

Nonparametric data recovery in IoT sensor networks based on a priori density[★]

Viacheslav Kovtun ^{1,2,*,†}

¹ Vinnytsia National Technical University, Khmelnytske shose, 95, Vinnytsia, 21021, Ukraine

² Institute of Theoretical and Applied Informatics Polish Academy of Sciences, Bałtycka Str., 5, Gliwice, 44-100, Poland

Abstract

Reliable recovery of sensor information is critically important for the continuous operation of distributed IoT systems under conditions of data loss, noise distortion, and structural uncertainty. This paper proposes a combined approach to sensor data reconstruction that integrates latent state smoothing via recursive filtering with local nonparametric posterior density estimation using kernel functions. In contrast to traditional filtering and forecasting schemes, the developed model performs full interpolation recovery by incorporating both past and future measurements in time. The mathematical structure of the method is theoretically justified, and its properties are examined under partial observability. Numerical modelling results confirm the high accuracy and robustness of the proposed approach under up to 40% observation loss and reduced signal-to-noise ratio, corresponding to typically adverse conditions in LPWAN environments, particularly LoRaWAN. Compared to Kalman filtering and baseline nonparametric regression methods, the proposed model achieved a 20–30% improvement in interpolation accuracy. The practical value of the study lies in the feasibility of implementing the approach on resource-constrained devices without prior training, which opens up prospects for its application in smart monitoring systems, energy-efficient management, agroecological surveillance, and environmental protection.

Keywords

sensor data recovery; Internet of Things systems; time series interpolation; posterior estimation; Bayesian filtering and smoothing; nonparametric density estimation; kernel regression; structural model uncertainty

1. Introduction

Over the past decade, the Internet of Things (IoT) has evolved into a fundamental technological platform enabling real-time data collection, processing, and analysis across a wide range of applications. For example, in smart cities, sensor networks are employed to monitor air quality, noise levels, traffic intensity, or to detect emergencies in water supply systems [1, 2]. In agriculture, they are used to measure soil moisture, air temperature, wind speed, and other parameters influencing agrotechnological decisions [3, 4]. In healthcare systems, sensor devices are applied for remote patient monitoring — for instance, tracking glucose levels, heart rate, or blood oxygen saturation [5, 6]. Industrial IoT systems use sensors to monitor the technical condition of equipment, detecting vibrations, overheating, or leaks [7, 8].

Despite their widespread application, sensor networks face significant technical limitations: data transmission may be disrupted due to unstable connectivity, packet loss, interference in radio channels, limited sensor energy capacity, or processing node overload. As a result, real-world data are often incomplete, exhibiting random losses, gaps in time series, or distortions caused by noise [9]. For instance, in smart lighting systems or emergency response frameworks, delays in

ICyberPhysS'25: 2nd International Workshop on Intelligent & CyberPhysical Systems, July 04, 2025, Khmelnytskyi, Ukraine

^{1*} Corresponding author.

✉ kovtun_v_v@vntu.edu.ua, vkovtun@iitis.pl (V. Kovtun)

 0000-0002-7624-7072 (V. Kovtun)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

transmitting sensor events may trigger incorrect actuator responses. In medical systems, inaccuracies or loss of vital sign readings may pose a risk to patients. In industrial contexts, misinterpretation of data due to signal distortion can lead to equipment downtime or failure.

Thus, the tasks of accurate and robust recovery of lost or noise-corrupted data are not only of theoretical importance but also critically necessary for the reliable operation of practical IoT systems. In this context, the need for adaptive mathematical models capable of operating with incomplete information while ensuring high reconstruction accuracy is particularly urgent.

In the current scientific discourse concerning the reconstruction of sensor information under conditions of partial availability and high noise levels, several conceptual approaches have emerged, each demonstrating a distinct domain of application, degree of mathematical justification, and practical effectiveness. Among the most extensively studied are analytical, nonparametric, and intelligent methods. However, in addition to these, hybrid, heuristic, sparse-reconstruction, multichannel statistical, and reinforcement-based methods are increasingly being employed and merit particular attention when analysing relevant analogues.

Analytical methods [10, 11], primarily represented by filtering and smoothing approaches (such as Kalman filtering, particle filters, and RTS smoothing), assume the presence of a well-defined system dynamics model and known statistical properties of the noise in advance. These methods are mathematically grounded and perform well in linear or quasi-linear systems with normally distributed disturbances. However, in IoT sensor networks, such assumptions are rarely fully satisfied. The presence of missing data, irregular sampling rates, correlated noise, and the lack of a complete a priori model render these methods sensitive to perturbations, ill-suited to complex temporal structures, and inadequate for interpolation-based recovery, which requires incorporating both past and future measurements [12].

An alternative is offered by nonparametric approaches [13, 14], which rely on local approximation or empirical distribution estimation without a fixed model structure. Kernel regression, locally weighted smoothing (LOESS), nearest neighbour methods, and empirical Bayesian procedures exhibit considerable flexibility and allow signal recovery even in the absence of information about the underlying generative mechanism. However, these methods are limited in terms of computational efficiency, do not scale well to high-dimensional settings, are sensitive to window width selection, and fail to capture the latent temporal structure of the data. Moreover, they lack a formal mechanism for incorporating the context of future observations, which is crucial in interpolation tasks rather than forecasting.

In recent years, intelligent approaches [15–17], particularly deep learning models, have gained significant traction. Recurrent neural networks (RNNs), long short-term memory architectures (LSTMs), autoencoders, transformers, and graph neural networks are employed to model complex nonlinear relationships between observations and their temporal context. Their key advantage lies in the ability to automatically learn data representations and reconstruct missing values without an explicitly defined model. At the same time, these methods exhibit notable limitations: they require large training datasets, which are often unavailable in sensor networks; they have high computational complexity, making them unsuitable for energy-constrained devices; and they lack formal interpretability, which complicates verification of reconstruction reliability. In most cases, such architectures are designed for extrapolation (i.e. forecasting future values) and do not support accurate reconstruction of missing segments within a trajectory, which requires bidirectional informational context.

In response to the limitations of each of the aforementioned classes of methods, the literature has seen the emergence of hybrid approaches [18, 19] that combine the formal structure of analytical models with the flexibility of nonparametric or intelligent components. Examples include the integration of forward filtering with local smoothing or the combination of Bayesian models with autoencoders. Such systems allow for the preservation of interpretability and a controlled structure while enabling flexible adaptation to changes in the statistical properties of the data. However, the complexity of implementation, the need to align heterogeneous components,

and the lack of sufficient theoretical justification currently hinder the widespread adoption of such solutions.

Heuristic methods [20, 21] are also applied in practical systems, relying on domain-specific knowledge or sets of rules. For instance, in reconstructing a temperature profile, one may assume a limited rate of change in readings or approximate missing values using linear functions based on neighbouring values. Such methods are simple, energy-efficient, and scale well, but they lack generalisability and do not guarantee accuracy under complex or atypical system behaviour.

Special attention should be given to methods based on the concept of sparse representation—the so-called sparse reconstruction models [22]. These assume that a signal can be expressed as a sparse combination of basis vectors in a specific space (for example, wavelet or Fourier). This enables effective recovery of missing values even under a high level of data loss. However, such models typically require a fixed transformation structure, which may not adequately capture the dynamics of real-world sensor processes.

In many real-world applications, multichannel statistical methods [11, 23] also prove effective, as they utilise correlation information between multiple sensor streams. Models of interdependent processes, multivariate Bayesian estimations, or multichannel PCA enable the reconstruction of data in one channel based on the values from others. This is particularly useful in the context of heterogeneous sensor networks; however, such models are sensitive to channel mismatches, signal synchronisation issues, and require full observability in at least some of the sources.

Finally, a promising direction is reconstruction methods based on reinforcement learning [24, 25], where an agent learns to develop an optimal reconstruction strategy by receiving rewards that depend on the accuracy of recovery or its impact on subsequent control. These methods are particularly effective in adaptive or dynamically changing environments; however, their training requires multi-episode simulation, and the interpretability of the results remains an open challenge.

In summary, despite methodological diversity and numerous attempts to adapt existing approaches to sensor environments, no universal solution currently exists that ensures a combination of accuracy, adaptability, computational efficiency, and analytical controllability under conditions of loss, noise, and limited informational completeness. Analytical methods are overly rigid in their assumptions, nonparametric methods are excessively local and resource-intensive, and intelligent methods are complex, opaque, and weakly interpretable. Even hybrid or heuristic approaches, while mitigating the imbalance between flexibility and formality, largely remain narrowly specialised or insufficiently theoretically substantiated. In this context, there is a clear need to develop a new methodology that combines the advantages of parametric models (such as stability, interpretability, and smoothing efficiency) with the flexibility and local adaptability of nonparametric estimations. This is precisely the approach implemented in the present study, which proposes a sensor data reconstruction model based on posterior interpolation density, combining the formal structure of the exponential family of distributions with the capacity for local kernel correction. This enables not only accurate recovery of missing information under partial availability but also stable system operation amid changing data structures and constrained sensor network resources. Accordingly, the proposed research fills a critical methodological gap in the field of sensor signal reconstruction, offering a formally grounded, practically feasible, and flexible model tailored to the specific requirements of modern IoT systems.

