SUCCESS-6G: EXTEND

WP3 Deliverable E8

V2X radio interface enhancements

Project Title:	SUCCESS-6G: EXTEND

Title of Deliverable:	V2X radio interface enhancements
Status-Version:	Final-1.0
Delivery Date:	30/09/2025

Contributors:	Charalampos Kalalas, Roshan Sedar (CTTC)
Lead editor::	Charalampos Kalalas (CTTC)
Reviewers:	Miquel Payaro (CTTC)
Keywords:	V2X, missing data, imputation, reconstruction

Document Revision History

Version	Date	Description
v0.1	10 July 2025	Initial content addded
v0.2	31 August 2025	Main content addded
v0.3	10 September 2025	Revision and additional content added
v0.4	25 September 2025	Final editions
v1.0	30 September 2025	Approved form of Deliverable E8

Acknowledgement
The research conducted by SUCCESS-6G - TSI-063000-2021-39/40/41 receives funding from the Ministerio de Asuntos Económicos y Transformación Digital and the European Union-NextGenerationEU under the framework of the "Plan de Recuperación, Transformación y Resiliencia" and the "Mecanismo de Recuperación y Resiliencia".

[1] http://creativecommons.org/licenses/by-nc-nd/3.0/deed.en_US

Executive summary

Dependable measurement data are essential for the accuracy and integrity of vehicular state estimation by the maintenance center, which performs condition monitoring tasks. However, vehicular networks are often subject to missing sensor observations due to -among others- channel stochasticity, hardware failures, and security attacks. In this deliverable, we study the problem of missing data in the vehicular measurement streams. We discuss the mechanisms that causally induce occlusions and investigate the ability of various imputation methods to fit the observed data at the aggregation point, and to impute missing values by extracting knowledge from the spatiotemporal synergy among the ambient vehicular measurement space. A rigorous assessment of various missing data configurations based on empirical evaluations reveals meaningful performance trends for model fitting and recovery of incomplete information. Our results demonstrate that rSLDS-based imputation can exploit the spatiotemporal relationships among measurements to make accurate inferences about missing V2X data. By mitigating the impact of unreliable wireless channels at the data layer, imputation offers an AI-driven approach to enhancing reliability, efficiency, and overall performance. Its integration into both existing and future V2X frameworks represents a proactive step aligned with ongoing standardization efforts toward intelligent and resilient vehicular communication systems.

Table of Contents

1 Introduction

The emerging potential of Vehicle-to-Everything (V2X) communication is a cornerstone of intelligent transportation systems, enabling cooperative, connected, and automated mobility (CCAM). V2X networks are designed to support ultra-reliable, low-latency, and high-throughput communications. However, in realistic network environments, the wireless medium is inherently unreliable. Signal interference, multipath fading, shadowing effects, and network congestion lead to packet losses, partial data reception, and temporal inconsistencies. These impairments collectively compromise the accuracy of data-driven inference mechanisms that depend on consistent, high-fidelity information streams.

More specifically, in condition monitoring systems, the acquisition of dependable V2X data is essential for the accuracy and integrity of vehicular state estimation by the maintenance center [1]. Data aggregation points located at the network edge combine vehicular measurement trajectories captured at different locations and time instances to describe the evolution of vehicular state and model the rich interactions between quantities that co-evolve in time. However, vehicular networks are often subject to missing sensor observations due to the inherent unreliability of the wireless channel, hardware/equipment failures, security attacks, etc. Incompleteness in the aggregated data unavoidably affects the downstream processing tasks, leading to incomplete vehicular state knowledge, posing risks in effective decision-making.

