is the one that yields the best fit to the empirical distribution of TX₁-RX₁, whereas the MG with K equal to $K_{\Gamma} = 20$ is the one that yields the best fit to the empirical distribution of TX₂₈-RX₂.

Figure 3 depicts the statistical characterization of the TX₄-RX₁ and TX₁₆-RX₁ links. In Fig. 3a, the KL values of GMs and MGs with different K for both TX₄-RX₁ and TX₁₆-RX₁ are presented. For the case of GM modeling, it is observed that, for both TX₄-RX₁ and TX₁₆-RX₁ as K increases KL is reduced. Up to $K = 7$, KL presents a significant variation for both links. However, for $K \in [8, 20]$, the resulting KL values stabilize. From Table 1, the minimum KL value for both the TX₄-RX₁ and TX₁₆-RX₁ links corresponds to a GM with $K = 15$, whereas $K = 1$ leads to the worst fit. For the case of MG modeling, it is observed that, for both TX₄-RX₁ and TX₁₆-RX₁ as K increases KL is reduced. For TX₄-RX₁, up to $K = 5$, KL shows a significant variation, whereas for $K \in [6, 20]$, the KL values stabilize. Also, according to the KL metric the MG that performs the best

Link	d (m)	KL	\hat{R} (dB)	K	Link	d (m)	KL	\hat{R} (dB)	K
Tx ₁ – Rx ₁	13.72	0.191	–12.61	1	Tx ₂₄ – Rx ₂	57.11	0.267	–10.93	1
Tx ₁ – Rx ₁	–	0.037	–15.41	4	Tx ₂₄ – Rx ₂	–	0.038	–13.99	17
Tx ₁ – Rx ₁	–	0.036	–15.35	11	Tx ₂₄ – Rx ₂	–	0.037	–13.98	18
Tx ₁₃ – Rx ₁	16.41	0.097	–14.7	1	Tx ₁₅ – Rx ₁	58.68	0.597	–11.44	2
Tx ₁₃ – Rx ₁	–	0.044	–16.64	8	Tx ₁₅ – Rx ₁	–	0.025	–18.17	17
Tx ₁₃ – Rx ₁	–	0.045	–16.71	11	Tx ₂₅ – Rx ₂	57.71	0.932	–7.25	2
Tx ₂₁ – Rx ₂	19.71	0.238	–11.68	1	Tx ₂₅ – Rx ₂	–	0.106	–13.4	18
Tx ₂₁ – Rx ₂	–	0.154	–12.4	4	Tx ₂₅ – Rx ₂	–	0.126	–13.48	12
Tx ₈ – Rx ₁	23.1	0.109	–13.89	2	Tx ₁₈ – Rx ₂	59.21	0.102	–13.69	1
Tx ₈ – Rx ₁	–	0.058	–15.38	20	Tx ₁₈ – Rx ₂	–	0.058	–15.15	6
Tx ₂ – Rx ₁	27.73	0.144	–14.44	1	Tx ₁₈ – Rx ₂	–	0.058	–15.22	16
Tx ₂ – Rx ₁	–	0.026	–17.47	15	Tx ₄ – Rx ₁	64.46	0.317	–13.72	1
Tx ₂ – Rx ₁	–	0.026	–17.5	19	Tx ₄ – Rx ₁	–	0.019	–19.28	15
Tx ₂₀ – Rx ₂	34.1	0.152	–14.11	1	Tx ₄ – Rx ₁	–	0.019	–19.37	20
Tx ₂₀ – Rx ₂	–	0.063	–15.59	7	Tx ₃₂ – Rx ₂	67.23	0.42	–10.51	1
Tx ₂₀ – Rx ₂	–	0.062	–15.59	9	Tx ₃₂ – Rx ₂	–	0.05	–14.12	13
Tx ₃₀ – Rx ₂	37.25	0.219	–13	1	Tx ₃₂ – Rx ₂	–	0.046	–14.08	20
Tx ₃₀ – Rx ₂	–	0.159	–13.6	8	Tx ₁₁ – Rx ₁	73.6	0.145	–14.29	1
Tx ₃₀ – Rx ₂	–	0.158	–13.57	12	Tx ₁₁ – Rx ₁	–	0.025	–18.42	4
Tx ₁₄ – Rx ₁	38.15	0.061	–15.85	1	Tx ₅ – Rx ₁	78.47	0.195	–14.97	1
Tx ₁₄ – Rx ₁	–	0.018	–18.03	7	Tx ₅ – Rx ₁	–	0.035	–17.85	14
Tx ₁₄ – Rx ₁	–	0.018	–18.09	10	Tx ₅ – Rx ₁	–	0.035	–17.88	20
Tx ₂₇ – Rx ₂	38.6	0.263	–13.68	1	Tx ₁₆ – Rx ₁	81.02	0.797	–10.69	1
Tx ₂₇ – Rx ₂	–	0.064	–15.61	17	Tx ₁₆ – Rx ₁	–	0.015	–18.07	15
Tx ₉ – Rx ₁	40.02	0.331	–12.89	1	Tx ₁₆ – Rx ₁	–	0.015	–18.12	19
Tx ₉ – Rx ₁	–	0.063	–16.1	5	Tx ₂₃ – Rx ₂	81.09	1.593	–6.74	1
Tx ₉ – Rx ₁	–	0.063	–16.11	11	Tx ₂₃ – Rx ₂	–	0.156	–10.83	20
Tx ₁₉ – Rx ₂	42.35	0.454	–12.53	2	Tx ₁₇ – Rx ₁	94.66	0.026	–18.85	20
Tx ₁₉ – Rx ₂	–	0.024	–18.61	20	Tx ₁₂ – Rx ₁	99.61	0.283	–13.16	1
Tx ₃ – Rx ₁	45.19	0.39	–13.69	2	Tx ₁₂ – Rx ₁	–	0.065	–15.38	18
Tx ₃ – Rx ₁	–	0.024	–18.57	18	Tx ₁₂ – Rx ₁	–	0.064	–15.37	20
Tx ₃ – Rx ₁	–	0.024	–18.62	19	Tx ₂₂ – Rx ₂	110.1	0.528	–9.54	1
Tx ₃₁ – Rx ₂	53.21	0.116	–13.44	1	Tx ₂₂ – Rx ₂	–	0.138	–12	12
Tx ₃₁ – Rx ₂	–	0.048	–14.78	8	Tx ₂₂ – Rx ₂	–	0.135	–12	19
Tx ₃₁ – Rx ₂	–	0.046	–14.74	9	Tx ₁₈ – Rx ₁	127.86	0.391	–12.17	1
Tx ₁₀ – Rx ₁	57.11	0.187	–14.35	1	Tx ₁₈ – Rx ₁	127.86	0.064	–15.03	10
Tx ₁₀ – Rx ₁	–	0.015	–18.94	12	Tx ₁₈ – Rx ₁	127.86	0.064	–15.05	20
Tx ₁₀ – Rx ₁	–	0.015	–18.98	13	–	–	–	–	–
Tx ₂₈ – Rx ₂	19.86	0.123	–13.82	4	Tx ₂₈ – Rx ₂	19.86	0.122	–13.78	9

