analytics

MDPI

Article

Configuration-Aware Bayesian Shelf Inference for Mobile RFID Library Inventory

Sherzod Mukhammadjonov, Marat Rakhmatullayev * and Husniya Boysunova

Department of Information and Library Systems, Tashkent University of Information Technologies Named After Muhammad al-Khwarizmi, Tashkent 100084, Uzbekistan; sh.muxammadjonov@tuit.uz (S.M.); husniya.librarian@gmail.com (H.B.)

* Correspondence: marat56@mail.ru

Abstract

Mobile RFID inventory in libraries must be planned and evaluated under noisy observations, configuration-dependent read regimes, and incomplete supervision. This paper presents an uncertainty-aware analytics framework for robot-assisted RFID inventory using the public RFID Location dataset. The framework has three phases. Phase 1 converts irregular list-encoded logs into atomic RFID events and quantifies how operating configuration changes read density and signal variability. Phase 2 performs map-constrained Bayesian shelf inference by synchronizing RFID reads with robot trajectory and antenna geometry and by fusing RSSI and carrier phase over feasible shelf candidates. Phase 3 translates posterior spread and non-convergence into proxy review workload and cost, enabling configuration comparison and certainty–throughput trade-off analysis when strict EPC-to-item linkage is unavailable. Across 688,073 aligned RFID observations, the pipeline produces 18,190 posterior tag estimates from five inventory runs. The empirical results show strong run dependence: the best run achieves a mean posterior spread of 0.906 m with a convergence rate of 0.553, whereas a degraded run reaches only 0.004 convergence with a mean spread above 2.1 m. Because EPC-to-item linkage is unavailable, these values are posterior concentration and workload indicators rather than ground-truthed localization-accuracy metrics. A saved phase-weight ablation further shows that adding phase information substantially sharpens posterior concentration relative to an RSSI-only baseline. Under the proxy workload model, `autonomous-S1-P30` provides the most favorable balance among posterior certainty, scan effort, and implied review burden.

Keywords: RFID analytics; mobile robotics; library inventory; Bayesian inference; uncertainty quantification; operational decision support

Check for updates

Academic Editor: Carson K. Leung

Received: 4 May 2026
Revised: 14 June 2026
Accepted: 15 June 2026
Published: 17 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

RFID is widely used in library circulation and inventory because it supports non-line-of-sight identification and much faster shelf scanning than barcode-based workflows [1,2]. Yet practical mobile inventory remains difficult. Dense shelving, partial shielding, metallic structures, multipath propagation, and reader motion all affect both how often tags are observed and how stable those observations remain [3,4]. In robot-assisted inventory, these sensing effects interact with trajectory geometry and antenna viewpoint, so the quality of the evidence cannot be inferred from scan time alone.

Recent RFID localization and inventory-robot research reinforces this point. Newer systems combine RSSI and phase fingerprints, multi-frequency or multi-view acquisition,

synthetic-aperture processing, and autonomous mobile inventory platforms to improve localization or inventory throughput [5–11]. These developments raise the standard for evaluation: a useful library-inventory study must explain what supervision is available, what operating configurations are represented, and which claims can be tested from the released data.

This creates two connected research problems. The first is an inference problem: RFID measurements are indirect, noisy, and configuration-dependent. RSSI is easy to collect but highly sensitive to indoor propagation effects, while carrier phase offers richer spatial information at the cost of phase wrapping, hardware offsets, and motion sensitivity [12–16]. The second is an evaluation problem: many RFID localization studies assume item-level ground truth and report absolute RMSE-style accuracy, but real operational datasets often do not expose the explicit EPC-to-item linkage needed for strict per-tag scoring.

The public RFID Location dataset is a representative but limited example of this setting [17]. It provides RFID observation logs, robot trajectories, reader-to-antenna mappings, static antenna transforms, an occupancy map, and baseline shelf coordinates for a mobile library inventory robot. The dataset was released in 2018, so it should not be interpreted as a complete representation of current RFID hardware or all contemporary mobile-robot inventory systems. Its value for the present study is narrower: it remains one of the few public library-scale datasets that exposes synchronized RFID, trajectory, antenna, map, and shelf metadata. However, it does not provide a public EPC-to-item linkage table that would support direct item-level accuracy evaluation. This missing linkage is not merely a nuisance; it changes what can be defended scientifically. In this setting, the central questions become: which operating configurations produce more informative observations, how strongly does posterior belief concentrate under map-constrained fusion, and what operational burden remains when the posterior does not converge cleanly?

These questions are important for deployment-oriented inventory robotics. Prior work has shown that autonomous inventory robots must be assessed not only by sensing quality but also by how well they support repeatable operation, route design, and downstream auditing efforts [18–21]. Likewise, smart-library systems that combine RFID with other sensing modalities still report degraded performance under dense tag populations, collisions, and distorted phase observations [22]. The practical need is therefore not another idealized localization benchmark, but an analysis framework that can compare sensing regimes and operating points under realistic supervisory limits.

This paper addresses that need by framing mobile RFID library inventory as an uncertainty-aware analytics problem under incomplete supervision. The proposed framework has three phases. Phase 1 converts irregular RFID logs into atomic observations and quantifies how operating configuration changes evidence density and signal variability. Phase 2 performs map-constrained Bayesian shelf inference by synchronizing RFID reads with robot trajectories and antenna geometry and by updating posterior belief over feasible shelf candidates using RSSI and phase evidence. Phase 3 transforms posterior spread and non-convergence into proxy workload and cost metrics so that configurations can be compared in terms of certainty, throughput, and implied review burden. Rather than claiming ground-truthed item-level localization accuracy that cannot be verified from the public release, the paper focuses on posterior uncertainty concentration, failure-mode exposure, and operating-point selection.

The main contributions of this paper are fourfold:

1. We provide a reproducible ingestion and normalization procedure for irregular RFID observation logs, preserving multi-read structure while converting the public dataset into 688,073 aligned atomic observations.

2. We characterize configuration-dependent signal behavior in terms of per-tag read density and signal variability, showing that evidence quantity and evidence stability do not improve together across operating modes.
3. We develop a map-constrained Bayesian shelf-inference pipeline that fuses synchronized RSSI and phase observations with robot trajectory and antenna geometry to produce shelf-level posterior estimates with explicit uncertainty and convergence diagnostics.
4. We introduce a proxy operational evaluation that translates posterior spread and non-convergence into workload and cost indicators, enabling deployment-oriented configuration comparison under incomplete supervision.

The remainder of this paper is organized as follows. Section 2 reviews the relevant literature. Section 3 introduces the dataset and formalizes the problem setting. Section 4 details the proposed system architecture. Section 5 presents the experimental results, Section 6 discusses the findings and limitations, and Section 7 concludes the paper.

2. Related Work

Prior work related to this study falls into four groups: library RFID inventory studies, recent RFID localization and sensor-fusion methods, mobile inventory robots, and operational evaluation of inventory systems.

2.1. Library RFID and Smart-Library Inventory

Library RFID performance is highly context-dependent. Even handheld inventory studies report that read rate varies with tag placement, book geometry, and scanning motion rather than with RFID deployment alone [3]. Smart-library robot systems further show that dense shelving introduces inter-tag coupling, collisions, and incomplete phase observations, so observation quality is heterogeneous even within the same environment [22]. Recent smart-library and book-inventory studies confirm that RFID continues to be adopted for library security, circulation, and inventory automation, but they also emphasize system integration and deployment constraints rather than pure localization accuracy [23–25]. These studies motivate our Phase 1 emphasis on characterizing the sensing regime itself rather than assuming that all collected reads are equally informative.

2.2. RFID Localization, Fusion, and Recent Benchmarks

RSSI-based methods remain attractive because RSSI is readily available, but they are sensitive to indoor propagation effects and therefore often require probabilistic filtering, fingerprinting, or optimization instead of direct range inversion [12,26,27]. Phase-based methods can provide stronger spatial discrimination, but only with careful handling of phase periodicity, hardware offsets, motion, and ambiguity [14–16,28–30]. Recent work has therefore moved toward joint RSSI–phase fusion, multi-frequency or multi-view acquisition, and synthetic-aperture formulations [5–8]. These directions are consistent with the present paper's use of RSSI for coarse discrimination and phase for posterior sharpening, while also showing why configuration and trajectory geometry must be reported explicitly.

Machine learning localization has also become more prominent. Recent RFID fingerprint-fusion work shows that learned models can be effective when labeled training data are available [7]. However, such methods usually require supervised labels or stable fingerprints, which are not available in the public RFID Location dataset used here. For this reason, the present study reports an RSSI-only baseline and a phase-weight sensitivity analysis, but it does not claim a direct numerical comparison against supervised deep-learning systems that require missing EPC-to-item ground truth.

2.3. Mobile RFID Inventory Robots

Mobile RFID inventory is increasingly studied as a robotics problem rather than only as a reader-design problem. Prior work on stocktaking robots, product maps, and retail inventory metrics has shown that performance depends on coverage, route design, and operational repeatability [18–20,31]. Recent systems extend this trend through hybrid warehouse robots, autonomous UHF RFID-equipped robots, and digital-twin populations using mobile RFID platforms [9–11]. Plug-and-play inventory robots further show that autonomous waypoint and itinerary generation can approach the performance of human-designed routes [21]. This literature is relevant because it shifts attention from isolated localization accuracy toward repeatable deployment performance and configuration selection.

2.4. Operational and Cost-Aware Evaluation

Operational inventory studies increasingly evaluate RFID systems through throughput, review workload, and cost-related indicators, not only through localization error [32–34]. At the same time, RFID read-rate optimization remains an active concern because antenna placement, transmit settings, tag orientation, and surrounding materials can alter the number and quality of reads available to the inference layer [35]. These studies support the decision to include a proxy workload and cost layer in Phase 3. They also motivate the sensitivity analysis reported in this paper because cost coefficients and review times are site-specific rather than universal.

The gap addressed by the present work is therefore specific. Existing studies typically assume one of three conditions that do not hold simultaneously in our setting: reliable item-level supervision, a primary interest in metric localization accuracy, or an evaluation focused on coverage rather than inference uncertainty. By contrast, our goal is to analyze a public robot-assisted library dataset in which configuration metadata, robot trajectories, and map constraints are available, but direct EPC-to-item linkage is not. The contribution of this paper is to combine configuration-aware signal analysis, map-constrained Bayesian shelf inference, an RSSI-only baseline, phase-weight sensitivity analysis, and proxy operational evaluation in a single uncertainty-aware analytics pipeline. The novelty is not a new RFID sensing primitive; it is a reproducible framework for comparing operating conditions and downstream review burden when complete supervision is unavailable.

3. Dataset and Problem Setting

3.1. Dataset Contents and Constraints

The experiments in this study are based on the public RFID Location dataset, which contains five inventory runs collected by a mobile robot equipped with RFID readers and multiple antennas in a real library environment. According to the dataset record, the library contains approximately 7000 tagged books with associated shelf-location information in the baseline metadata [17]. The public release includes raw RFID observation logs, robot trajectory files, run metadata, reader-port to antenna mappings, static robot-to-antenna transforms, a 2D occupancy map, and a baseline shelf-location table.

