Maximum Likelihood Sensor Node Localization using Received Signal Strength in Multi-media with Multi-path Characteristics

Herman Sahota and Ratnesh Kumar, Fellow, IEEE

Abstract—Sensors are key to situation-awareness and response, and need to maintain time and position information to tag their measurement-data. While local clocks can be used for time-stamping, geo-tagging can be challenging for sensors with no access to GPS, such as the underground environment in precision agriculture. We study the problem of sensor node localization for a hybrid wireless sensor network for precision agriculture, with satellite nodes located above ground and sensor nodes located underground. This application is quite unique in possessing multi-media and multi-path features. We use received signal strength of signals transmitted between neighboring sensor nodes and between satellite nodes and sensor nodes as a means to perform the ranging measurement. The localization problem is formulated as that of estimating the parameters of the joint distribution of the received signal strength at all nodes in the network. First, we arrive at path loss and fading models for various multi-media and multi-path communication scenarios in our network to model the received signal strength in terms of the propagation distance and hence, the participating nodes’ location coordinates. We account for various signal degradation effects such as fading, reflection, transmission, and interference between two signals arriving along different paths. Then, we formulate a maximum likelihood optimization problem to estimate the nodes’ location coordinates using the derived statistical model. We also present a sensitivity analysis of the estimates with respect to soil permittivity and magnetic permeability.

I. INTRODUCTION AND RELATED WORK

Precision agriculture refers to the use of information and control technologies in agriculture. Agricultural inputs such as irrigation, fertilizers, pesticides, etc. are applied in precise quantities as determined by modeling of crop growth patterns to maximize the crop yield and to minimize the impact on the environment. Fertilizer uptake within a field depends on factors such as variability in plant population, nitrogen mineralization from organic matter, water stress, soil properties, pests, etc. These factors vary in space and time. Fertilizers must be applied according to the needs of the crop. If under-applied, the crop yield suffers while if over-applied, contamination of ground and surface water resources may take place leading to issues such as aquatic hypoxia. Furthermore, there is a considerable energy cost associated with the production of nitrogen based fertilizers. Also, understanding the process of carbon sequestration, which is thought of as a method to control global warming, could be aided by the data collection from soil sensing, and its further analysis.

Research in our group has led to the development of low-cost and portable network analyzer capable of in-situ soil dielectric spectroscopy for measuring soil conductance and capacitance over a spectrum of frequencies ranging from 1MHz to 40MHz with 90% accuracy [1]; see Figure 2. The sensor is equipped with a small meta-materials based antenna for wireless interfacing, which also doubles as the sensing element for dielectric measurements. Data collected from such sensors can then be used for making resource allocation decisions, namely, irrigation and fertilization. Research in our group

The research was supported in part by the National Science Foundation under the grants NSF-ECCS-0926029, NSF-CCF-1331390, and NSF-ECCS-1509420.
has also developed sensor networking protocols for energy-efficient operation (synchronization, scheduling and routing) enabled by multiple power modes for both transmission and reception. These schemes and the resulting energy gains are reported in [2].

Determining location information, i.e., the \((x, y, z)\)-positioning of sensors is valuable in geo-tagging sensor data, and also for performing sensor management including synchronization, scheduling, routing and battery and fault management. The focus of the current paper is to develop a sound mathematically-grounded sensor localization strategy for a hybrid network of sensors deployed in multiple media. Future work will then take such geo-tagged and time-stamped sensor data to combine with agro-geo-ecological models to predict crop demands and yields and enable the agricultural inputs prescription recommendations.

In [3], authors use canonical duality theory for solving sensor network localization originally formulated as a non-convex quadratic minimization problem. The dual problem turns out to be concave maximization over a convex set in symmetrical matrix space. The authors also suggest ideas for solving the general NP-hard problems in case the canonical dual problem has no solution in its positive definite domain. A two-objective evolutionary algorithm is proposed in [4] which takes into account localization accuracy as well as the topological constraints induced by connectivity considerations. [5] proposes a self localization system that selects and maintains a relative coordinate system for a sensor network with non-stationary nodes. The relative coordinate system is defined by a set of stable nodes that are less likely to disappear/fail. The method of selection of this group of nodes is also presented. Multiple local coordinate systems are stitched together by performing coordinate translation to arrive at a network wide coordinate system. [6] suggests removing the anchor nodes the signal from which has high measurement error as well as those anchors that are not expected to increase the estimate’s accuracy due to their positions. This can help reducing the computation and communication load on the energy constrained sensor nodes. [7] presents the implementation of two methods - Gauss Newton and Particle Swarm Optimization - for solving the optimization problem for localization in micro controller. The method uses both deployment information (obtained using a pedometer attached to a human deployment agent) and communication range to arrive at location estimates. Effectiveness of the two approaches under different operating conditions is also studied. In [8], authors study the challenges imposed by emerging applications (Unmanned vehicle with WSN navigation and cyber transportation systems) on localization. [9] presents a comparative study of different variants of Particle Swarm Optimization methods used in sensor network localization under different node topologies. In [10], authors present the development of two approaches for localization for a wireless sensor network deployed in an indoor environment: a model learning based approach and triangulation. The system is also tested using a real hardware implementation. [11] tackles the problem at the level of the physical signal design. The authors present the development of a signaling scheme incorporating both data transmission and localization. The effect of bandwidth reduction and decoding errors, due to simultaneous data transmission, on location estimates is studied.