The current V2X radio interfaces, as defined by standards such as 3GPP Release 17/18 for C-V2X and IEEE 802.11bd, have made significant progress in supporting advanced vehicular communication use cases. Nonetheless, both technologies face persistent limitations when operating in highly dynamic, interference-rich environments—conditions typical of urban mobility scenarios and dense vehicular traffic. Traditional error-control mechanisms (e.g., HARQ, retransmissions, and channel coding) provide some resilience against transmission losses but are insufficient when data degradation occurs in structured, context-dependent patterns (e.g., temporally correlated packet losses or geographically localized interference). These conditions result in missing or incomplete data at higher network layers, undermining the performance of data-driven services such as condition monitoring, predictive maintenance, or distributed sensor fusion.

To address this challenge, there is a need for data-driven radio interface enhancements that explicitly integrate missing-data imputation techniques into the V2X communication framework. By leveraging machine learning and statistical modeling, imputation methods can infer and reconstruct missing information, effectively mitigating the impact of channel imperfections at the data-processing layer rather than the physical layer alone. We provide the details in the following sections.

2 The Problem of Missing Data in V2X Systems

Aggregating measurement streams is an essential task for the conversion of raw V2X data into meaningful information. This process determines the integrity of the transmitted data and the resilience of the acquisition infrastructure. Nevertheless, a key challenge in achieving efficient data fusion and subsequent knowledge extraction lies in the *completeness* of the aggregated information. In practice, the emergence of missing data in the fused measurement streams is inevitable [2]. Missing information can be generally attributed to the following factors [3]:

1. *Hardware failures*: Hardware component malfunctions, such as synchronization failures or inaccuracies in vehicular sensor readings, may result in persistent missing observations for one or multiple state variables. Moreover, the intermittent availability of energy for sensors operating with renewable sources may cause interruptions in their operation, resulting in gaps in the data acquisition process. In the case of interconnected systems, hardware failures may inadvertently occur in a cascade, where neighboring sensors become progressively and rapidly compromised. Dealing with cascade data occlusions with temporal dependence is often challenging, impeding the efficacy of the reconstruction techniques.

2. *Software failures and user data entry errors*: Software failures can always cause data loss, as can user data entry errors in the case of manual entry of data values. For example, erroneous configuration values for various sensors can result in wrong default or threshold values in case of outliers.

3. *Connectivity issues*: The imperfections of the underlying communication infrastructure in vehicular systems constitute an inseparable aspect of the data acquisition procedure. In fact, the shared wireless medium is inherently unreliable due to signal interference, network congestion, and multipath effects. Such adverse conditions may result in connectivity outages and packet losses, possibly across consecutive time steps. Overall, the induced signal distortion leads to aggregated data inconsistencies and partial system observability, which, in turn, may adversely affect the inference methods at the edge.

4. *Security attacks*: The pervasive digitalization of vehicles expands the attack surface and introduces vulnerabilities and threat vectors, opening up entirely new questions from the perspectives of security and privacy. Across all stages of the data acquisition chain, several entry points become available for potential adversaries to exploit and execute malicious attacks. For example, zero-injection measurements in the form of systematic modification of monitoring information may perniciously affect the normal operation of the vehicular system.

3 Missing-Data Imputation

To overcome data incompleteness in vehicular systems, a plausible strategy would involve the straightforward exclusion of unobserved samples. Although effective when dealing with a limited number of incomplete samples, this method could introduce bias and lead to the loss of valuable contextual information in vehicular condition monitoring systems. Consequently, a more pragmatic approach to managing missing values is through *imputation*, where the missing values are substituted with one or a set of estimations.

Missing-data imputation represents a paradigm shift from traditional redundancy-based error correction. Instead of retransmitting or over-provisioning network resources, imputation leverages the inherent structure and correlation in vehicular data streams, such as spatial proximity, temporal continuity, and cross-sensor dependencies, to estimate missing or corrupted information. This approach aligns with current trends in AI-native communications envisioned by 3GPP's "AI/ML for the Air Interface" initiatives and ongoing work within ETSI TC ITS and IEEE 802.11bd Next Generation V2X (NGV2X) discussions.

