INHERENT LIMITATIONS OF PARAMETER ESTIMATION OF A TEMPO-SPATIAL FIELD USING AN ARRAY OF HETEROGENEOUS SENSORS

Aviv Regev, Jonatan Ostrometzky†, Senior Member, IEEE, and Hagit Messer*, Life Fellow, IEEE*

*School of Electrical Engineering, Tel Aviv University, Israel
†Digital Sciences for High-tech, Faculty of Engineering, Tel Aviv University, Israel

ABSTRACT

Based on an analysis of the Fisher Information Matrix (FIM), this paper presents a study of the inherent limitations in parameter estimation of a localized tempo-spatial field, characterized by a parametric model. The problem is motivated by the need to retrieve rain fields in rural areas, where sudden flash floods are a life-threatening hazard. We first identify the minimum number of sensors necessary for estimating all of the unknown parameters, where two types of sensors are considered: time projection sensors - characterized as a point in space, line-projection in time (such as rain gauges), and spatial projection sensors - characterized as a point in time, line projection in space (such as wireless microwave links). We show that a single spatial projection sensor with one or more sensor of any type are required. Then, we show that the Cramer-Rao bound of each of the unknown parameters is characterized by a U-shape curve as a function of the observation period. By studying the condition number of the FIM we identify the sufficient conditions for the estimation errors to be small (i.e., at the bottom of the U-shape). We demonstrate the results of our analysis with different combinations of sensors.

***Index Terms*—** Parameter Estimation, Fisher Information Matrix, Rain Monitoring.

I. INTRODUCTION

How many sensors are needed to identify a localized Spatio-temporal physical field (such as a rain-field, an irradiance field, etc.)? The answer obviously depends on the sensors' type, the field's characteristics, and the sensor's coverage relative to the field's location and movement. Willing to examine the above question, we present a study that deals with the ability to identify a spatio-temporal localized field by a limited number of sensors. We consider temporal measurements from two types of sensors: i) Time Projection Sensor (TPS), where a measurement is an integration over a time interval for a given point in space; ii) Spatial Projection Sensor (SPS), where a measurement is an integration along a line in space in a given time index.

An example of such an application is the case of rain retrieval in rural areas, which is characterized by a sparse coverage of near-ground rain monitoring instruments. Beyond economic drivers like agriculture and livestock farming, frequent unpredicted flash floods might cause damage and even fatalities in those areas. In this example, the TPSs used are rain gauges (RGs), and available commercial microwaves links (CMLs) in the area are the SPSs. RGs detect the accumulated amount of rainfall, over a given time interval, at a fixed point in which they are installed. CMLs can be used as opportunistic sensors for near-ground rain monitoring as first introduced in 2006 [1]. A CML's signal is attenuated due to precipitation along its path, and based on the difference between the measured transmitted signal level (TSL) and the received signal level (RSL), one can relate the CML-path attenuation with the fallen precipitation.

CMLs are widely used in cellular networks for backhaul communication, and therefore are widely spread world-wide. The dense coverage of CMLs in comparison with other monitoring instruments has a clear advantage for weather monitoring purposes and is therefore addressed in many past studies. Some studies focused on the reconstruction of spatial rain fields using CML measurements as a stand alone rainfall monitoring tool [2]–[4], while others combined CMLs with RGs [5]–[7]. Other studies treated the CML as a virtual RG [8], [9]. Furthermore, monitoring the dynamic rain field via CMLs was also proposed [10]–[12].

In this paper, we adopt a parameter estimation approach, previously suggested in [7], [13], where a rain cell is modeled by a parametric model, whose parameters are to be estimated. We adopt a specific model [14], where we assume that a cell has a Gaussian shape, referred to as a Gaussian cell (GC) [15]. We extend the model by introducing additional velocity parameters $\{v_x, v_y\}$, which are required to introduce time dependency.[1] We study a scenario in which a parameter vector $\boldsymbol{\theta}$ is to be estimated from a time series of measurements taken by either TPS or SPS. Based on the chosen model we determine the minimal number of sensors required to identify the field. The sufficient conditions are determined by the number of samples and the sampling rate, which are evaluated by analyzing the Fisher Information

[1] Although this study is based on the GC model, it can be applied to any parametric model, where a field is presented as $R(x, y, t) = \phi(x, y, t; \boldsymbol{\theta})$ where ϕ is a known function and $\boldsymbol{\theta}$ is an unknown parameter vector.

