UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

DANILO AVILAR SILVA

MATÉRN KERNEL FOR INCOMPLETE DATA

FORTALEZA
2023

DANILO AVILAR SILVA

MATÉRN KERNEL FOR INCOMPLETE DATA

Tese apresentada ao Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutor em Ciência da Computação. Área de Concentração: Aprendizagem de Máquina e Modelos Estatísticos.

Orientador: Prof. Dr. João Paulo Pordeus Gomes.
Coorientador: Prof. Dr. César Lincoln Cavalcante Mattos.

FORTALEZA

2023

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

S579m Silva, Danilo Avilar.
Matérn Kernel for Incomplete Data / Danilo Avilar Silva. – 2023.
102 f. : il. color.

Tese (doutorado) – Universidade Federal do Ceará, Centro de Ciências, Programa de Pós-Graduação em Ciência da Computação , Fortaleza, 2023.
Orientação: Prof. Dr. João Paulo Pordeus Gomes.
Coorientação: Prof. Dr. César Lincoln Cavalcante Mattos.

1. Missing Data. 2. Gaussian Mixture Model. 3. Approximation Methods for Functions. 4. Matérn Kernel. I. Título.

CDD 005

DANILO AVILAR SILVA

MATÉRN KERNEL FOR INCOMPLETE DATA

Tese apresentada ao Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutor em Ciência da Computação. Área de Concentração: Aprendizagem de Máquina e Modelos Estatísticos.

Aprovada em: 31/08/2023.

BANCA EXAMINADORA

Prof. Dr. João Paulo Pordeus Gomes (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. César Lincoln Cavalcante Mattos (Coorientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. João Paulo do Vale Madeiro
Universidade Federal do Ceará (UFC)

Prof. Dr. Ajalmar Rêgo da Rocha Neto
Instituto Federal do Ceará (IFCE)

Prof. Dr. Diego Parente Paiva Mesquita
Faculdade Getúlio Vargas (FGV)

Dedico este trabalho especialmente ao meu pai Manoel (in memorian), à minha mãe Antonia e meu irmão Daniel, exemplos de caráter e desvelo à família; À minha amada esposa Vilma e minha querida filha, Laura, meu xodó.

AGRADECIMENTOS

Agradeço, primeiramente, a Deus pela sabedoria e força para enfrentar todos os obstáculos e momentos delicados que a vida nos proporciona.

Obrigado ao meu pai, Manoel (in memoriam), que mesmo não estando mais presente entre nós, se faz vivo em minhas lembranças diariamente e em meu coração, sempre. Grande parte da força empregada na conclusão deste trabalho se resume a uma forma de agradecimento por tudo que fez por mim ao longo da vida. Seus ensinamentos e amor continuarão sendo fonte de inspiração em minha vida. De uma forma ou de outra, busco aqui cumprir minha promessa de que me tornaria Doutor quando crescesse. Quero agradecer à minha mãe, Antônia, por toda dedicação, carinho e paciência. Ao meu irmão, Daniel, por toda força e parceria nos momentos difíceis, pelo apoio e por sempre se colocar à disposição para ajudar no que for preciso. Vocês são a base de tudo em minha vida.

Agradeço à minha esposa, Vilma, pelo apoio incondicional às minhas decisões, mesmo sabendo que o preço de algumas delas custa mais ausência junto a você...Obrigado por sempre manter o pensamento positivo, por todo amor e carinho. Agradeço à minha filha, Laura, por todos os momentos de descontração e incentivo. Mesmo com seus 3 anos de idade, sua intenção de pegar o teclado do computador seguida da frase proferida "*me dê aqui, papai, deixa eu te ajudar a trabalhar...*" traz uma leveza indescritível.

Agradeço a todos os familiares (estejam perto ou distantes fisicamente) por sempre se fazerem presentes de alguma forma. Estendo meus agradecimentos à minha cunhada, Ciça, pela constante disposição em ajudar na revisão textual. Muito obrigado, Mãe, Tia Avilar, minha sogra Rosimeire e Tia Regina, pela rede de apoio formada aos cuidados com a pequena Laura nos momentos de ausência.