Table 1. Fitting accuracy metrics of GMs with different values of K. Significant values are in bold

fit to TX₄–RX₁ is the one with K = 20 and KL = 0.003, whereas the worst fit is for the MG with K = 1 and KL = 2.152. By taking this into account, and according to the KL results for the GMs of Table 1, the MG yields a better fit to the empirical channel gain distribution of this link. For TX₁₆–RX₁, up to K = 7 the KL results of MGs vary significantly, whereas, for K ∈ [8, 20], the KL values of the MGs tend to stabilize. Also, according to the KL metric the MG that yields the best fit for this link is the one with K = 20 and KL = 0.009, whereas the worst fit is obtained for K = 1 and KL = 1.856. From the KL values for the GMs of Table 1, it is observed that the MG achieves a better fit to the empirical channel gain measurements of TX₁₆–RX₁ in comparison with the GM. However, from Fig. 3a it is deduced that, both the GMs and MGs for the TX₄–RX₁ and TX₁₆–RX₁ link yield similar KL values for K ∈ [8, 20]. In more detail, the good fit achieved by both the GMs and MGs to the empirical distributions of the investigated links is demonstrated by means of Fig. 3e and f. Meanwhile, Fig. 3b shows the RMSE metric results of the GMs and MGs for different values of K for both of the TX₄–RX₁ and TX₁₆–RX₁ links. For both TX₄–RX₁ and TX₁₆–RX₁ it is observed that as K increases the RMSE of the GMs is improved. However, for both the links the RMSE values for K ≤ 10 showcase significant variation. Table 1 reveals that, for both of the links K = 1, yields the worst fitting accuracy performance, in terms of RMSE. Meanwhile, for TX₄–RX₁ the GMs with K = 15, K = 17, and K = 20 yield R equal to –19.28, –19.33, and –19.37 dB respectively. For

TX₁₆–RX₁ the resulting RMSE values almost stabilize for $K \geq 15$. For example, $K = 15$, $K = 19$, and $K = 20$ yield $\hat{R} = -18.07$, $\hat{R} = -18.12$, and $\hat{R} = -18.11$ dB, respectively.

From Fig. 3b, it is observed that for the TX₄–RX₁ link the RMSE results of the MGs vary significantly for $K \leq 8$ and improve with the increase of K . Meanwhile, based on the RMSE metric the MG with $K = 1$ yields the worst fit, whereas the best fit is achieved for $K = 19$ with $\hat{R} = -22.87$. Furthermore, the RMSE metric results shown in Table 1 for the GMs and Fig. 3b demonstrate the better fitting accuracy of MGs compared to GMs for the empirical channel gain distribution of TX₄–RX₁. For the TX₁₆–RX₁ link as Fig. 3b illustrates the RMSE values of the MGs tend to stabilize for $K \geq 11$. The best fit for the link according to the RMSE is achieved for $K = 20$ with $\hat{R} = -20.33$ dB, whereas the worst for $K = 1$ with $\hat{R} = -8.57$ dB. Also, according to the RMSE values of the GMs for the TX₁₆–RX₁ link of Table 1 and Fig. 3b, the MGs yield a better fit to the empirical channel gain measurements of this link. Fig. 3c and d present the fitting accomplished by the analytical PDFs and CDFs of GMs with different K to the empirical channel gain distributions of the links TX₄–RX₁ and TX₁₆–RX₁. The blue circles represent the empirical channel gain distributions of the investigated links, while the continuous and dashed lines stand for the fitted analytical GMs of for TX₄–RX₁ and TX₁₆–RX₁, respectively. By taking into account the KL and RMSE values of Table 1 and by observing Fig. 2c and d, it can be ascertained that the increase of K leads to analytical GM expressions with improved fit to the empirical PDF and CDF. Moreover, it is obvious that a single Gaussian distribution (i.e. $K = 1$) can not accurately describe the empirical data. Figures 3e and f illustrate the fitting achieved by the analytical PDFs and CDFs of the GMs and MGs with different K to the empirical channel gain distributions of the TX₄–RX₁ and TX₁₆–RX₁ links. In these figures, the blue circles stand for the empirical channel gain PDFs and CDFs of the TX₄–RX₁ and TX₁₆–RX₁. The continuous red and green lines denote the best fit achieved by the analytical GM to the empirical data of TX₄–RX₁ according to the RMSE and KL metrics, respectively, while the corresponding dashed lines denote the best fitting GM curves to TX₁₆–RX₁. Meanwhile, the curves marked with the red crosses and cyan dots indicate the analytical MGs that yield the best according to the RMSE and KL metrics to the empirical distribution of TX₄–RX₁ link with K equal to $K_{\Gamma} = 19$ and $K_{\Gamma} = 20$, respectively, while the red crosses with $K_{\Gamma} = 20$ denote the MG that yields the best fit to TX₁₆–RX₁ according to both metrics. Figures 3e and f illustrate that both the GMs and MGs can yield a good fit to the data and can be both considered for the THz small-scale fading channel modeling.