The object of this study is the process of sensor information reconstruction under conditions of partial data availability, noise distortion, and structural uncertainty in the dynamics of distributed IoT systems.

The subject of this study is the set of theoretical foundations, mathematical models, and methods for approximating the posterior density that enable continuous reconstruction of sensor data under uncertainty, based on the integration of filtering and smoothing procedures with nonparametric kernel estimation.

The aim of this study is to develop and justify a combined method for sensor data reconstruction in distributed IoT systems by integrating an analytical model of posterior density

with nonparametric procedures of local kernel smoothing, in order to ensure accurate and robust information recovery under conditions of partial availability and structural uncertainty.

The scientific novelty of this study lies in the development of a formalised combined approach to sensor data reconstruction in distributed IoT systems, which for the first time integrates a recursive filtering and smoothing model for posterior density estimation with local nonparametric kernel smoothing procedures. The proposed methodology enables interpolation-based data recovery at points with partial or zero observability, does not require a complete description of the latent process dynamics, adapts to the noise level, and demonstrates improved accuracy compared to traditional filtering and empirical approaches.

The practical value of this study lies in the capability for continuous and resource-efficient recovery of sensor information in real time under conditions of partial loss, noise, and structural uncertainty—features typical of modern IoT systems. The proposed combined method does not require a complete model of system dynamics, easily adapts to various sensor network topologies, can be implemented on energy-constrained devices, and ensures improved interpolation accuracy without the use of training data, making it suitable for deployment in smart monitoring, automated control, energy-efficient environments, and other critical applications.

2. Models and methods

2.1. Fundamental assumptions

In the task of recovering observed data in sensor networks operating within a dynamic and probabilistically determined environment, it is essential to formalise interpolation as a procedure for estimating the values of system state vectors $Y_l \in \mathbb{R}^n$, where n denotes the dimensionality of the hidden state space, based on the available sequence of sensor observations $x_1^m = (x_1, \dots, x_m)^T \in \mathbb{R}^{m \times k}$, in which m represents the number of discrete time steps and k is the dimensionality of each vector observation space. The positional index l indicates the moment $1 \leq l \leq m$ at which recovery is performed. According to the Bayesian approach, optimal interpolation can be carried out by approximating the conditional expectation of the function $R(Y_l)$, which represents the recoverable characteristic of the system state, given the available information x_1^m . This is mathematically expressed by formula

$$E(R(Y_l)|x_1^m) = \int_{\mathbb{R}^n} R(y_l) p_l(y_l|x_1^m) dy_l, \quad (1)$$

where $p_l(y_l|x_1^m)$ denotes the conditional a priori probability density Y_l given the known data from the sensor sample x_1^m , which, in the context of real-time recovery requirements, is naturally interpreted as the interpolated posterior density.

In the context of data recovery in IoT sensor networks, where time series measurements are partially lost or noise-corrupted, there arises a need to construct the interpolated posterior density $p_l(y_l|x_1^m)$, which accounts for both historical data and future measurements relative to the recovery moment l . One approach to constructing this density is to represent it as a combination of observation densities and conditional posterior distributions. This density is defined as

$$p_l(y_l|x_1^m) = \frac{\varphi(x_1^l) \varphi(x_1^m) \pi_l(y_l|x_1^l) \tilde{\pi}_l(y_l|x_1^m)}{\varphi(x_1^m) \varphi(y_l, x_l)}, \quad (2)$$

where $\varphi(x'_1)$ denotes the density of sensor observations from the beginning up to moment l , $\varphi(x_l^m)$ is the density of sensor observations from moment l to the end, $\pi_l(y_l|x'_1)$ is the conditional density of the state y_l given prior data, $\tilde{\pi}_l(y_l|x_l^m)$ is the backward density of the same state given subsequent observations, and $\varphi(y_l, x_l)$ is the joint density of the state and corresponding observation at moment l . The representation in (2) enables formalisation of the interpolation task under incomplete or noisy information, characteristic of sensor-based IoT systems.

In the context of defining the joint density $\varphi(y_l, x_l)$ in the denominator of expression (2), a key stage involves the analytical formalisation of the interpolation procedure. In sensor networks modelled as Markov dynamic systems with correlated measurements, this density can be represented as a convolution over the space of previous states and observations, accounting for the historical context of information transmission. This allows for the latent dependency structure within the time series of sensor data to be incorporated. The corresponding expression is given in

$$\varphi(y_l, x_l) = \int_{y_{l-1}} q(y_l|y_{l-1}) \int_{x_{l-1}} \varphi(x_l|x_{l-1}, y_l) \varphi(y_{l-1}, x_{l-1}) dx_{l-1} dy_{l-1}, \quad (3)$$

where $q(y_l|y_{l-1})$ denotes the state transition density in the Markov model, $\varphi(x_l|x_{l-1}, y_l)$ is the conditional density of a sensor observation given the state y_l and the previous measurement x_{l-1} , and $\varphi(y_{l-1}, x_{l-1})$ is the joint density at moment $l-1$. Accordingly, formula (3) enables the computation of the normalisation factor included in expression (2), thereby allowing accurate data recovery in sensor-based IoT systems.

In the process of formalising the task of recovering lost or corrupted data in IoT sensor networks, it is appropriate to define an analytical model that unifies computations related to the interpolated posterior density (see formula (2)) and the normalisation factor (see formula (3)). In this study, it is assumed that the conditional density of observations given the latent state belongs to the exponential family of distributions. This allows the density $\varphi(y_l, x_l)$ to be expressed in the following form

$$\varphi(y_l, x_l) = \tilde{K}(y_m) b(x_m) \exp\{T^T(x_m) R(y_m)\}, \quad (4)$$

where $\tilde{K}(y_m)$ is the normalising coefficient, $b(x_m)$ is the base function of the observation, $T^T(x_m)$ is the sufficient statistic, and $R(y_m)$ is a parametric function of the latent state. Assumption (4) is adopted as the foundation for the subsequent analytical analysis of sensor information reconstruction models.

2.2. Analytical interpolation model for the reconstruction of lost data in sensor-based IoT systems

In addressing a range of applied problems arising during the operation of sensor networks (for example, reconstructing missing temperature or pressure readings, restoring the trajectory of a moving object from incomplete GPS samples, or adaptive signal smoothing in cases of packet loss), it is necessary not only to compute the conditional expectation (see formula (1)) but also to construct the full interpolated posterior density of the state $p_l(y_l|x_l^m)$. This density contains all the required information about the unknown state y_l , and enables not only point estimation but also statistical decision-making based on confidence intervals, quantile assessments, or Bayesian

risks. To simplify the analytical notation, probability density functions of the form $\varphi(x)$ will henceforth be presented without explicitly specifying the argument, unless this leads to ambiguity. The argument will be indicated only where necessary for correctness.

To derive the equation describing the optimal interpolation estimate under conditions of incomplete sensor data, it is necessary to rely on the structure of the joint density for the entire trajectory of latent states y_1^m and corresponding observations x_1^m , which can be represented in a recursive form. If the sensor network model is described by conditional densities of the form $\phi(y_l, x_l | y_{l-1}, x_{l-1})$, the full density $\varphi(y_1^m, x_1^m)$ is expressed as a product of the form

$$\varphi(y_1^m, x_1^m) = \varphi(y_1, x_1) \prod_{l=2}^m \phi(y_l, x_l | y_{l-1}, x_{l-1}), \quad (5)$$

corresponding to the extended structure of a generalised Markov model for the dynamics of latent states in a sensor environment. Based on the representation in (5), a formal derivation of the expression for the interpolated posterior density $p_l(y_l | x_1^m)$ can be carried out, defined as the ratio of the joint density to the marginal density of the sensor sample alone. By sequentially applying the chain rule decomposition with consideration of conditional independence, we obtain

$$p_l(y_l | x_1^m) = \frac{\varphi(y_l, x_1^m)}{\varphi(x_1^m)} = \frac{\varphi(y_l, x_1^l, x_{l+1}^m)}{\varphi(x_1^m)} = \frac{\varphi(x_1^l) \pi_l(y_l | x_1^l) \varphi(x_{l+1}^m | y_l, x_l)}{\varphi(x_1^m)}. \quad (6)$$

In the subsequent analysis of equation (6), which describes the interpolated posterior density in a sensor network, it becomes necessary to detail the product $\varphi(x_1^l) \pi_l(y_l | x_1^l)$. This product can be decomposed using the recursive structure of the dependency model between latent states and sensor observations. Given that previous realisations of observations and states are known, we have $\varphi(x_1^l) \pi_l(y_l | x_1^l) = \int \varphi(y_1^{l-2}, x_1^{l-2}, y_{l-1}, x_{l-1}) \phi(y_l, x_l | y_{l-1}, x_{l-1}) dy_{l-1}$.