Matrix (FIM). We follow the methodology suggested in [16] where analysis of the FIM is used to identify inherent limitations in an estimation problem. By defining "identifiability" as a situation in which all parameters of the field can be estimated with a reasonable estimation error, we analyse the condition number of the FIM to find sufficient conditions for identifiability.

The rest of the paper is organized as follows: Section II describes the mathematical formalism. Section III presents the necessary conditions to identify the GC according to the number of available sensors. In Section IV, we discuss the sufficient conditions for the presented problem. Lastly, Section V concludes the paper.

II. MATHEMATICAL FORMALISM

II-A. The Proposed Model

The parametric model of the field is based on a two-dimensional Gaussian-shape function (that represents a rain cell), characterized by 8 parameters: R_0, the peak level (i.e., the maximal rain-intensity value in mm/h); $\{\mu_{0_x}, \mu_{0_y}\}$ - the initial position of the center of the cell, in Cartesian coordinates; $\{\sigma_x^2, \sigma_y^2\}$ (in km^2) represent the width of the cell; and $\{v_x, v_y\}$ (in km/h), that represent the velocity (assumed constant during the observation period) of the cell's movement. The parameter ρ controls relation between the x and the y coordinates, and thus, characterizes the shape of the cell. Our model of a localized spatio-temporal field is therefore given by:

$$\begin{aligned} R(x,y,t;\boldsymbol{\theta}) = R_0 \cdot \exp[-\frac{1}{2(1-\rho^2)} \\ \cdot(\frac{(x-(\mu_{0_x}+v_x t))^2}{\sigma_x^2} + \frac{(y-(\mu_{0_y}+v_y t))^2}{\sigma_y^2} \\ -\frac{2\rho(x-(\mu_{0_x}+v_x t))(y-(\mu_{0_y}+v_y t))}{\sigma_x\sigma_y})] \end{aligned} \tag{1}$$

where $\boldsymbol{\theta} = [R_0, \mu_{0_x}, \mu_{0_y}, \sigma_x^2, \sigma_y^2, \rho, v_x, v_y]^T$ is the vector of the unknown parameters. Under this assumed model, the field can be evaluated for any set of coordinates $\{x, y\}$ for any time-index t, once $\boldsymbol{\theta}$ is given.

II-B. The Available Measurements

The measurements are collated to a vector, consists of two types of measurements whose components are given by either (2) for the case of time projection sensors and by (3) for the case of spatial projection sensors:

$$z_{T_i}^n = \int_{t_{T_i}^n-\Delta T}^{t_{T_i}^n} g_T(x_i, y_i, t)dt + W_{T_i}^n \tag{2}$$

$$z_{S_i}^n = \int_L g_S(x(l), y(l), t_{S_i}^n)dl + W_{S_i}^n \tag{3}$$

where $z_{T_i}^n, z_{S_i}^n$ are the n^{th} measurement (taken at the time indexes $t_{T_i}^n, t_{S_i}^n$) from the i^{th} TPS and SPS, respectively. $W_{T_i}^n \sim (0, \sigma_{T_i}^2), W_{S_i}^n \sim (0, \sigma_{S_i}^2)$ are assumed as additive zero mean Gaussian noise processes.[2] ΔT is the temporal span of the TPS, and (x_i, y_i) is the i^{th} TPS's location. $g_T(), g_S()$ can be any known functions. For the case of rain retrieval, these functions are $g_T() = R(x,y,t;\boldsymbol{\theta})$, as a RG measures the rain directly and $g_S()$ is a power law that relates the rain-induced signal attenuation along the propagation path with the rain rate, given by [17]:

$$g_T(x,y,t;\boldsymbol{\theta}) = aR^b(x,y,t;\boldsymbol{\theta}) \tag{4}$$

where $\{a, b\}$ are the power-law coefficients, that depend on the frequency and polarization of the signal as well as on the rain drop size distribution. These coefficients are assumed to be known and can be found in the literature [18].