Recent publications reveal several examples motivating the need for self localization as a key feature to enhance the functionality of different layers of the sensor network system. For example, location based routing techniques proposed for wireless sensor networks require location data as an integral input, as seen in [12], [13]. Also, in our precision agriculture application, the routing strategy for the wireless sensor network, as seen in [14], [15] and [2], chooses the next hop node from neighboring nodes based on their geographical proximity to the sink. Several security schemes also rely on location awareness of the sensor nodes. For example, in [16], authors propose an end-to-end data security mechanism which involves binding the secret keys stored at a node to its location.

In [17], authors study sensor node localization from a statistical signal processing standpoint. They use statistical models to describe the three common measurement technologies used for ranging in sensor node localization: time of arrival (ToA), angle of arrival (AoA) and received signal strength (RSS). The models are used to derive Cramer-Rao lower bounds on the location estimates’ variances. In [18], authors address node localization problems from a theoretical standpoint. They derive the conditions for unique localizability of a network and also study the computational complexity of network localization.

Localization approaches can be broadly categorized as anchor-free or anchor-based. Anchor-free approaches do not rely on any node in the network having prior location information. Several such localization schemes have been proposed in literature where nodes estimate their locations in a relative coordinate system formed by a few chosen nodes in the system. In [19], authors formulate localization as several local optimization problems each of which is solved at a cluster head node for building the local coordinate system of nodes in the cluster. Nodes estimate distances to their neighbors using time of arrival measurements. During the second stage, cluster heads collaborate to build a global map of the network by combining the estimated local maps.

On the other hand, anchor-based node localization relies on a number of anchor nodes in the network that are assumed to have prior location information. In [20], authors present an anchor-based range-free localization scheme for a sensor network in an environment with obstacles. Range-free localization techniques do not use ranging measurements such as time of arrival, signal strength and angle of arrival. Instead, alternative data such as hop count can be used for approximating the distance between two nodes once the hop length is estimated using anchor nodes. As another example, centroid based schemes can be used to estimate location coordinates given a set of anchor nodes in proximity. In [20], authors argue that for a sensor network deployment in an environment with obstacles leading to a concave network shape, the shortest packet delivery path distance may not be approximated by the geographical Euclidean distance. They
propose a scheme to identify line of sight anchors from non-line of sight anchors in order to ignore the hop count measurements from the latter set of anchors. In [21], authors propose an anchor-based scheme based on multidimensional scaling (MDS) [22] that uses connectivity information of the neighboring nodes to localize a network of $n$ nodes with a computational complexity of $O(n^3)$. A distributed variant is also presented that builds local maps of nodes and then patches them together to form a global map. The distributed approach is more robust for irregular networks because connectivity information for far away nodes is ignored taking advantage of the fact that smaller local maps are more accurate. Least squares optimization is incorporated to further improve the accuracy at the expense of additional computation.

Range-based localization techniques require special hardware at each node to measure a ranging signal and estimate the distance to the transmitting node. The advantage is increased localization accuracy compared to range-free localization. An example is a controlled event driven localization system where detectable events are generated using an event disseminator. Nodes report the detection of the events along with the timestamps to a central location server which computes the position of the nodes based on its own knowledge of the event propagation delays in the sensor field. In [23], authors present the implementation results of such a system with light as the detected event. The detection time at a sensor node is used to arrive at a spatial relationship between the node and the event generator. In [24], authors present the design of a localization system using uncontrolled events, which is more suitable for an outdoor environment where controlled dissemination of events may not be readily feasible. In this method, a small number of anchor nodes in the field are first used to estimate the generation parameters of the uncontrolled events. Thereafter, controlled event localization techniques are used to estimate the nodes’ unknown location coordinates.

In [25], authors present an application of cooperative localization techniques to UWB (ultra wide bandwidth) wireless networks. The authors quantify the performance of several algorithms based on the ranging models available for UWB technology. They also present a localization algorithm by mapping a statistical model for graphical inference onto the network topology. In [26], authors present a sequential Monte Carlo localization method for sensor networks with mobile nodes that uses both range measurements and hop distance as well as mobility information of the nodes.

Applications exist with varying degrees of localization accuracy needed. A trade-off is made among the localization hardware complexity, algorithm complexity, deployment costs and localization accuracy. In [27], authors present a study of the error inducing parameters in sensor node localization. The authors derive the Cramer-Rao bound for Gaussian measurement error in multi-hop localization systems using distance and angular measurements. They study the effect of the parameters such as measurement technology accuracy, node density, beacon node concentration, and beacon uncertainty on the localization error.