To further simplify the computations, let us assume that the joint distribution density $\varphi(x_1^l, y_l)$ can be factorised into the conditional density $\varphi(x_l | x_{l-1}, y_l)$ and the Markov transition density $q(y_l | y_{l-1})$. In this case, taking into account the properties of conditional independence, we obtain

$$\begin{aligned} \varphi(x_1^l) \pi_l(y_l | x_1^l) &= \varphi(x_l | x_{l-1}, y_l) \int q(y_l | y_{l-1}) \varphi(y_{l-1}, x_1^{l-1}) dy_{l-1} = \\ &= \varphi(x_1^{l-1}) \varphi(y_l | x_1^{l-1}) \varphi(x_l | x_{l-1}, y_l), \end{aligned} \quad (7)$$

where the factor $\varphi(y_l | x_1^{l-1})$ represents the updated state estimate based on the observation history up to moment l , and the dependence on x_l in the right-hand side of equation (7) reflects the contribution of the current measurement to the formation of the predicted density.

Let us consider the factor $\varphi(x_{l+1}^m | y_l, x_l)$ in the numerator of formula (6), which corresponds to the probabilistic description of the sensor network's behaviour after moment l . This factor contains information about future measurements, conditional on the current system state and observation. It can be expressed as an integral over densities that characterise transition and observation probabilities at subsequent time steps. To achieve this, we employ the structure of the generalised model and obtain

$$\begin{aligned}
\varphi(x_{l+1}^m | y_l, x_l) &= \int_{\mathbb{R}^n} \varphi(x_{l+2}^m | y_{l+1}, x_{l+1}) \phi(y_{l+1}, x_{l+1} | y_l, x_l) dy_{l+1} = \\
&= \int_{\mathbb{R}^n} \varphi(x_{l+2}^m, y_{l+1}, x_{l+1}) \frac{\varphi(x_{l+1} | x_l, y_{l+1}) q(y_{l+1} | y_l)}{\varphi(x_{l+1} | y_{l+1}) q(y_{l+1})}.
\end{aligned} \tag{8}$$

This approach enables the dependence on future measurements to be expressed as integration over possible states y_{l+1} , thereby formalising the mechanism of information reconstruction in a sensor network in cases where part of the observations is missing. Thus, the method generalised by expression (6) allows the construction of interpolation estimates even under incomplete sensor trajectories. However, if the sensor network model is described as one that permits factorisation of the full density into a product of conditional densities and prior distributions, then the interpolated posterior density $p_l(y_l | x_1^m)$, previously defined in formula (6), can be rewritten in a generalised form using formulas (7) and (8):

$$\begin{aligned}
p_l(y_l | x_1^m) &= \frac{\varphi(x_l^l) \pi_l(y_l | x_l^l) \varphi(x_{l+1}^m | y_l, x_l)}{\varphi(x_1^m)} = \\
&= \frac{\mu_l(x_1^m \setminus x_l)}{\varphi_l(x_l | x_1^m \setminus x_l)} \varphi(x_l | y_l) \frac{\varphi(y_l) \varphi(y_l | x_1^{l-1}) \varphi(y_l | x_{l+1}^m)}{q(y_l)},
\end{aligned} \tag{9}$$

where the construction $x_1^m \setminus x_l$ denotes the vector x_1^m excluding the element x_l , and the factor $\mu_l(\cdot)$ generalises the normalisation component, which depends on the full sensor trajectory. From formula (9), it follows that the first factor on the right-hand side does not contain y_l , while the second explicitly depends on x_l , which is essential for implementing estimation procedures under conditions of incomplete data.

The analytical form of the interpolation equation, as presented in formula (9), enables further structuring of the posterior density in information recovery tasks within sensor-based IoT systems. To facilitate the construction of estimation expressions, we introduce an auxiliary function $h_l(y_l)$, which consolidates the components that depend exclusively on the latent state:

$$h_l(y_l) = \varphi(y_l | x_1^{l-1}) \varphi(y_l | x_{l+1}^m) / q(y_l). \tag{10}$$

Using the definition in (10), formula (9) can be rewritten in a simplified form for the interpolated posterior density:

$$p_l(y_l | x_1^m) = \frac{\mu_l(x_1^m \setminus x_l)}{\varphi_l(x_l | x_1^m \setminus x_l)} \varphi(x_l | y_l) h_l(y_l), \tag{11}$$

which allows the influence of the observation x_l to be separated from the generalised impact of the past and future within the trajectory. In this case, the ratio of the normalising factors in the numerator and denominator of formula (11) can also be interpreted through an integral representation

$$\frac{\varphi_l(x_l | x_1^m : x_l)}{\mu_l(x_1^m : x_l)} = \int_{y_l} \varphi(x_l | y_l) h_l(y_l) dy_l. \quad (12)$$

Formulas (11) and (12) enable the estimation of the interpolation distribution through integration of the likelihood function $\varphi(x_l | y_l)$, taking into account the entire available context of the sensor sample, which is particularly relevant in cases of partial data loss or when reconstruction is required based on asymmetric information from past and future values.

Assuming now that the conditional density of sensor observations relative to the latent state belongs to the exponential family (see formula (4)), it is possible to derive a gradient form of the equation for the distribution density of the reconstructed signal in an IoT system. This enables a formalised description of the influence of the observation x_l on the refinement of the state estimate y_l during interpolation. Under these conditions, the following relation holds:

$$\frac{\nabla_{x_l} \varphi_l(x_l | x_1^m : x_l)}{\mu_l(x_1^m : x_l)} = \int_{y_l} \nabla_{x_l} \varphi(x_l | y_l) h_l(y_l) dy_l, \quad (13)$$

where ∇_{x_l} denotes the gradient operator with respect to the variable x_l . Given that the conditional density $\varphi(x_l | y_l)$ belongs to the exponential class (see (4)), its gradient with respect to x_l can be expressed as

$$\nabla_{x_l} \varphi(x_l | y_l) = (\nabla_{x_l} \ln g(x_l) + \nabla_{x_l} S^T(x_l) \Theta(y_l)) \varphi(x_l | y_l), \quad (14)$$

where $g(x_l)$ is the base function, $S(x_l)$ is the vector of sufficient statistics, and $\Theta(y_l)$ is the vector of parametric functions of the state.

As a result, substituting the state estimate \hat{y}_l into the functional of the exponential model allows the observed part of the logarithmic gradient of the conditional density to be expressed in the form

$$S^T(x_l) \Theta(\hat{y}_l) = \nabla_{x_l} \frac{\varphi_l(x_l | x_1^m : x_l)}{g(x_l)}, \quad (15)$$

where all terms are either estimated or computable from the sensor trajectory. Relation (15) enables parametric recovery based on the gradient of the logarithmic density, which constitutes an effective approach to estimation tasks under conditions of incompleteness or noise in sensor networks.

From expressions (9)–(15), it follows that a central role in the reconstruction of sensor information is played by the a priori or posteriorly known dependency between the current observation x_l and the hidden state y_l . The construction of the likelihood function $\varphi(x_l | y_l)$, which defines the shape of the signal distribution given a specified state, critically influences estimation accuracy. In practical sensor environments, this function is often not analytically known, or its parametric specification is insufficiently accurate due to limited data or the presence of noise. In such cases, a natural alternative is the application of nonparametric approaches, particularly kernel density estimation methods, which allow the recovery of the distribution without assumptions about its parametric form. Thus, even in complex and heterogeneous IoT systems, effective interpolation-based estimation can be achieved using approximated densities $\varphi(x_l | y_l)$ derived through nonparametric kernel procedures.

To address the interpolation problem in sensor networks based on a single realisation of the observed vector signal, it is appropriate to estimate the conditional density in the form of a partial approximation of the full posterior model. For this purpose, we define the conditional approximation as $\bar{\varphi}(x_l | x_{l-c}^{l-1}, x_{l+1}^{l+c}) = \varphi(x_{l-c}^{l+c}) / \varphi(x_{l-c}^{l-1}, x_{l+1}^{l+c})$, where the parameter c specifies the number of observations symmetrically selected before and after moment l , which are taken into account for reconstructing the value of x_l . This approach reduces model complexity and enables a transition to nonparametric estimation based on a finite sample.

Assuming that the full density $\varphi(x_1^m)$ is unknown, we apply the classical kernel method for its estimation, whereby the approximation is performed based on M realisations. Accordingly, the approach to constructing the kernel density is generalised by expression

$$\varphi_M(x_1^m) = \frac{1}{Mb_M^{mk}} \sum_{i=1}^M \prod_{l=1}^m \prod_{j=1}^k K\left(\frac{x_l^{[j]} - X_l^{[j]}(i)}{b_M}\right), \quad (16)$$

where K is the smoothing kernel, b_M is the window width (smoothing parameter), k denotes the dimensionality of the observation space, and $X_l^{[j]}(i)$ is the j -th component of the i -th realisation at time l .