Figure 4 presents the statistical characterization of TX₂₅–RX₂ link. In more detail, Fig. 4a shows the KL achieved by GMs and MGs with different K . For the GM it is observed that as K increases the KL improves. The value of $K = 5$ yields $KL = 0.151$, which is the first local minimum. Meanwhile, for $K \geq 9$ the KL stabilizes to almost the optimum value. For example, GMs with $K = 9, 14, 18$, and 20 result to $KL = 0.131, 0.117, 0.106$, and 0.112 , respectively. Meanwhile, according to Table 1, $K = 2$ yields the maximum value of KL and hence the worst fit. Moreover, from Fig. 4a it is observed that the MGs have similar performance with the GMs in terms of fitting when the KL metric is employed. The best fit of the MG is achieved for $K = 20$, where $KL = 0.016$. The similar fitting performance of GM and MG can also be observed in Fig. 4c and d. In Fig. 4b the RMSE for GMs and MGs with different K is presented. In more detail, for the GMs the first local minimum is obtained for $K = 4$ and is $\hat{R} = -11.18$ dB, while the second local minimum results for $K = 5$ and is $\hat{R} = -13.1$ dB. Moreover, for $K \geq 10$ the RMSE almost stabilizes to the optimum value. For example, the GMs with $K = 10, 12$, and 20 yield $\hat{R} = -13.47, -13.48$, and -13.4 dB, respectively. Similar observations for the RMSE results of the MGs can be extracted as those for the GMs. However, according to this metric the MGs perform significantly better in terms of fitting for $K \geq 13$. The best fit is accomplished for the MG with $K = 20$, where $\hat{R} = -16.28$ dB.

In Fig. 4c and d the fitting achieved by the analytical PDF and CDF GM and MG expressions with different values of K to the empirical channel gain distribution of TX₂₅–RX₂ are presented. In more detail, the blue circles stand for the empirical distribution of the investigated link, whereas the continuous red, green and magenta lines indicate the GM with K equal to 4, 12 and 20, respectively. Also, the dashed black lines denote the analytical MG expressions obtained for K equal to $K_{\Gamma} = 20$, which denotes the best fitting MG based on both metrics. Figure 4c and d illustrate that the best fit to the empirical data is accomplished by the GM with $K = 18$, which is in accordance with the KL metric results. Also, it can be concluded that, in the case an empirical PDF with multiple peaks the increase of K , leads to a GM with a higher fitting accuracy performance. In this sense, the GM with $K = 4$ performs the worst fit. As an example, for $K = 4$ the metrics are $KL = 0.233$ and $\hat{R} = -11.18$ dB.

Figure 5 presents the statistical characterization of TX₁₇–RX₁ link. In more detail, Fig. 5a shows the KL achieved by GMs and MGs with different K . It is observed that, for the GMs as K increases the KL improves. In more detail, the GM with $K = 5$ yields $KL = 0.139$, which is the local minimum of the KL metric. Meanwhile, for $K \geq 11$ the KL results are almost equal. For example, for $K = 11, 15$ and 20 the resulting KL is equal to $0.029, 0.03$, and 0.026 , respectively. Furthermore, based on Table 1, the GM with $K = 2$ performs the worst fit in terms of the KL metric. Meanwhile, from Fig. 5a it is similarly observed that the increase of K improves the fitting accuracy of the MGs to the empirical channel gain data. Also, the KL results tend to stabilize for $K \geq 10$ and the best fit for the MG is accomplished for $K = 11$, where $KL = 0.181$. It should be noted that, according to the GMs KL metric of Table 1 the GM performs a better fit to the empirical data when compared to MG in terms of the KL metric. This significant difference is illustrated in Fig. 5c. In Figure 5b the RMSE for GMs and MGs with different K is presented. It is observed that, for the GMs the first and second RMSE local minima are $\hat{R} = -14.77$ and -17.89 dB, which are obtained for a GM with $K = 5$ and 9 , respectively. The minimum RMSE according to Table 1 is accomplished for the GM with $K = 20$. Meanwhile, as Fig. 5b illustrates as K increases the RMSE of the MGs improves and then deteriorates. This indicates that for the TX₁₇–RX₁ link increasing the number of Gamma mixtures does not improve the fitting performance. The best fit in terms of the RMSE metric for the MG is achieved for $K = 11$, where $\hat{R} = -12.85$ dB. Both the KL and RMSE metrics shown in Fig. 5a and b denote that for an empirical distribution with multiple peaks the GM can yield a better fit in comparison with the MG.

Figure 5c and d illustrate the fitting achieved by the analytical PDF and CDF GMs and MGs of different K to the empirical channel gain measurements of TX₁₇–RX₁. In these figures, the blue circles stand for the empirical

distributions, whereas the continuous red, green and magenta lines indicate the analytical PDFs and CDFs of the GMs with $K = 4, 12$ and 20 , respectively. Moreover, the dashed black lines stand for the MG obtained with K equal to $K_{\Gamma} = 11$, which denotes the best fitting MG based on both metrics. Figure 5c and d demonstrate that the GM with $K = 20$ yields the best fit. This can be verified by the KL and RMSE metric results of Table 1. Furthermore, it can be concluded that, in order to analytically describe an empirical distribution presenting multiple peaks a GM with a greater K is needed. As Figure 5c demonstrates the GM with $K = 4$ achieves the worst fit to the empirical data.

Discussion

The majority of the THz small-scale fading channel modeling works employ analytical distributions, such as Nakagami- m , Rayleigh, Rice, $\alpha - \mu$, and Weibull^{7,27–29}. However, these distributions are capable of only describing single-peak fading channels. In this work, the suitability of modeling single and multiple peaks PDFs of outdoor THz channels in terms of GMs is investigated. Also, MGs are fitted to the empirical channel gain measurements of some indicative links. It is observed that, for both the cases of single and multiple peaks, empirical channel gain distributions the increase of K yields GMs and MGs that better fit the data. Accordingly, this is verified by the results of the KL and RMSE fitting accuracy metrics. In more detail, for all of the investigated links, for the lower values of K , the KL and RMSE fitting accuracy performance deteriorates. For most of the links, low values of K tend to yield significant variations to the KL and RMSE. On the other hand, for all of the examined links, as K increases beyond a specific value, the KL and RMSE fitting accuracy results tend to stabilize. This elucidates that, for any given link, the best fit is accomplished by a GM or MG with a particular value of K or higher. Hence, further increasing K is expected to make only a slight difference on the fitting performance of the GMs to the empirical distributions. Moreover, from the analytical GM and MG distributions illustrated in Figs. 2, 3, 4 and 5 and according to equations (4) and (7) the defining parameter for an analytical GM or MG to present significant peaks is the weight parameter w of its Gaussian or Gamma distribution coefficients. In more detail, as an example, for analytical GM distributions such as those presented in Figs. 2(c) and 3(c) for each K the differences of the w parameters are not significant. On the other hand, for analytical GM distributions such as those that are shown in Figs. 4(c) and 5(c), especially by increasing K there are w values that are greater compared to the rest. As a result, the corresponding Gaussian distribution coefficient with such a w , is more prominent in defining the peak amplitudes of the total GM. To demonstrate this, Table 2 presents the w parameter values for the $Tx_1 - Rx_1$ and $Tx_{25} - Rx_2$ links. Moreover, the fitting accuracy statistics for the MGs employed in this work, verified that the MGs can model the small-scale fading amplitude of THz links. By comparing the fitting accuracy of the MGs and GMs for some indicative TX-RX links, it is observed that they both achieve a good fit to the empirical channel gain measurements. This observation verifies the previous technical works, where both the GMs and MGs were found suitable for THz channel modeling²⁵. Meanwhile, the MG yields a better fit than the GM for the majority of the investigated links. However, the fitting accuracy of the GM is superior than that of the MG for links with multiple peaks with severe changes of amplitude. In more detail, as Fig. 5c illustrates and based on the KL and