The dataset also indicates whether each run was performed in autonomous or manual mode, together with the RFID session setting and transmit power. In the released data, this yields three effective operating configurations:

- `autonomous-S1-P30`,
- `autonomous-S2-P30`,
- `manual-S1-P30`.

These configurations are important because the proposed analysis is not limited to tag estimation alone; it also examines how the sensing regime itself changes across operating modes. In particular, the dataset allows comparison of observation density,

signal variability, convergence behavior, and implied operational burden across different robot inventory settings.

A central constraint of the public release is that, although `location_baseline.csv` provides baseline shelf coordinates indexed by `item_reference`, it does not provide an explicit EPC-to-item linkage table for the observed RFID tags. Consequently, strict item-level evaluation metrics such as per-EPC RMSE, MAE, or exact shelf-assignment accuracy cannot be computed directly from the released files alone. Rather than introducing unverifiable assumptions, this work treats the dataset as an incomplete-supervision benchmark and evaluates the system through posterior concentration, convergence behavior, and operational proxy measures [17].

3.2. Configuration Coverage and External Validity

The dataset supports only a limited set of configuration contrasts. Specifically, all released runs use the same nominal transmit power (30 dBm) and the same antenna hardware, while the available variation is concentrated in inventory mode, RFID session, and the executed scan trajectory. Therefore, this paper does not claim to evaluate the full design space of transmit power, antenna placement, robot velocity, and scanning strategy. Instead, it evaluates the configuration diversity that is actually present in the public data and treats unobserved operating dimensions as external-validity limitations.

This distinction is important for interpreting the results. The analysis can show that `autonomous-S1-P30`, `autonomous-S2-P30`, and `manual-S1-P30` produce different evidence density, signal variability, posterior concentration, and review workload. It cannot establish how a different transmit power, a redesigned antenna rig, or a newly collected route family would behave. A contemporary validation campaign should therefore add factorial variation in transmit power, antenna height and side, robot velocity, shelf aisle geometry, and scanning strategy; the present paper provides the analysis pipeline and diagnostic metrics needed for such an extension.

3.3. Problem Setting

Let

$$\mathcal{X} = \{\mathbf{x}_j\}_{j=1}^{M} \tag{1}$$

denote the finite set of feasible shelf candidates derived from the baseline shelf coordinates after map-based filtering. For a given EPC e, we collect a sequence of RFID observations

$$Z_e = \{z_i\}_{i=1}^{N_e}, \qquad z_i = (r_i, \phi_i, t_i, f_i, a_i), \tag{2}$$

where r_i is RSSI, ϕ_i is carrier phase, t_i is timestamp, f_i is carrier frequency, and a_i is antenna identity. Each observation is aligned with the corresponding robot trajectory and antenna pose information obtained from the dataset metadata and transform files.

Let

$$c_r = (\texttt{inventory_type}, \texttt{RFID_session}, \texttt{RFID_power}) \tag{3}$$

denote the configuration of run r. The objective is to estimate, for each EPC, a posterior distribution over feasible shelf candidates conditioned on its synchronized RFID observations and the run configuration. In compact form, this can be written as

$$p(\mathbf{x}_j \mid Z_e, c_r) \propto p_0(\mathbf{x}_j \mid c_r) \prod_{i=1}^{N_e} p(z_i \mid \mathbf{x}_j, \pi_i, c_r), \tag{4}$$

where π_i denotes the synchronized antenna pose for observation z_i, and $p_0(\mathbf{x}_j \mid c_r)$ is a map-constrained prior over feasible shelf candidates.

Because direct EPC-to-item supervision is unavailable, the primary goal of this work is not ground-truthed item-level localization accuracy, but uncertainty-aware shelf inference under realistic operational constraints. Accordingly, the analysis focuses on three questions: how RFID signal behavior changes across configurations, how strongly the resulting posterior distributions concentrate under Bayesian fusion, and how residual uncertainty and non-convergence translate into proxy review workload for deployment. The detailed measurement models, fusion equations, and operational cost definitions are introduced in Section 4.

4. System Architecture

The proposed framework consists of three connected phases, as illustrated in Figure 1. The phases correspond to three levels of analysis. Phase 1 transforms irregular raw logs into configuration-level signal descriptors. Phase 2 converts synchronized RFID evidence into posterior distributions over feasible shelf candidates. Phase 3 converts posterior uncertainty into proxy workload and cost indicators that support operating-point comparison. This progression is important for the current dataset because the available evidence supports uncertainty-aware decision analytics more directly than strict item-level localization accuracy.

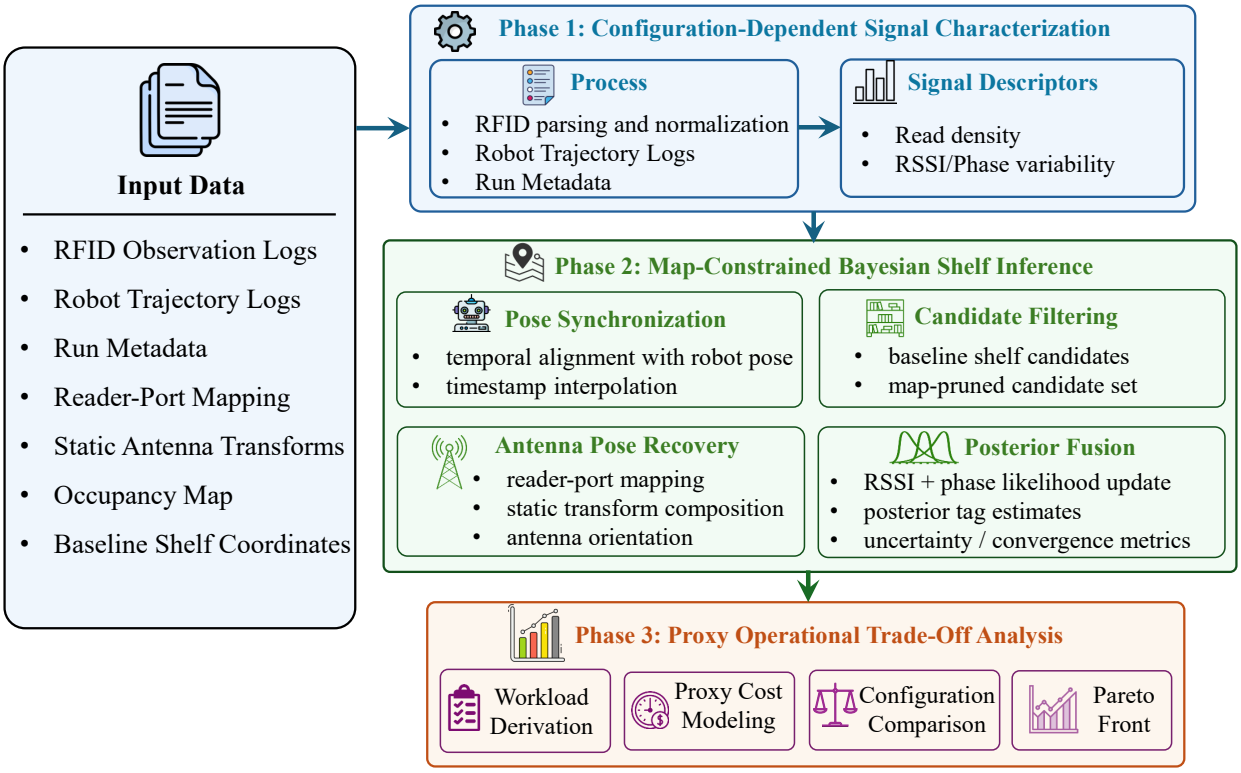


Figure 1. Overview of the proposed three-phase analytics framework. The input layer integrates RFID observations, robot trajectories, run metadata, antenna mappings, and transforms the occupancy map and baseline shelf coordinates. Phase 1 produces configuration-aware signal descriptors, Phase 2 performs map-constrained Bayesian shelf inference, and Phase 3 converts posterior spread and non-convergence into proxy workload, cost, and trade-off summaries.

4.1. Configuration-Dependent Signal Characterization

The first phase converts heterogeneous RFID logs into a clean atomic-read table and summarizes how signal behavior changes across operating configurations. This step is necessary because the observation files are not organized as one physical read per row. Instead, several fields, including timestamps, antenna ports, RSSI values, phases,

and carrier frequencies, may appear as list-encoded entries within a single CSV row. As a result, one row can represent either a single RFID event or multiple logically paired events. Before any inference is performed, these entries must be separated while preserving the correspondence among their signal fields.

The preprocessing pipeline parses list-like strings, aligns multi-valued columns row-wise, expands the aligned values into atomic read rows, converts timestamps and signal fields to numeric form, and infers the timestamp unit before expressing all times in seconds. To verify the integrity of this step, the implementation also records ingestion statistics such as the number of rows before parsing, the number of rows after expansion, the number of rows dropped after numeric filtering, and the number of rows with mismatched array lengths.

Let the i-th atomic RFID read be represented as

$$z_i = (t_i, r_i, \phi_i, f_i, a_i, e_i, c_i), \tag{5}$$

where t_i is the timestamp, r_i is RSSI, ϕ_i is carrier phase, f_i is carrier frequency, a_i is antenna identity, e_i is EPC, and c_i is the run configuration. The configuration vector is

$$c_i = (\texttt{inventory_type}, \texttt{RFID_session}, \texttt{RFID_power}). \tag{6}$$

After normalization, observations are grouped by configuration and tag identity. For a given EPC e under configuration c, the corresponding read set is

$$\mathcal{Z}_{e,c} = \{z_i \mid e_i = e,\ c_i = c\}. \tag{7}$$

From this set, Phase 1 computes per-tag signal descriptors that summarize both evidence quantity and evidence stability. Specifically, the mean RSSI, RSSI variance, phase variance, and temporal read density are defined as

$$\mu_r(e,c) = \frac{1}{|\mathcal{Z}_{e,c}|} \sum_{z_i \in \mathcal{Z}_{e,c}} r_i, \tag{8}$$

$$\sigma_r^2(e,c) = \mathrm{Var}(\{r_i \mid z_i \in \mathcal{Z}_{e,c}\}), \tag{9}$$

$$\sigma_\phi^2(e,c) = \mathrm{Var}(\{\phi_i \mid z_i \in \mathcal{Z}_{e,c}\}), \tag{10}$$

$$\rho(e,c) = \frac{|\mathcal{Z}_{e,c}|}{\max(t_{\max}(e,c) - t_{\min}(e,c), \epsilon)}, \tag{11}$$

where $\epsilon > 0$ is a minimum-duration safeguard used to avoid unstable density values when the observed time span is very short.

These descriptors capture a practical point that is central to the rest of the paper: more frequent reading does not necessarily imply cleaner reading. A configuration can generate many observations while still exhibiting large variability, and a sparse configuration can appear stable while providing too little evidence for strong posterior concentration. Accordingly, configurations are compared not only in terms of throughput, but also in terms of signal consistency.

The per-tag descriptors are then aggregated at the configuration level using summary statistics such as the mean, median, and standard deviation of the quantities in (8)–(11). These configuration-level summaries serve two purposes. First, they provide a descriptive comparison of sensing conditions across operating modes. Second, they supply interpretable signal-scale information that can guide the likelihood settings used later in Phase 2. Algorithm 1 summarizes the Phase 1 procedure.