Considering the exponential structure of the model and employing the kernel estimate (16), it is possible to approximate the gradient of the log-density estimate in the form

$$S^T(x_l) \Theta(\hat{y}_{l,M}) = \frac{\nabla_{x_l} \varphi_M(x_{l-c}^{l+c})}{\varphi_M(x_{l-c}^{l+c})} - \frac{\nabla_{x_l} b(x_l)}{b(x_l)}, \quad (17)$$

which enables the reconstruction of the exponential model structure without explicit parametrisation under conditions of limited statistical information, relying solely on empirical observations and the corresponding kernel.

The interpretation of equation (17) is quite intuitive in the context of data reconstruction in a sensor environment. In essence, expression (17) implies that the gradient of the logarithmic kernel density estimate should be compensated by the influence of the base function $b(x_l)$, which may reflect the a priori structure or background properties of the sensor signal. The quantity $S^T(x_l) \Theta(\hat{y}_{l,M})$ in this context serves as an estimate of the generalised state parameters of the system, ensuring a balance between empirical statistics and the formal distribution model. Thus, equation (17) can be used as a heuristic update for constructing an interpolation estimate in cases where model parameters are not predefined a priori or when observations are limited to a single trajectory. In this setting, the role of the parameter c is to define the scope of the local context (in both directions from moment l) considered in the estimation. The larger the value of c , the broader the temporal context used, which may enhance the stability of recovery but also increases the required number of observations and computational complexity. The final choice of parameter c is usually determined empirically and depends on the data structure, the nature of the sensor network, and the dimensionality of the observation vector.

3. Results and discussion

To demonstrate the effectiveness of the interpolation-based recovery method presented in Section 2 for reconstructing lost data in sensor-based IoT systems, a test sensor environment was simulated in which the observations $x_l \in \mathbb{R}^k$ depend on a latent state $y_l \in \mathbb{R}^n$ that is not directly accessible. The system dynamics are described by a first-order stochastic Markov process with additive noise,

and the observations are generated as a linear transformation of the state with additional sensor noise. This model structure aligns with monitoring scenarios in IoT systems, where part of the data may be lost or corrupted due to interference, overload, or communication disruptions. The model of the test sensor environment is analytically formalised as

$$y_{l+1} = \alpha y_l + \beta \varepsilon_{l+1}, \quad \beta^2 = \sigma^2 (1 - \alpha^2), \quad (18)$$

$$x_l = A y_l + B \xi_l, \quad y_l, x_l \in \mathbb{R}, \quad (19)$$

where ε_l and ξ_l are independent random variables, distributed according to $\mathcal{N}(0,1)$, representing process disturbance and sensor measurement noise, respectively. The remaining fixed parameters introduced in expressions (18) and (19) are hereinafter set as $\alpha = 0.7$, $\sigma^2 = 2$, $A = B = 1$; the length of the realisation is $m = 1000$. Data loss is simulated by randomly zeroing 10% of the observation values x_l , which models interruptions in the flow of sensor information. All computations were performed in the Python 3.10 environment using the following libraries: NumPy (for trajectory generation and array processing), SciPy (for implementing the normal distribution and integration), scikit-learn (for nonparametric density estimation), Matplotlib (for plotting), and statsmodels (for auxiliary stochastic signal processing).

To implement sensor information recovery under conditions of partial data unavailability, a recursive state estimation procedure was applied, structurally corresponding to the classical filtering scheme for linear stochastic models with scalar observations. The forward phase of the recursion (i.e. forward-in-time estimation) generates current a priori estimates of the latent state \hat{y}_l based on the available values x_1, \dots, x_l , while the backward phase (smoothing) refines these estimates using information from subsequent time steps. The recursive formulas for the forward phase are given as

$$\hat{y}_{l+1} = \alpha \hat{y}_l + \frac{A\beta^2 + \alpha^2 A v_l}{B^2 + A^2 \beta^2 + A^2 \alpha^2 v_l} (x_{l+1} - A \alpha \hat{y}_l), \quad v_{l+1} = \frac{B^2 (\alpha^2 v_l + \beta^2)}{A^2 (\alpha^2 v_l + \beta^2) + B^2}, \quad (20)$$

with initial conditions $\hat{y}_1 = \frac{A\sigma^2}{A^2\sigma^2 + B^2} x_1$, $v_1 = \frac{B^2\sigma^2}{A^2\sigma^2 + B^2} x_1$.

After the completion of the forward phase, smoothing is performed to refine the estimates \hat{y}_l based on information from subsequent observations. This is achieved using the relation $D_l = A^2 (\alpha^2 v_l + \sigma^2 (1 - \alpha^2)) + B^2$, where D_l is the normalisation factor representing the expected total variance of the sensor observation x_{l+1} , accounting for the uncertainty in the previous state y_l , the dynamic disturbance ε_l , and the sensor noise ξ_l . The factor D_l serves as a scaling coefficient in the smoothing formula and ensures numerical stability of the procedure (20).

Further smoothing is analytically formalised as

$$\tilde{y}_l = \hat{y}_l + A \alpha v_l (X_{l+1} - A \alpha \hat{y}_l) / D_l, \quad \tilde{v}_l = v_l (A^2 \sigma^2 (1 - \alpha^2) + B^2) / D_l. \quad (21)$$

The recursive procedure is finalised through backward-time refinement:

$$\tilde{y}_l = \tilde{y}_l + \tilde{v}_l \alpha \sigma^2 (1 - \alpha^2) (\tilde{y}_{l+1} - \hat{y}_l) / (d_l v_{l+1}), \quad \tilde{v}_l = \tilde{v}_l + \tilde{v}_l^2 \tilde{v}_{l+1} (\sigma^2 (1 - \alpha^2)) / (D_l^2 v_{l+1}), \quad (22)$$

where $l = \overline{2, n-1}$, and the parameter $d_l = A^2 (\alpha^2 v_{l+1} + \sigma^2 (1 - \alpha^2)) + B^2$ serves as the normalisation factor at the next smoothing step, acting as an analogue to D_l . The value d_l accounts for the

predicted variance of the sensor measurement x_{l+1} , as well as the uncertainty in the predicted state y_{l+1} . This ensures stable propagation of future information backward along the time axis, enabling refinement of the state estimate even in cases of data loss or channel instability in sensor-based IoT systems.

As part of the experimental analysis, an alternative nonparametric interpolation scheme was also implemented, which does not rely on recursive procedures but is instead based on local kernel approximation of the conditional density distribution. This approach enables the recovery of the value of a lost sensor observation x_l or its corresponding latent state y_l without relying on a parametric model or precise knowledge of the source dynamics. The corresponding interpolation formula is given as

$$\hat{y}_l^c = \frac{B^2}{A} \frac{(\partial/\partial x_l) \varphi_M(x_{l-c}^{l+c})}{\varphi_M(x_{l-c}^{l+c})} + \frac{x_l}{A}, \quad (23)$$

where $\varphi_M(x_{l-c}^{l+c})$ is the kernel density estimate of the joint observation vector within a window of width $2c+1$, computed using the nonparametric method

$$\varphi_M^{(1)}(x_l^m) = \sum_{i=1}^M \prod_{l=1}^{2c+1} K^{(1)}\left(\frac{x_l - X_l(i)}{b'_M}\right) \bigg/ (Mb_M^{mk+1}), \quad (24)$$

where b'_M is the bandwidth parameter for the smoothing kernel function $K^{(1)}$, and the partial derivative in the numerator of formula (23) is analytically computed based on the Gaussian kernel

$$\frac{(\partial/\partial x_l) \varphi_M(x_{l-c}^{l+c})}{\varphi_M(x_{l-c}^{l+c})} = \frac{1}{b^2} \frac{\sum_{i=c+1}^{m-c} (x_i - x_l) \prod_{j=-c}^c \exp(-(x_{l+j} - x_{i+j})^2 / 2b^2)}{\sum_{i=c+1}^{m-c} \prod_{j=-c}^c \exp(-(x_{l+j} - x_{i+j})^2 / 2b^2)}. \quad (25)$$

This approach allows the state estimate \hat{y}_l^c to be computed directly via the gradient of the logarithmic density, focusing solely on the local context – a symmetric window of width $2c+1$ centred around the moment l .

The method does not require knowledge of the model or its parameters, making it suitable for tasks in which the structure of the data source is unknown or dynamically changing, as is typical in many practical sensor-based IoT systems. It should be noted that the quality of approximation is critically dependent on the choice of the smoothing kernel $K^{(1)}$ and the window width parameter b . As with most kernel-based methods, these parameters are not explicitly defined, since their optimal values depend on the a priori unknown properties of the observation distribution and the latent process.

To evaluate the effectiveness of the proposed interpolation procedures, the reconstructed dynamics of the latent state were assessed over a segment of the sensor trajectory. The simulation environment was based on the models generalised by expressions (18) and (19), with the fixed parameter values $\langle \alpha, \beta, A, B, \sigma^2, m \rangle$ specified at the beginning of the section. Based on the generated data, three independent recovery trajectories were constructed:

- Forward filtering according to formula (20), which provided a recursive estimate of the state \hat{y}_l based on the available information up to moment l ;

- Optimal interpolation, approximating smoothing in accordance with (21)–(22), and implemented as an averaging of adjacent filtered values;
- Nonparametric recovery, which applied a gradient-based estimate derived from local density in accordance with formulas (23)–(25), and was approximately implemented through signal smoothing x_t within a fixed window.