$Tx_{25} - Rx_2$							
$Tx_1 - Rx_1$	$K = 2$	$K = 4$	$K = 11$	$K = 20$	$K = 4$	$K = 12$	$K = 18$
w_1	0.35	0.069	0.02	0.01	0.2	0.06	0.05
w_2	0.65	0.23	0.03	0.01	0.22	0.06	0.05
w_3	–	0.34	0.03	0.03	0.27	0.07	0.05
w_4	–	0.36	0.05	0.03	0.31	0.07	0.05
w_5	–	–	0.09	0.04	–	0.08	0.05
w_6	–	–	0.1	0.04	–	0.08	0.05
w_7	–	–	0.12	0.05	–	0.08	0.05
w_8	–	–	0.13	0.05	–	0.09	0.05
w_9	–	–	0.13	0.05	–	0.09	0.05
w_{10}	–	–	0.16	0.06	–	0.1	0.05
w_{11}	–	–	0.16	0.06	–	0.11	0.05
w_{12}	–	–	–	0.06	–	0.11	0.06
w_{13}	–	–	–	0.06	–	–	0.06
w_{14}	–	–	–	0.06	–	–	0.06
w_{15}	–	–	–	0.06	–	–	0.06
w_{16}	–	–	–	0.06	–	–	0.06
w_{17}	–	–	–	0.06	–	–	0.06
w_{18}	–	–	–	0.06	–	–	0.06
w_{19}	–	–	–	0.07	–	–	–
w_{20}	–	–	–	0.07	–	–	–

Table 2. Weight parameters, w for the GMs with K equal to 2, 4, 11, 12, 18, and 20 for the $Tx_1 - Rx_1$ and $Tx_{25} - Rx_2$ links.

RMSE fitting accuracy tests, the GM yields an accurate fit with $K = 20$ to the empirical PDF of $\text{TX}_{17}\text{-RX}_1$. On the other hand, the MG fails to yield a good fit to the data for $K \leq 20$, where according to both the metrics the best fit of the MG is accomplished for $K = 11$. As a consequence, the resulting analytical PDF and CDF MG expressions do not fit at all the empirical ones of the $\text{TX}_{17}\text{-RX}_1$ link. Finally, as a future work we intend to use more outdoor THz wireless measurements and compare the fitting achieved by Gaussian and Gamma mixtures to the empirical channel distributions.

Methods

Preprocessing of the measurement data. The wireless communication channel is expressed in terms of the product of one deterministic and one stochastic coefficient. The deterministic part encapsulates the large-scale effects of the propagation, i.e., the pathloss. The large-scale fading phenomena are time-invariant and remain unchanged during the wireless signal propagation. On the other hand, the stochastic channel coefficient expresses the small-scale fading characteristics of the channel, which are time and frequency dependent. The study of the small-scale fading behavior of RF wireless signals is of great importance, because it can cause unpredicted deep fades to the received signal power. As a consequence, to perform small-scale fading characterization of the channel, the effect of pathloss should be eliminated. The channel sounding performed in the outdoor campus measurements provides power angular delay profiles (PADPs) for each of the Tx–Rx links. For any given link, the PADPs are expressed

$$\text{PADP}(\phi, t) = \sum_{i=1}^I G P_i \delta(\phi - \phi_i) \delta(t - t_i), \quad (1)$$

where ϕ_i , P_i and t_i stand for the azimuth angle at the Rx, the propagation delay gain and time of the i -th propagation path, respectively. The parameter G , known as the broadside angle, denotes the combined gains of the Tx and Rx antennas, while $\delta(\cdot)$ and I are the Dirac delta function and the total number of multipath components of a link, respectively. Subsequently, in order to eliminate the deterministic phenomenon of pathloss, by employing (1) to each link, the link pathgain measurements are normalized to unity as

$$\zeta_i^2 = \frac{P_i}{\frac{\sum_{i=1}^I P_i}{I}}. \quad (2)$$

Incrementing a link channel realizations. The inherent high frequencies of the THz band lead to much higher propagation losses in comparison with the lower mmWave and ultra-high-frequency (UHF) bands^{15,17,46}. The THz free space pathloss even at distances of a few meters and a low transmission frequency can be severe. As an example, for an operational frequency of 140 GHz and a communication distance of 1 m the free space pathloss can be in the excess of 80 dB^{17,47}. Moreover, the atmospheric water vapor causes severe attenuation to the propagating THz signal^{7,15}. Also, the wavelength of the emitted THz signal can be much smaller compared to the size of obstacles laid within the propagation environment⁴⁸. As a consequence, the refraction and reflection losses of the THz band are significantly stronger, when compared to lower frequency bands^{46,49–51}. This leads to a significant reduction of the number of dominant rays, since the THz signal power is drastically weakened, when it is reflected or scattered two or more times^{48,49}. In this sense, the ability of the THz electromagnetic wave to propagate through blockages is nearly lost, due to the severe penetration loss. As a result, the ability of THz signals to diffract around obstacles is significantly reduced. The aforementioned remarks elucidate that, the THz band yields non-rich multipath environments, when compared for example to the mmWave band. However, still there are surfaces that can act as scatterers for propagating wireless THz signals^{16,17,20,29,46}. This leads to the existence of reflected NLoS multipath components carrying a significant amount of power, which are capable of being detected by the Rx. Nevertheless, the amount of measured multipath components, utilized in our analysis, is still not adequately enough to perform small-scale fading statistics analysis for a THz wireless channel. This limitation is surpassed by generating different realizations of the transfer function. This is accomplished by changing the phases of the measured multipath components of a link^{7,45}. The random phases are assumed to be stochastic and are given by a uniform distribution in the interval $[0, 2\pi]$. This assumption is based on the contribution of Molisch et. al, which was based on the principle that the aggregated phases of different paths in an environment of moving scatterers followed a uniform distribution⁴⁵. Hence, from the electromagnetic theory point of view, this is extracted by taking into account the phase shift due to the Doppler effect and it stands in any propagation environment where motion is present. The channel coefficient of the system can be obtained as^{7,45}