To the best of our knowledge, none of the existing work in literature has addressed sensor node localization for a network where the nodes are placed in different physical media. In this article, we present our approach to sensor node localization for such a network in which the sensor nodes are buried underground while the anchor nodes (a.k.a satellite nodes) are located above ground.

The following are the contributions of this article:

- Statistical 3-D localization framework based on received signal strength (RSS) measurements for multiple media (air and soil) and multiple reflected paths in a lossy medium (soil).
- Rician fading models for multi-media (air to soil) and multi-path (soil to soil) communication.
- Simulation results that implement the proposed localization scheme in Python, including error analyses.
- Sensitivity analyses of the estimated location coordinates with respect to various parameters of interest.

Section II presents our localization framework. Section V presents the simulation results, including sensitivity analysis in Section V-B. Section VI presents our concluding remarks. Section VII presents the future directions for this work.

II. RSS-BASED LOCALIZATION MODELING

A wireless signal experiences fading due to a combination of large-scale and small-scale effects [28]. The former is due to large-scale uncertainties such as obstacles, weather etc., whereas the latter is due to small-scale uncertainties resulting in multi-path interference. Large-scale power fading is known to cause received power to be lognormally distributed; whereas small-scale power faded received signal power follows a non-central $\chi^2$ distribution with 2 degrees of freedom (equivalently, the amplitude has a Rician distribution).

Accordingly, the signal power, $R_{mn}$, received at node $n$ and transmitted from node $m$, has the following combined distribution:

$$f_{R_{mn}}(r_{mn}) = \int_0^\infty f_{R_{mn}}^S(r_{mn}; p_{mn}, \kappa_{mn}) f_{R_{mn}}^L(p_{mn}; \mu_{mn}, \sigma_{mn}) dp_{mn},$$  \hspace{1cm} (1)

where $f_{R_{mn}}^S(r_{mn}; p_{mn}, \kappa_{mn})$ is the non-central $\chi^2$ pdf modeling the effect of small scale fading with mean received power $p_{mn}$ from the specular (line-of-sight) and scatter (non-line-of-sight) components, and $\kappa_{mn}$ is the specular power to scatter power ratio; and $f_{R_{mn}}^L(p_{mn}; \mu_{mn}, \sigma_{mn})$ is the lognormal pdf modeling the effect of large scale fading with parameters $\mu_{mn}$ and $\sigma_{mn}$. The distributions are expressed as follows, where the notation $I_0(\cdot)$ denotes the modified Bessel function of the first kind and order 0:

$$f_{R_{mn}}^S(r_{mn}; p_{mn}, \kappa_{mn}) = \frac{1 + \kappa_{mn}}{p_{mn}} e^{-\left(\frac{\kappa_{mn} + 1}{p_{mn}}\right)r_{mn} + \kappa_{mn}} I_0 \left(\sqrt{\frac{4(\kappa_{mn} + 1)\kappa_{mn}r_{mn}}{p_{mn}}}\right)$$  \hspace{1cm} (2)

$$f_{R_{mn}}^L(p_{mn}; \mu_{mn}, \sigma_{mn}) = \frac{1}{p_{mn}\sqrt{2\pi\sigma_{mn}^2}} e^{-\frac{(\ln p_{mn} - \mu_{mn})^2}{2\sigma_{mn}^2}}.$$  \hspace{1cm} (3)
Lognormal fading occurs over longer time-scales causing slower variations in the signal while multi-path fading occurs over much shorter time scales [28]. In a short time frame when the nodes communicate to gather the received signal strength data, the large-scale fading effect can be ignored by treating its distribution to have zero-variance ($\sigma_{mn}=0$); so the distribution $f_{R_{mn}}(p_{mn};\mu_{mn},0)$ becomes an impulse centered at the mean value, $p_{mn}=e^{\mu_{mn}}$. In this case, the received power's pdf simplifies to:

$$f_{R_{mn}}(r_{mn}) = f_{R_{mn}}^{S}(r_{mn};p_{mn},\kappa_{mn}), \quad (4)$$

which is the fading model that we rely on for the short time periods when the localization measurements are taken.