The fundamental problem of missing-data imputation among signals captured from a physical process has been extensively studied in the literature. A comprehensive study of likelihood-based imputation methods with general applicability, such as expectation maximization (EM) and data augmentation, is provided in [4]. While computationally efficient, conventional imputation methods, such as linear, cubic, or nearest neighbor interpolation, fail to incorporate cross-correlation relationships in the ambient measurement space. This motivates the adoption of imputation methods based on singular value decomposition (SVD). The key characteristic of such methods lies in their capacity to infer linear relationships among measurement streams, thereby enabling the reconstruction of missing values in a data stream based on observations from the remaining ones. In [5], the missing values were inferred by applying an iterative scheme based on low-rank decompositions. It was noted, however, that SVD-based approaches do not inherently preserve temporal smoothness and become sensitive to transient content and outliers present in the measurements [6].

Dynamical systems that are capable of exploiting spatiotemporal correlations among measurement streams can be employed in an iterative strategy to estimate missing values [7]. Formally, dynamical systems concern the analysis, prediction, and understanding of the behavior of systems of differential equations or iterative mappings that describe the evolution of the state of a system. This formulation is general enough to cover a wide range of phenomena, including those observed in vehicular systems. Since it is plausible that the time series of measurements may not be sufficiently described by a single linear dynamical system (LDS) as shown in [7], in our previous work [8], we employed more sophisticated models, such as the switching linear dynamical system (SLDS) and the recurrent switching linear dynamical system (rSLDS) [9], to address underfitting.

In [8], missing value imputation is characterized by the following steps. First, missing values are reconstructed using linear interpolation, and parameters are initialized with respect to their prior distributions. Then, belief propagation is used to compute the posterior expectations of latent variables, considering both the aggregation of observed and interpolated data, along with prior parameter values. In the third step, the estimated parameter values are updated through maximum likelihood by using the expected values of latent variables. Finally, imputation is carried out by computing the conditional expectation of the missing values with respect to the values of observed variables, the posterior expectations of latent variables, and the updated parameter values. The iteration proceeds by updating the posterior expectations of latent variables, incorporating the aggregation of observed measurements and the newly imputed values of missing variables, until convergence is achieved.

Within the context of V2X systems, imputation techniques can be integrated into multiple layers of the protocol stack:

- At the radio interface layer, where physical and MAC layer metrics (RSSI, SINR, CQI) inform imputation models of likely data quality.
- At the network and application layers, where contextual vehicular data (e.g., position, speed, heading angle) can support the estimation of missing perception or control messages.
- At edge computing nodes, where aggregated data from multiple vehicles can be fused and reconstructed collaboratively, improving both reliability and situational awareness across the monitoring infrastructure.

This multi-layer integration enables a cross-domain mitigation of radio-induced imperfections, enhancing the dependability of V2X-based inference systems without requiring fundamental redesigns of the underlying physical layer.

4 Imputation Performance

4.1 Dataset Description

The VeReMi dataset [10] includes 19 anomaly types (misbehavior attacks) and models two road traffic densities: high-density (37.03 vehicles/km2) and low-density (16.36 vehicles/km2). A log file per vehicle is generated, which contains basic safety messages (BSM) transmitted by neighboring vehicles over its entire trajectory. Each attack type dataset contains a ground truth file to record the observed behavior of all participating vehicles. BSMs constitute a three-dimensional vector for position, speed, acceleration, and heading angle features. Figure 1 depicts a raw sample of BSM data for a single vehicle.