$$h = \sum_{i=1} \zeta_i \exp(-j2\pi f t_i) \exp(j\psi_i), \quad (3)$$

where $\psi_i \sim U(0, 2\pi)$ represents the random phase of the i -th multipath component. Moreover, by assuming that the amplitude of the channel coefficients does not change dramatically among the progressing t_i , i.e., the channel can be considered as flat-fading then, $t_i = 0$ ⁴⁵. Also, the term $U(\cdot, \cdot)$ is the uniform distribution operator⁵².

Expectation-maximization based fitting approach. *The gaussian and gamma mixture models.* The THz small-scale fading phenomenon has been the epicenter of many recent channel modeling studies^{7,25,27,51}. Moreover, it has been experimentally observed that there are wireless THz propagation scenarios, where the small-scale fading channel amplitude shows significant fluctuations²⁵. In this sense, the commonly used ana-

lytical distributions that are only capable of fitting single peak distributions are now inadequate to describe the small-scale fading amplitude of such THz channels. However, by considering small-scale fading THz and lower frequency studies, mixture distributions such as Gaussian and Gamma can be employed instead^{25,35,38,53}.

The GMs have been extensively employed to describe the small-scale fading channel amplitude of RF wireless channels^{25,35,38}. The PDF of the GM is defined as

$$f_{gm}(x) = \sum_{i=1}^K w_i \frac{\exp\left(-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right)}{\sqrt{2\pi}\sigma_i}, \quad (4)$$

where K and w_i denote the number of GM components and the weight of the i -th mixture component, respectively. The parameters μ_i and σ_i stand for the mean and standard deviation of the i -th GM component, respectively. Also, $w_i \in [0, 1]$ and

$$\sum_{i=1}^K w_i = 1. \quad (5)$$

The CDF of the GM is expressed as

$$F_{gm}(x) = \frac{1}{2} \sum_{i=1}^K w_i \operatorname{Erfc}\left(\frac{\mu_i - x}{\sqrt{2}\sigma_i}\right), \quad (6)$$

where $\operatorname{Erfc}(\cdot)$ is the complementary error function⁴¹. Moreover, of note is the fact that the K Gaussian distributions that comprise equation (4) are mutually independent. Hence, the GM is not only a favorable distribution for modeling significant empirical distribution amplitude fluctuations, but also it can offer analytical tractability. The latter is of great importance, when the performance analysis of a wireless system must be evaluated. Also, it should be noted that since this work employs pathloss measurements the x instance of a GM is always non-negative, hence for the PDF of equation (4) $x \in [0, \infty)$.

The MGs have been employed in various channel modeling works in lower frequency bands and the THz band as well^{25,54,55}. The PDF of the MG is defined as

$$f_{\Gamma m}(x) = \sum_{i=1}^K w_i \frac{1}{b_i^{a_i} \Gamma(a_i)} x^{a_i-1} \exp\left(-\frac{x}{b_i}\right), \quad (7)$$

where a_i and b_i stand for the shape and scale parameters of the i -th MG component. Also, according to the definition of equation (7) $x \in [0, \infty)$ and the operator $\Gamma(\cdot)$ denotes the gamma function⁴¹. The CDF of the MG is defined as

$$F_{\Gamma m}(x) = \sum_{i=1}^K w_i \frac{1}{\Gamma(a_i)} \gamma\left(a_i, \frac{x}{b_i}\right), \quad (8)$$

where $\gamma(\cdot, \cdot)$ stands for the lower incomplete gamma function⁴¹.

The expectation maximization algorithm. The weights and the parameters of the Gaussian distributions that compose the GM with the best possible fit to the empirical data must be identified by employing an appropriate method. The EM algorithm is such a method. The EM is a machine learning approach that simplifies maximum-likelihood-estimate (MLE) problems and is vastly used in calculating the parameters of mixture models^{25,35}.

The EM is a two step algorithm. It consists of the expectation (E) and the maximization (M) steps⁴⁰. To operate the EM algorithm, the K number of mixtures and the vector $\mathbf{y} = (y_1, \dots, y_n)$ of the n channel gain measurements of a link are required as inputs. Subsequently, the mixtures parameters are updated at the M-step during the $m + 1$ iteration of the EM algorithm until the convergence criterion is met. Otherwise the EM terminates, when a predefined number of repetitions is reached. The converge criterion is defined as

$$|L^{[m+1]} - L^{[m]}| > \varepsilon, \quad (9)$$

where ε stands for the desired convergence value. The term $L^{[m]}$ signifies the MLE log-likelihood at the m -th iteration of the EM algorithm and can be obtained as

$$L^{[m]} = \frac{1}{n} \sum_{i=1}^n \ln \left(\sum_{j=1}^K w_j^{[m]} \phi\left(y_i \mid \mu_j^{[m]}, \sigma_j^{[m]}\right) \right), \quad (10)$$

where $j \in [1, K]$, $i \in [1, n]$ and $\ln(\cdot)$ stands for the natural logarithm. The term $\phi\left(y_i \mid \mu_j^{[m]}, \sigma_j^{[m]}\right)$ is the Gaussian distribution of the j -th mixture component at the m -th iteration of the EM, which has mean and standard deviation $\mu_j^{[m]}$ and $\sigma_j^{[m]}$, respectively. Meanwhile, the E-step of the EM is implemented as

$$\gamma_{ij}^{[m]} = \frac{w_j^{[m]} \phi\left(y_i \mid \mu_j^{[m]}, \sigma_j^{[m]}\right)}{\sum_{l=1}^K w_l^{[m]} \phi\left(y_i \mid \mu_l^{[m]}, \sigma_l^{[m]}\right)}. \quad (11)$$