Algorithm 1 Configuration-Aware Signal Characterization

Require: Observation files $\{\mathcal{O}_r\}_{r=1}^{R}$, configuration file $\mathcal{I}$
Ensure: Atomic read table and per-configuration signal summaries
for each run file $\mathcal{O}_r$ **do**
    parse list-encoded fields into arrays
    align multi-valued columns row-wise
    expand aligned arrays into atomic read rows
    convert timestamps and signal fields to numeric values
    infer timestamp unit and convert time to seconds
    assign run identifier r
end for
concatenate all atomic reads into one table
merge atomic reads with run metadata from $\mathcal{I}$
for each configuration c and EPC e **do**
    compute $\mu_r(e,c)$, $\sigma_r^2(e,c)$, $\sigma_\phi^2(e,c)$, and $\rho(e,c)$
end for
for each configuration c **do**
    aggregate per-tag descriptors into configuration-level summaries
end for

4.2. Map-Constrained Bayesian Shelf Inference

Phase 2 converts asynchronous RFID detections into shelf-level posterior estimates by combining robot trajectory, antenna geometry, and map constraints. Because the public dataset does not provide direct EPC-to-item linkage, the goal of this phase is not to claim verified item-level localization accuracy. Instead, it is to estimate which shelf candidates are most plausible for each EPC and to quantify how strongly the posterior concentrates as evidence accumulates.

For each RFID read, the robot pose is first synchronized with the trajectory by timestamp interpolation. The physical antenna used for that read is then identified from the reader-port mapping, and its pose in the map frame is recovered by composing the interpolated robot pose with the static robot-to-antenna transform. This produces the antenna position associated with each observation, which is the geometric reference used in the inference stage.

Candidate shelf locations are derived from the baseline location table and filtered using the occupancy map so that posterior mass is assigned only to map-consistent non-free shelf candidates. Let the feasible candidate set be

$$\mathcal{X} = \{\mathbf{x}_j \in \mathcal{X}_{\text{raw}} \mid \Omega(\mathbf{x}_j) = 1\}, \tag{12}$$

where $\mathcal{X}_{\text{raw}}$ denotes the raw shelf coordinates and $\Omega(\mathbf{x})$ is the occupancy-based feasibility operator.

The implementation uses the `occupied_or_unknown` map mode. Each shelf coordinate is projected into the ROS occupancy grid image using the map resolution (0.02 m), origin, and occupancy thresholds from `map.2-d.yaml`. Candidate points outside the map are removed, and points inside cells whose occupancy probability is below the free-space threshold (`free_thresh` $= 0.196$) are treated as clearly free space and pruned. Points in occupied or unknown cells are retained; the stricter `occupied_only` mode is not used in the reported experiments. This choice is intentionally conservative because shelves may be represented as occupied or partially unknown structures in the map. Nevertheless, the filter can introduce bias if the occupancy map is stale, misregistered, or if a true shelf coordinate falls in a cell labeled as free. For this reason, the candidate reduction is reported explicitly, and the map filter is interpreted as a constraint for posterior concentration rather than as ground-truth validation of item position.

The prior is intentionally conservative. In the absence of EPC-to-item linkage, there is no defensible per-tag shelf prior, so the initial posterior is uniform over the feasible candidate set or over the local computational subset used for a tag:

$$p_0(\mathbf{x}_j) = \begin{cases} 1/|\widetilde{\mathcal{X}}_e|, & \mathbf{x}_j \in \widetilde{\mathcal{X}}_e, \\ 0, & \text{otherwise}, \end{cases} \tag{13}$$

where $\widetilde{\mathcal{X}}_e \subseteq \mathcal{X}$ denotes the nearest candidate subset around the mean observed antenna position for EPC e. This prior encodes only map feasibility and computational locality; it does not inject unobserved item identity information.

For each EPC e, the synchronized observations are processed sequentially to update a discrete posterior over the feasible shelf candidates. The posterior update is written as

$$p_i(\mathbf{x}_j) \propto p_{i-1}(\mathbf{x}_j)\, p(z_i \mid \mathbf{x}_j, \pi_i, c_r), \tag{14}$$

where z_i is the i-th RFID observation, π_i is the synchronized antenna pose, and c_r is the configuration of the corresponding run. The likelihood combines RSSI and phase evidence, with the phase contribution allowed to be either fixed or adaptively weighted. In this way, the model uses RSSI for coarse discrimination and phase for finer spatial refinement while still accounting for the circular nature of phase measurements.

The RSSI component uses a standard log-distance attenuation model:

$$\hat{r}_{ij} = r_0 - 10n \log_{10}\big(\max(d_{ij}, d_{\min})\big), \tag{15}$$

$$L_{ij}^{(r)} = \exp\left[-\frac{1}{2}\left(\frac{r_i - \hat{r}_{ij}}{\sigma_r}\right)^2\right], \tag{16}$$

where d_{ij} is the distance from the synchronized antenna pose to candidate $\mathbf{x}_j$, r_0 is the nominal RSSI at 1 m, n is the path-loss exponent, $d_{\min}$ prevents singular behavior at very small distances, and σ_r controls the width of the RSSI likelihood. In the experiments, these values are fixed before evaluation as $r_0 = -46$ dBm, $n = 2.2$, $d_{\min} = 0.2$ m, and $\sigma_r = 4.5$ dB.

The phase component is modeled on the wrapped two-way propagation phase:

$$\hat{\phi}_{ij} = \big(4\pi d_{ij}/\lambda_i\big) \bmod 2\pi, \qquad \Delta\phi_{ij} = \operatorname{atan2}\big(\sin(\phi_i - \hat{\phi}_{ij}), \cos(\phi_i - \hat{\phi}_{ij})\big), \tag{17}$$

$$L_{ij}^{(\phi)} = \exp\left[-\frac{1}{2}\left(\frac{\Delta\phi_{ij}}{\sigma_\phi}\right)^2\right], \tag{18}$$

where λ_i is the wavelength associated with the carrier frequency of observation i. Because the public data do not provide a hardware phase-offset calibration, the phase likelihood is deliberately broad ($\sigma_\phi = 0.8$ rad) and reliability-weighted rather than treated as an absolute range measurement.

The combined likelihood is

$$p(z_i \mid \mathbf{x}_j, \pi_i, c_r) \propto L_{ij}^{(r)} \left(L_{ij}^{(\phi)}\right)^{w_i}, \tag{19}$$

where w_i is the phase weight. The main experiments use a fixed value $w_i = 0.25$. The saved sensitivity analysis also evaluates $w_i \in \{0, 0.1, 0.25, 0.5, 1.0\}$ and an adaptive reliability setting in which local circular phase dispersion reduces the effective weight. Thus, the RSSI-only case is not a separate method but the special case $w_i = 0$ of the same Bayesian update.

To reduce computational cost, the implementation evaluates the posterior on a local subset of candidates near the mean antenna position for a given EPC. This subset is a

computational approximation to the full map-feasible candidate set and is kept large enough in the reported experiments (`candidate-k` = 500) to preserve multiple plausible shelf alternatives.

After all reads for an EPC have been assimilated, the system extracts the posterior mean, MAP candidate, entropy, and covariance-based posterior spread. Posterior spread is summarized as

$$u_{e,i} = \sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_{e,i})}, \tag{20}$$

where $\boldsymbol{\Sigma}_{e,i}$ is the posterior covariance after the i-th update. This quantity is expressed in meters and is therefore directly comparable across runs and configurations.

A tag is considered converged when its posterior spread falls below a fixed threshold:

$$u_{e,i} \leq \tau_{\sigma}. \tag{21}$$

The first read index satisfying (21) is recorded as the reads-to-convergence value. Together, these outputs allow Phase 2 to measure not only where the posterior is centered, but also how quickly and how confidently it stabilizes. Algorithm 2 summarizes the Phase 2 procedure.

Algorithm 2 Map-Constrained Bayesian Shelf Inference

Require: Atomic RFID observations, trajectory files, antenna mappings, static transforms, occupancy map, baseline shelf coordinates
Ensure: Posterior estimates and posterior-spread metrics for each EPC
for each RFID read z_i **do**
    interpolate the robot base pose at time t_i
    resolve the physical antenna identity from the reader-port mapping
    compose transforms to recover the antenna pose $\mathbf{s}_i$
end for
construct $\mathcal{X}_{\mathrm{raw}}$ from baseline shelf coordinates
apply occupancy filtering to obtain $\mathcal{X}$
for each EPC e **do**
    collect synchronized reads $\mathcal{Z}_e$
    initialize the posterior over $\mathcal{X}$ or its local approximation $\widetilde{\mathcal{X}}_e$
    for each read $z_i \in \mathcal{Z}_e$ **do**
        compute candidate distances d_{ij}
        evaluate RSSI and phase likelihoods
        update and normalize the posterior
        compute posterior spread $u_{e,i}$
        if $u_{e,i} \leq \tau_{\sigma}$ for the first time **then**
            record reads-to-convergence
        end if
    end for
    output posterior mean, MAP candidate, entropy, effective support, and posterior spread
end for

4.3. Proxy Operational Evaluation

Phase 3 translates posterior outputs into deployment-oriented indicators for configuration comparison. Because the public dataset does not provide direct EPC-to-item linkage, unresolved estimates are not interpreted as hidden classification errors. Instead, large posterior spread and non-convergence are treated as indicators of likely follow-up effort. This allows the evaluation to remain operationally meaningful even when strict per-EPC ground truth is unavailable.

For each run, the system counts how many estimated tags remain uncertain under a fixed posterior-spread threshold and how many fail to satisfy the convergence criterion introduced in Phase 2. These quantities are also normalized by the number of estimated tags so that runs with different throughput can be compared on a per-tag basis.

The main workload proxy for run r is defined as

$$N_{\text{proxy},r} = N_{\text{uncertain},r} + \lambda N_{\text{nonconv},r}, \tag{22}$$

where $N_{\text{uncertain},r}$ is the number of tags whose posterior spread exceeds the uncertainty threshold, $N_{\text{nonconv},r}$ is the number of non-converged tags, and λ is a penalty factor that assigns greater weight to unresolved cases.

The corresponding proxy operational cost is

$$C_r^{\text{proxy}} = C_{\text{robot}} T_{\text{scan},r} + C_{\text{human}} N_{\text{proxy},r} t_{\text{review}}, \tag{23}$$

where $T_{\text{scan},r}$ is the scan duration, t_{review} is the assumed review time per flagged case, and C_{robot} and C_{human} denote robot-time and human-time cost coefficients, respectively.

In this framework, τ_u, τ_σ, λ, C_{robot}, C_{human}, and t_{review} are scenario parameters rather than learned quantities. Their values are chosen before comparative evaluation, and the resulting configuration ranking should therefore be interpreted as a trade-off under the present proxy assumptions rather than as a universal cost ranking. Because published cost data for robot-assisted RFID library inventory remain scarce, Section 5.3 reports sensitivity analyses for workload thresholds and review-cost assumptions instead of treating a single coefficient set as empirical ground truth.