For subsequent imputation analysis, we have considered the log file for a single vehicle and kept only the genuine information by properly removing the misbehaving attack data, since the attack detection and classification are considered irrelevant tasks to our problem. Synthetic dropouts are then used to generate missing data by selecting space-time points for occlusion. This step is performed using the library *pyampute*, where a multivariate amputation procedure is implemented, enabling the introduction of different missingness patterns. In our experiments, we considered both noncontiguous and contiguous missingness patterns; noncontiguous occlusions are generated by uniformly selecting space–time points for dropout, whereas contiguous occlusions are generated by dropping consecutive measurements of varying time length starting at a random point in time.

	type	sendTime	sender	senderPseudo	messageID	pos	spd	acl	hed
17	4	25210.186332	33	10332	21562	[1393.9276845310885, 1203.692849621629, 0.0]	[0.049400297340067005, -0.686074278731542, 0.0]	[0.166603725521922, -2.313798731172836, 0.0]	[0.063269582720791, -0.9979964728907291, 0.0]
19	4	25210.436332	33	10332	21595	[1393.9276845310885, 1203.692849621629, 0.0]	[0.049400297340067005, -0.686074278731542, 0.0]	[0.166603725521922, -2.313798731172836, 0.0]	[0.063269582720791, -0.9979964728907291, 0.0]
21	4	25210.686332	33	10332	21622	[1393.9276845310885, 1203.692849621629, 0.0]	[0.049400297340067005, -0.686074278731542, 0.0]	[0.166603725521922, -2.313798731172836, 0.0]	[0.063269582720791, -0.9979964728907291, 0.0]
25	4	25210.936332	33	10332	21696	[1393.9276845310885, 1203.692849621629, 0.0]	[0.049400297340067005, -0.686074278731542, 0.0]	[0.166603725521922, -2.313798731172836, 0.0]	[0.063269582720791, -0.9979964728907291, 0.0]
27	4	25211.186332	33	10332	30134	[1394.1720072035407, 1201.94700381985, 0.0]	[0.183983214273645, -2.555169745474803, 0.0]	[0.158360369223177, -2.199314281648395, 0.0]	[0.063269582720943, -0.99799647289072, 0.0]
...	...	...	...	...	...	...	...	...	...
7165	4	25368.936332	33	10332	715829	[127.9440058255349, 885.9631063084606, 0.0]	[-8.275102433876064, -0.48628168363595903, 0.0]	[4.492259181435928, 0.263999685982995, 0.0]	[-0.9703792835172421, 0.24158651891312902, 0.0]
7179	4	25369.186332	33	10332	716850	[122.01936718688863, 885.6254381946134, 0.0]	[-3.793979120628686, -0.21788847407861903, 0.0]	[4.49260375347267, 0.25802779102927603, 0.0]	[-0.9731995855753991, 0.22996209825940903, 0.0]
7192	4	25369.436332	33	10332	717806	[122.01936718688863, 885.6254381946134, 0.0]	[-3.793979120628686, -0.21788847407861903, 0.0]	[4.49260375347267, 0.25802779102927603, 0.0]	[-0.9731995855753991, 0.22996209825940903, 0.0]
7205	4	25369.686332	33	10332	718725	[122.01936718688863, 885.6254381946134, 0.0]	[-3.793979120628686, -0.21788847407861903, 0.0]	[4.49260375347267, 0.25802779102927603, 0.0]	[-0.9731995855753991, 0.22996209825940903, 0.0]
7228	4	25369.936332	33	10332	720286	[122.01936718688863, 885.6254381946134, 0.0]	[-3.793979120628686, -0.21788847407861903, 0.0]	[4.49260375347267, 0.25802779102927603, 0.0]	[-0.9731995855753991, 0.22996209825940903, 0.0]

Figure 1: BSM data for a specific vehicle (ID:33) in VeReMi

4.2 Results

For performance comparison, we considered the following four data imputation schemes: *i*) a reconstruction method that uses a Gaussian mixture model (GMM) regressor with parameters obtained by an EM algorithm [11]; *ii*) a method based on SVD, which transforms the weighted low-rank approximation problem into a maximum-likelihood problem with missing values and approximates them by computing the SVD iteratively [5]; *iii*) a k-means clustering imputation approach where missing values are imputed by the corresponding value from the centroid of the nearest cluster [12]; and *iv*) an rSLDS-based imputation method with Bayesian parameter learning using blocked Gibbs sampling [8]. The performance of those imputation schemes is contingent upon carefully selecting suitable values for their hyperparameters. Considering computational resources, hyperparameters were judiciously tuned by carrying out a limited hyperparameter search following the configurations outlined in each work.