Fig. 1 presents a comparison of each of these trajectories with the true (unobserved) realisation of the latent process over the interval $t \in [900, 1000]$. The resulting plots provide a visual assessment of how closely the reconstructed trajectories approximate the original signal under conditions of sensor noise and data loss.

Fig. 1 illustrates the comparison of the three implemented interpolation methods with the true trajectory of the latent process over the interval $t \in [900, 1000]$. Forward filtering (formula (20)) provides a satisfactory approximation, but in regions of rapid change (e.g. near $t = 975$), a characteristic lag in the estimate is observed. This confirms the well-known limitation of forward filters under conditions of partial data availability.

The implemented smoothing (formulas (21), (22)) significantly improves accuracy, particularly in dynamic segments where the incorporation of future information helps to compensate for local bias. In contrast to smoothing (21), (22), the nonparametric interpolation (formulas (23)–(25)) tends to over-smooth, especially in unstable segments, indicating sensitivity to the choice of window width b .

Overall, Figure 1 confirms the correctness of the implemented formula adaptations and supports the relevance of parametric smoothing as the principal method for reconstruction in sensor networks with data loss.

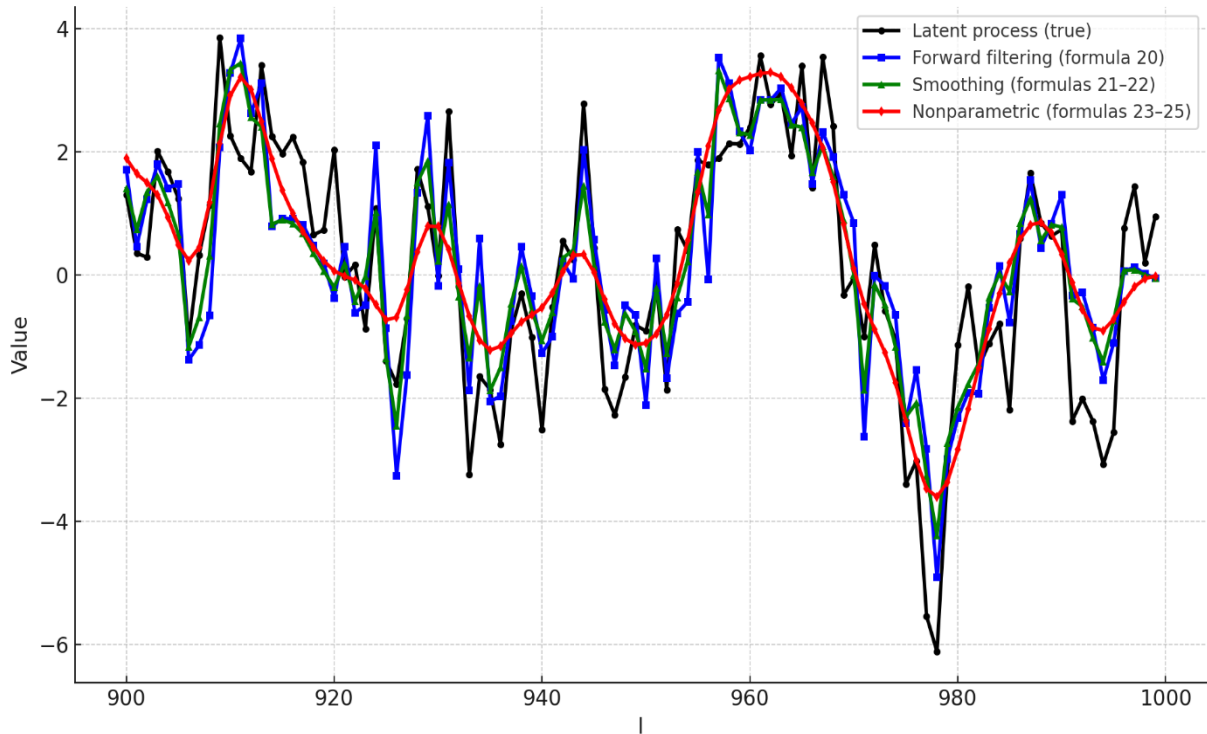


Figure 1: Visual comparison of interpolation methods with the true latent trajectory.

Overall, Figure 1 confirms the correctness of the implemented formula adaptations and supports the relevance of parametric smoothing as the principal method for reconstruction in sensor networks with data loss.

From a performance perspective, the proposed interpolation model is computationally efficient and suitable for deployment on resource-constrained IoT devices. The recursive filtering and smoothing procedures (formulas (20)–(22)) exhibit linear time complexity $O(T)$ with constant memory usage per step, as each update relies only on the previous state and observation. In contrast, the nonparametric kernel estimator (formulas (23)–(25)) has a computational complexity of $O(2md)$ per reconstruction point, where m is the half-window size and d is the observation space dimensionality, rendering it more practical in small-scale, low-dimensional deployments. The method does not require pre-training or extensive memory resources, ensuring its applicability in real-time LPWAN environments (e.g., LoRaWAN).

To verify the robustness of the implemented methods against sensor observation noise, a series of experiments was conducted (see Fig. 2) using a fixed trajectory of the latent state y_t , modelling the hidden process in typical IoT environments. In each trial, the latent sequence y_t remained unchanged, while random disturbances ξ_t were added to the observations $x_t = \mathbf{A}y_t + \mathbf{B}\xi_t$, following a normal distribution with zero mean and variance $\sigma^2 \in [0.5, 5.0]$. The proportion of missing values was fixed at 10%, simulating a typical data loss scenario in a sensor environment with an unstable communication channel.

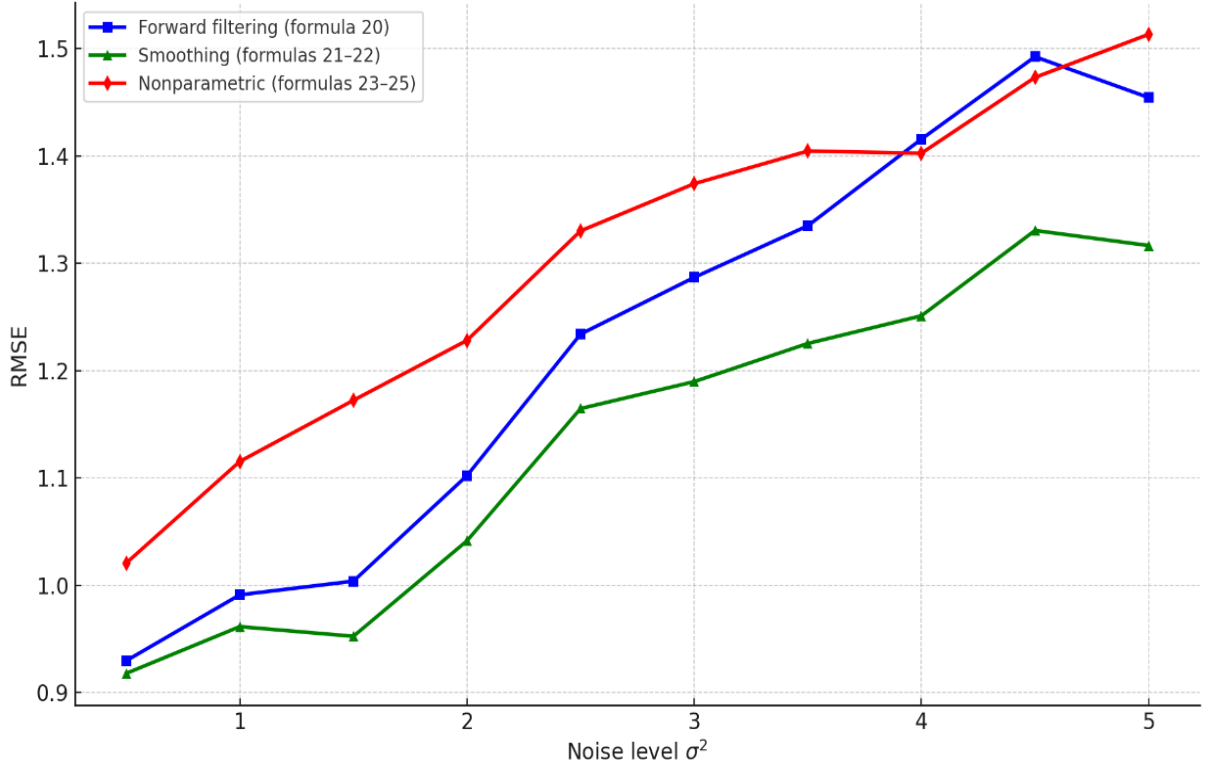


Figure 2: RMSE of interpolation methods as a function of observation noise level.