At the configuration level, run-level metrics are averaged over all runs sharing the same operating mode:

$$\bar{m}(c) = \frac{1}{|R(c)|} \sum_{r \in R(c)} m(r), \tag{24}$$

where $R(c)$ is the set of runs executed under configuration c and $m(\cdot)$ denotes any run-level metric, such as convergence rate, posterior spread, or proxy cost per estimated tag.

Finally, Phase 3 constructs a Pareto view over scan duration and posterior spread. This produces a non-dominated set of operating points rather than a single one-dimensional ranking, which is more appropriate when speed and posterior certainty must be balanced jointly. Algorithm 3 summarizes the Phase 3 procedure.

Algorithm 3 Proxy Operational Evaluation

Require: Phase 2 posterior estimates, run summaries, configuration metadata, thresholds τ_σ and τ_u, penalty factor λ, cost parameters
Ensure: Run-level and configuration-level proxy trade-off metrics
for each run r **do**
  count estimated tags $|\mathcal{E}_r|$
  count tags with posterior spread above τ_u
  count non-converged tags under threshold τ_σ
  compute proxy workload $N_{\text{proxy},r}$
  compute proxy cost C_r^{proxy}
  compute normalized workload and cost per estimated tag
end for
aggregate run-level metrics by configuration using (24)
construct the Pareto front on scan duration versus posterior spread
export run-level, configuration-level, and Pareto summaries

5. Results

This section reports empirical results for three related questions: how operating configuration changes the observation regime, how those sensing regimes affect posterior concentration under map-constrained shelf inference, and how residual uncertainty translates into deployment-oriented workload. Because the public release does not provide direct EPC-to-item linkage, the reported quantities should be interpreted as posterior spread, convergence, and operational indicators rather than as ground-truthed item-level localization accuracy.

5.1. Configuration-Dependent Signal Behavior

Phase 1 first establishes that the raw logs can be transformed into a stable atomic-read dataset. Across the five runs, the ingestion pipeline expands 672,731 raw observation rows into 688,073 atomic RFID observations while preserving field correspondence across list-encoded entries. No rows are lost after numeric filtering, and only five observations require pose extrapolation after synchronization. The preprocessing report also shows that irregular list structure is non-trivial: 8682 source rows contain mismatched array lengths before alignment. This validates the decision to treat preprocessing as part of the contribution rather than as a routine implementation detail.

After normalization, the three effective operating configurations show clear differences in evidence density and signal variability. Table 1 reports the resulting configuration-level descriptors.

Table 1. Phase 1 configuration-level signal characterization results.

Configuration	n_{tags}	Mean RSSI Var	Mean Phase Var	Mean Density (Hz)
Autonomous-S1-P30	7354	20.16	2590.78	0.044
Autonomous-S2-P30	7446	7.12	2615.39	0.003
Manual-S1-P30	6863	16.92	2636.41	0.087

Three patterns are important. First, `autonomous-S2-P30` yields the lowest mean RSSI variance, suggesting the most stable amplitude regime, but it is also by far the sparsest configuration. Second, `manual-S1-P30` produces the highest read density, yet this denser evidence stream does not coincide with the lowest variability. Third, `autonomous-S1-P30` lies between these extremes in evidence density while exhibiting the largest average RSSI variance. In other words, configuration choice cannot be reduced to maximizing reads or minimizing variance alone; the sensing regime changes both evidence quantity and evidence reliability.

Figure 2 summarizes the configuration-level differences in mean RSSI variance and mean read density. The relationship is clearly non-monotonic: the densest configuration is not the most stable one, and the most stable one is also the least informative in terms of evidence volume. This distinction matters because Phase 2 depends on both properties. Sparse evidence may be insufficient for strong posterior concentration, whereas dense but noisy evidence can drive less selective likelihood updates.

Figure 3 presents the same relationship as a trade-off surface. The main conclusion is that Phase 1 justifies a configuration-aware downstream analysis: none of the three configurations dominates simultaneously in evidence density and evidence stability, so later stages must evaluate how these sensing regimes translate into posterior concentration and review burden.

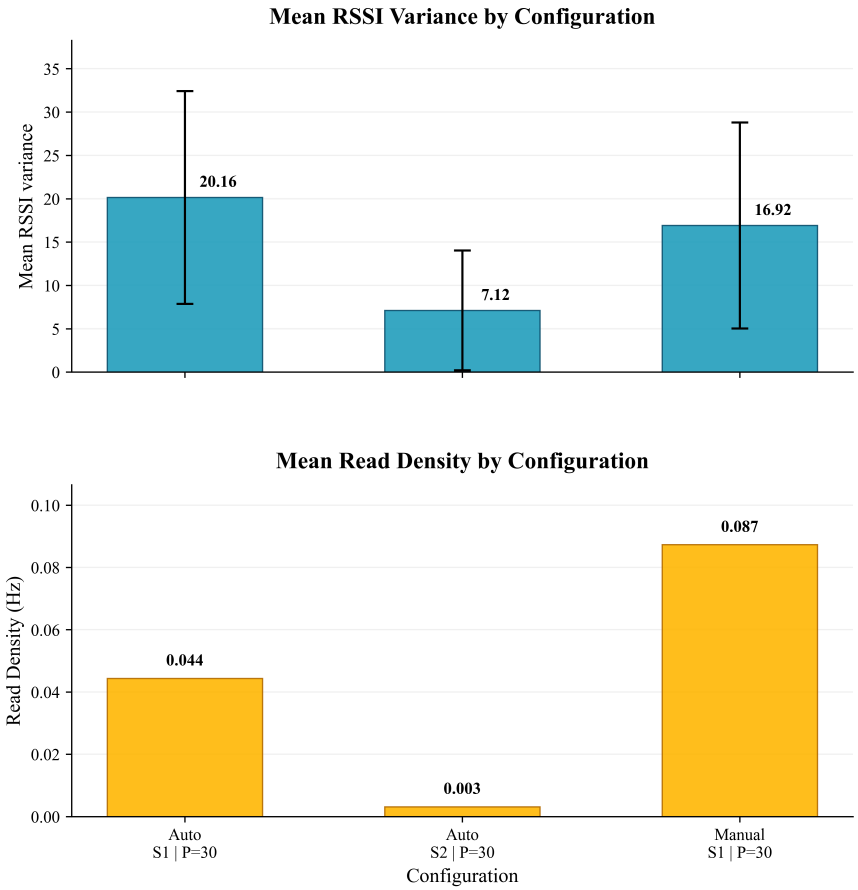


Figure 2. Configuration-level summary of Phase 1 results, comparing mean RSSI variance and mean read density across the three operating configurations.

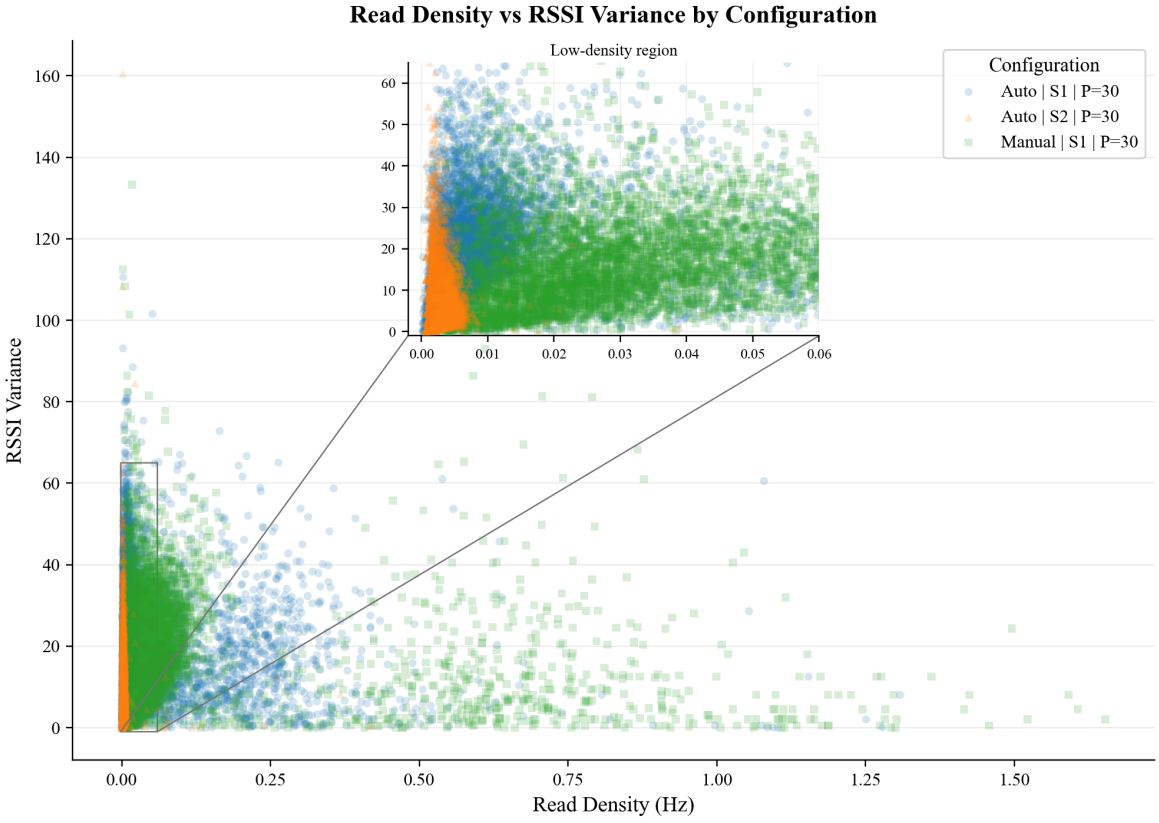


Figure 3. Trade-off between read density and RSSI variance across the three operating configurations.

Overall, the Phase 1 results support the configuration-aware formulation adopted in this paper. They show that the raw RFID sensing process changes materially across operating modes and that these changes affect both the amount and the quality of the evidence available for downstream inference.

5.2. Map-Constrained Bayesian Shelf Inference Results

Phase 2 evaluates the posterior behavior of the map-constrained shelf-inference pipeline after temporal synchronization, antenna-pose recovery, and candidate-space filtering. Across all five runs, the pipeline processed 688,073 aligned RFID observations and produced 18,190 posterior tag estimates. Candidate filtering reduced the raw shelf set from 7423 baseline points to 2749 map-consistent candidates, eliminating approximately 63.0% of the raw candidate positions before posterior updating. This pruning is important because it restricts posterior mass to map-consistent shelf candidates and makes uncertainty measures easier to interpret operationally. Because the filter can bias the posterior if the occupancy map is wrong or misregistered, Section 4 now specifies the exact `occupied_or_unknown` criterion and treats the resulting estimates as map-constrained posterior concentration metrics rather than as independent proof of true item position.

Table 2 summarizes the run-level outputs in terms of estimated tag count, mean posterior spread, convergence rate, uncertain-tag rate, and scan duration.

Table 2. Phase 2 run-level posterior concentration and convergence results.