The effectiveness of imputation (reconstruction) is first evaluated in terms of the mean squared error (MSE), defined as the average of the squared differences between the real and reconstructed missing measurements. To reduce random effects, we repeated each simulation 100 times and reported the average MSE. The average MSE is less sensitive to outliers or random fluctuations, offering a more reliable and representative measure of missing-data reconstruction and model fitting performance.

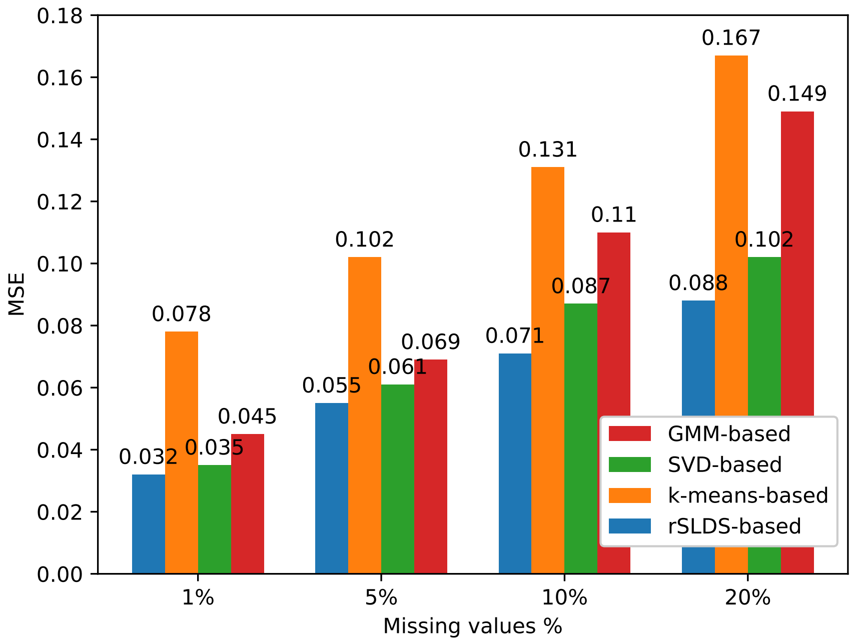

Figure 2: Imputation performance for noncontiguous occlusions.

In the case of noncontiguous occlusions with a varying percentage (1%, 5%, 10%, and 20%) of missing measurements, Figure 2 illustrates the imputation performance in terms of the MSE. For all schemes, it can be observed that, as expected, the imputation performance registers a decline with a growing percentage of missing entries. Notably, rSLDS-based imputation is capable of drawing useful knowledge from the observed measurement streams to make valid inferences for the missing data and achieve the lowest MSE levels. As such, rSLDS-based imputation is shown to be more robust to missingness. This is in stark contrast to the level of degradation registered by the k-means-based and GMM-based methods, which can be attributed to the inability of such schemes to adequately capture the governing nonlinear dynamics, thereby limiting their effectiveness in estimating missing values.

To further shed light on the imputation performance, Figure 3 illustrates the R^2 score for different missingness percentages. The R^2 score depicts how well imputed data align with ground truth. In particular, a decreasing R^2 shows how well the structure is preserved as missingness increases. Compared to other imputation methods, the superiority of rSLDS-based imputation is notable, as it maintains a high R^2 score, thereby preserving the structure of the original data.

