IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium | 979-8-3503-2010-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/IGARSS52108.2023.10281671

MINIMIZING ESTIMATION ERROR VARIANCE USING A WEIGHTED SUM OF SAMPLES FROM THE SOIL MOISTURE ACTIVE PASSIVE (SMAP) SATELLITE

Mohammad Koosha, Nicholas Mastronarde

University at Buffalo, The State University of New York

ABSTRACT

The National Aeronautics and Space Administration's (NASA) Soil Moisture Active Passive (SMAP) is the latest passive remote sensing satellite operating in the protected L-band spectrum from 1.400 to 1.427 GHz. SMAP provides global-scale soil moisture images with point-wise passive scanning of the earth's thermal radiations. SMAP takes multiple samples in frequency and time from each antenna footprint to increase the likelihood of capturing RFI-free samples. SMAP's current RFI detection and mitigation algorithm excludes samples detected to be RFI-contaminated and averages the remaining samples. But this approach can be less effective for harsh RFI environments, where RFI contamination is present in all or a large number of samples. In this paper, we investigate a bias-free weighted sum of samples estimator, where the weights can be computed based on the RFI's statistical properties.

Index Terms— Passive Remote Sensing, SMAP, Soil Moisture, Radio Frequency Interference (RFI), Quadratic Programming, Mean Square Error (MSE).

1. INTRODUCTION

Passive radiometry is the crucial foundation for passive remote sensing with a wide range of applications from climatology to earth sciences. Empirical field experiments and data from multiple passive remote sensing satellites indicate increasing Radio Frequency Interference (RFI) across the globe, despite international radio regulations for preserving protected RFI-free passive sensing frequency bands [1]. This alarming increase in the RFI level experienced by passive radiometry satellites calls for improvements in RFI mitigation techniques.

Although satellite-borne passive remote sensing has been done in various frequency bands, including the C-band and X-band, the new generation of passive radiometry satellites, starting with the European Space Agency's (ESA) Soil Moisture and Ocean Salinity (SMOS) satellite [2], operate in the protected portion of the L-band from 1.400 to 1.427 GHz. The National Aeronautics and Space Administration's (NASA) Soil Moisture Active Passive (SMAP) is one of the latest in a series of remote sensing satellites that provides high-resolution images of global soil moisture with passive scanning of earth's thermal radiations in this band [3].

SMAP's digital backend takes multiple measurements for each antenna footprint and sends them to Earth for ground-based processing. An RFI detection and mitigation algorithm removes samples detected as being RFI-contaminated from the pool of samples of each antenna footprint, and averages the remaining samples [4]. Since no sample can be precisely declared RFI-free, we investigate a probabilistic approach where each sample has an assigned RFI distribution parameterized by its mean and variance. In contrast to SMAP's approach, we investigate a weighted sum of samples where each sample is assigned a weight. Assuming the means and covariance matrix of the RFI values are known, we model the estimation error variance as a quadratic function and calculate the weight for each sample. We show that this estimator is bias-free and minimizes the estimation error variance.

The paper is organized as follows: Section 2 outlines SMAP's radiometer design and RFI detection algorithms. Section 3 presents our proposed RFI minimization method. Section 4 showcases simulation results. Section 5 concludes and discusses future work.

2. PRELIMINARIES

SMAP captures high-resolution images of earth's soil-moisture content with periodic global coverage every two to three days [5]. SMAP has a 6-meter-wide conically-scanning golden mesh reflector with a 3-dB antenna beam width of 2.4° that projects a footprint of roughly 40×40 km^2 on earth with an incident angle of approximately 40°. An Ortho-Mode Transducer (OMT) feedhorn collects the reflected radiations from the mesh reflector and duplexes them separately into vertical and horizontal polarizations [6] (Fig. 1).

In the next phase, for each polarization, SMAP's Radiometer Digital Electronics (RDE) sample the received signal in a 24 MHz full-band channel centered at 1.413 GHz. The sampled signal then goes through a poly-phase filter bank with 16 sub-bands of 1.5 MHz. The RDE provides 8-time samples of 1.2 ms integration time for each antenna footprint at each subband. As illustrated in Fig. 1, the final products

The work of M. Koosha and N. Mastronarde was supported in part by the NSF under Award #2030157.