Run	Estimated Tags	Mean Posterior Spread (m)	Convergence Rate	Uncertain Rate	Scan Duration (s)
1	3812	0.906	0.553	0.224	2133.4
2	950	2.105	0.004	0.871	2915.2
3	1170	1.365	0.387	0.365	2108.1
4	6797	1.260	0.415	0.395	2065.6
5	5461	1.622	0.243	0.577	1702.2

The run-level results show substantial variability in posterior behavior. Run 1 exhibits the strongest posterior profile, with the lowest mean spread (0.906 m) and the highest convergence rate (0.553). Run 4 provides a different but still competitive operating point: it produces the largest number of estimated tags while maintaining moderate posterior spread and convergence, making it attractive from a throughput perspective. Run 5 is the fastest run, but its posterior quality is weaker, with a larger spread and a substantially larger unresolved fraction than Run 1. Taken together, these differences indicate that scan duration alone does not determine posterior quality. Faster runs may be operationally efficient, but they can also yield less concentrated posteriors and therefore greater downstream review burden.

Because the number of independent inventory runs is small, the run-level means in Table 2 should not be interpreted as a 30–50 trial statistical study. At the tag-estimate level; however, the posterior-spread means are stable within each run because each mean is computed over hundreds to thousands of tags. The corresponding approximate 95% confidence intervals for mean posterior spread are 0.883–0.928 m for Run 1, 2.074–2.136 m for Run 2, 1.332–1.399 m for Run 3, 1.240–1.280 m for Run 4, and 1.601–1.644 m for Run 5. These intervals quantify tag-level estimation stability, while the limited number of runs remains an external-validity limitation discussed in Section 6.

To further assess statistical reliability without pretending to create new physical inventory runs, we added a bootstrap analysis over the Phase 2 tag-level estimates. For each reported bootstrap budget, the analysis resamples tag estimates with replacement within each recorded run, recomputes configuration-level posterior spread, convergence,

and proxy workload for all three configurations, and records the rank-1 configuration. Table 3 summarizes representative budgets for `autonomous-S1-P30`, `autonomous-S2-P30`, and `manual-S1-P30`. In the reported budgets, `autonomous-S1-P30` remains the best-ranked configuration, while the other two configurations remain consistently higher in proxy workload. This improves the statistical reliability of the reported ranking at the tag-estimate level, but it is not a substitute for collecting 30–50 new independent physical inventory runs.

Table 3. Bootstrap reliability analysis over tag-level posterior estimates for all operating configurations. The table reports representative bootstrap budgets and summarizes the stability of the configuration ranking.

Bootstrap Replicates	Configuration	Workload/Tag Mean (95% CI)	Rank-1 Probability	Physical Runs
10	`autonomous-S1-P30`	0.820 (0.809–0.831)	1.00	2
10	`manual-S1-P30`	1.333 (1.320–1.345)	0.00	1
10	`autonomous-S2-P30`	1.426 (1.406–1.448)	0.00	2
50	`autonomous-S1-P30`	0.826 (0.806–0.844)	1.00	2
50	`manual-S1-P30`	1.333 (1.317–1.350)	0.00	1
50	`autonomous-S2-P30`	1.422 (1.401–1.450)	0.00	2
100	`autonomous-S1-P30`	0.825 (0.812–0.843)	1.00	2
100	`manual-S1-P30`	1.334 (1.316–1.350)	0.00	1
100	`autonomous-S2-P30`	1.423 (1.398–1.449)	0.00	2

The clearest failure case is Run 2. It combines the longest scan duration, the largest mean posterior spread (2.105 m), the lowest convergence rate (0.004), and the highest uncertain-tag rate (0.871). This pattern indicates a genuine failure regime rather than a simple lack of scan time. Additional acquisition time does not rescue inference when the synchronized evidence remains weak or geometrically uninformative. To investigate this failure more precisely, Table 4 reports read-density, antenna-pose coverage, and phase-residual diagnostics computed from the pose-aligned observation table.

Table 4. Failure-mode diagnostics for Run 2 compared with the other runs. The percentage below eight reads is computed before the Phase 2 minimum-read filter. Antenna-pose area is the bounding-box area covered by pose-aligned antenna positions. Phase residual circular standard deviation is computed against the final posterior mean estimates.

Run	Median Reads/Tag	Tags Below 8 Reads (%)	Antenna-Pose Area (m^2)	Phase Residual Circ. Std. (rad)	Convergence Rate
1	21	30.4	1.09	2.104	0.553
2	4	87.3	2.27	3.011	0.004
3	4	84.3	18.27	2.824	0.387
4	29	8.1	1.12	2.268	0.415
5	19	23.1	0.91	2.593	0.243

The diagnostics identify three contributing factors for Run 2. First, it is extremely sparse at the tag level: 87.3% of observed tags have fewer than eight reads, and the median observed tag has only four reads. Second, the estimated tags in Run 2 are supported by a narrow evidence budget, with a median of nine usable reads after filtering. Third, compared with Run 3, which has similarly sparse S2 reads but much better convergence, Run 2 covers a smaller antenna-pose area and shows the largest circular phase-residual spread. The root cause is therefore best described as a combination of sparse repeated observations, weaker geometric leverage, and low phase coherence, rather than as a synchronization or parsing failure. This interpretation is consistent with the implementation diagnostics: all 688,073

observations received antenna poses, no reader-port mappings were missing, and only five observations required pose extrapolation.

Figure 4 provides a run-level overview of throughput and convergence behavior. The figure makes clear that the five runs occupy different operating regions: some favor posterior concentration, some favor throughput, and one exhibits near-complete non-convergence.

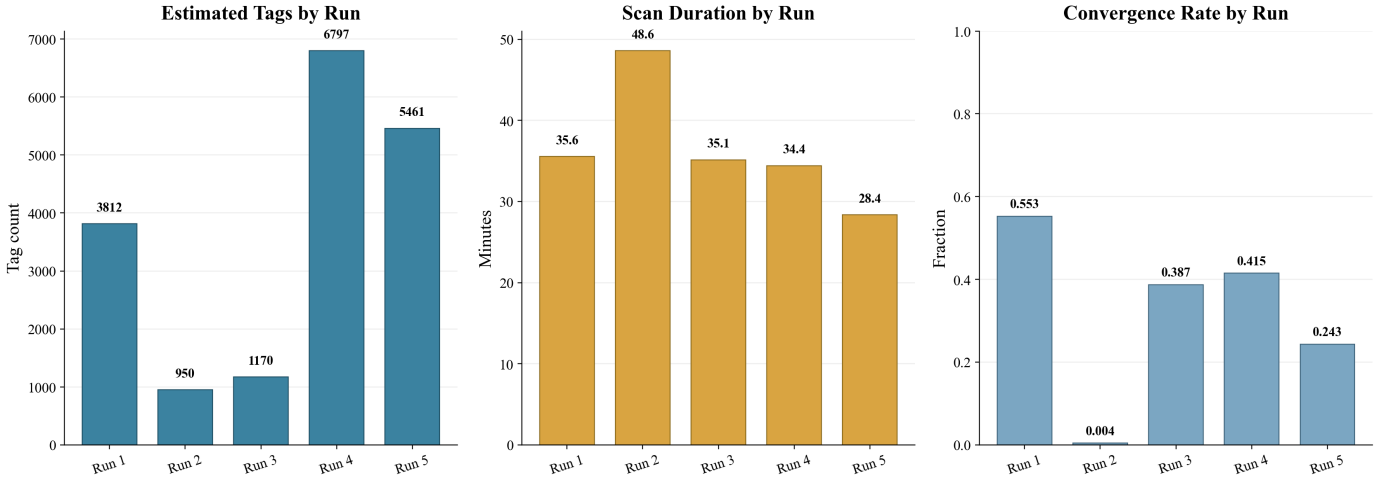


Figure 4. Phase 2 run overview showing estimated tags, scan duration, and convergence rate.

Figure 5 shows the distribution of posterior spread by run. The contrast between Run 1 and Run 2 is especially informative: Run 1 concentrates much more posterior mass into low-spread estimates, whereas Run 2 remains broadly dispersed. This figure captures the central meaning of posterior concentration in the present study: not verified point accuracy, but the extent to which accumulated evidence narrows the posterior over feasible shelf candidates.

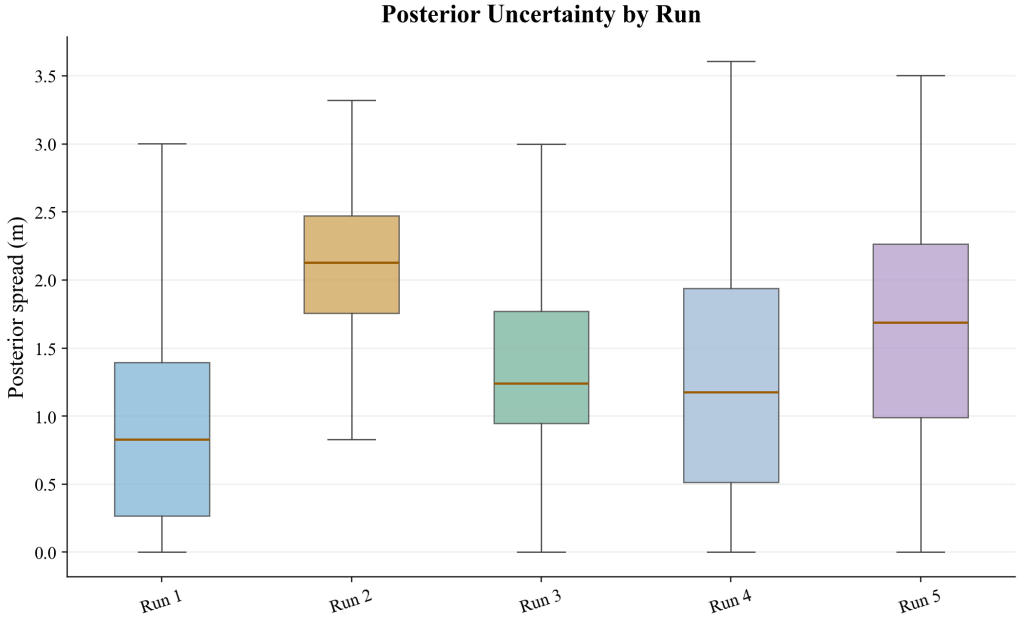


Figure 5. Posterior spread distribution by run.

The trade-off between scan duration and posterior spread in Figure 6 further shows that longer runs are not necessarily better. Run 2 again occupies the least desirable region of this space, while Runs 1, 4, and 5 occupy different but more favorable trade-off positions.

This reinforces the conclusion that the quality of synchronized evidence matters at least as much as total acquisition time.