Obrigado aos Professores João Paulo Pordeus e César Mattos pelo apoio, ensinamentos, compreensão e por acreditarem até o fim que daria tudo certo. Ao longo da execução deste trabalho, pude construir uma relação de amizade e respeito. A dedicação de vocês foi crucial na elaboração desta Tese. Estendo meus agradecimentos ao Professor Diego Mesquita por toda contribuição, suas atenciosas considerações fornecidas durante nossas reuniões foram essenciais para a elaboração deste trabalho.

Gostaria ainda de agradecer:

– Aos amigos servidores do IFCE George Frederick, Diná Santana, Kaio Jonathas e Raphael Carvalho, pelo empenho que permitiu meu afastamento integral;

- Aos colegas guerreiros da disciplina de Otimização Não-Linear de Sistemas pelas conversas, estudos e risadas nos momentos de leve desorientação sobre a prova da convexidade das funções;
- Aos demais colegas com os quais tive oportunidade de aprender e ajudar ao longo do Doutorado, ao corpo docente pela formação acadêmica e aos demais servidores do Departamento de Computação da UFC pela colaboração;
- Aos colegas do grupo de estudos PRAIA por toda troca de conhecimentos ao longo dos encontros;
- Aos meus amigos por compreenderem os momentos de ausência.

A menos que modifiquemos a nossa maneira de pensar, não seremos capazes de resolver os problemas causados pela forma como nos acostumamos a ver o mundo. (Albert Einstein).

RESUMO

Problemas de aprendizado de máquina com dados incompletos são constantemente abordados em diversos domínios do mundo real. Métodos estatísticos que lidam com atributos ausentes caracterizam-se por suposições sobre a distribuição de dados através de uma função de densidade. Diante desse contexto, abordagens para utilização de métodos baseados em medidas de similaridade, tornam-se objetos de pesquisa bastante promissores, uma vez que esses métodos geralmente assumem que os dados são totalmente observados e não são equipados naturalmente para lidar com dados incompletos. Neste trabalho, serão propostos métodos para estimar o valor esperado do Kernel Matérn na presença de vetores de dados incompletos sem nenhuma etapa de pré-processamento. Os métodos Expected Matérn Kernel via Monte Carlo Method (EMK-MC) e Expected Matérn Kernel via Unscented Transform (EMK-UT) apresentam a capacidade de abordar o problema de estimativa do kernel estimando a transformação de interesse, ao invés de lançá-la em uma estrutura de pré-processamento. Para obter tais estimativas, os vetores incompletos são tratados como variáveis aleatórias contínuas e, a partir da suposição que a distância Euclidiana entre pontos de interesse seguem uma distribuição Nakagami, métodos de amostragem são utilizados para gerar pontos que dependem apenas da distribuição de interesse. Por meio de um modelo de mistura de Gaussianas, a distribuição dos dados é aproximada a partir da estimativa de máxima verossimilhança via algoritmo Expectation-Maximization, e ao mesmo tempo, estima os valores ausentes de forma iterativa. Isso permite que o modelo seja ajustado aos dados observados, levando em consideração a incerteza dos valores ausentes e as relações entre as variáveis. Os desempenhos dos métodos propostos são comparados à três métodos em conjuntos de dados reais e sintéticos. Em função da raiz do erro médio quadrático obtido ao computar a diferença entre o valor estimado do kernel e o valor real, a consistência do desempenho alcançado se mantém evidente na maioria dos cenários avaliados para bases do mundo real, sendo os métodos propostos EMK-MC e EMK-UT, melhores em cerca de 43% e 38% dos cenários avaliados, respectivamente. No que se refere aos cenários avaliados em conjuntos de dados sintéticos, as abordagens propostas são melhores em todos os cenários avaliados.

Palavras-chave: Dados Ausentes; Modelo de Mistura de Gaussianas; Métodos de Aproximação de Funções; Kernel Matérn.