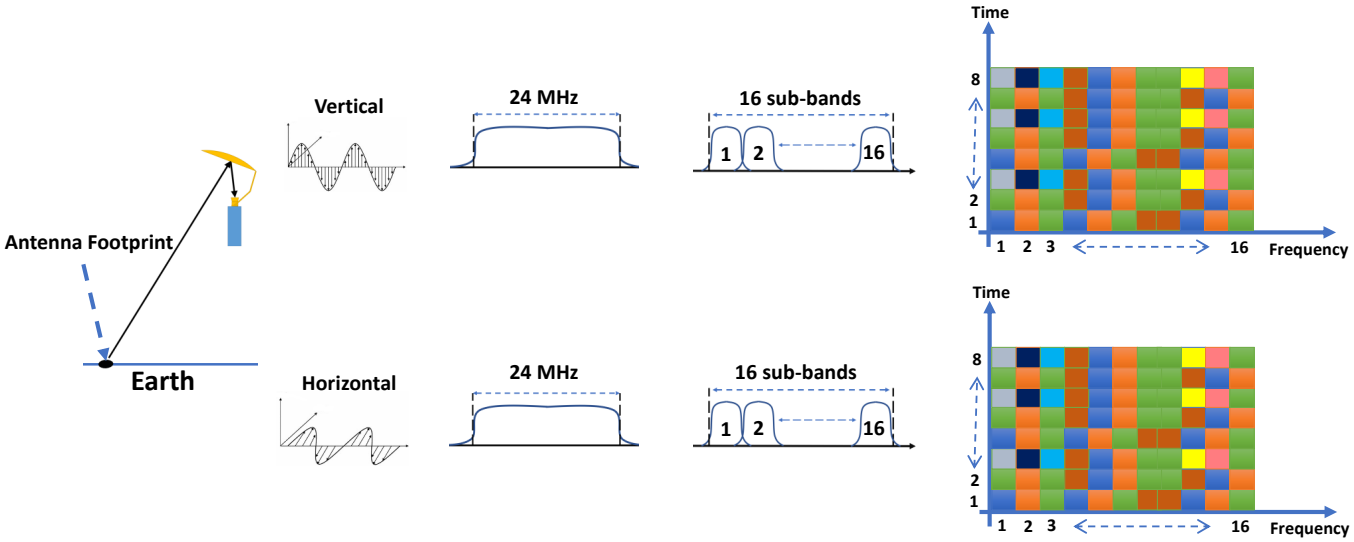

Fig. 1. SMAP's digital back-end

are two 16×8 frequency-time spectrograms for both vertical and horizontal polarizations [6]. This sampling strategy increases the probability that SMAP captures RFI-free samples under sporadic interference in both frequency and time.

SMAP's superiority is its advanced Digital Signal Processing (DSP) backend, which enables ground-based science RFI detection and removal algorithms. Along with frequency-time spectrograms, SMAP provides 2^{nd}, 3^{rd}, and 4^{th} raw moments of both full-band and frequency-time samples of spectrograms, along with the 3^{rd} and 4^{th} stokes parameters of both sub-bands and full-bands. This auxiliary information provides the basis for rendering anomaly detection algorithms on the received data and excluding potentially RFI-contaminated frequency-time samples [6].

Each frequency-time sample has to go through 9 different RFI detection algorithms of pulse, cross-frequency, kurtosis, and polarization-anomaly types. According to [4], most of these 9 RFI detection methods operate by comparing the deviation of a particular measurement T from its expected mean m to a threshold, where the threshold is a multiple β of the expected standard deviation σ of the measurement, with detection declared for measurements satisfying

$$|T - m| \geq \beta\sigma. \tag{1}$$

The aggregate detection techniques can detect narrow-band, wide-band pulse, and continuous wave RFI. Each RFI detection algorithm assigns each frequency-time sample an RFI flag. In a maximum Probability of Detection (PD) approach, each frequency-time sample has to pass all 9 RFI detection algorithms without any "on" RFI flag. In the last phase, the frequency-time samples declared RFI-free are averaged [4].

In recent years, advanced techniques for detecting RFI in SMAP sampling structures have emerged. For instance, Alam et al. [1] proposed a cutting-edge supervised deep learning method utilizing convolutional neural networks. This method effectively identifies RFI-contaminated pixels in SMAP's frequency-time spectrograms, providing a promising alternative to traditional detection approaches. However, existing literature predominantly focuses on RFI detection rather than optimizing the recovery and estimation of soil moisture.