Upon the completion of the E-step, the EM algorithm implements the M-step. The M-step provides the updated values of the distribution parameters of the j -th mixture at the $m + 1$ step of the algorithm, which for the particular case of a GM are calculated as in equations (12)–(14)

$$w_j^{[m+1]} = \frac{1}{n} \sum_{i=1}^n \gamma_{ij}^{[m]}, \quad (12)$$

$$\mu_j^{[m+1]} = \frac{\sum_{i=1}^n \gamma_{ij}^{[m]} y_i}{\sum_{i=1}^n \gamma_{ij}^{[m]}}, \quad (13)$$

$$\sigma_j^{[m+1]} = \sqrt{\frac{\sum_{i=1}^n \gamma_{ij}^{[m]} (y_i - \mu_j^{[m+1]})^2}{\sum_{i=1}^n \gamma_{ij}^{[m]}}}. \quad (14)$$

The convergence of the EM algorithm depends on K and the initialization values of the mixtures parameters that are provided as inputs. Several methods are available to provide initialization values for the mixtures parameters. One of the most common is to employ the K -nearest-neighbour (KNN) algorithm⁵⁶.

Evaluation of the fitting. *The kolmogorov-smirnov test.* The KS goodness of fit test is defined as⁴¹

$$\max(|F_{emp}(x) - F_{gm}(x)|) \leq \sqrt{-\frac{1}{2N} \ln\left(\frac{A}{2}\right)}, \quad (15)$$

where $F_{emp}(x)$ and N stand for the empirical values of the channel gain CDF of the examined link and the number of discrete samples of $F_{emp}(x)$, respectively. The parameter $F_{gm}(x)$ denotes the analytical CDF of the examined analytical distribution, while $A = 5\%$ is the selected significance level.

Kullback-leibler divergence test. The KL divergence test is defined as the distance between the empirical PDF $f_{emp}(x)$ and the analytical PDF $f_{gm}(x)$ of the examined distribution i.e.,⁴²

$$KL = -\sum_{i=1}^N f_{emp}(x_i) \ln\left(\frac{f_{gm}(x_i)}{f_{emp}(x_i)}\right) \quad (16)$$

The closer the value of equation (16) to 0 the better is the fit of the analytical fading distribution to the empirical channel gain distribution.

The root mean square error. The RMSE is defined as⁴³

$$\hat{R} = \sqrt{\frac{1}{N} \sum_{i=1}^N (f_{emp}(x_i) - f_{gm}(x_i))^2}. \quad (17)$$

The lower the value of \hat{R} the better the fit of the analytical $f_{gm}(x)$ PDF to the empirical distribution. Also, it should be noted that the RMSE results are commonly presented in dB scale.

Data availability

The initial data are owned by Aalto University Finland. Any researcher affiliated to one of the ARIADNE project partners is allowed to access and use the shared data for research purposes. The shared data must however not be made accessible to any person not affiliated with an ARIADNE project partner. All the processed data present in this work accompany this manuscript as a supplementary material.

Received: 4 November 2022; Accepted: 15 April 2023

Published online: 19 April 2023

References

- Boulogeorgos, A.-A.A., Jornet, J. M. & Alexiou, A. Directional terahertz communication systems for 6G: Fact check. *IEEE Veh. Techn. Mag.* **16**, 68–77. <https://doi.org/10.1109/MVT.2021.3113883> (2021).
- Alwis, C. D. *et al.* Survey on 6G frontiers: Trends, applications, requirements, technologies and future research. *IEEE Open J. Commun. Soc.* **2**, 836–886. <https://doi.org/10.1109/OJCOMS.2021.3071496> (2021).