In general, the received signal has in-phase as well as quadrature-phase envelopes $X_{in}$ and $X_{quad}$ both of which are Gaussian random variables. This follows from the application of the Central Limit Theorem, as the received signal is the superposition of a large number of scattered components traveling along a multitude of paths [29]. Thus, the overall signal envelope $X = X_{in} + jX_{quad}$ is complex gaussian distributed. If the signal travels over a direct and a reflected path, the received signal envelope, $X + X^{(r)}$, is also complex gaussian distributed; where $X^{(r)} = X_{in}^{(r)} + jX_{quad}^{(r)}$ corresponds to the reflected path. Then, the overall average power, $p_{mn}$, of the received signal is given by Equation (3), where (3) $\rightarrow$ (4) follows from assuming that $X$ and $X^{(r)}$ are independent; $V_{X}e^{j\phi_X}$ and $V_{X^{(r)}}e^{j\phi_{X^{(r)}}}$ represent the specular (line-of-sight) amplitudes of $X$ and $X^{(r)}$, respectively, in polar form. (4) follows from the following: the specular component power, $V_{X_{in}}$, is related to the total average power $P_{RX}$ by the Rician factor ($\kappa$) as: $V_{X_{in}}^{2} = \frac{1}{\kappa + 1}P_{RX}$. Similarly, $V_{X^{(r)}}^{2} = \frac{\kappa^{(r)}}{\kappa_{m} + 1}P_{X^{(r)}}$.

$$\mathbb{E}[(X + X^{(r)})^{2}] = \mathbb{E}[(X + X^{(r)})(X + X^{(r)})^{*}]$$
$$= \mathbb{E}[XX^{*}] + \mathbb{E}[X^{(r)}X^{(r)*}] + \mathbb{E}[XX^{(r)*}]$$
$$= P_{X} + 2P_{X^{(r)}} + \mathbb{E}[X]\mathbb{E}[X^{(r)*}] + \mathbb{E}[X^{(r)*}]\mathbb{E}[X]$$
$$= P_{X} + P_{X^{(r)}} + V_{X}V_{X^{(r)}}e^{j\phi_{X}^{(r)} - j\phi_{X}}$$
$$= P_{X} + P_{X^{(r)}} + 2V_{X}V_{X^{(r)}}e^{j\phi_{X}^{(r)} - j\phi_{X}}$$
$$= P_{X} + P_{X^{(r)}} + 2V_{X}V_{X^{(r)}}$$
$$= P_{X} + P_{X^{(r)}} + 2\sqrt{P_{X}P_{X^{(r)}}}$$
$$= P_{X} + P_{X^{(r)}} + 2\sqrt{P_{X}P_{X^{(r)}}}$$

Note that while the received signal envelope $X + X^{(r)}$ is complex Gaussian, its amplitude has a Rician distribution and its power has a non-central $\chi^{2}$ distribution with 2 degrees of freedom. In the special case that there is no line of sight path (so the received signal has no “specular” component and only the “scatter” component), the amplitude becomes Rayleigh distributed (as a special case of Rician) and the power becomes exponentially distributed (as a special case of non-central $\chi^{2}$), given as follows:

$$f_{R_{mn}}(r_{mn}) = f_{R_{mn}}^{S}(r_{mn};p_{mn},0) = \frac{1}{p_{mn}}e^{-p_{mn}r_{mn}}. \quad (9)$$

III. EXPRESSING RSS PARAMETERS USING LOCATIONS

In this section we express the parameters of the probability distribution of the received signal power in terms of the location coordinates of the sensor nodes in order to formulate the estimation problem for the location coordinates. The received power pdf given by Equation (4) has two parameters; the average received power $p_{mn}$ and the Rician K-factor $\kappa_{mn}$. Existing works study the variation of the Rician K-factor with distance [30] and its estimation [31]. However, we assume it to be a known constant for the purposes of localization for our application. Thus, we express $p_{mn}$ in terms of the location coordinates of the sender and receiver nodes.

It is known that the mean signal power $p_{mn} = e^{\mu_{mn}}$ decays with distance along a single path within a single lossless medium following the power law [32]:

$$p_{mn}(d) = \eta d^{-k} \quad (10)$$

where $d$ is the distance between the two nodes $m$ and $n$, $\eta$ is a constant and $k$ is the path loss exponent, both of which are functions of the medium. For example, in free space:

$$k = 2 \quad \text{and} \quad \eta = \frac{p_{m}G_{m}G_{n}\lambda^{2}}{(4\pi)^{2}}, \quad (11)$$

where $p_{m}$ is the sender power, $G_{m}$ and $G_{n}$ are the sender and receiver antenna gains, and $\lambda$ is the wavelength. For simulation purposes, we use $\eta = p_{m}\lambda^{2}/(4\pi)^{2}$ (ie. $G_{m} = G_{n} = 1$ and $k = 2$). A lossy dielectric medium such as soil has additional attenuation due to conductivity losses [32]. The wave propagation equation for such a medium is given by:

$$E(r,t) = E_{0}e^{-\alpha r}cos(\omega t - \beta r), \quad (12)$$

where $E$ is the electric field at time $t$ at a distance $r$ from the source which transmits at amplitude $E_{0}$. The complex propagation constant of such a lossy dielectric medium is $\alpha + j\beta$, where:

$$\alpha = \omega \sqrt{\frac{\mu'\epsilon'}{2}} \left[ \sqrt{1 + \left( \frac{\epsilon'}{\epsilon'} \right)^{2}} - 1 \right], \quad (13)$$

and

$$\beta = \omega \sqrt{\frac{\mu'\epsilon'}{2}} \left[ \sqrt{1 + \left( \frac{\epsilon'}{\epsilon'} \right)^{2}} + 1 \right], \quad (14)$$

where $\mu$ is the permeability, and $\epsilon'$ and $\epsilon''$ are the real and imaginary parts of the complex permittivity of the medium.