3. PROPOSED METHOD

Most RFI detection techniques used by ground-based science algorithms rigidly throw away samples above specific predefined threshold values, and the remaining samples are crudely averaged. This approach is suboptimal from an estimation theory perspective because (a) the final product is a biased estimate and (b) the error variance of the final estimate is not minimized. These two problems arise because most samples cannot be precisely declared RFI-free, especially those that are marginally below or above the threshold values used by ground-based RFI detection techniques. As noted in [7], RFI appears as a random additive value in each sample, possibly with a Generalized Extreme Value (GEV) probability density function (pdf). Instead of rigidly excluding samples and averaging the remaining samples, we investigate a weighted sum of samples approach where the sample weights are determined according to the RFI's mean and covariance matrix.

As noted in Section 2, SMAP provides a total of $2 \times 16 \times 8 = 256$ samples from each antenna footprint. Each sample p_i can be expressed as the sum of the earth's thermal emissions in the antenna's footprint (p_{soil} Watts) and an additive RFI random variable (P^i_{RFI} Watts) with mean μ_i, i.e.,

$$p_i = P^i_{RFI} + p_{soil} \quad \text{Watts}, \quad i = 1, .., 256. \tag{2}$$

We write the measured samples in vector form as $\mathbf{p} = [p_1, ..., p_i, ..., p_{256}]^T$, the RFI experienced by each sample in vector form as $\mathbf{P}_{RFI} = [P^1_{RFI}, ..., P^i_{RFI}, ..., P^{256}_{RFI}]^T$, and the mean RFI for each sample in vector form as $\boldsymbol{\mu} = [\mu_1, ..., \mu_i, ..., \mu_{256}]^T$. We assign each sample p_i a weight of α_i to make a weighted sum average. We write the weight vector as $\boldsymbol{A} = [\alpha_1, ..., \alpha_i, ..., \alpha_{256}]^T$. We define $\bar{p}$ as an unbiased estimator of p_{soil} using a weighted sum of samples:

$$\bar{p} = \boldsymbol{A}^T(\boldsymbol{p} - \boldsymbol{\mu}) = \sum_{i=1}^{256} \alpha_i(p_i - \mu_i) \tag{3}$$

$$= \sum_{i=1}^{256} \alpha_i(P^i_{RFI} - \mu_i) + \sum_{i=1}^{256} \alpha_i p_{soil}$$

$$= \boldsymbol{A}^T(\boldsymbol{P}_{RFI} - \boldsymbol{\mu}) + p_{soil} \tag{4}$$

$$\text{subject to:} \quad \mathbf{1}^T\boldsymbol{A} = 1, \tag{5}$$

where $\mathbf{1} = [1, ..., 1]^T$ is a 256-element vector of ones. The estimator $\bar{p}$ in (3) is an unbiased estimator since $\mathbb{E}[\bar{p}] = p_{soil}$. To increase the accuracy of estimator $\bar{p}$ in (2), we must minimize the **estimation error** $Err = \bar{p} - p_{soil}$. Since $\mathbb{E}[Err] = 0$, smaller values of the estimation error variance σ^2_{Err} means that error values are concentrated closer to zero. Using (4) we write the **estimation error variance** σ^2_{Err} as follows:

$$\sigma^2_{Err} = \mathbb{E}\big[(\bar{p} - p_{soil})^2\big] \tag{6}$$

$$= \mathbb{E}\big[\boldsymbol{A}^T(\boldsymbol{P}_{RFI} - \boldsymbol{\mu})(\boldsymbol{P}_{RFI} - \boldsymbol{\mu})^T\boldsymbol{A}\big]$$

$$= \boldsymbol{A}^T\mathbb{E}\big[(\boldsymbol{P}_{RFI} - \boldsymbol{\mu})(\boldsymbol{P}_{RFI} - \boldsymbol{\mu})^T\big]\boldsymbol{A}$$

$$= \boldsymbol{A}^T\boldsymbol{\Sigma}\boldsymbol{A} \tag{7}$$

where $\boldsymbol{\Sigma}$ is the covariance matrix of $\boldsymbol{P}_{RFI}$, i.e.,

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2_{1,1} & \sigma^2_{1,2} & \cdots & \sigma^2_{1,256} \\ \sigma^2_{2,1} & \sigma^2_{2,2} & \cdots & \sigma^2_{2,256} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma^2_{256,1} & \cdots & \cdots & \sigma^2_{256,256} \end{bmatrix}.$$