3. Boulogeorgos, A.-A.A. *et al.* Terahertz technologies to deliver optical network quality of experience in wireless systems beyond 5G. *IEEE Commun. Mag.* **56**, 144–151 (2018).
4. Sardeddeen, H., Alouini, M.-S. & Al-Naffouri, T. Y. An overview of signal processing techniques for terahertz communications. *Proc. IEEE* **109**, 1628–1665. <https://doi.org/10.1109/JPROC.2021.3100811> (2021).
5. Sardeddeen, H., Saeed, N., Al-Naffouri, T. Y. & Alouini, M.-S. Next generation terahertz communications: A rendezvous of sensing, imaging, and localization. *IEEE Commun Mag* **58**, 69–75. <https://doi.org/10.1109/MCOM.001.1900698> (2020).
6. Trelvakis, S. *et al.* Localization as a key enabler of 6G wireless systems: A comprehensive survey and an outlook. *Techriv*, <https://doi.org/10.36227/techriv.21822096.v1>
7. Papatotiriou, E. N., Boulogeorgos, A.-A.A., Haneda, K., de Guzman, M. F. & Alexiou, A. An experimentally validated fading model for THz wireless systems. *Sc. Rep.* **11**, 1–14 (2021).
8. Abbasi, N. A. *et al.* Double directional channel measurements for THz communications in an Urban environment. *IEEE Int. Conf. Commun. (ICC)* <https://doi.org/10.1109/ICC40277.2020.9148631> (2020).
9. Xing, Y. & Rappaport, T. S. Propagation measurements and path loss models for sub-THz in urban microcells. *IEEE Int. Conf. Commun.* <https://doi.org/10.1109/ICC42927.2021.9500385> (2021).
10. Gougeon, G., Corre, Y. & Aslam, M. Z. Ray-based deterministic channel modelling for sub-THz band. In *IEEE 30th Int. Symp. Per. Ind. Mob. Radio Commun. (PIMRC)*, 1–6, <https://doi.org/10.1109/PIMRCW.2019.8880819> (2019).
11. Ju, S. & Rappaport, T. S. Sub-terahertz spatial statistical MIMO channel model for urban microcells at 142 GHz. *IEEE Global Commun. Conf. (GLOBECOM)* <https://doi.org/10.1109/GLOBECOM46510.2021.9685929> (2021).
12. Abbasi, N. A. *et al.* Thz band channel measurements and statistical modeling for urban D2D environments. *IEEE Trans. on Wir. Commun.* <https://doi.org/10.1109/TWC.2022.3184929> (2022).
13. Chen, Y. & Han, C. Time-varying channel modeling for low-terahertz urban vehicle-to-infrastructure communications. *IEEE Global Commun. Conf. (GLOBECOM)* <https://doi.org/10.1109/GLOBECOM38437.2019.9013865> (2019).
14. Federici, J. F., Ma, J. & Moeller, L. Review of weather impact on outdoor terahertz wireless communication links. *Nano Commun. Net.* **10**, 13–26 (2016).
15. Kokkonen, J., Janne, L. & Markku, J. A line-of-sight channel model for the 100–450 gigahertz frequency band. *EURASIP J. on Wir. Commun. and Net.* (2021).
16. Nguyen, S. L. H., Järveläinen, J., Karttunen, A., Haneda, K. & Putkonen, J. Comparing radio propagation channels between 28 and 140 GHz bands in a shopping mall. In *12th Eur. Conf. on Ant. and Prop. (EuCAP 2018)*, 1–5, <https://doi.org/10.1049/cp.2018.0874> (2018).
17. Nguyen, S. L. H., Haneda, K., Järveläinen, J., Karttunen, A. & Putkonen, J. Large-scale Parameters of Spatio-temporal short-range indoor backhaul channels at 140 GHz. In *2021 IEEE 93rd Veh. Techn. Conf. (VTC 2021-Spring)*, 1–6, <https://doi.org/10.1109/VTC2021-Spring51267.2021.9448958> (2021).
18. Kokkonen, J., Lehtomäki, J. & Juntti, M. Simple molecular absorption loss model for 200–450 Gigahertz frequency band. In *2019 Eur. Conf. Net. Commun. (EuCNC)*, 219–223, <https://doi.org/10.1109/EuCNC.2019.8801950> (2019).
19. Kokkonen, J., Lehtomäki, J. & Juntti, M. Simplified molecular absorption loss model for 275–400 gigahertz frequency band. In *12th Eur. Conf. Antennas Propag. (EuCAP)* (London, UK, 2018).
20. Pometcu, L. & D’Errico, R. An indoor channel model for high data-rate communications in D-band. *IEEE Access* **8**, 9420–9433. <https://doi.org/10.1109/ACCESS.2019.2960614> (2020).
21. Afsharinejad, A., Davy, A., Jennings, B., Rasmann, S. & Brennan, C. A path-loss model incorporating shadowing for THz band propagation in vegetation. *IEEE Glob. Commun. Conf. (GLOBECOM)*, 1–6, <https://doi.org/10.1109/GLOCOM.2015.7417038> (San Diego, CA, USA, 2015).
22. Boronin, P., Petrov, V., Moltchanov, D., Koucheryavy, Y. & Jornet, J. M. Capacity and throughput analysis of nanoscale machine communication through transparency windows in the terahertz band. *Nano Commun. Netw.* **5**, 72–82 (2014).
23. Afsharinejad, A., Davy, A., Jennings, B. & Brennan, C. An initial path-loss model within vegetation in the THz band. In *9th Eur. Conf. Ant. Prop. (EuCAP)*, 1–5 (Lisbon, Portugal, 2015).
24. Boulogeorgos, A.-A.A., Papatotiriou, E. N. & Alexiou, A. Analytical performance assessment of THz wireless systems. *IEEE Access* **7**, 11436–11453. <https://doi.org/10.1109/ACCESS.2019.2892198> (2019).
25. Tekbiyik, K., Ekti, A. R., Kurt, G. K., Görçin, A. & Yarkan, S. Modeling and analysis of short distance sub-terahertz communication channel via mixture of gamma distribution. *IEEE Trans. Veh. Techn.* **70**, 2945–2954. <https://doi.org/10.1109/TVT.2021.3063209> (2021).
26. Boulogeorgos, A.-A. A., Papatotiriou, E. N. & Alexiou, A. analytical performance evaluation of THz wireless fiber extenders. In *IEEE 30th An. Int. Symp. Per., Indoor Mob. Radio Commun. (PIMRC)*, 1–6 (Istanbul, Turkey, 2019).
27. Ekti, A. R. *et al.* Statistical modeling of propagation channels for terahertz band. In *IEEE Conf. Standards Commun. Netw. (CSCN)*, 275–280, <https://doi.org/10.1109/CSCN.2017.8088634> (Helsinki, Finland, 2017).
28. Kim, S. & Zajić, A. Statistical modeling of THz scatter channels. In *9th Eur. Conf. Antennas Propag. (EuCAP)*, 1–5 (Lisbon, Portugal, 2015).
29. Kim, S. & Zajić, A. Statistical modeling and simulation of short-range device-to-device communication channels at sub-THz frequencies. *IEEE Trans. Wireless Commun.* **15**, 6423–6433, <https://doi.org/10.1109/TWC.2016.2585103> (2016).
30. Papatotiriou, E. N., Boulogeorgos, A.-A. A., Guzman, M. F. D., Haneda, K. & Alexiou, A. A new look to THz wireless links: Fading modeling and capacity assessment. In *Per. Ind. Mob. Commun. (PIMRC)* (IEEE, 2021).
31. Papatotiriou, E. N., Boulogeorgos, A.-A. A. & Alexiou, A. fading modeling in indoor THz wireless systems. In *IEEE Int. Balkan Conf. Commun. Netw. (BalkanCom)*, 161–165, <https://doi.org/10.1109/BalkanCom53780.2021.9593181> (2021).
32. Chauhan, P. S., Kumar, S., Soni, S. K., Upadhdhaya, V. K. & Pant, D. Average channel capacity over mixture gamma distribution. In *Int. Conf. El. Electr. Engin. (ICE3)*, <https://doi.org/10.1109/ICE348803.2020.9122966> (2020).
33. Kumar, S., Soni, S. K. & Jain, P. Performance of MRC receiver over Hoyt/lognormal composite fading channel. *Int. J. Electr.* **105**, 1433–1450. <https://doi.org/10.1080/00207217.2018.1460870> (2018).
34. Kumar, S. Energy detection in hoyt/gamma fading channel with micro-diversity reception. *Wir. Per. Commun.* **101**, 723–734 (2018).
35. Selim, B., Alhoussein, O., Muhaidat, S., Karagiannidis, G. K. & Liang, J. Modeling and analysis of wireless channels via the mixture of gaussian distribution. *IEEE Trans. Veh. Techn.* **65**, 8309–8321. <https://doi.org/10.1109/TVT.2015.2503351> (2016).
36. Du, H. *et al.* Performance and optimization of reconfigurable intelligent surface aided THz communications. *IEEE Trans. Commun.* **70**, 3575–3593. <https://doi.org/10.1109/TCOMM.2022.3162645> (2022).
37. Romero-Jerez, J. M., Lopez-Martinez, F. J., Peña-Martín, J. P. & Abdi, A. Stochastic fading channel models with multiple dominant specular components. *IEEE Trans. Veh. Techn.* **71**, 2229–2239. <https://doi.org/10.1109/TVT.2022.3141949> (2022).
38. Alhoussein, O. *et al.* A generalized mixture of gaussians for fading channels. In *2015 IEEE 81st Veh. Techn. Conf. (VTC Spring)*, 1–6, <https://doi.org/10.1109/VTCSpring.2015.7145607> (2015).
39. Valverde, G., Tortós, J. Q. & Terzija, V. Comparison of gaussian mixture reductions for probabilistic studies in power systems. In *2012 IEEE Power Energy Soc. General Meet.*, 1–7, <https://doi.org/10.1109/PESGM.2012.6345346> (2012).
40. Dempster, A. P., Laird, N. M. & Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Stat. Soc. Ser. B (Meth.)* **39**, 1–22 (1977).
41. Papoulis, A. & Pillai, S. Probability, random variables, and stochastic processes. McGraw-Hill Ser. Electr. Eng. Commun. signal process. (Tata McGraw-Hill, 2002).