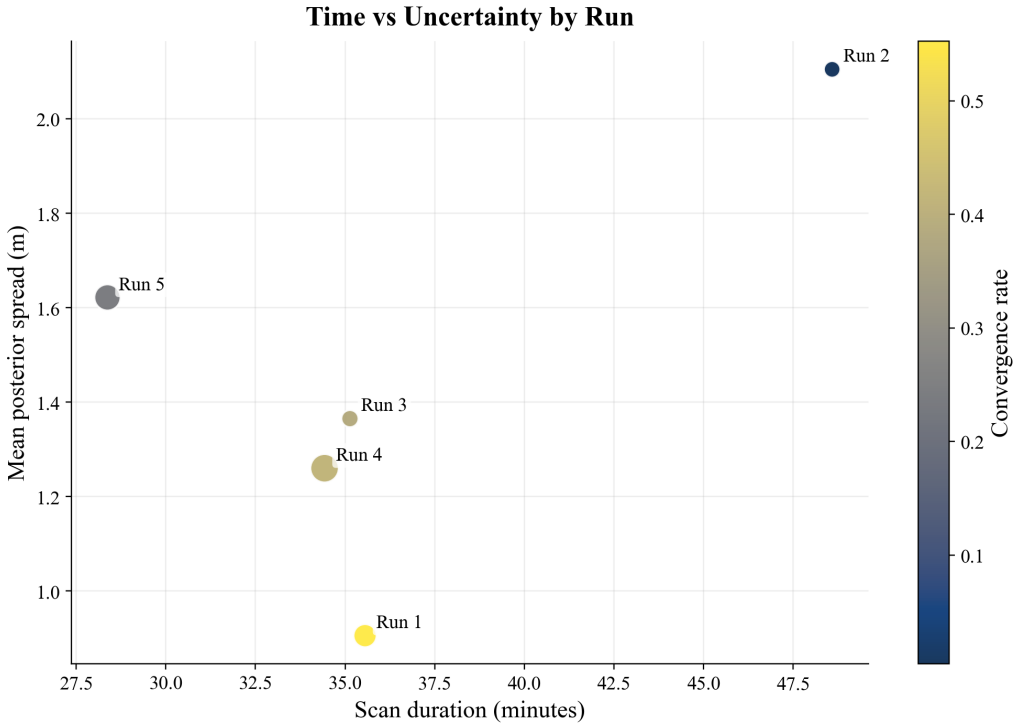


Figure 6. Run-level trade-off between scan duration and mean posterior spread.

To isolate the contribution of phase information, we additionally analyzed a saved robustness subset containing 3445 estimated tags from Run 1. The subset was generated from the phase-weight experiment outputs using identical synchronization, candidate filtering, and scan duration while varying only the phase contribution in the posterior update. Under this design, the setting `fixed_w0` acts as an RSSI-only baseline, whereas the remaining settings introduce progressively stronger phase influence. The results are summarized in Table 5.

Table 5. Phase-weight sensitivity study on a 3445-tag robustness subset from Run 1.

Setting	Mean Posterior Spread (m)	Convergence Rate
`fixed_w0`	1.378	0.068
`fixed_w0.25`	0.908	0.543
`fixed_w0.5`	0.705	0.734
`fixed_w1`	0.522	0.864
`adaptive_w1_win8`	0.674	0.740

The baseline comparison is decisive. Relative to the RSSI-only setting `fixed_w0`, the fully phase-enabled fixed model `fixed_w1` reduces weighted mean posterior spread from 1.378 m to 0.522 m and raises convergence from 0.068 to 0.864. Even intermediate phase weights improve both criteria monotonically. The adaptive strategy remains competitive, but it does not surpass the strongest fixed setting on this subset. The most defensible interpretation is therefore not that adaptive weighting is ineffective, but that the present adaptive reliability model still requires refinement. What is already clear from the saved experiment outputs is that phase information materially improves posterior concentration relative to an RSSI-only baseline in this environment.

The adaptive strategy likely underperforms the best fixed phase weight because its reliability estimate is based only on local circular dispersion of the observed phase sequence. This rule can be overly conservative in Run 1, where phase is globally informative despite local wrapping and motion-induced fluctuations. It also does not model antenna-specific phase offsets, RSSI-dependent phase reliability, candidate-dependent phase residuals, or whether phase changes are consistent with the robot's motion geometry. A stronger adaptive model should therefore weight phase using posterior innovation or residual consistency, per-antenna calibration terms, signal-strength-dependent reliability, and motion-aware phase coherence rather than local phase dispersion alone.

Taken together, the Phase 2 results show that the proposed inference pipeline can produce meaningful posterior concentration under favorable observation regimes, but that its behavior is strongly conditioned by run quality. This is consistent with the uncertainty-aware framing of the paper: the main value of the method lies in distinguishing favorable and unfavorable inference regimes, rather than in overclaiming uniform item-level localization performance.

5.3. Proxy Operational Evaluation Results

Phase 3 evaluates the practical implications of posterior spread by converting uncertain and non-converged estimates into proxy workload and cost terms. The reported values use the base analysis settings saved with the experiment outputs: uncertainty threshold $\tau_u = 1.5\,\text{m}$, convergence threshold $\tau_\sigma = 0.5\,\text{m}$, non-convergence penalty $\lambda = 1.0$, review time of 45 s per flagged case, robot cost of 18 currency units per hour, and human review cost of 12 currency units per hour. These values are not claimed as universal; they define a concrete comparative scenario under which configuration ranking can be interpreted.

Table 6 summarizes the configuration-level trade-off analysis. The reported metrics include mean posterior spread, convergence rate, proxy workload per estimated tag, proxy cost per estimated tag, and mean scan time.

Table 6. Phase 3 configuration-level proxy operational trade-offs.

Configuration	Mean Posterior Spread (m)	Conv. Rate	Proxy Workload/Tag	Proxy Cost/ Tag	Mean Scan Time (s)
Autonomous-S1-P30	1.083	0.484	0.826	0.126	2099.5
Autonomous-S2-P30	1.735	0.196	1.422	0.225	2511.6
Manual-S1-P30	1.622	0.243	1.334	0.202	1702.2

The configuration-level comparison indicates that, under the present proxy assumptions, `autonomous-S1-P30` provides the most favorable overall trade-off between posterior concentration and operational efficiency. Although it is not the fastest configuration, it achieves the smallest mean posterior spread, the highest convergence rate, the lowest proxy workload per estimated tag, and the lowest proxy cost per estimated tag. In contrast, `manual-S1-P30` remains attractive when scan time is the dominant priority, but it incurs a larger unresolved workload and higher cost per estimated tag. The weakest overall profile is observed for `autonomous-S2-P30`, which combines sparse observations, weaker convergence, larger posterior spread, and the largest workload and cost burden among the three configurations.

The sensitivity to the proxy workload thresholds was evaluated with three terminal-threshold scenarios in Table 7. In this post hoc analysis, $N_{\text{uncertain}}$ is recomputed from the final posterior spread using τ_u, while terminal non-convergence is approximated by final posterior spread above τ_σ. This does not re-estimate the first read-to-convergence time for each alternative τ_σ, because the saved outputs do not contain the full per-read

posterior-spread trajectory for every tag. It does; however, test whether the configuration ranking is sensitive to reasonable changes in τ_u, τ_σ, and λ. Across strict, base, and lenient settings, `autonomous-S1-P30` remains the lowest-workload configuration.

Table 7. Sensitivity of proxy workload ranking to uncertainty threshold τ_u, terminal convergence threshold τ_σ, and non-convergence penalty λ.

Scenario	Configuration	τ_u (m)	τ_σ (m)	λ	Workload/Tag
Strict	`autonomous-S1-P30`	1.0	0.40	1.5	1.596
Strict	`manual-S1-P30`	1.0	0.40	1.5	2.104
Strict	`autonomous-S2-P30`	1.0	0.40	1.5	2.330
Base	`autonomous-S1-P30`	1.5	0.50	1.0	1.009
Base	`manual-S1-P30`	1.5	0.50	1.0	1.458
Base	`autonomous-S2-P30`	1.5	0.50	1.0	1.598
Lenient	`autonomous-S1-P30`	2.0	0.75	0.5	0.454
Lenient	`manual-S1-P30`	2.0	0.75	0.5	0.775
Lenient	`autonomous-S2-P30`	2.0	0.75	0.5	0.845

To test whether this ranking is an artifact of one coefficient set, Table 8 reports a cost sensitivity analysis. The low-review scenario uses a lower human review rate and shorter review time, the base scenario matches Table 6, and the high-review scenario increases robot cost, human review cost, review time, and the non-convergence penalty. The absolute cost values change, but the ranking remains stable: `autonomous-S1-P30` has the lowest cost per estimated tag in all three scenarios. This result should be interpreted as robustness of the proxy comparison, not as a claim that the selected coefficients are universal empirical library costs [32,33].

Table 8. Sensitivity of proxy cost per estimated tag under alternative review-cost assumptions.

Scenario	Configuration	Proxy Workload/Tag	Proxy Cost/Tag
Low review	`autonomous-S1-P30`	0.826	0.057
Low review	`autonomous-S2-P30`	1.422	0.107
Low review	`manual-S1-P30`	1.334	0.090
Base	`autonomous-S1-P30`	0.826	0.126
Base	`autonomous-S2-P30`	1.422	0.225
Base	`manual-S1-P30`	1.334	0.202
High review	`autonomous-S1-P30`	1.084	0.364
High review	`autonomous-S2-P30`	1.824	0.623
High review	`manual-S1-P30`	1.712	0.573

These results are important because they show that the most favorable deployment mode is not the one with the shortest scan time or the lowest RSSI variance in isolation. Instead, the strongest operating mode is the one that provides the best joint outcome once evidence density, posterior concentration, and likely follow-up burden are considered together. This is precisely the type of comparison that is needed when direct item-level correctness is unavailable but configuration selection still matters operationally.

Figure 7 presents the Pareto frontier defined over two competing objectives: minimizing scan duration and minimizing mean posterior spread. Runs 5, 4, and 1 form the non-dominated set because none of them is jointly outperformed in both objectives. Specifically, Run 5 provides the shortest scan duration but with higher posterior spread, Run 1

provides the lowest posterior spread at the cost of longer scan time, and Run 4 occupies an intermediate trade-off position. By contrast, Runs 3 and 2 are dominated operating points: each is outperformed by at least one other run in both duration and uncertainty, with Run 2 representing the clearest degraded case. This Pareto analysis is more informative than a single scalar ranking because it exposes the operational trade-off structure directly and allows different libraries to prioritize either speed or certainty according to deployment needs.