II-C. The Fisher Information Matrix

For the estimation problem described above, and assuming that the additive noise values are independent and identically distributed (iid), the FIM, $I_{\boldsymbol{\theta}}(\boldsymbol{\theta})$, can be expressed by [19]:

$$I_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \sum_{i=0}^{N-1} \frac{1}{\sigma_i^2} \cdot \nabla_{\boldsymbol{\theta}} f_i \cdot \nabla_{\boldsymbol{\theta}} f_i^T \tag{5}$$

Where f_i is the i^{th} entry of the mean of the measurement vector (a deterministic, assumed known function of the unknown parameter vector and the known parameters of the sensors), σ_i^2 is its i^{th} variance (assumed known) ,and $\nabla_{\boldsymbol{\theta}}$ is the gradient operator with respect to the $\boldsymbol{\theta}$.

III. NECESSARY CONDITIONS

In order to achieve full identifability, a given set of measurements should allow estimating a unique parameter vector. Thus, if two parameter vectors are shown to correspond to the same measurement-set, the lack of identifability is proven. Willing to prove this for the case of a pair of TPSs or a single SPS, we incorporate the GC model (eq. (1)) into the TPSs (eq. (2)) and the SPSs (eqs. (3) and (4)) based measurements, yielding[3]:

$$\begin{aligned} z_{T_i}^n = \int_{t_{T_i}^n-\Delta T}^{t_{T_i}^n} R_0 \cdot \exp[-\frac{1}{2(1-\rho^2)} \\ \cdot(\frac{(x_i-(\mu_{0_x}+v_x t))^2}{\sigma_x^2} + \frac{(\mu_{0_y}+v_y t)^2}{\sigma_y^2} \\ +\frac{2\rho(x_i-(\mu_{0_x}+v_x t))(\mu_{0_y}+v_y t)}{\sigma_x\sigma_y})]dt \\ +W_{T_i}^n; i = 1, 2 \end{aligned} \tag{6}$$

[2]For the case of e.g. an intensity field this assumption is theoretically incorrect as the measured values are non-negative. However, it is a valid approximation in scenarios with high signal to noise ratio.

[3]For simplicity, we assume that the sensors are placed on the x-axis.

$$z_{S_1}^n = \int_0^L R_0^b \cdot \exp[-\frac{b}{2(1-\rho^2)} \cdot (\frac{(x-(\mu_{0_x}+v_x t_{S_i}^n))^2}{\sigma_x^2} + \frac{(\mu_{0_y}+v_y t_{S_i}^n)^2}{\sigma_y^2} + \frac{2\rho(x-(\mu_{0_x}+v_x t_{S_i}^n))(\mu_{0_y}+v_y t_{S_i}^n)}{\sigma_x \sigma_y})]dx + W_{S_1}^n \quad (7)$$

where, as before, $\boldsymbol{\theta} = [R_0, \mu_{0_x}, \mu_{0_y}, \sigma_x^2, \sigma_y^2, \rho, v_x, v_y]^T$

For $\boldsymbol{\theta_C} = [R_0, \mu_{0_x}, C\mu_{0_y}, \sigma_x^2, C^2\sigma_y^2, \rho, v_x, Cv_y]^T, \forall C \in R$, the measurements are identical, so there is an inherent ambiguity. Note that the same ambiguity occurs for any given set of SPSs and TPSs that lie on the same line. A visual example of this ambiguity for measurements produced by a single CML is illustrated in Fig. 1.

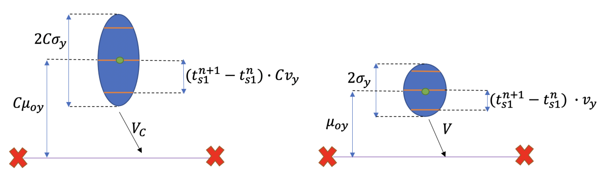

Fig. 1. Two different GCs (blue ellipses) moving through a CML (i.e., - an SPS - magenta lines bounded by **X**). The green dots mark the centers of the GCs, and the orange lines show the GCs' path integration that are to be sampled by the SPSs. These path integration are identical, illustrating the ambiguity of the parameter estimation.