42. MacKay, D. J. C. *Information theory, inference, and learning algorithms* (Cambridge University Press, 2003).
43. Bickel, P. J. & Doksum, K. A. *Mathematical statistics: basic ideas and selected topics, volumes I-II package, vol. I-II* (Chapman and Hall/CRC, New York, 2015), first edn.
44. de Guzman, M. F., Koivumäki, P. & Haneda, K. Double-directional multipath data at 140 GHz derived from measurement-based ray-launcher. In *2022 95th Veh. Tech. Conf. (VTC2022-Spring)*, 1–6 (Helsinki, Finland, 2022).
45. Molisch, A. F., Steinbauer, M., Toeltsch, M., Bonek, E. & Thoma, R. S. Capacity of MIMO systems based on measured wireless channels. *IEEE J. Select. Area Commun.* **20**, 561–569. <https://doi.org/10.1109/49.995515> (2002).
46. Wu, Y., Kokkonen, J., Han, C. & Juntti, M. Interference and coverage analysis for terahertz networks with indoor blockage effects and line-of-sight access point association. *IEEE Trans. Wir. Commun.* **20**, 1472–1486. <https://doi.org/10.1109/TWC.2020.3033825> (2021).
47. Jornet, J. M. & Akyildiz, I. F. Channel modeling and capacity analysis for electromagnetic wireless nanonetworks in the terahertz band. *IEEE Trans. Wir. Commun.* **10**, 3211–3221 (2011).
48. Stratidakis, G., Pappasotiropoulos, E. N., Konstantinis, H., Boulogeorgos, A.-A. A. & Alexiou, A. Relay-based blockage and antenna misalignment mitigation in THz wireless communications. In *2nd 6G Wir. Sum. (6G SUMMIT)*, 1–4 (2020).
49. Han, C., Bicen, A. O. & Akyildiz, I. F. Multi-ray channel modeling and wideband characterization for wireless communications in the terahertz band. *IEEE Trans. Wir. Commun.* **14**, 2402–2412. <https://doi.org/10.1109/TWC.2014.2386335> (2015).
50. Kokkonen, J., Lehtomäki, J. & Juntti, M. Stochastic geometry analysis for mean interference power and outage probability in THz networks. *IEEE Trans. Wir. Commun.* **16**, 3017–3028. <https://doi.org/10.1109/TWC.2017.2673844> (2017).
51. Khalid, N., Abbasi, N. A. & Akan, O. B. Statistical characterization and analysis of low-THz communication channel for 5G Internet of Things. *Nano Commun. Netw.* **22**, 100258. <https://doi.org/10.1016/j.nancom.2019.100258> (2019).
52. Goldsmith, A. *Wireless communications* (Cambridge university press, 2005).
53. Al-Hmood, H. A mixture gamma distribution based performance analysis of switch and stay combining scheme over alpha -kappa shadowed fading channels. In *2017 An. Conf. New Trends Inf. Commun. Techn. Appl. (NTICT)*, 292–297. <https://doi.org/10.1109/NTICT.2017.7976096> (2017).
54. Wiper, M., Insua, D. R. & Ruggeri, F. Mixtures of gamma distributions with applications. *J. Comput. Graph. Stat.* **10**, 440–454 (2001).
55. da Silva, V. R. & Yongacoglu, A. EM algorithm on the approximation of arbitrary pdfs by gaussian, gamma and lognormal mixture distributions. In *2015 7th IEEE Latin-Am. Conf. Commun. (LATINCOM)*, 1–6 (IEEE, 2015).
56. Vegas-Sanchez-Ferrero, G. *et al.* Gamma mixture classifier for plaque detection in intravascular ultrasonic images. *IEEE Trans. Ultras. Ferroel. Freq. Control* **61**, 44–61. <https://doi.org/10.1109/TUFFC.2014.6689775> (2014).

Acknowledgements

This work has received funding from the European Commission Horizon 2020 research and innovation programme ARIADNE under grant agreement No. 871464.

Author contributions

E. N. P. for all the measured links processed the received power measurements, extracted the empirical distributions, performed the fitting of the analytical Gaussian Mixture distributions to the empirical ones. A.-A. A. B. performed the fitting of the analytical Mixture Gamma distributions to the empirical ones. E. N. P., A.-A. A. B. and A. A. conducted the literature review of the existing works in the field of study. A.-A. A. B. assisted in the organization of the Results section. All authors aided in the justifications of the method section and reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-33598-x>.

Correspondence and requests for materials should be addressed to E.N.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023