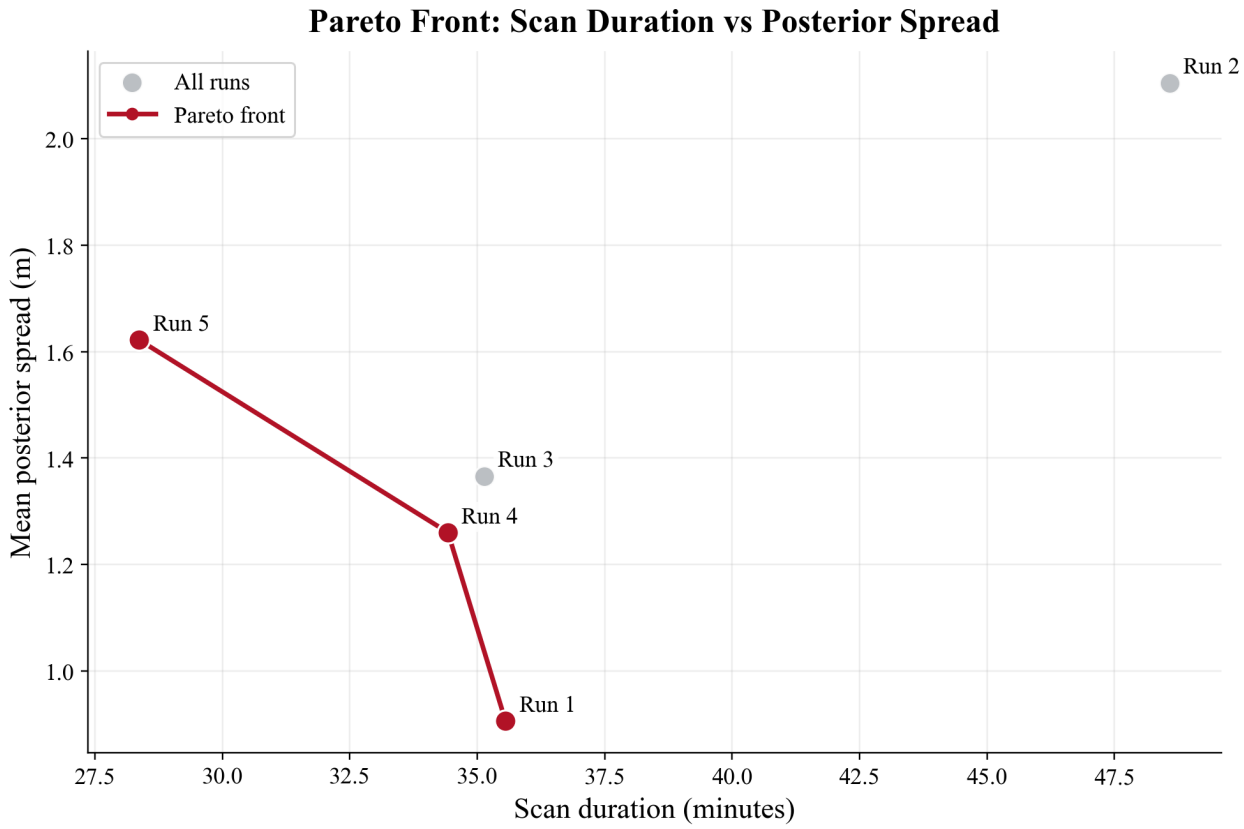


Figure 7. Pareto frontier for scan duration versus mean posterior spread. Runs 5, 4, and 1 are the non-dominated operating points, while Runs 3 and 2 are dominated.

Figure 8 provides a complementary view by decomposing the proxy operational cost by run. The figure makes clear that the total burden is not determined by scan time alone. Runs with weak convergence or broad posterior spread can accumulate greater implied review cost even if they are relatively short. This is precisely the value of the proxy evaluation: in the absence of EPC-linked ground truth, unresolved posterior spread remains operationally meaningful because it corresponds to additional inspection effort.

Overall, the Phase 3 results translate the outputs of Bayesian shelf inference into deployment-oriented terms. They show that uncertainty-aware inference is valuable not only because it produces sharper posteriors but also because it supports rational configuration selection when scan speed, throughput, and likely review burden must be balanced under incomplete supervision.

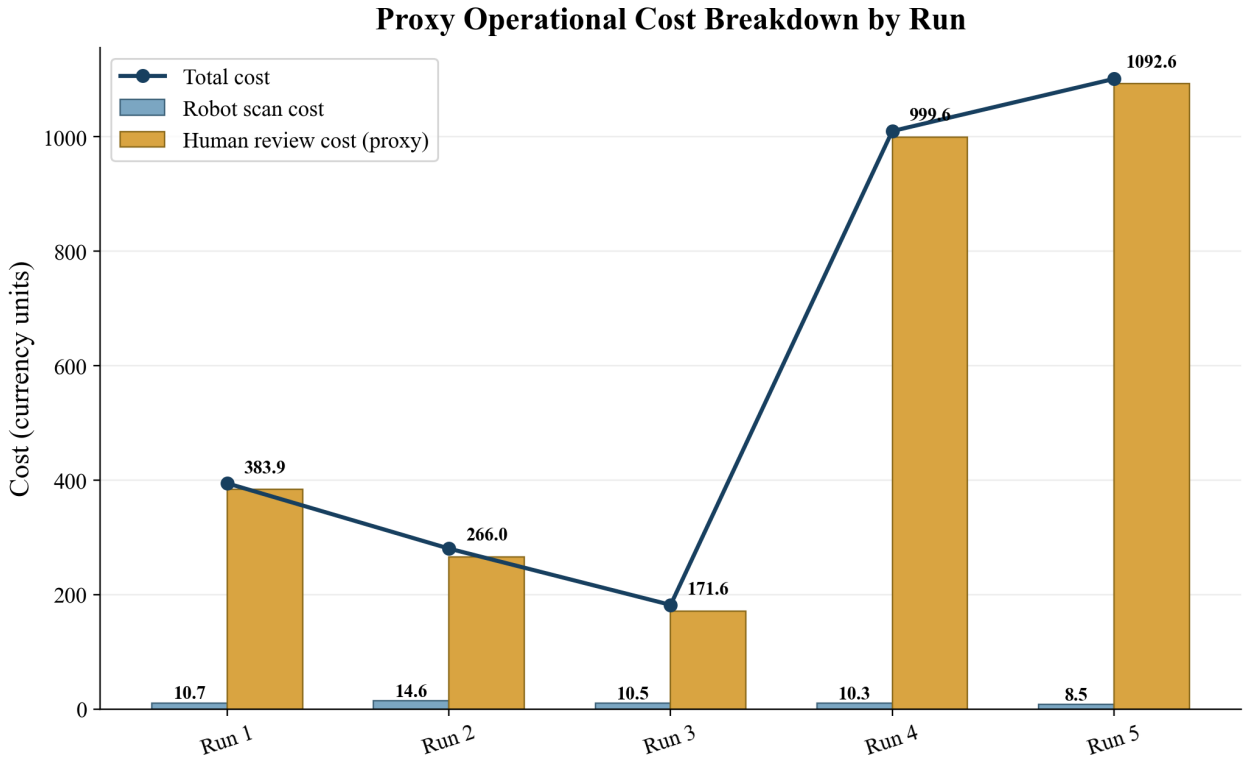


Figure 8. Proxy operational cost breakdown by run, combining scan-time cost and review-workload cost.

6. Discussion, Limitations, and Future Work

6.1. Interpretation of the Main Findings

Mobile RFID inventory quality is governed by a coupled interaction among evidence density, signal variability, antenna-pose geometry, and posterior convergence. These factors do not vary monotonically across operating modes: denser observations are not necessarily cleaner, lower raw RSSI variance does not automatically yield the sharpest posterior concentration, and longer scans do not guarantee better inference. The strongest practical outcome is therefore a decision-analytic result rather than a metric-accuracy result: among the tested modes, `autonomous-S1-P30` provides the most favorable balance between posterior certainty, convergence behavior, and implied review effort under the adopted proxy assumptions.

This framing clarifies the manuscript's validity domain. The paper should be read as an uncertainty-aware analytics study for mobile RFID inventory under incomplete supervision, not as a claim of state-of-the-art item-level localization accuracy. Within that scope, the main strengths are the reproducible preprocessing pipeline, the map-constrained posterior inference procedure, the explicit uncertainty outputs, the RSSI-only baseline, the phase-weight sensitivity analysis, and the operational comparison layer that turns unresolved posterior behavior into decision-support metrics.

6.2. Physical Factors Affecting RFID Inventory Accuracy

The observed run dependence is consistent with known physical limitations of passive UHF RFID in dense library environments. Shelving material affects both attenuation and reflection. Metallic shelf frames, bookends, and nearby fixtures can create strong multipath and shadowing, while wooden or composite shelving typically produces less severe reflection but still changes the local propagation path. Book density and tag placement also matter: tightly packed books reduce tag visibility, change tag orientation relative to the

antenna, and can shield tags behind other items. These effects alter both read probability and RSSI variance, so a high read count is not automatically equivalent to high-quality evidence [3,35].

Environmental obstacles further complicate mobile scanning. A robot may observe the same shelf from slightly different antenna poses across repeated passes, but human traffic, cart placement, shelf-end structures, or aisle geometry can reduce the angular diversity needed for reliable phase-based discrimination. This is particularly relevant to the Run 2 diagnostics in Table 4: sparse repeated reads, a smaller antenna-pose coverage area than Run 3, and high phase-residual dispersion are all plausible symptoms of limited geometric leverage and degraded propagation. The present dataset does not include direct annotations for shelf material, local book packing density, or temporary obstacles, so these effects cannot be isolated experimentally here. They should nevertheless be treated as primary design variables in future data collection.

6.3. Limitations

The main limitations are as follows. First, the public dataset is from 2018 and should not be treated as a comprehensive benchmark for current RFID hardware, antenna designs, or autonomous inventory robots. Its value is that it publicly exposes synchronized RFID observations, robot trajectories, antenna mappings, map data, and shelf coordinates at library scale. Second, the public release does not provide direct EPC-to-item linkage for strict per-tag RMSE evaluation, so the present claims remain comparative and uncertainty-based rather than externally verified against true item identities.

Third, configuration diversity is limited. The released runs vary inventory mode, RFID session, and trajectory, but all use 30 dBm nominal transmit power and the same antenna hardware. Therefore, the paper cannot claim to evaluate alternative power levels, antenna placements, robot velocities, or scanning strategies beyond those already present in the data. Fourth, the saved phase-weight robustness study is restricted to a 3445-tag subset from Run 1. It is sufficient to expose a meaningful RSSI-only versus phase-enabled contrast, but not sufficient to establish a universally optimal phase-weighting strategy. Fifth, the proxy workload and cost model depend on fixed thresholds and scenario coefficients. The sensitivity analysis in Table 8 shows that the ranking is stable across three plausible scenarios, but site-specific empirical cost data would be needed for a true economic claim.

Finally, the number of independent inventory runs is small. The tag-level confidence intervals and bootstrap reliability analysis reported in Section 5.2 quantify stability of posterior-spread and ranking estimates within the available data, but they do not replace a 30–50 run physical experimental campaign. The conclusions should therefore be interpreted as evidence from a public, limited-run dataset with strong tag-level resampling stability rather than as a statistically exhaustive deployment trial.

6.4. Future Work

The most valuable next step is a contemporary validation campaign with explicit EPC-to-item linkage and a factorial configuration design. Such a campaign should vary transmit power, RFID session, antenna placement, robot velocity, aisle trajectory, and scanning strategy while recording shelf material, book density, and temporary obstacles. This would allow direct estimation of item-level RMSE, shelf-assignment accuracy, and configuration interactions that cannot be recovered from the current public data.

Additional methodological extensions are also clear. The Bayesian model should be evaluated against supervised deep learning, SLAM-RFID, synthetic aperture, and optimization-based baselines when matching labels are available. The likelihood model should be expanded to include calibrated antenna radiation patterns, per-antenna phase

offsets, and material-aware attenuation terms. The operational layer should be calibrated using site-specific staff time, robot operating cost, and audit records from real library operations. Together, these steps would strengthen external validity while preserving the uncertainty-aware, deployment-oriented perspective developed here.

7. Conclusions

This paper presented an uncertainty-aware analytics framework for robot-assisted RFID library inventory under incomplete supervision. Using a public library dataset, we showed that operating configuration materially changes both the density and variability of RFID observations. We then introduced a map-constrained Bayesian shelf-inference pipeline that synchronizes RFID reads with robot trajectory and antenna geometry to produce shelf-level posterior distributions with explicit uncertainty. Finally, we translated posterior spread and non-convergence into proxy review workload and cost, enabling deployment-oriented comparison when direct EPC-to-item ground truth is unavailable.