This ambiguity is resolved in scenarios in which measurements from more than one SPS or two TPSs are available as long as they are not co-linear, or if some entries of the parameter vector $\boldsymbol{\theta}$ are considered known. For instance, avoiding this ambiguity can be achieved in the case of GC known to be symmetric, where $\sigma_x^2 = \sigma_y^2 = \sigma^2, \rho = 0$, or if side information about the GC's movement (i.e., its velocity) is provided. In those scenarios, the model is characterized by $\boldsymbol{\theta} = [R_0, \mu_{0_x}, \mu_{0_y}, \sigma^2, v_x, v_y]^T$ and $\boldsymbol{\theta} = [R_0, \mu_{0_x}, \mu_{0_y}, \sigma_x^2, \sigma_y^2, \rho]^T$, respectively. Nonetheless, the identification remains unattainable if only a single TPS is available. In this case, it can be shown, following the same settings of eq. (6), and by relying on the fact that a TPS measuring a moving field is practically equivalent to an SPS lying along its movement direction, that $\boldsymbol{\theta}$ can be represented as $\boldsymbol{\theta_C} = [R_0 \cdot e^{\frac{-C}{2\sigma^2}}, \mu_{0_x}, \pm\sqrt{\mu_{0_y}^2 - C}, \sigma^2, v_x, v_y]^T$ given that the GC is symmetrical, or as $\boldsymbol{\theta_C} = [R_0, \mu_{0_x}, C\mu_{0_y}, \sigma_x^2, C^2\sigma_y^2, \rho]^T$ under the assumption that the velocity of the GC movement is known. The results from this Section are summarized in Table I.

IV. SUFFICIENT CONDITIONS

In cases where the FIM is full rank, but its condition number (CN), defined as the ratio between the largest and the smallest eigenvalue, is very large, the estimation error of at least some of the parameters will be very large. We refer to such a case as "practically unidentifiable". In this section we analyse cases where the necessary conditions for identifiablity are met, but the condition number of the FIM is large, to identify sufficient conditions for practical identifiability. Our methodology is based on setting a scenario for which the problem is fully identifiable (so the FIM is full rank, and its condition number is below a pre-determined threshold) and a study of how the CN changes as a function of the system parameters - in particular: (i) the total observation period (OP) and (ii) the number of samples (N) from each sensor. We found out that for all fully identifiable scenarios, the CN as a function of OP for a given N (or vice versa) has a u-shape characteristic [20] (see Fig. 2). The conditions which guarantee the CN to be below a given threshold are considered as the sufficient conditions for identifiabily. In the sequel we demonstrate this methodology with a real life example.

Table I. Necessary conditions to achieve identifiability according to the number of the available sensors. "V" indicates attainable identification and "X" indicates the opposite.

	1 TPS	1 SPS or 2 TPS	>1 SPS or >2 TPS
No Prior Data	X	X	V
Velocity or symmetric known	X	V	V

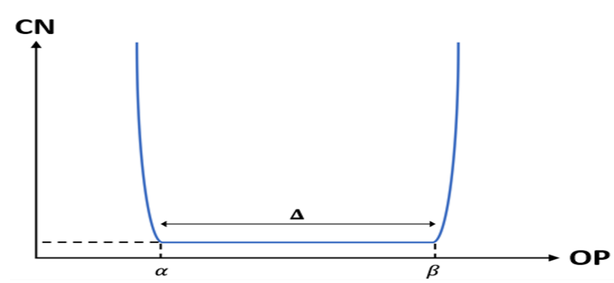

Fig. 2. A schematic U-shape function indicating that for values of OP inside the U, the CN is low and is roughly constant so that one can estimate the model's parameters accurately. If the OP is outside the U shape, the CN is large, indicating taht some of the parameters are practically unidentifable. $[\alpha, \beta]$ is the interval of identifiability, and Δ is its width.

IV-A. The Tested Scenario

Our tested scenario is presented in Fig. 3. The scenario is based on actual CMLs and RGs located in the south of Israel, in an area susceptible to flash-floods. The rain GC is approximated based on an actual rain-event detected in the area by a weather-radar [20]. We consider three RGs (TPS) and two CMLs (SPS) as depicted in Fig. 3. The observation period (OP - defined as the time duration between the first and the last samples that were taken in a given scenario) is two hours, the sampling time is 10 minutes for the RGs and 15 minutes for the CML (which are typical sampling rates for such instrument in this region), and the noise's variance is considered to be equal between all measurements. FIMs for the nominal case in which three RGs, two CMLs, or one CML and one RG are used. In all this cases the necessary conditions are met and the FIMs are full rank, while with any single CML or a pair of RGs the FIMs are singular.