Hence, in the more general setting where $\alpha \neq 0$, Equation [10] takes the form:

$$p_{mn}(d) = \eta d^{-k}e^{-2\alpha d}. \quad (15)$$
In the following discussion, we extend this mathematical notation to the case of multiple paths as well as multiple media. In the notation that follows, we introduce the superscripts \( \alpha \) and \( \sigma \), respectively, when referring to air and soil as the medium.

1) Multi-path extension: soil to soil communication: When the sender node \( m \) as well as the receiver node \( n \) are within soil as shown in Figure [3] there exist the direct and the reflected paths and consequently, following Equation (8), the average received power is given by:

\[
p_{mn} = p_{mn}(d_{mn}) + p_{mn}(d_{mn}^{(r)}) + p_{mn}(d_{mn}^{(s)}) + p_{mn}(d_{mn}^{(t)})
\]

where

\[
\rho = \frac{\sqrt{\mu_a/\mu_s}}{\sqrt{\mu_a/\mu_s} + \sqrt{\mu_s/\mu_s}}
\]

is the reflection coefficient, and \( d_{mn} \) and \( d_{mn}^{(r)} \) are the distances of the direct and reflected paths, respectively, which are given by (see Figure [3]):

\[
d_{mn} = (x_m - x_n)^2 + (y_m - y_n)^2 + (z_m - z_n)^2, \quad \text{and} \quad d_{mn}^{(r)} = \sqrt{(x_m - x_n)^2 + (y_m - y_n)^2 + (z_m - z_n)^2 + z_{mn}^2}
\]

Equation (16), while the total average received scatter power is simply the sum of the scatter powers for the two paths:

\[
p_{mn}(d_{mn}) + p_{mn}(d_{mn}^{(r)})
\]

Since

\[
\frac{\kappa_{mn}}{\kappa_{mn} + 1} + \frac{p_{mn}(d_{mn}^{(s)})}{p_{mn}(d_{mn}^{(s)}) + 1}.
\]

Next, we derive the average received power for signal propagation from a satellite node in air to a sensor node buried in soil.

2) Multi-media extension: air to soil communication: When the sender node \( m \) is in air while the receiver node \( n \) is in soil, the average received power is given by:

\[
p_{mn} = \eta(d_{mn}^{(a)}) - \kappa_{mn}(d_{mn}^{(s)}) - k_{mn}^{(a)} - 2\alpha_{mn}^{(s)}
\]

where \( \tau = 1 - \rho \) is the transmission coefficient (\( \rho \) is given by Equation (17)), and \( d_{mn}^{(a)} \) and \( d_{mn}^{(s)} \) are, respectively, the distances traveled in the air and soil, which are computed as the solutions of the following two equations:

\[
d_{mn}^{(a)} = \sqrt{(x_m - x_n)^2 + (y_m - y_n)^2 + \frac{z_m}{\mu_a}}
\]

\[
d_{mn}^{(s)} = \sqrt{(x_m - x_n)^2 + (y_m - y_n)^2 + \frac{z_n}{\mu_s}}
\]

Also, the parameter \( \kappa_{mn} \), measuring the ratio of the specular to scatter power in the received signal can be derived by noting that the total average received power is \( p_{mn} \) as calculated in Equation (20), while the total average received scatter power is simply the sum of the scatter powers for the two paths:

\[
\kappa_{mn} = \frac{p_{mn}(d_{mn}^{(s)})}{p_{mn}(d_{mn}^{(s)}) + 1}.
\]
Thus, we have expressed the parameters of the distribution of the received signal strength for both signal propagation scenarios in our application in terms of the distance and hence, the coordinates of the respective sender and receiver nodes.

IV. MLE-BASED LOCALIZATION FROM RSS MEASUREMENTS

Now, we formulate a maximum likelihood problem to estimate the location coordinates of sensor nodes given the measurements of the received signal powers at the nodes from their neighboring underground as well as above-ground satellite nodes.