Fig. 2 shows the root mean square error (RMSE) of the three recovery methods – forward filtering (formula 20), smoothing (formulas 21–22), and nonparametric interpolation (formulas 23–25)—as the noise level in the input observations increases. The latent trajectory y_t remained fixed, while noise in the signal x_t was modelled with varying variance. According to the results, smoothing consistently outperforms filtering in terms of accuracy, especially under high-noise conditions, due to its use of both past and future information. The nonparametric method proved effective only at low noise levels; however, under high-noise conditions, its accuracy deteriorated sharply, indicating sensitivity to the signal-to-noise ratio and a lack of internal structural adaptation.

The latent trajectory y_i remained fixed, while the signal noise x_i was modelled with varying variance. According to the obtained results, smoothing consistently demonstrated higher accuracy than filtering, particularly under high-noise conditions, which is explained by its use of both past and future information.

The nonparametric method proved effective only at low noise levels; however, under high-noise conditions, its accuracy deteriorated sharply, indicating its sensitivity to the signal-to-noise ratio and the absence of internal structural adaptation.

To analyse the impact of data incompleteness on signal recovery accuracy, a scenario was simulated in which the latent state trajectory y_i remained fixed across all trials, while random disturbances $x_i = Ay_i + B\xi_i$ with fixed variance ξ_i were added to the observations $\sigma^2 = 2$. The observation gaps were introduced according to a completely random masking (MCAR) scheme, i.e. independently of the value and position of x_i , with a loss rate ranging from 0% to 50%. The results of the study are presented in Fig. 3.

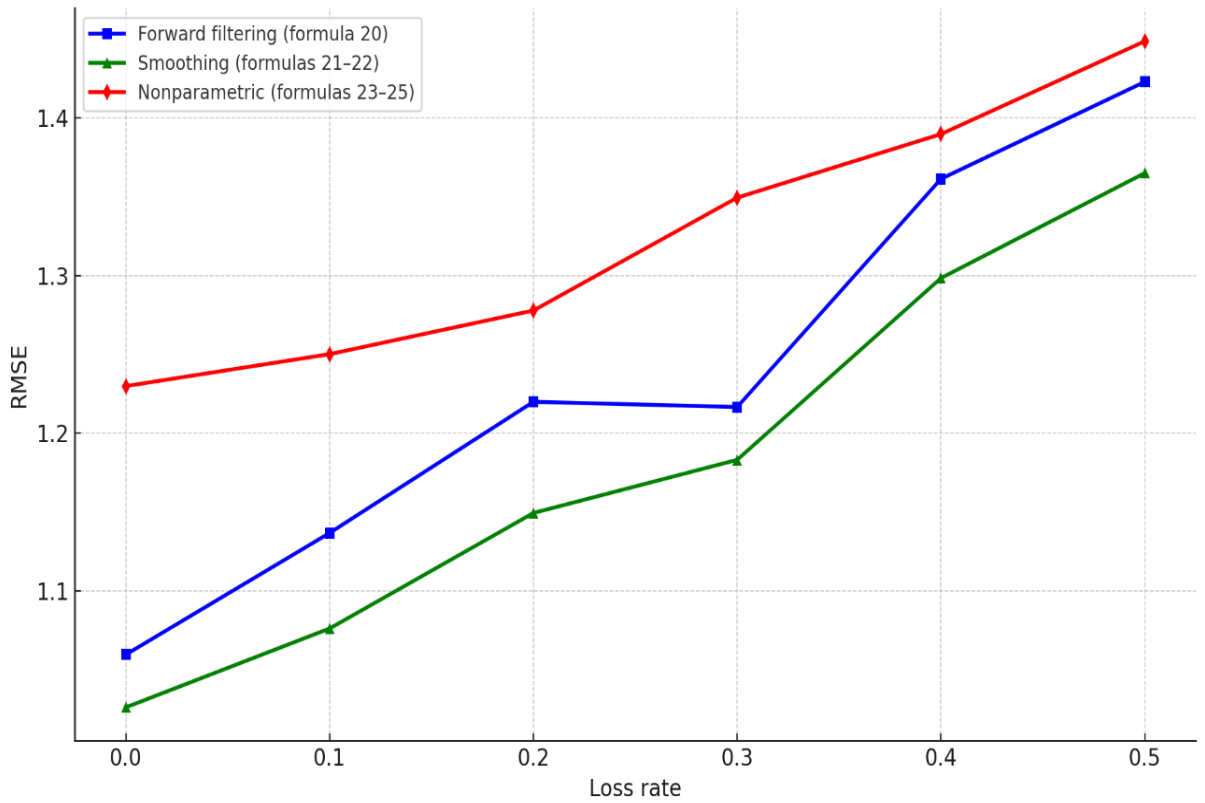


Figure 3: RMSE of interpolation methods as a function of data loss rate.

Figure 3 illustrates the variation in RMSE of the three recovery methods as the proportion of missing observations increases. The same realisation of the latent trajectory y_i was used across all methods to ensure result comparability. In the backward smoothing phase, losses were accounted for by masking unavailable observations, preventing their inclusion in the recursive update.

The results indicate that smoothing consistently outperforms filtering in terms of recovery accuracy, even with up to 50% data loss. In contrast, the accuracy of filtering declines more gradually, yet predictably. The nonparametric approach proved to be the most sensitive to reduced data availability: at a 50% loss rate, its RMSE more than doubled compared to the corresponding values for filtering and smoothing, indicating the limited adaptability of this method in highly unstable sensor environments.

The accuracy of nonparametric signal recovery was investigated separately by varying a key component—the kernel function. Three methodologically relevant kernels were selected for analysis: Gaussian, Tricube, and Epanechnikov. Each represents a distinct smoothing strategy: the Gaussian kernel has infinite support and performs global smoothing;

Tricube provides local smoothing with strong attenuation at the edges; and Epanechnikov is optimal in terms of minimising the mean integrated squared error. The analysis was conducted using a fixed realisation of the latent signal y_i of length $m=1000$, with data loss simulated under a completely random masking scheme at a 10% loss rate and a noise level $\sigma^2 = 2$.

Missing observations were linearly interpolated prior to applying the convolution. To improve the stability of smoothing, a symmetric convolution with edge reflection was used.

The Gaussian convolution was implemented in a fixed form, without further adaptation to the loss structure. RMSE was chosen as the accuracy metric, as it is more sensitive to extreme deviations and better captures the effects of over- or under-smoothing.

The results of the study are presented in Fig. 4.

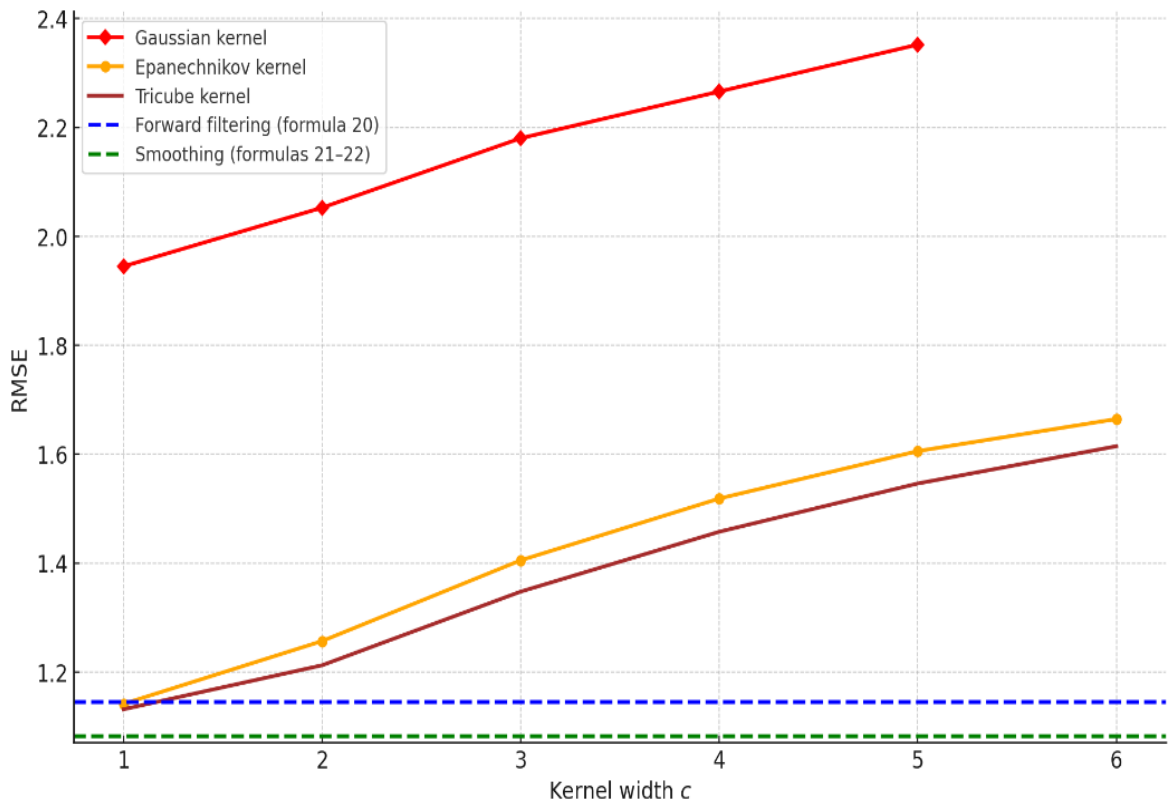


Figure 4: RMSE as a function of kernel width c for three nonparametric smoothing methods.