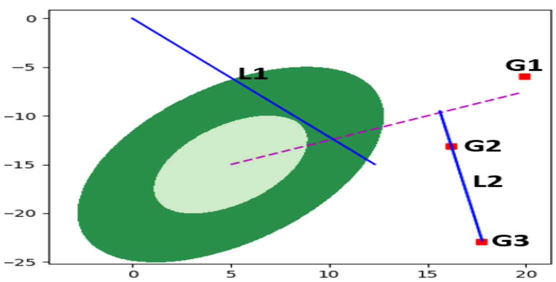

Fig. 3. The initial state of a GC (in green) that moves along the CMLs (blue lines) and the RGs (in red), in the direction of the dashed magenta line. In the end of the observation period (two hours) the GC's center location is at the end of the dashed magenta line.

IV-B. Identifiability as a Function of the System Parameters

While the scenario parameters are given, the system parameters can be designed, based on the user requirements. Such parameters include the number of samples (from each monitoring instrument) N_{T_i}, N_{S_i}, and the OP. For a fixed number of samples the desired OP has upper and lower limits. While extremely short OP will result in a dense samples consisting of similar information regarding the GC, very long OP dictates that some of the measurements are sampled at times in which the GC center is far from the sensors (with respect to the GC width - $\{\sigma_x, \sigma_y\}$), causing the sensors to miss the GC. Therefore, by examining the FIM's CN as a function of the OP, it can be seen that there is a limited section of the OP where the identification is attainable. Fig. 4 presents the interval of the OP that allow identifiability of the scenario of Fig. 3, for given sets of sensors-availability, and for different number of samples (the sampling rate and the number of samples are identical for all sensors). It shows that for all cases where the necessary conditions for identifiability hold, there is a range of system parameters (observation period, *OP*, and the number of samples, N) that provide sufficient condition too. The maximal OP (β in Fig. 2) grows roughly linearly with the number of samples, N, while the minimal value of OP (α in Fig. 2) remains relatively constant. The type, number, and mainly the location of the available sensors relative to the spatio-temporal field directly affect the range of OP that is sufficient for identifiability for a given N. Therefore, no general claim about the superiority of a certain scenario can be made.

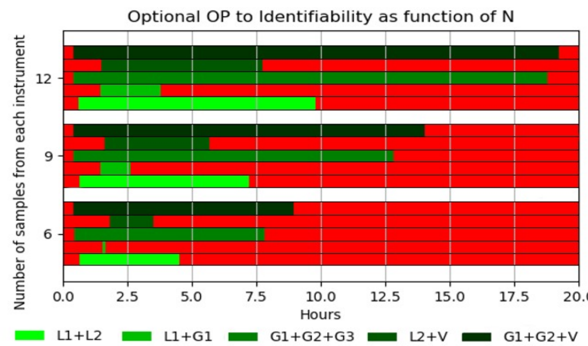

Fig. 4. Optional OP interval as a function of N for different combinations of instruments (as detailed at the bottom). *V* indicates the (assumed) known velocity. The green sections indicate identifiability sections while the red indicates weak identifiability.

V. CONCLUSION

This paper discusses the necessary and sufficient conditions required for the identification of a tempo-spatial field from a limited number of TPSs and SPSs. Based on a parameter model of the field we have identified necessary conditions for identifiablity, while a numerical study of the FIM's has revealed the required sufficient conditions. The necessary conditions are summarized in Table I. In practice, it is required to design the system as such that the system parameters will result in an OP and the number of samples that provide enough measurements with sufficient information. Further research is needed, however, to study the accuracy of the estimation in the case of identification and the sensitivity of the results to miss-modeling, where the actual cell does not fit the assumed parametric model. Lastly, it is worth noting that for the application of flash flood prediction one is interested in the estimation of the accumulated rain in a given area and time. The results of this paper can be used for providing such estimates [20]. However, note that accumulated rain can be estimated even if the necessary conditions of Table I are not met.

VI. ACKNOWLEDGEMENTS

This work was supported in part by the German Research Foundation (DFG) through the project "Integrating Microwave CML Data for Analysis of Precipitation in Complex Terrain: Theoretical Aspects and Hydrometeorological Applications (IMAP). We would also like to thank Shani Gat for many fruitful conversations.