Let \( N \) denote the set of node pairs \((m, n)\) that communicate to gather the received signal power measurements. The log-likelihood of the measured received signal strength for all the sender-receiver pairs \((m, n) \in N\) can be expressed using Equation (2) as follows:

\[
L(\Theta | R) = - \sum_{(m,n) \in N} \left\{ \ln(p_{mn}) - \ln(1 + \kappa_{mn}) + \left( \frac{r_{mn}^{\kappa_{mn} + 1}}{p_{mn}^{\kappa_{mn}}} + \kappa_{mn} \right) - \ln \left[ I_0 \left( \frac{4(\kappa_{mn} + 1)\kappa_{mn} r_{mn}}{\kappa_{mn}} \right) \right] \right\}
\]

(26)

where \( R \) is the set of received signal strength measurements \( \{r_{mn} | (m, n) \in N\} \) and \( \Theta \) is the set of unknown parameters to be estimated (i.e., the location coordinates). Also, the received signal strength \( r_{mn} \) is measured and hence known, whereas the set of unknown parameters consists of \( p_{mn} \) and \( \kappa_{mn} \) which in turn are functions of the unknown location coordinates that must be estimated.

The average powers \( p_{mn} \)'s are expressed in terms of the location coordinates of the nodes \( m \) and \( n \) following Equations (16) \textendash (19) and (21) \textendash (25), while the Rician factors \( \kappa_{mn} \)'s are expected to be constants, independent of the locations, over the duration of the round and can be estimated at the beginning of each round. An MLE approach will proceed by writing the parameters \( p_{mn} \) and \( \kappa_{mn} \) appearing in the likelihood function of Equation (26) as the corresponding functions of the location coordinates of the sensor and satellite nodes as derived in Section III and then optimizing the likelihood function with respect to the location variables.

Method of Moments based Approximation: Alternatively, one can use the approach based on the Method of Moments that avoids solving the optimization problem of maximum likelihood for non-central \( \chi^2 \) distribution, and instead reduces the parameter estimation problem to solving a set of algebraic equations in which the first few moments of the distribution at hand are simply equated to their respective sample-based values.

### Table I: Real and imaginary parts of relative permittivity of soil for different moisture contents.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transmit power (Satellite node)</td>
<td>( P^{(s)} )</td>
<td>30 dBm</td>
</tr>
<tr>
<td>Transmit power (Sensor node)</td>
<td>( P^{(s)} )</td>
<td>30 dBm</td>
</tr>
<tr>
<td>Thermal noise</td>
<td>( N_0 )</td>
<td>-110 dBm</td>
</tr>
<tr>
<td>Path loss factor (air)</td>
<td>( k_a )</td>
<td>2</td>
</tr>
<tr>
<td>Path loss factor (soil)</td>
<td>( k_s )</td>
<td>2</td>
</tr>
<tr>
<td>Relative permeability (soil)</td>
<td>( \mu_s )</td>
<td>1.0084</td>
</tr>
<tr>
<td>Frequency</td>
<td>( f )</td>
<td>433 MHz</td>
</tr>
<tr>
<td>Wavelength</td>
<td>( \lambda )</td>
<td>0.7 m</td>
</tr>
<tr>
<td>Number of readings per node pair (RSS)</td>
<td>( N_{RSS} )</td>
<td>100</td>
</tr>
<tr>
<td>Number of readings per node pair (ToA)</td>
<td>( N_{ToA} )</td>
<td>100</td>
</tr>
<tr>
<td>Signal duration</td>
<td>( T_s )</td>
<td>1ms</td>
</tr>
</tbody>
</table>

### Table II: Parameters used in localization simulations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transmit power (Satellite node)</td>
<td>( P^{(s)} )</td>
<td>30 dBm</td>
</tr>
<tr>
<td>Transmit power (Sensor node)</td>
<td>( P^{(s)} )</td>
<td>30 dBm</td>
</tr>
<tr>
<td>Thermal noise</td>
<td>( N_0 )</td>
<td>-110 dBm</td>
</tr>
<tr>
<td>Path loss factor (air)</td>
<td>( k_a )</td>
<td>2</td>
</tr>
<tr>
<td>Path loss factor (soil)</td>
<td>( k_s )</td>
<td>2</td>
</tr>
<tr>
<td>Relative permeability (soil)</td>
<td>( \mu_s )</td>
<td>1.0084</td>
</tr>
<tr>
<td>Frequency</td>
<td>( f )</td>
<td>433 MHz</td>
</tr>
<tr>
<td>Wavelength</td>
<td>( \lambda )</td>
<td>0.7 m</td>
</tr>
<tr>
<td>Number of readings per node pair (RSS)</td>
<td>( N_{RSS} )</td>
<td>100</td>
</tr>
<tr>
<td>Number of readings per node pair (ToA)</td>
<td>( N_{ToA} )</td>
<td>100</td>
</tr>
<tr>
<td>Signal duration</td>
<td>( T_s )</td>
<td>1ms</td>
</tr>
</tbody>
</table>

In our setting, following the properties of non-central \( \chi^2 \) distribution, the first and second moments of \( R_{mn} \) are given by [34]:

\[
\mathbb{E}[R_{mn}] = p_{mn},
\]

\[
\mathbb{E}[R_{mn}^2] = p_{mn}^2 \left[ 1 + \frac{1}{(1 + \kappa_{mn})^2} \right].
\]

These are equated to the corresponding sample-based estimates of the moments, and then we can algebraically solve the equations to obtain the estimates of the unknown parameters \( p_{mn} \)'s and \( \kappa_{mn} \)'s, and consequently the location coordinates. The demonstration of method-of-moments based approach is a direction for future research as noted in the Conclusion section.