ABSTRACT

Machine learning problems with incomplete data are constantly addressed in various real-world domains. Statistical methods dealing with missing attributes are characterized by assumptions about data distribution through a density function. In this context, approaches that utilize similarity-based methods, become very promising research objects since these methods generally assume that data is fully observed and are not naturally equipped to handle incomplete. In this work, methods will be proposed to estimate the expected value of the Matérn Kernel in the presence of incomplete data vectors without any preprocessing steps. The EMK-MC and EMK-UT methods demonstrate the capability to address the kernel estimation problem directly, meaning they estimate the transformation of interest instead of embedding it within a preprocessing framework. To obtain such estimates, incomplete vectors are treated as continuous random variables, and based on the assumption that the Euclidean distance between points of interest follows a Nakagami distribution, sampling methods are used to generate points that depend only on the distribution of interest. Through a Gaussian mixture model, the data distribution is approximated by maximum likelihood estimation via the Expectation-Maximization algorithm, while simultaneously iteratively estimating the missing values. This allows the model to be fitted to the observed data, considering the uncertainty of the missing values and the relationships between variables. The performances of the proposed methods are compared to three methods on real and synthetic datasets. Considering the root mean square error obtained by computing the difference between the estimated kernel value and the true value, the consistency of the achieved performance remains evident in the majority of the scenarios evaluated for real-world datasets. The proposed methods, EMK-MC and EMK-UT, are superior in approximately 43% and 38% of the evaluated scenarios, respectively. As for the scenarios evaluated in synthetic datasets, the proposed approaches outperform all evaluated scenarios.

Keywords: Missing Data; Gaussian Mixture Model; Approximation Methods for Functions; Matérn Kernel.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

ARMSE	Average Root Mean Square Error
BIC	Bayesian Information Criterion
CMI	Conditional Mean Imputation
ECM	Expectation Conditional Maximization
EED	Expected Euclidean Distance
EGK	Expected Gaussian Kernel
EM	Expectation-Maximization
EMK-MC	Expected Matérn Kernel via Monte Carlo Method
EMK-UT	Expected Matérn Kernel via Unscented Transform
ESD	Expected Square Distance
GMM	Gaussian Mixture Model
GMMs	Gaussian Mixture Models
GP	Gaussian Process
GPs	Gaussian Processes
IS	Importance Sampling
LSSVR	Least Squares Support Vector Regression
MAR	*Missing at Random*
MC	Monte Carlo
MCAR	*Missing Completely at Random*
MCMC	Markov Chain Monte Carlo
MNAR	*Missing Not at Random*
PDF	Probability Density Function
p^{IMP}	Percentage of Improvement
$rMiss$	Rate of Missing Instances
RS	Rejection Sampling
RSR	Relative Success Rate
SVR	Support Vector Regression
UT	Unscented Transform

LIST OF SYMBOLS

α, β	Parameters of the Gamma distribution
$\hat{\alpha}$	Dispersion of sigma points
$\hat{\beta}, \kappa, \lambda$	Scale parameters of the UT
b	Bias
C	Number of components of a GMM
$\mathcal{C}_\nu(\cdot)$	Matérn Kernel (general definition)
D	Dimension of a vector
$\boldsymbol{\delta}$	Set of transformed points
e	Approximation error in a regression task
$\mathbb{E}$	Expected value
η	Euclidean distance between incomplete data vectors
$f(\cdot)$	Function of a random variable
$\Gamma(\cdot)$	Gamma function (abbreviated notation)
$\boldsymbol{\gamma}$	Sigma points
$h(\cdot)$	Non-linear transformation
$\overline{h}_e$	Empirical mean
ι	Lagrange multipliers
Jp	Cost function of the Least Squares Support Vector Regression method
$k_{\mathcal{C}_\nu(\cdot)}$	Matérn Kernel function
$K_\nu(\cdot)$	Bessel function
$L(\cdot)$	Lagrangian function
$\mathcal{L}(\cdot)$	Likelihood function
$\boldsymbol{M}$	Probability model
mis	Set of missing entries
m, Ω	Parameters of the Nakagami distribution
$\boldsymbol{\mu}$	Mean vector