To minimize the estimation error variance, the sample weight vector $\boldsymbol{A}$ must be selected to minimize σ^2_{Err}. In (7), σ^2_{Err} is a quadratic function, and the weights vector $\boldsymbol{A}$ can be acquired using Quadratic Programming (QP) convex optimization [8]. The general problem can be stated as follows:

$$\arg\min_{\boldsymbol{A}} \quad \boldsymbol{A}^T\boldsymbol{\Sigma}\boldsymbol{A} \tag{8}$$

$$\text{subject to:} \quad \mathbf{1}^T\boldsymbol{A} = 1.$$

Equation (8) has a solution for positive definite (PD) matrix $\boldsymbol{\Sigma}$. For the particular case of diagonal $\boldsymbol{\Sigma}$, the problem turns into a Mean Square Error (MSE) problem [9]. However, a diagonal covariance matrix means that RFI contaminating samples are completely independent; thus, off-diagonal elements are all zero. Using Lagrange multiplier λ, we can write the general solution to (8) as (9) [8]:

$$\begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{A} \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}. \tag{9}$$

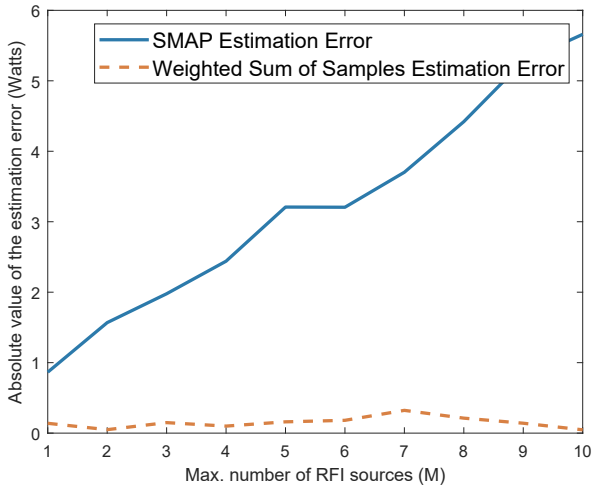

Fig. 2. Comparison of estimation errors.

4. SIMULATION RESULTS

In this section, we examine the performance of our proposed weighted sum of samples method compared to SMAP's RFI mitigation technique in terms of estimation error and estimation error variance.

For illustration, we imagine a maximum of M RFI sources in the environment. Each RFI source has independent sporadic emissions both in time and frequency. Consequently, each SMAP sample p_i can be impacted by $k_i = 1, 2, \ldots, M$ RFI sources. For each sample p_i, k_i is randomly and uniformly chosen between 1 and M. Each RFI source impacting sample p_i has a normally distributed amplitude. Consequently, P^i_{RFI} (power in Watts) is the sum of k_i independent squared normal random variables, i.e., it has a chi-squared distribution $\chi^2(k_i)$ with k_i degrees of freedom. For each value of $M \in \{1, 2, \ldots, 10\}$, we generate 256 independent RFI values such that $P^i_{RFI} \sim \chi^2(k_i)$ for $i = 1, 2, \ldots, 256$ and $k_i \in \{1, 2, \ldots, M\}$. We then use these values to estimate p_{soil} using SMAP's approach and our proposed weighted sum of samples approach.

Our primary requirement on the SMAP side is that we know the sample mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$, but we do not know the exact RFI realizations $P^i_{RFI} = p^i_{RFI}$. With the assumption that $P^i_{RFI} \sim \chi^2(k_i)$, this requirement is equivalent to knowing the number of active RFI sources k_i. Specifically, $\mu_i = \mathbb{E}(P^i_{RFI}|k_i) = k_i$ and $\sigma^2_{i,i} = 2k_i$. Since

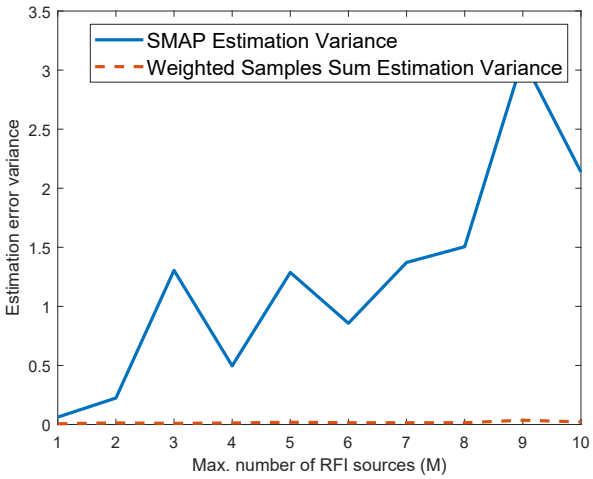

Fig. 3. Estimation error variance.

each RFI source impacts p_i samples independently (equivalently, each k_i is independent), the covariance terms $\sigma^2_{i,j} = 0$ for all $i \neq j$. The mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ can be determined accordingly.