V. SIMULATION RESULTS

We simulate a network of 25 sensor nodes in a square field of size 150 m \( \times \) 150 m to validate and evaluate the performance of our localization framework. The sensor nodes are randomly deployed within the square field at a depth between 3 \( \text{cm} \) to 5 cm. Our localization model discussed in Section II depends on the soil permittivities which in turn are functions of the moisture content of the soil. We assume a clay loam soil with a clay content of 20%. For this soil type, the relative real and imaginary values of the permittivity are computed [35] for various values of volumetric water contents as given in Table I. Other parameters used in the simulations are summarized in Table II. The relative magnetic permeability of soil is close to 1 for soils not containing significant iron [36]. Thermal noise is assumed to be \(-110\) dBm based on standard receiver sensitivity, which is also the case for our receivers [37]. Thus, applying Equations (11) \textendash (15), (16) \textendash (21), with the parameters of Tables I and II, the underground transmission range is approximately 34 m for dry soil at a transmission power level of 30 dBm leading to an average of 3 neighbors per sensor node, whereas the air-to-soil transmission range is approximately 1000 m.
We use a hierarchical and iterative approach for estimating the location coordinates of the sensor nodes, applying first the air-to-soil models on the signals transmitted between the satellite nodes and sensor nodes to arrive at a first set of estimates, followed by the combination of air-to-soil and soil-to-soil models, adding signal data among neighboring sensor nodes to arrive at more refined and accurate estimates of the coordinates. The second stage is applied iteratively with each new iteration initialized with estimates obtained from the preceding iteration, until the estimates converge within a tolerance. The convergence tolerance for $X$ and $Y$ estimates is chosen to be 5 cm, while for $Z$ estimate it is chosen to be 1 cm.

Figure 5 presents the localization estimates obtained using the received signal strength model for a Rician K-factor $(\kappa)$ of 20. We have also verified the localization framework for other values of $\kappa$ (40, 60, 80, 100). Each simulation for estimating the nodes’ coordinates for the network of chosen size took about 30 minutes on a laptop computer with a 2.4 GHz dual-core processor and 8 GB 1067 MHz DDR3 DRAM. A visual inspection of the results shows that our method successfully estimates the location coordinates of the sensor nodes, with the error margins as described below.

![Localization estimates](image)

**Fig. 5: Localization using received signal strength model for Rician factor $\kappa = 20$.**

We use a minimal setup of our network to perform variance analysis of our location estimates: three sensor nodes deployed randomly within the field but within the radio communication ranges of each other and four satellite nodes so that we can closely capture the real world deployment scenario where each sensor node has 2 – 3 neighboring sensor nodes. Then, we compute the sample standard deviation of the location estimate of one of the sensor nodes. Figure 7 shows the sample standard deviation of the estimates for a sample size of 1000 for received signal strength based localization, which we choose as a metric for the error margin of the estimates. We also study the variation of the standard deviation of the estimates with the change in the moisture content (leading to change in $\epsilon'$ and $\epsilon''$ as per Table 1). The figures are omitted from this paper for space considerations. $X$ and $Y$ coordinates are estimated with a Rician K-factor $(\kappa)$ level of 20 (which amounts to standard deviation of about 0.4 m). Note as $\kappa$ increases to 100, the standard deviation reduces to about 0.18 m. Similarly, the standard deviation of the $Z$ coordinate estimate decreases with an increase in $\kappa$. The standard deviation in $X$ and $Y$ estimates is relatively insensitive to change in the soil permittivity. However, the $Z$ estimate has the least standard deviation for $\epsilon' = 24.5$ and $\epsilon'' = 3.86$, corresponding to a moisture content of 40%. The standard deviation of the $Z$ coordinate estimate decreases with the increase in the moisture content of the soil. This behavior can be explained as follows: with the increase in the moisture content the attenuation of the signal increases, decreasing the average received power. From Equations (27) and (28), the variance of the received signal strength also reduces, reducing the variance of the estimate.

The simulation results demonstrate that our technique is efficient, scalable and reliable. Although the results are presented for a field size of 150 m $\times$ 150 m, the technique is equally applicable to larger fields.

### A. Simulation Software

Data gathering phase

- Numpy Random module
- Observed received signal strength values

Estimation phase

- Scipy Optimize module
- Estimates node location coordinates

**Fig. 6: Simulation framework.**

Our modeling framework requires sampling of the received signal strength values measured between pairs of sender and receiver nodes. Since there is no explicit need for measuring the timing of various events for the simulations (although in deployment, scheduling of the various transmissions needs to be performed to ensure there are no collisions between different signals), we do not require an elaborate simulation approach such as that of event driven simulations. We model signal strength values by sampling from their probability distribution...
difference between the observed value of the respective parameter and its true value expressed as a percentage of the true value. The sensitivities with respect to the different parameters vary. A 10% error in the measured $\epsilon_s'$ caused the largest shift in the estimated location of 0.024 m; whereas the same error in $\epsilon_s'$ or $\mu_s$ causes a shift of 0.012 m.