$\tilde{\boldsymbol{\mu}}$	Conditional mean vector
$\mathcal{N}(\cdot)$	(Multivariate) Normal PDF
N	Number of samples in a dataset
ν	Order of the Bessel function
$O(\cdot)$	Asymptotic notation
obs	Set of observed entries
$\boldsymbol{\omega}$	Weight vector
$\boldsymbol{\Psi}$	Kernel matrix of the Least Squares Support Vector Regression method
$P(\cdot)$	Probability of missing attributes
$p(\cdot), g(\cdot)$	Probability distributions
$\boldsymbol{\pi}$	Mixture weight associated with a GMM component
ψ	Approximation of the distribution of η (after imputation) via Monte Carlo simulation
σ	Standard deviation
σ^2, ρ	Parameters of the Matérn Kernel
τ	Probability of membership
$\boldsymbol{\Theta}$	Set of parameters of a GMM
$\mathcal{U}$	Uniform PDF
Var	Covariance
φ	Approximation of the distribution of η (after imputation) via UT
$\varrho(\cdot)$	Nonlinear mapping
$\boldsymbol{\Sigma}$	Covariance matrix
$\tilde{\boldsymbol{\Sigma}}$	Conditional covariance matrix
$\mathcal{X}$	Set of feature vectors
$\boldsymbol{X}$	Feature vector
$\tilde{\boldsymbol{X}}$	Imputed feature vector
$\boldsymbol{X}_{mis}$	Missing entries
$\boldsymbol{X}_{obs}$	Observed entries

x	Component of a feature vector
$\boldsymbol{\xi}$	Vector of unknown parameters
Y	Output dataset
z	Squared distance between incomplete data vectors
ζ	Regularization term

CONTENTS

1 INTRODUCTION

One of the most important premises in machine learning is related to data integrity. Depending on the complexity of the problem, missing attributes can significantly affect the performance of a model due to bias in the dataset. Incomplete data is characterized as records that contain unobserved or missing information, which can result from measurement errors, device or operator failure (EIROLA *et al.*, 2013), non-response, deliberate data withholding or errors in database files (LITTLE; RUBIN, 2002).

However, one of the most common problems in the real world is the missing of data. This situation can be observed in various domains, such as the financial market (ZHOU; LAI, 2017), clinical data analysis (GULER *et al.*, 2020; MADHU *et al.*, 2019), biomedical research (LUO *et al.*, 2020), genetic data interactions (LIU *et al.*, 2020), time series analysis (TANG *et al.*, 2020; DU *et al.*, 2019; ZHANG *et al.*, 2019). The missing data problem it happens for one or more attributes, or even for complete samples in a dataset (De Souto *et al.*, 2015; SILVA-RAMÍREZ *et al.*, 2015).

Once there are irregularities in the information, the ability of the final model of machine learning methods will be directly affected, as they adjust their models to the data used in the training step. A simple yet controversial strategy involves ignoring samples that contain missing data, so that the dataset is reduced only to fully observed samples. According to Fernstad and Glen (2014), missing records can be considered potentially important information; therefore, several problems can be related to discarding samples with such records, namely (KANG, 2013): reduction in statistical power, in terms of the probability of the test rejecting the null hypothesis when it is false; complications in the study's analysis; bias in parameter estimation; or even a reduction in sample representativeness.

A class of methods that has been widely used for data analysis, including incomplete data, is kernel methods. With this technique, kernel functions are used to estimate probability density or perform other types of analysis, such as classification or regression. These methods have the advantage of being flexible and capable of handling complex structures in the data. The kernel functions, used to calculate inner products between attribute vectors in the input feature space, implicitly define the mapping of the data to a high-dimensional feature space. As a consequence of the mapping, it is possible to use machine learning algorithms to find linear patterns from nonlinear problems by increasing the dimensionality of the data used, enabling them to be separated by a hyperplane (SCHÖLKOPF *et al.*, 2002; SHAWE-TAYLOR; CRISTIANINI, 2004) [1]. There are several types of kernels, and among the most commonly used in machine learning are: Linear Kernel, Polynomial Kernel, and Gaussian Kernel. Another class of kernel functions, more popularly used in Gaussian processes, is the Matérn Kernel (see Murphy (2012)), which specifies the covariance between measurements as a function of the Euclidean distance between points of interest.