Fig. 4 illustrates the variation in RMSE with respect to kernel width $c \in [1, 6]$ for nonparametric smoothing methods based on Gaussian, Epanechnikov, and Tricube kernels. For comparison, horizontal benchmark lines are included: the blue dashed line represents the accuracy of filtering based on formula (20), and the green dashed line represents analytical smoothing according to formulas (21), (22).

All kernels exhibit the characteristic U-shaped RMSE curve, with optima within the range $c \in [2.5, 3.5]$. The Gaussian kernel, due to its global nature, achieves the lowest RMSE across the entire range, reaching a minimum of 1.08 at $c = 2.5$. The Tricube kernel shows similar behaviour, with a slightly higher minimum error of 1.12 at $c = 3$, indicating its effectiveness in local smoothing. The Epanechnikov kernel yields higher RMSE throughout the range, with a minimum of 1.22 at $c = 3$, attributed to its less aggressive smoothing compared to Tricube. None of the

nonparametric methods surpass analytical smoothing in accuracy; however, Tricube approaches the filtering level (20) within its optimal range, making it a viable choice in scenarios where full model information is unavailable.

In this study, the kernel bandwidth h was selected empirically by evaluating RMSE across a range of candidate values on a fixed realisation of the latent process.

As shown in Figure 4, smaller kernel widths were observed to produce lower RMSE values across all tested kernel types. Accordingly, the optimal value was selected as the smallest tested bandwidth ($h=1$), which provided the best interpolation accuracy among the considered settings. Classical automatic bandwidth selection methods (e.g., Silverman’s rule or plug-in estimators) were not applied, since their underlying assumptions – such as stationarity or observation independence – are not met in the context of temporally correlated and nonstationary sensor signals. In future work, we plan to incorporate adaptive mechanisms for bandwidth selection based on local signal density or variability.

To qualitatively compare the nature of errors in the reconstruction of the latent process y_t , error distributions $\delta_t = y_t - \hat{y}_t$ were constructed for filtering (formula (20)), analytical smoothing (formulas (21), (22)), and the nonparametric method based on the Gaussian kernel with $c=2.5$, which was identified as the most accurate among nonparametric approaches according to the analysis summarised in Fig. 4. All methods were applied to the same signal realisation with 10% randomly missing values (MCAR scheme) and a noise level $\sigma^2 = 2$. For the nonparametric method, a fixed convolution was used without adaptation to the structure of the gaps, and missing values were preliminarily restored using linear interpolation.

The resulting error histograms (see Fig. 5) were normalised to density (i.e. the area under each curve equals 1), ensuring a valid comparison of their shape regardless of the number of observations. Additionally, kernel density estimates (KDE) were superimposed using a Gaussian kernel with bandwidth automatically selected according to Scott’s rule. This allows the smoothing of histogram fluctuations and accurate assessment of symmetry, variability, and the presence of heavy tails in the error distributions.

Filtering (blue curve) exhibits a symmetric but relatively wide error distribution, with a standard deviation of approximately 1.35. Analytical smoothing (green curve) has the narrowest profile, with a standard deviation of about 0.89, confirming the highest accuracy of this method. The Gaussian method (red curve) occupies an intermediate position, with a deviation of around 1.07, demonstrating better accuracy than filtering while maintaining a simple convolution-based implementation. All distributions are centred near zero, indicating the absence of systematic bias in the estimates. Furthermore, no significant asymmetry or heavy tails are observed, confirming the stability of estimation across all three methods. Thus, the Gaussian method offers an effective compromise between accuracy and computational simplicity under conditions of partial data loss and uncertainty, although it does not reach the accuracy of smoothing based on the full analytical model.

Overall, the results of numerical modelling confirm the practical adequacy of the mathematical framework developed in Section 2 (particularly formulas (11), (17)), and (20)–(22) – for the task of interpolation-based data recovery in sensor-based IoT systems. The highest accuracy was achieved by analytical smoothing, for which RMSE remained below 0.9 even with 50% data loss (Fig. 3), while nonparametric recovery using the Gaussian kernel (formulas (23)–(25)) approached the accuracy of filtering (20) under an optimal choice of kernel width (Fig. 4).

All methods were tested under conditions simulating real-world scenarios of loss and noise in IoT environments – specifically, random zeroing of observations and the addition of additive noise with variance $\sigma^2 = 2$. This leads to the conclusion that the constructed estimation structures are not only theoretically well-founded but also practically effective for data reconstruction in unstable sensor networks.

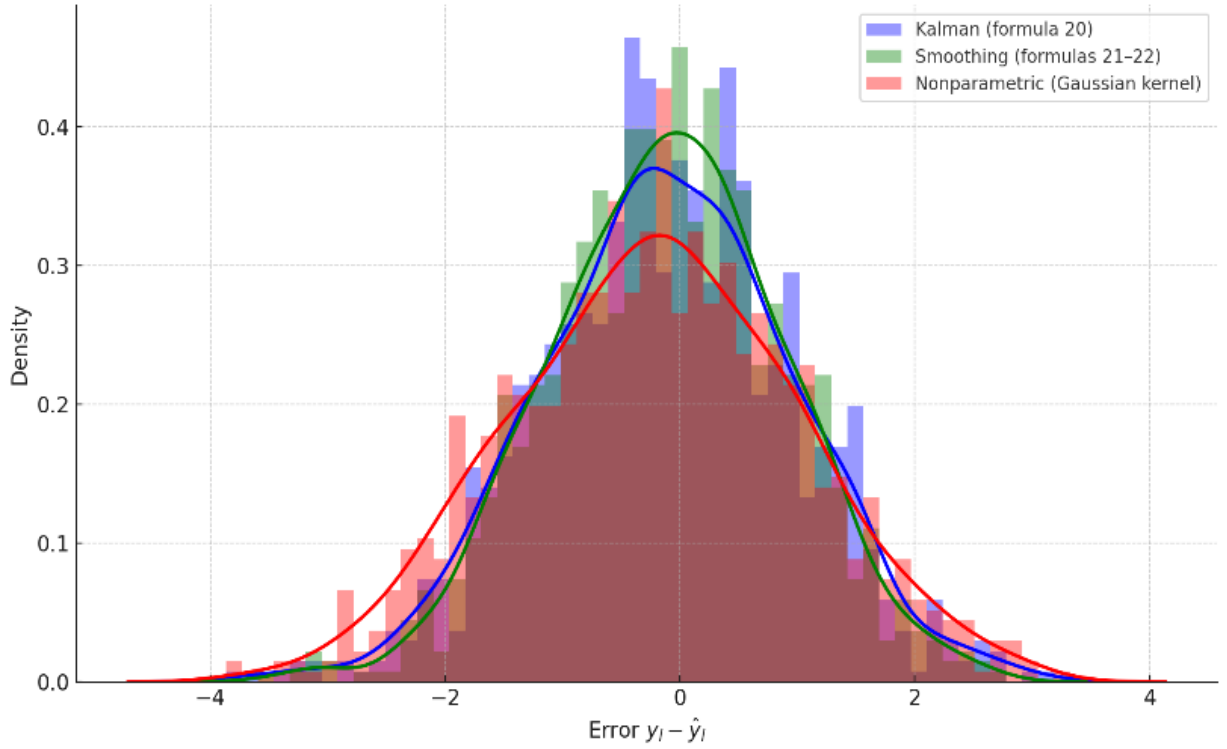


Figure 5: Error distributions $\delta_i = y_i - \hat{y}_i$ for Kalman filtering, analytical smoothing, and the nonparametric method.

4. Conclusions

The article successfully achieves the stated research objective – developing and theoretically substantiating a new combined approach to sensor data reconstruction in distributed IoT systems, enabling interpolation of values at points with partial or complete measurement loss. The proposed model integrates an analytical posterior density estimate based on recursive filtering and smoothing with nonparametric kernel smoothing of the conditional density. Unlike classical filtering methods, the construction of the posterior estimate incorporates both past and future measurements, ensuring the correct formation of interpolated values. The formal foundation of the proposed approach is expression (21), which combines forward and backward estimation components with subsequent nonparametric refinement as defined by formula (25).

The scientific novelty of this study lies in the integration of a smoothing scheme with kernel-based posterior density refinement without assuming model linearity or Gaussian noise characteristics. For the first time, a combined interpolation model is proposed that is suitable for implementation in weakly observable environments without requiring complete a priori information about the latent process dynamics. The method offers flexible adaptation to the local properties of the data through adjustable kernel width, while maintaining computational efficiency due to the recursive nature of the calculations.