The main conclusion is that deployment quality cannot be inferred from scan speed or raw signal stability alone. Instead, the most useful operating mode is the one that yields the best joint balance among evidence density, posterior concentration, convergence, and downstream review burden. In the present dataset, `autonomous-S1-P30` provides that balance most consistently under the adopted proxy assumptions and remains the best-ranked configuration in the threshold and cost-sensitivity analyses. More broadly, the paper argues that mobile RFID inventory should be evaluated as an uncertainty-aware operational analytics problem when perfect supervision is unavailable. The claim is intentionally bounded: contemporary item-level benchmarking, broader configuration testing, and empirical cost calibration require new data with EPC-to-item linkage and controlled experimental variation.

Author Contributions: Conceptualization, S.M. and M.R.; Methodology, S.M. and M.R.; Software, S.M.; Validation, S.M., M.R. and H.B.; Formal analysis, S.M. and H.B.; Investigation, H.B.; Resources, M.R. and H.B.; Data curation, H.B.; Writing—original draft, S.M.; Writing—review & editing, S.M. and H.B.; Supervision, M.R.; Project administration, M.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are openly available in [RFID Location dataset] at [https://zenodo.org/records/1215660] (accessed on 3 May 2026).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Yusuf, M.; Akintunde, A.; Habeeb, S.; Quadri, A. Radio Frequency Identification (RFID) based library management system. *Int. J. Technol. Syst.* **2023**, *8*, 21–35. [CrossRef]
2. Suhaimi, M.M.; Mohamed, Z.; Khusaini, N.S. Effectiveness of RFID smart library management system. *J. Mech. Eng. (JMechE)* **2023**, *12*, 133–152. [CrossRef]
3. Abcouwer, K.; van Loon, E. Library inventory using a RFID wand: Contribution of tag and book specific factors on the read rate. *Libr. Hi Tech* **2021**, *39*, 368–379.
4. Hamad, F.; Al-Fadel, M.; Fakhouri, H. The provision of smart service at academic libraries and associated challenges. *J. Librariansh. Inf. Sci.* **2023**, *55*, 960–971.
5. Kleniatis, A.; Dimitriou, A.; Bletsas, A. Device-Free Localization of Multiple Humans with Passive RFID and Joint RSSI-Phase Techniques. In *Proceedings of the 2024 IEEE International Conference on RFID (RFID)*; IEEE: New York, NY, USA, 2024. [CrossRef]
6. Martino, G.; Bevacqua, M.T.; Catarinucci, L.; Merenda, M. Towards Multi-Frequency and Multi-View Localization via UHF-RFID Passive Tags. In *Proceedings of the 2024 IEEE International Conference on RFID Technology and Applications (RFID-TA)*; IEEE: New York, NY, USA, 2024. [CrossRef]

7. Wang, S.; Wang, S.; Feng, Y.; Huang, W.; Jiang, S.; Zhang, Y. RP-Fusion: Robust RFID Indoor Localization via Fusion RSSI and Phase Fingerprint. In *Proceedings of the 2024 27th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*; IEEE: New York, NY, USA, 2024. [CrossRef]
8. Motroni, A.; Tavanti, E.; Nepa, P. Mobile Robot Trajectory Reconstruction with UHF-RFID Synthetic Aperture Localization. In *Proceedings of the 2025 IEEE International Conference on RFID Technology and Applications (RFID-TA)*; IEEE: New York, NY, USA, 2025. [CrossRef]
9. Alajami, A.A.; Perez, F.; Pous, R. The Design of an RFID-Based Inventory Hybrid Robot for Large Warehouses. In *Proceedings of the 2024 9th International Conference on Control and Robotics Engineering (ICCRE)*; IEEE: New York, NY, USA, 2024. [CrossRef]
10. DiBattista, M.A.; Frericks, J.; Garcia, C.I. SCOUT: An Autonomous UHF RFID-Equipped Robot Dog for Flexible Inventory Monitoring. *Manuf. Lett.* **2025**, *44*, 1525–1532. [CrossRef]
11. Pous, R.; Alajami, A.; Hernandez, L. Populating the Digital Twin of a Retail Store Using an RFID Autonomous Mobile Robot. In *Proceedings of the 2025 10th International Conference on Smart and Sustainable Technologies (SpliTech)*; IEEE: New York, NY, USA, 2025. [CrossRef]
12. Zafari, F.; Papapanagiotou, I.; Hacker, T.J. A novel Bayesian filtering based algorithm for RSSI-based indoor localization. In *Proceedings of the 2018 IEEE International Conference on Communications (ICC)*; IEEE: New York, NY, USA, 2018; pp. 1–7.
13. Shangguan, L.; Yang, Z.; Liu, A.X.; Zhou, Z.; Liu, Y. Relative localization of {RFID} tags using {Spatial-Temporal} phase profiling. In Proceedings of the 12th USENIX Symposium on Networked Systems Design and Implementation (NSDI 15), Oakland, CA, USA, 4–6 May 2015; pp. 251–263.
14. Qiu, L.; Huang, Z.; Zhang, S.; Jing, C.; Li, H.; Li, S. Multifrequency phase difference of arrival range measurement: Principle, implementation, and evaluation. *Int. J. Distrib. Sens. Netw.* **2015**, *11*, 715307. [CrossRef]
15. Magnago, V.; Palopoli, L.; Buffi, A.; Tellini, B.; Motroni, A.; Nepa, P.; Macii, D.; Fontanelli, D. Ranging-free UHF-RFID robot positioning through phase measurements of passive tags. *IEEE Trans. Instrum. Meas.* **2019**, *69*, 2408–2418.
16. Fu, Y.; Wang, C.; Liu, R.; Liang, G.; Zhang, H.; Ur Rehman, S. Moving object localization based on UHF RFID phase and laser clustering. *Sensors* **2018**, *18*, 825. [CrossRef] [PubMed]
17. Morenza-Cinos, M.; Casamayor-Pujol, V. *RFID Location Dataset;* Zenodo: Geneva, Switzerland, 2018. [CrossRef]
18. Gareis, M.; Parr, A.; Trabert, J.; Mehner, T.; Vossiek, M.; Carlowitz, C. Stocktaking robots, automatic inventory, and 3D product maps: The smart warehouse enabled by UHF-RFID synthetic aperture localization techniques. *IEEE Microw. Mag.* **2021**, *22*, 57–68. [CrossRef]
19. Motroni, A.; Bernardini, F.; Vaiani, S.; Buffi, A.; Nepa, P. Performance assessment of a UHF-RFID robotic inventory system for industry 4.0. In *Proceedings of the 2022 16th European Conference on Antennas and Propagation (EuCAP)*; IEEE: New York, NY, USA, 2022; pp. 1–5.
20. Gastón, B.; Casamayor-Pujol, V.; López-Soriano, S.; Pous, R. A metric for assessing, comparing, and predicting the performance of autonomous RFID-based inventory robots for retail. *IEEE Trans. Ind. Electron.* **2021**, *69*, 10354–10362. [CrossRef]
21. Lopez-Soriano, S. Plug-and-Play Inventory Robots: Autonomous Itinerary Planning through Autonomous Waypoint Generation. *IEEE Internet Things J.* **2023**, *11*, 1711–1718.
22. Zhang, J.; Liu, X.; Gu, T.; Zhang, B.; Liu, D.; Liu, Z.; Li, K. An RFID and computer vision fusion system for book inventory using mobile robot. In *Proceedings of the IEEE INFOCOM 2022—IEEE Conference on Computer Communications*; IEEE: New York, NY, USA, 2022; pp. 1239–1248.
23. Wang, H.; Yang, Y. Design of Intelligent Inventory Robot Control System and Its Application Practice in Library Book Inventory. In *Proceedings of the 5th International Conference on Computer Information and Big Data Applications*; Association for Computing Machinery: New York, NY, USA, 2024. [CrossRef]
24. Peng, X. Smart Library RFID Security Algorithm Based on Intelligent Perception System. *Procedia Comput. Sci.* **2024**, *247*, 841–848. [CrossRef]
25. Abdo, K.W. Radio Frequency Identification (RFID) Implementation in an IoT Smart Library. *J. Inf. Syst. Eng. Manag.* **2024**, *9*, 28480. Available online: https://www.jisem-journal.com/download/radio-frequency-identification-rfid-implementation-in-an-iot-smart-library-14925.pdf (accessed on 3 May 2026). [CrossRef]
26. Shirazi, A.B.; Fraanje, R.; Coenen, J. Robust Passive UHF RFID Tag Localisation by Intersecting RSSI Indexed Antenna Sensitivity Regions. *Procedia Comput. Sci.* **2026**, *277*, 1238–1247. [CrossRef]
27. Lv, M.; Wang, Z.; Huang, Y. DR-ILS: High-Precision RFID Indoor Localization via Differential RSSI and Iterated Local Search. In *Proceedings of the 2026 International Conference on Communication Networks and Machine Learning (CNML)*; IEEE: New York, NY, USA, 2026. [CrossRef]
28. Khudoyberdiev, A.; Ryoo, J. MirrorVision: Light-Weight Floor Detection System for an Autonomous Robot in a Crowded Elevator. *J. Image Graph.* **2024**, *12*, 1–9. [CrossRef]

29. Wang, X.; Inserra, D.; Wen, G.; Nepa, P. A Synthetic Aperture Radar UHF RFID Localization Method Based on Phase Jumps. In *Proceedings of the 2024 IEEE 12th Asia-Pacific Conference on Antennas and Propagation (APCAP)*; IEEE: New York, NY, USA, 2024. [CrossRef]
30. Khudoyberdiev, A.; Kim, H.Y.; Ryoo, J. PLUS-CODE+: Zero-installment Rover Indoor Localization. *IEEE Sens. J.* **2025**, *25*, 23088–23104. [CrossRef]
31. Casamayor-Pujol, V.; Gastón, B.; López-Soriano, S.; Alajami, A.A.; Pous, R. A simple solution to locate groups of items in large retail stores using an RFID robot. *IEEE Trans. Ind. Inform.* **2021**, *18*, 2021. [CrossRef]
32. Bandara, I.; Simpson, O.; Sun, Y. Optimizing Efficiency Using a Low-Cost RFID-Based Inventory Management System. In *Proceedings of the 2024 International Wireless Communications and Mobile Computing (IWCMC)*; IEEE: New York, NY, USA, 2024. [CrossRef]
33. Deepika, M.; Vinoline, I.A. Incorporation of RFID Technology in the Inventory Management System to Optimize the Inventory Cost. *Indian J. Sci. Technol.* **2024**, *17*, 4819–4827. [CrossRef]
34. Bousselmi, S.; Gannouni, M.; Ouni, K. IoT Application for Smart Inventory Management System Based on RFID. In *Proceedings of the 2024 IEEE International Multi-Conference on Smart Systems and Green Process (IMC-SSGP)*; IEEE: New York, NY, USA, 2024. [CrossRef]
35. Jin, M. Research on Challenges and Solutions for Optimizing Read Rate in RFID Systems. *Appl. Comput. Eng.* **2025**, *117*, 120–125. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.