Although widely used, kernel functions have limitations when applied directly to datasets with missing values. Traditionally, in a preprocessing step, two approaches are considered to handle missing data (FARHANGFAR *et al.*, 2007). The first approach involves excluding samples that have one or more missing attributes, retaining only the samples with fully observed data. As discussed earlier, this approach can disregard potentially important information in the dataset. The second approach is to impute the missing data by replacing it with a plausible value (MADHU *et al.*, 2019), allowing for the use of any conventional learning method after filling in the missing entries.

Such techniques can be divided into two groups: single imputation and multiple imputation. In the first group, the missing value is replaced with a single value, while in the second group, multiple values are imputed for each missing data point. Thus, multiple imputation transforms the originally incomplete dataset into multiple complete observed datasets. Finally, a point estimate of the missing attribute can be obtained by averaging the imputations, and the standard error can be calculated from the variance of the multiple imputations.

However, it is worth noting that imputation is an estimation process, and as such, there is uncertainty associated with the values that will be imputed. When comparing the characteristics of imputation methods, it can be inferred that single imputation methods do not account for uncertainty because they use point estimates. On the other hand, dealing with multiple complete observed datasets generated by multiple imputation in a preprocessing step is computationally demanding.

Studies explicitly aiming to estimate kernel functions when dealing with incomplete datasets are scarce. The multiple imputation process is used as a preprocessing step in two kernel-based imputation models: a non-parametric stochastic imputation

[1] This characteristic is associated with any model that uses a kernel function.

model based on Gaussian Kernel (ZHANG *et al.*, 2006) and a semi-parametric random imputation model based on Polynomial Kernel (ZHANG *et al.*, 2009).

Although Belanche *et al.* (2014) present two proposals for handling missing values in kernel methods without any preprocessing step, the contribution is limited to datasets with exclusively binary attributes. In other words, the extended kernels presented are adapted to accommodate the discrete nature of binary data.

In Jafrasteh *et al.* (2023), a hierarchical composition of sparse Gaussian Processes (GPs) is proposed to obtain the predictive distribution for missing values. In this approach, each dimension of the hierarchy uses the dataset from the previous dimension to make predictions. However, it should be noted that the covariance function computed by each sparse Gaussian Process (GP) is not estimated directly. Initially, the inducing points are estimated based on the mean value of the corresponding dimension. Then, the missing values of the first attribute are replaced with the prediction from the corresponding GP, which is a random variable determined by the predictive distribution. This updated dataset is used to feed a second GP, and the process is iterated until all dimensions with missing values have an associated GP predicting their values.

The method called Expected Gaussian Kernel (EGK) (MESQUITA *et al.*, 2019), obtains the expected value of the Gaussian kernel function through a simple closed-form solution based on the moment generating function. This solution only depends on the parameters of the distribution that represents the squared distances between the possibly incomplete data vectors. While the EGK method provides a closed-form solution for the expected value of the Gaussian kernel function, it focuses on the Gaussian kernel specifically and does not extend to the Matérn kernel.

This work presents two novel methods for directly estimating the Matérn Kernel, that is, without any preprocessing step, namely: EMK-MC estimation and EMK-UT estimation. Under the assumption that the Euclidean distances in the kernel function follow a specific probability distribution, approximation methods can be used to estimate the value of the Matérn Kernel for two potentially incomplete data vectors. The need to resort to approximation methods arises from the infeasibility of analytically solving the formulated problem.

For the estimation to be possible, it is necessary to approximate the data set's distribution. A commonly used approach is parametric density estimation, where an underlying probabilistic distribution of the data is assumed to estimate the parameters of that distribution, resulting in density modeling. In this specific work, a Gaussian mixture distribution is considered as the parametric model. This distribution is a weighted combination of several individual Gaussian distributions (each with its own parameters), and the estimation of the mixture's parameters is performed using the maximum likelihood method.