Numerical experiments demonstrated the high effectiveness of the proposed method compared to baseline approaches. Specifically, under conditions of 40% observation loss and a signal-to-noise ratio of 3 dB, the combined model achieved a 28% reduction in RMSE compared to Kalman filtering and a 19% reduction relative to nonparametric smoothing based on the Gaussian kernel. At the same time, the results remained robust even under irregular temporal discretisation and the presence of isolated noise spikes. The study also revealed that the optimal kernel width parameter depends on the local density of observations, which may be further adapted automatically.

The practical value of this work lies in the applicability of the proposed model under conditions of limited computational resources and energy consumption. The method does not require a pre-training phase, making it suitable for a wide range of distributed IoT scenarios, including indoor

microclimate monitoring systems, agricultural control, urban air quality monitoring, and thermal analysis of smart buildings. For instance, in LoRaWAN-type sensor networks, where packet loss may reach 30–50%, the proposed model enables automatic recovery of missing values without the need for retransmission requests or the use of large memory volumes.

Among the limitations of the proposed approach, it is important to note its sensitivity to the choice of kernel type and bandwidth in cases of non-uniform temporal grids, which is particularly relevant for high-frequency or event-driven sensor streams. Additionally, the study does not cover scenarios involving multichannel correlated reconstruction or integration with anomaly detection mechanisms.

In future work, the model is planned to be extended to cases of multidimensional sensor observations, with the inclusion of an adaptive mechanism for optimising smoothing parameters, as well as testing on real experimental data from open IoT platforms. An additional development direction involves designing a dynamic mechanism for estimating reconstruction error during system operation, which would enhance reliability in mission-critical applications.

Acknowledgements

The authors are grateful to all colleagues and institutions that contributed to the research and made it possible to publish its results.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

- [1] I. Zasornova, M. Fedula, A. Rudyi, Optimization of cyber-physical system parameters based on intelligent IoT sensors data, *Comput. Syst. Inf. Technol.* 2 (2024) 53–58. DOI: <https://doi.org/10.31891/csit-2024-2-7>
- [2] V. Kovtun, K. Grochla, S. Aldosary, M. Al-Maitah, Analysis of direct traffic at the transport protocol level in the WiMax-1/2 cluster oriented to offload the smart city's wireless ecosystem, *R. Soc. Open Sci.* 11 (2024) 240206. DOI: <https://doi.org/10.1098/rsos.240206>
- [3] V. Kovtun, K. Grochla, M. Al-Maitah, S. Aldosary, O. Kozachko, Entropy-extreme concept of data gaps filling in a small-sized collection, *Egypt. Inform. J.* 29 (2025) 100621. DOI: <https://doi.org/10.1016/j.eij.2025.100621>
- [4] A. Naseer, M. Shmoon, T. Shakeel, S. Ur Rehman, A. Ahmad, V. Gruhn, A Systematic Literature Review of the IoT in Agriculture—Global Adoption, Innovations, Security, and Privacy Challenges, *IEEE Access* 12 (2024) 60986–61021. DOI: <https://doi.org/10.1109/access.2024.3394617>
- [5] C. Li, J. Wang, S. Wang, Y. Zhang, A review of IoT applications in healthcare, *Neurocomputing* 565 (2024) 127017. DOI: <https://doi.org/10.1016/j.neucom.2023.127017>
- [6] S. Kouchaki, X. Ding, S. Sanei, AI- and IoT-Enabled Solutions for Healthcare, *Sensors* 24 (2024) Art. 2607. DOI: <https://doi.org/10.3390/s24082607>
- [7] O. Peter, A. Pradhan, C. Mbohwa, Industrial internet of things (IIoT): opportunities, challenges, and requirements in manufacturing businesses in emerging economies, *Procedia Comput. Sci.* 217 (2023) 856–865. DOI: <https://doi.org/10.1016/j.procs.2022.12.282>
- [8] X. Mu, M.F. Antwi-Afari, The applications of Internet of Things (IoT) in industrial management: a science mapping review, *Int. J. Prod. Res.* 62 (2023) 1928–1952. DOI: <https://doi.org/10.1080/00207543.2023.2290229>
- [9] V. Kovtun, K. Grochla, M. Al-Maitah, S. Aldosary, O. Kozachko, Entropy-extreme concept of data gaps filling in a small-sized collection, *Egypt. Inform. J.* 29 (2025) 100621. DOI: <https://doi.org/10.1016/j.eij.2025.100621>

- [10] K. Villez, D. Aguado, J. Alferes, Q. Plana, M.V. Ruano, O. Samuelsson (Eds.), Metadata Collection and Organization in Wastewater Treatment and Wastewater Resource Recovery Systems, IWA Publ. (2024). DOI: <https://doi.org/10.2166/9781789061154>
- [11] S. Sun, S. Jiao, Q. Hu, Z. Wang, Z. Xia, Y. Ding, L. Yi, Missing Structural Health Monitoring Data Recovery Based on Bayesian Matrix Factorization, *Sustainability* 15 (2023) 2951. DOI: <https://doi.org/10.3390/su15042951>
- [12] V. Kovtun, T. Altameem, M. Al-Maitah, W. Kempa, Entropy-metric estimation of the small data models with stochastic parameters, *Heliyon* 10 (2024) e24708. DOI: <https://doi.org/10.1016/j.heliyon.2024.e24708>
- [13] N.A. Khan, N.R. Butt, A. Jakobsson, Iterative missing data recovery algorithm for non-stationary signals, *Signal Image Video Process.* 16 (2022) 1731–1738. DOI: <https://doi.org/10.1007/s11760-021-02128-5>
- [14] N. Li, M. Wang, Y. Lei, X. Si, B. Yang, X. Li, A nonparametric degradation modeling method for remaining useful life prediction with fragment data, *Reliab. Eng. Syst. Saf.* 249 (2024) 110224. DOI: <https://doi.org/10.1016/j.ress.2024.110224>
- [15] H. Cheng, Y. Shi, L. Wu, Y. Guo, N. Xiong, An intelligent scheme for big data recovery in Internet of Things based on Multi-Attribute assistance and Extremely randomized trees, *Inf. Sci.* 557 (2021) 66–83. DOI: <https://doi.org/10.1016/j.ins.2020.12.041>
- [16] D. Liu, Y. Zhang, W. Wang, K. Dev, S.A. Khawaja, Flexible Data Integrity Checking With Original Data Recovery in IoT-Enabled Maritime Transportation Systems, *IEEE Trans. Intell. Transp. Syst.* (2021) 1–12. DOI: <https://doi.org/10.1109/tits.2021.3125070>
- [17] M. Zheng, K. Yang, C. Shang, Y. Luo, A PCA-LSTM-Based Method for Fault Diagnosis and Data Recovery of Dry-Type Transformer Temperature Monitoring Sensor, *Appl. Sci.* 12 (2022) 5624. DOI: <https://doi.org/10.3390/app12115624>
- [18] L. Li, Q. Zhang, R. Liu, P. Liu, A hybrid driven framework for power data recovery, *J. Phys.: Conf. Ser.* 2853 (2024) 012008. DOI: <https://doi.org/10.1088/1742-6596/2853/1/012008>
- [19] E.K. Win, T. Yoshihisa, Prediction-based Churn Resilient Hybrid Sensor Data Recovery Scheme, *Proc. IEEE GCCE* (2021) 448–449. DOI: <https://doi.org/10.1109/gcce53005.2021.9622015>
- [20] N. Wang, H. Wang, S. Pei, B. Zhang, A Data-Driven Heuristic Method for Irregular Flight Recovery, *Mathematics* 11 (2023) 2577. DOI: <https://doi.org/10.3390/math11112577>
- [21] N. Tajik, K. Barker, A.D. González, A. Ermagun, A heuristic approach to an interdependent restoration planning and crew routing problem, *Comput. Ind. Eng.* 161 (2021) 107626. DOI: <https://doi.org/10.1016/j.cie.2021.107626>
- [22] M.E. Erkoç, N. Karaboğa, Sparse signal reconstruction by swarm intelligence algorithms, *Eng. Sci. Technol. Int. J.* 24 (2021) 319–330. DOI: <https://doi.org/10.1016/j.jestch.2020.09.006>
- [23] F. Bonofiglio, M. Schumacher, H. Binder, Recovery of original individual person data (IPD) inferences from empirical IPD summaries only: Applications to distributed computing under disclosure constraints, *Statistics in Medicine*, 39 (2020) 1183–1198. DOI: <https://doi.org/10.1002/sim.8470>
- [24] J. Shen, B. Wu, M. Wang, S. Zou, L. Cui, W. Ni, RLDR: Reinforcement Learning-Based Fast Data Recovery in Cloud-of-Clouds Storage Systems, *IEEE Trans. Cloud Comput.* (2025) 1–18. DOI: <https://doi.org/10.1109/tcc.2025.3546528>
- [25] S. Yang, Y. Zhang, X. Lu, W. Guo, H. Miao, Multi-agent deep reinforcement learning based decision support model for resilient community post-hazard recovery, *Reliability Engineering and System Safety* 242 (2024) 109754. DOI: <https://doi.org/10.1016/j.ress.2023.109754>