As a simplified RFI detection technique similar to SMAP's, we use the general test of equation (1) for all sample p_i for $i = 1, 2, \ldots, 256$. Similar to SMAP, we average all the samples passing the test. Consequently, the estimator for SMAP is $\bar{p} = \frac{1}{g}\sum_{i=1}^{g} p_i$, where g is the set of samples declared RF-free by the test in equation (1). In (1), we take $\beta = 1$ and m and σ are respectively the arithmetic mean and standard deviation of all 256 p_i samples.

Fig. 2 shows estimation error values for our proposed weighted sum of samples and SMAP's estimator. The horizontal axis represents M, i.e., the maximum number of RFI sources present in the environment. As observable in Fig. 2, our weighted sum of samples method has a much lower estimation error. Moreover, it can preserve low and nearly constant estimation error in harsher RFI environments, while SMAP's estimation error increases approximately linearly in the number of interference sources. Fig. 3 shows estimation error variances for our proposed weighted sum of samples method and SMAP's estimator. Our weighted sum of samples method has a much lower estimation error variance, especially in harsher RFI environments.

5. CONCLUSION & FUTURE WORK

This paper proposes an optimized soil moisture estimator based on SMAP's digital backend design. Instead of excluding RFI-contaminated samples and averaging the remaining ones, we investigate a probabilistic approach, where each sample is assigned a weight value in a weighted sample sum. The weights are determined through quadratic programming, using the mean values and covariance matrix of RFI. Simulations demonstrate this method's ability to make more accurate estimates. Assumptions are made regarding the knowledge of statistical properties such as the mean values and covariance matrix of the RFI distributions contaminating samples. Future work will investigate techniques for determining RFI probability distribution functions of samples, means, and covariance matrices. Additionally, real-world applications will be examined through a testbed incorporating a passive radiometer that is currently being calibrated and tested.

6. REFERENCES

[1] A. M. Alam, M. Kurum, and A. C. Gurbuz, "Radio frequency interference detection for SMAP radiometer using convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 15, pp. 10 099–10 112, 2022.

[2] Y. H. Kerr, P. Waldteufel, J.-P. Wigneron, J. Martinuzzi, J. Font, and M. Berger, "Soil moisture retrieval from space: The soil moisture and ocean salinity (SMOS) mission," *IEEE Trans. Geosci. and Remote Sens.*, vol. 39, no. 8, pp. 1729–1735, 2001.

[3] D. Entekhabi, S. Yueh, and G. De Lannoy, "SMAP handbook," 2014.

[4] A. Bringer, J. T. Johnson, Y. Soldo, D. M. Le Vine, P. Mohammed, S. Misra, P. de Matthaeis, and J. R. Piepmeier, "Properties of the RFI environment at 1400–1427 MHz as observed by the soil moisture active/passive mission microwave radiometer," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 14, pp. 7259–7267, 2021.

[5] M. Koosha and N. Mastronarde, "Opportunistic temporal spectrum coexistence of passive radiometry and active wireless networks," in *2022 IEEE Western New York Image and Signal Processing Workshop (WNYISPW)*. IEEE, 2022, pp. 1–4.

[6] J. Piepmeier, P. Mohammed, G. De Amici, E. Kim, J. Peng, C. Ruf, M. Hanna, S. Yueh, and D. Entekhabi, "Soil moisture active passive (SMAP) project algorithm theoretical basis document SMAP L1B radiometer data product: L1B_TB," Tech. Rep., 2016.

[7] P. N. Mohammed and J. R. Piepmeier, "Microwave radiometer RFI detection using deep learning," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 14, pp. 6398–6405, 2021.

[8] R. Fletcher, "A general quadratic programming algorithm," *IMA Journal of Applied Mathematics*, vol. 7, no. 1, pp. 76–91, 1971.

[9] M. Rousset, Y. Xu, and P.-A. Zitt, "Reducing variance by reweighting samples," 2019.