VII. REFERENCES

[1] Hagit Messer, Artem Zinevich, and Pinhas Alpert, “Environmental monitoring by wireless communication networks,” *Science*, vol. 312, no. 5774, pp. 713, may 2006.

[2] A Overeem, H Leijnse, and R Uijlenhoet, “Two and a half years of country-wide rainfall maps using radio links from commercial cellular telecommunication networks,” *Water Resources Research*, vol. 52, no. 10, pp. 8039–8065, 2016.

[3] O. Goldstein, H. Messer, and A. Zinevich, “Rain rate estimation using measurements from commercial telecommunications links,” *IEEE Transactions on Signal Processing*, vol. 57, pp. 1616–1625, 2009.

[4] Adam Eshel, Jonatan Ostrometzky, Shani Gat, Pinhas Alpert, and Hagit Messer, “Spatial reconstruction of rain fields from wireless telecommunication networks–scenario-dependent analysis of idw-based algorithms,” *IEEE Geoscience and Remote Sensing Letters*, 2019.

[5] Hagit Messer, “Multimodality for rainfall measurement,” in *International Conference on Latent Variable Analysis and Signal Separation*. Springer, 2017, pp. 333–343.

[6] Inbar Fijalkow, Elad Heiman, and Hagit Messer, “Parameter estimation from heterogeneous/multimodal data sets,” *IEEE Signal Processing Letters*, vol. 23, no. 3, pp. 390–393, 2016.

[7] Shani Gat and Hagit Messer, “A comparative study of the performance of parameter estimation of a 2-d field using line-and point-projection sensors,” in *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2019, pp. 146–150.

[8] Jonatan Ostrometzky and Hagit Messer, “Statistical signal processing approach for rain estimation based on measurements from network management systems,” in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 9026–9030.

[9] Martin Fencl, Jörg Rieckermann, Petr Sỳkora, David Stránskỳ, and Vojtěch Bareš, “Commercial microwave links instead of rain gauges: fiction or reality?,” *Water Science and Technology*, vol. 71, no. 1, pp. 31–37, 2015.

[10] Venkat Roy, Shahzad Gishkori, and Geert Leus, “Dynamic rainfall monitoring using microwave links,” *EURASIP Journal on Advances in Signal Processing*, vol. 2016, no. 1, pp. 1–17, 2016.

[11] Yoav Liberman, “Object tracking extensions for accurate recovery of rainfall maps using microwave sensor network,” in *2014 22nd European Signal Processing Conference (EUSIPCO)*. IEEE, 2014, pp. 1322–1326.

[12] M. Hadar, J. Ostrometzky, and H. Messer, “Parameter estimation of in-city frontal rainfall propagation,” in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 4910–4914.

[13] Elad Heiman and Hagit Messer, “Parameter estimation from multiple sensors with mixed resolution of quantization,” in *2014 IEEE 28th Convention of Electrical & Electronics Engineers in Israel (IEEEI)*. IEEE, 2014, pp. 1–5.

[14] C Capsoni and M D’Amico, “Morphological description of the rain structures in the padana valley,” in *Proc. ERAD*, 2004, pp. 541–544.

[15] Efrat Morin, David C Goodrich, Robert A Maddox, Xiaogang Gao, Hoshin V Gupta, and Soroosh Sorooshian, “Spatial patterns in thunderstorm rainfall events and their coupling with watershed hydrological response,” *Advances in Water Resources*, vol. 29, no. 6, pp. 843–860, 2006.

[16] Yosef Rockah and P Schultheiss, “Array shape calibration using sources in unknown locations–part I: Farfield sources,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 3, pp. 286–299, 1987.

[17] R. Olsen, D. Rogers, and D. Hodge, “The ar^b relation in the calculation of rain attenuation.,” *IEEE Transactions on Antennas and Propagation*, vol. 26, pp. 318–329, 1978.

[18] ITU-R.838, “Specific attenuation model for rain for use in prediction methods,” *ITU-R*, vol. 838-3, 1992-1999-2003-2005, 2005.

[19] S. M. Kay, “Fundamentals of statistical signal processing,” *PTR Prentice-Hall, Englewood Cliffs, NJ*, 1993.

[20] Aviv Regev, “Parameter estimation approach for accumulated rain estimation from few near-ground sensors,” *M.Sc. Thesis, Tel Aviv University*, 2021.