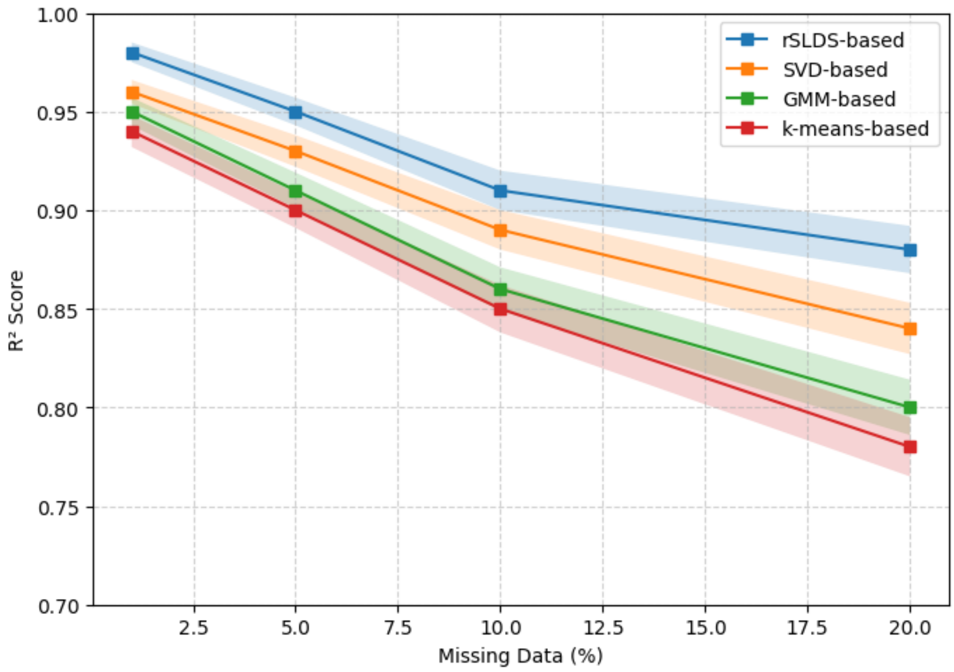

Figure 3: R^2 between imputed and true values for noncontiguous occlusions. The shaded regions show the confidence intervals across repetitions (e.g., multiple random missingness patterns).

The computational performance of the various imputation methods is compared in Figure 4. It can be observed that, with an increasing percentage of missingness, the computation time required for imputation increases, making certain methods, such as GMM-based imputation, scale poorly. It is also noted that the increased accuracy of rSLDS-based imputation comes with the inadvertent cost of elevated computational complexity (i.e., in terms of the computation time) compared with SVD-based imputation.

In the case of contiguous occlusions with varying occlusion lengths (i.e., in terms of consecutive missing time steps), similar insights can be drawn. As shown in Table 1, rSLDS-based imputation achieves the lowest MSE scores among the examined methods. In particular, even in the case of 50 consecutive missing entries, where the temporal information is limited to