Fig. 7: Sample standard deviation of estimates for RSS localization. Sample size = 1000.

as derived using our approach. For this purpose we use the random module from the numpy package of python. Next, during the optimization phase, we initialize the node locations to random values within the constraints of the field boundaries and use the observed values of the received signal strength (as sampled in the first phase of the simulations), to minimize the maximum likelihood function as derived in Equation (26) to estimate the true values of the nodes’ coordinates. For this purpose we use the optimize module from the scipy package of python.

B. Sensitivity analysis

Sensitivity analysis is used to determine the level of errors that may be introduced due to uncertainty in various parameters employed in the localization model. As an example, the soil permittivity is a parameter of our ranging model used in localization. Hence, localization performance depends on the accuracy with which the soil permittivity values ($\epsilon_s'$ and $\epsilon_s''$) are measured by our soil sensors. The uncertainty in their measurement may cause the mean of the estimators to be shifted relative to the actual coordinate values. We study the change in the estimated coordinates as a function of the difference between the measured $\mu_s$, $\epsilon_s'$ and $\epsilon_s''$ and their true values from Tables I and II. We use the same setup for sensitivity analysis as used in variance analysis. That is, a set of three sensor nodes are randomly deployed in the sensor field within the radio communication ranges of each other along with the four satellite nodes.

Figures 8, 9 and 10 present the sensitivities of the estimates (the shift in the estimate from its true value) to $\epsilon_s'$, $\epsilon_s''$ and $\mu_s$, respectively. The x axes in these figures represent the

VI. CONCLUDING REMARKS

We study node localization for a hybrid wireless sensor network system deployed across multiple physical media. As
in most applications of wireless sensor networks where sensed data is collected from different locations, location information is an integral part of the data for the application. Moreover, network layer routing can benefit from such information by routing data more efficiently to a destination node multiple hops away. Location information is also useful for network servicing operations such as battery replacement. We propose a cost-effective method based on receiving signal strength measurements for three-dimensional localization to compute the $X$, $Y$ coordinates along with the depth $Z$ of the sensor nodes. A group of nodes with known location coordinates, called satellite nodes, serve as the anchor nodes in our scheme. Received signal strength of the signal between neighboring underground sensor nodes and between satellite nodes and sensor nodes are used for estimating the inter-node distances.

To the best of our knowledge, past research has not considered localization in a sensor network where nodes are located in different physical media or on modeling the signal strength measurements for transmission across such multiple media. We present a model for the variation in received signal strength due to path loss for transmission in multiple media, multi-path fading, and reflection and refraction across soil-air interface. We use the non-central $\chi^2$ distribution to model the multi-path fading effects for two transmission scenarios: air to soil and soil to soil. The parameters of the distribution are governed by the distance between the receiver and the transmitter nodes, divergence and loss exponents, and Rician factor. We apply maximum likelihood optimization to the localization problem in a hierarchical and iterative manner to efficiently compute the location coordinates even for very large sensor networks.

An improvement in the running time complexity of the localization problem may be achieved by simplifying the optimization problem used for the estimation. For example, formulating the estimation as a method-of-moments problem is expected to speed up the computation of the estimates. However, the improvement in running time may come at the cost of the accuracy of the estimates. An error analysis of the estimates obtained using the method-of-moments approach should be performed.

Aside from developing accurate and efficient methods for resolving the proposed statistical estimation approach for sensor localization, we plan to develop a model-based framework for assimilation of sensor data towards agricultural prescription for irrigation and fertilization decision-making. Our approach would be to develop a framework (see Figure 11) which will allow us to integrate sensor data into site-specific fertilizer responses and to optimize the application of site specific fertilizer. Besides the sensor moisture and nutrient measurements, soil temperature and moisture (also from sensors), weather data (historical and forecasts) and crop management (planting, tillage, rotation, irrigation, etc.) will be assimilated into the model-based simulator APSIM (Agricultural Production System Simulator, www.apsim.info) to predict the yield responses. The model, as constrained by the data, will predict two important processes: nitrogen leaching (losses) versus organic-N mineralization (for crop uptake) and using these will determine yield-response curves to agricultural inputs. The economic optimum resource application rate will be achieved when the marginal cost for resources matches the marginal return on the yield.

**ACKNOWLEDGMENT**

The authors would like to thank Dr. Ahmed E. Kamal for his generous contribution in the technical discussions and for reviewing the work.

**REFERENCES**


[35] V. L. Mironov. Spectral dielectric properties of moist