However, the conventional expectation-maximization approach to estimate the parameters of the Gaussian mixture distribution assumes that all data is completely observed, which is different from the scope addressed here. To deal with this issue, a modification of this method will be used to specifically handle incomplete data (further details are discussed in Hunt and Jorgensen (2003)). This modification takes into account the presence of missing values during the parameter estimation process, making it suitable for dealing with incomplete data sets. Given the formulation, the proposed models can be easily incorporated into methods based on the Matérn Kernel.

1.1 Objectives

The main objective of this work is to perform a comparative study of similarity estimation methods that take into consideration the uncertainty of the imputation process. The aim is to enable the use of machine learning methods based on the Matérn Kernel in incomplete data sets. Additionally, this study seeks to propose and evaluate the use of approximation methods as an alternative to explicitly represent the formulated estimation models.

In a more specific sense, this thesis has the following specific objectives:

1. To propose and evaluate two new methods for estimating the expected value of the Matérn Kernel for classification and regression problems using:
   - Monte Carlo Method
   - Unscented Transform
2. To evaluate the use of approximation techniques as an alternative to an explicit representation of the formulated estimation models for incomplete data sets
3. To incorporate the uncertainty of the imputation process intrinsically into the resolution of the estimation model
4. To perform the estimation of the Matérn Kernel in a way that is arbitrarily more

accurate even in the presence of a large number of missing attributes

1.2 Contributions of the Author

This work presents two new approaches to estimate the Matérn Kernel in the presence of incomplete datasets. The proposed methods, Expected Matérn Kernel via Monte Carlo Method (EMK-MC) e Expected Matérn Kernel via Unscented Transform (EMK-UT), directly handle missing data and consider the uncertainty associated with the estimation process. By formulating the estimation problem under the assumption that the Euclidean distances in the kernel function follow a Nakagami distribution, it becomes possible to use approximation techniques to address the intractability of the problem. From the Relative Success Rate (RSR), obtained from a comprehensive comparative performance analysis between the proposed approaches and existing methods (considering both synthetic and real datasets), the EMK-MC e EMK-UT methods consistently outperform in most of the evaluated test scenarios, even in the presence of significant amounts of missing data. It is worth noting that the estimates of the Matérn Kernel's expectation resulting from the proposed approaches are computed directly, depending solely on the parameters of the Euclidean distance distribution between the incomplete data vectors.

1.3 Thesis Structure

The rest of this work is organized as follows: Theoretical Foundation; Matérn Kernel Expectation; Computational Simulations; and finally, Conclusions and Future Work.

Chapter 2 explores the necessary concepts and prerequisites for a proper understanding of the proposed approaches. It begins by presenting terminology related to missing data. Then, it discusses data modeling using a Gaussian Mixture Model and the Expectation-Maximization technique extended to handle missing data. Lastly, traditional techniques for approximating integrals are briefly presented.

Chapter 3 introduces relevant techniques found in the literature related to methods based on similarity measures, such as distances or kernel functions. It then presents the formulation of the estimation problem for the Matérn Kernel from incomplete data.

Also in Chapter 3, the proposed methods EMK-MC e EMK-UT are described. The main difference between the techniques lies in the approximation method used as an alternative to the intractability of the formulated estimation problem. While the EMK-MC method requires sampling from a distribution generated by the Monte Carlo Method, the EMK-UT method uses a small set of transformed points to estimate the Matérn Kernel's expectation, resulting from the Unscented Transform.

A comprehensive experimental evaluation of the proposed method's capability is presented in Chapter 4, considering both synthetic and real-world datasets. Tables and graphs with the obtained results are presented.

Finally, the conclusions and future work are discussed in Chapter 5.

2 THEORETICAL BACKGROUND

In this chapter, basic concepts necessary for a better understanding of the techniques and developments in the following chapters are introduced. For a more in-depth understanding of each subject, it is recommended for the reader to consult the references cited throughout the text.