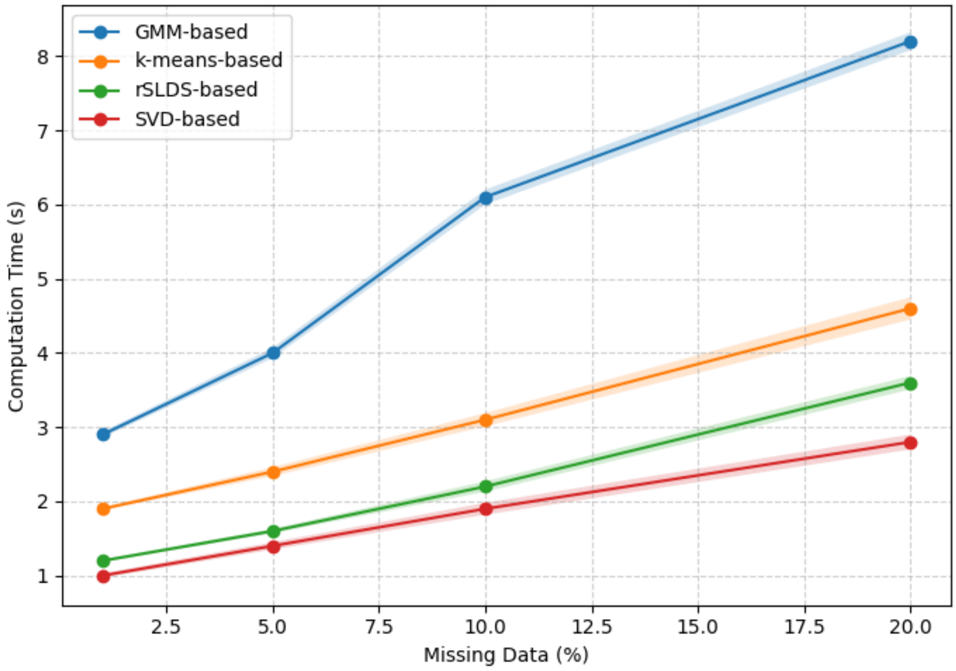

Figure 4: Computation time vs. missing data for noncontiguous occlusions. The shaded regions show uncertainty or variability (e.g., over multiple runs).

relatively distant time steps, the performance of the MSE does not escalate to prohibitive levels as observed with the k-means-based and GMM-based methods. Similar to the trend reported in Figure 4, the increased accuracy of rSLDS-based imputation comes at the expense of increased computation time compared with SVD-based imputation. Notably, the elevated computation times compared to the case of noncontiguous occlusions may pose a challenge for real-time vehicular applications with stringent latency requirements.

Imputation method	**Metric**	**Occlusion length (time steps)**				
		10	**20**	**30**	**40**	**50**
GMM-based [11]	MSE	0.122	0.167	0.253	0.338	0.41
	Time (s)	35.16	44.25	56.37	68.91	86.67
SVD-based [5]	MSE	0.086	0.11	0.147	0.172	0.205
	Time (s)	10.41	12.02	14.95	22.91	29.34
k-means-based [12]	MSE	0.146	0.198	0.385	0.567	0.662
	Time (s)	20.46	28.38	34.57	49.16	64.21
rSLDS-based [8]	MSE	0.067	0.093	0.118	0.133	0.156
	Time (s)	11.16	15.61	18.88	26.67	37.21

Table 1: Imputation performance and computation times for contiguous occlusions

5 Alignment with Standardization Activities

The proposed imputation enhancements align closely with ongoing standardization trends emphasizing data resilience and AI integration in next-generation vehicular communications:

- 3GPP Release 18/19 (C-V2X Evolution) includes studies on AI-native functionalities for radio resource management and reliability enhancement. The inclusion of imputation within the radio interface can be positioned as a complementary mechanism for data reliability assurance.
- IEEE 802.11bd and future NGV2X specifications are expanding support for machine-learning-assisted link adaptation and reliability prediction, providing a framework that can incorporate data imputation for upper-layer robustness.
- ETSI ITS-G5 and ETSI EN 302 890-2 emphasize cooperative awareness and perception data sharing. Imputation can directly enhance the cooperative perception messages and collective environment model frameworks by reconstructing missing sensory data.
- ISO TC204 WG16 promotes interoperability and cross-technology data consistency. Imputation aligns with these goals by ensuring consistent data availability across heterogeneous communication interfaces.

Thus, the integration of data-driven imputation within the V2X radio interface is not a divergence from standardization but a natural evolution aligned with the industry's movement toward AI-empowered and context-aware communication systems. Incorporating imputation techniques at the radio interface level is expected to yield substantial performance benefits, such as

- Reduction of effective data loss rates without increasing spectral overhead.
- More robust operation of cooperative perception and decision-making algorithms in the presence of imperfect network conditions.
- Reduced need for retransmissions or redundancy mechanisms, optimizing spectrum and energy utilization.
- Compatibility with existing V2X protocol stacks and ease of integration into emerging 5G-advanced and 6G vehicular frameworks.

6 Final remarks

In summary, this deliverable introduces a targeted enhancement to the V2X radio interface through the adoption of missing-data imputation. Using an open-source dataset, and creating synthetic dropouts, we have evaluated the reconstruction error for different imputation methods and missing data configurations. The objective was to assess the ability of imputation to mine measurement streams under incomplete received data. Performance assessment revealed that rSLDS-based imputation is capable of extracting knowledge from the spatiotemporal synergy among the respective measurements to make valid inferences for the missing V2X data.

By addressing the limitations imposed by unreliable wireless channels at the data layer, imputation provides a complementary, AI-enabled pathway to improved reliability, efficiency, and performance. The integration into existing and emerging V2X frameworks represents a forward-looking approach that aligns with the standardization community's efforts toward intelligent, resilient vehicular communication systems.

References

[1] P. Mulinka, C. Kalalas, M. Dzaferagic, I. Macaluso, D. G. Rojas, P. J. Nardelli, and N. Marchetti, "Information processing and data visualization in networked industrial systems," in *2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2021, pp. 1–6.

[2] Y. Liu, T. Dillon, W. Yu, W. Rahayu, and F. Mostafa, "Missing Value Imputation for Industrial IoT Sensor Data With Large Gaps," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 6855–6867, 2020.

[3] D. Gutierrez-Rojas, C. Kalalas, I. Christou, G. Almeida, E. Eldeeb, S. Bakri, N. Marchetti, J. M. S. Sant'Ana, O. L. Alcaraz López, H. Alves, C. Papadias, M. Haroon Tariq, and P. H. J. Nardelli, "Detection and classification of anomalies in wsn-enabled cyber-physical systems," *IEEE Sensors Journal*, vol. 25, no. 4, pp. 7193–7204, 2025.

[4] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*. Hoboken, NJ: John Wiley & Sons, 2002.

[5] N. Srebro and T. Jaakkola, "Weighted Low-Rank Approximations," in *Proceedings of the 20th International Conference on Machine Learning, ICML 2003*, vol. 3, 2003, pp. 720–727.

[6] H.-F. Yu, N. Rao, and I. S. Dhillon, "Temporal Regularized Matrix Factorization For High-Dimensional Time Series Prediction," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, ser. NIPS'16. Red Hook, NY, USA: Curran Associates Inc., 2016, p. 847–855.

[7] L. Li, J. McCann, N. S. Pollard, and C. Faloutsos, "DynaMMo: Mining and Summarization of Coevolving Sequences with Missing Values," in *Proc. of the 15th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD)*, 2009, pp. 507–515.

[8] T. A. Alexopoulos, C. Kalalas, and G. N. Korres, "On the Imputation of Power System Measurement Streams with Imperfect Wireless Communication," in *2020 2nd Global Power, Energy and Communication Conference (GPECOM)*, 2020, pp. 302–307.

[9] S. Linderman, M. Johnson, A. Miller, R. Adams, D. Blei, and L. Paninski, "Bayesian Learning and Inference in Recurrent Switching Linear Dynamical Systems," in *Proc. of the 20th International Conference on Artificial Intelligence and Statistics*, 2017, pp. 914–922.

[10] J. Kamel, M. Wolf, R. W. van Der Heijden, A. Kaiser, P. Urien, and F. Kargl, "Veremi extension: A dataset for comparable evaluation of misbehavior detection in vanets," in *2020 IEEE International Conference on Communications (ICC)*, Dublin, Ireland, Jun 2020.

[11] D. S. Silva and C. V. Deutsch, "Multivariate Data Imputation Using Gaussian Mixture Models," *Spatial Statistics*, vol. 27, pp. 74–90, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2211675316301300

[12] B. Patil, R. Joshi, and D. Toshniwal, “Missing Value Imputation Based On K-Means Clustering With Weighted Distance,” in *Communications in Computer and Information Science*. Berlin: Springer, 2010, pp. 600–609.