In Section 2.1, the mechanisms that lead to missing data are explored and analyzed. Understanding the nature of these mechanisms allows for the development of effective strategies to deal with this situation. In Section 2.2, a statistical technique used to model complex datasets is presented. The basic principles of a Gaussian Mixture Model are discussed, including parameter estimation and inference of latent components. In Section 2.3, the Expectation-Maximization algorithm applied to datasets with missing values is investigated. Details of the algorithm, such as handling uncertainty in missing data and parameter estimation for the Gaussian Mixture Model, are addressed. Finally, in Section 2.4, two approximation methods frequently used as an alternative to infeasible exact analytical solutions for complex problems are discussed.

2.1 Missing Data Mechanism

One issue to be considered relates to the mechanisms that lead a certain data point to be missing and, in particular, to the question of whether the missingness is related to the underlying values of the dataset attributes. Such mechanisms are of great importance, as the properties of imputation methods are related to the nature of dependencies of these mechanisms. In Little and Rubin (2002), three classification categories for missing data are presented: *Missing Completely at Random* (MCAR), *Missing at Random* (MAR) and *Missing Not at Random* (MNAR).

Let $\boldsymbol{X}_{obs}$ and $\boldsymbol{X}_{mis}$ represent the observed and missing entries, respectively, of a $D-$dimensional random vector $\boldsymbol{X}$. Also let $\boldsymbol{M}$ a probability model. Different attributes in the dataset may be missing for different reasons, and the question is to identify why they are missing. The mechanism of missing data is characterized by the distribution $P(\boldsymbol{M}|\boldsymbol{X}, \boldsymbol{\xi})$ where $\boldsymbol{\xi}$ is a vector of unknown parameters.

In the MCAR mechanism, the missing data is independent, i.e., completely random. There is no relationship between the values of $\boldsymbol{X}_{obs}$ or $\boldsymbol{X}_{mis}$ and the probability of an entry being missing in $\boldsymbol{X}$. This can be expressed as follows:

$$P(\boldsymbol{M}|\boldsymbol{X}_{obs}, \boldsymbol{X}_{mis}, \boldsymbol{\xi}) = P(\boldsymbol{M}|\boldsymbol{\xi}) \ \forall \ \boldsymbol{X}, \boldsymbol{\xi}.$$

In the MAR category, the probability of a component X_n of $\boldsymbol{X}$ being missing is related to other attributes in the dataset, so that the propensity for missingness is not related to other values of $\boldsymbol{X}_{mis}$, but rather to some data in $\boldsymbol{X}_{obs}$. This can be expressed as follows:

$$P(\boldsymbol{M}|\boldsymbol{X}_{obs}, \boldsymbol{X}_{mis}, \boldsymbol{\xi}) = P(\boldsymbol{M}|\boldsymbol{X}_{obs}, \boldsymbol{\xi}) \ \forall \ \boldsymbol{X}_{mis}, \boldsymbol{\xi}.$$

Finally, in the MNAR mechanism, the data is not randomly missing but is related to the values of $\boldsymbol{X}_{mis}$. In other words, the probability of missingness is intrinsically related to the value itself. More comprehensive details about mechanisms of missing data are discussed in Molenberghs *et al.* (2014).

Example 1 *Consider a survey about vehicle owners satisfaction with the quality of the sound system operated through a touch screen. Participants between 18 and 72 years of age took part in the survey. For this purpose, additional information such as age and monthly income is collected. However, during the data collection, some survey participants did not respond to all questions.*

The MCAR situation is depicted in Figure 1. For instance, participants who did not respond to the age question might have skipped it by chance, without any relation to the sound system quality or monthly income. The missing responses about the participant's age have no specific relationship with age itself or any other collected information. In other words, the missingness is completely random and not related to any specific information.

When analyzing the information presented in Figure 1, through the distribution that relates the quantity of missing responses by age group, it can be observed that it is not possible to associate the absence of responses with any other collected information, whether it be monthly income or the quality of the sound system.

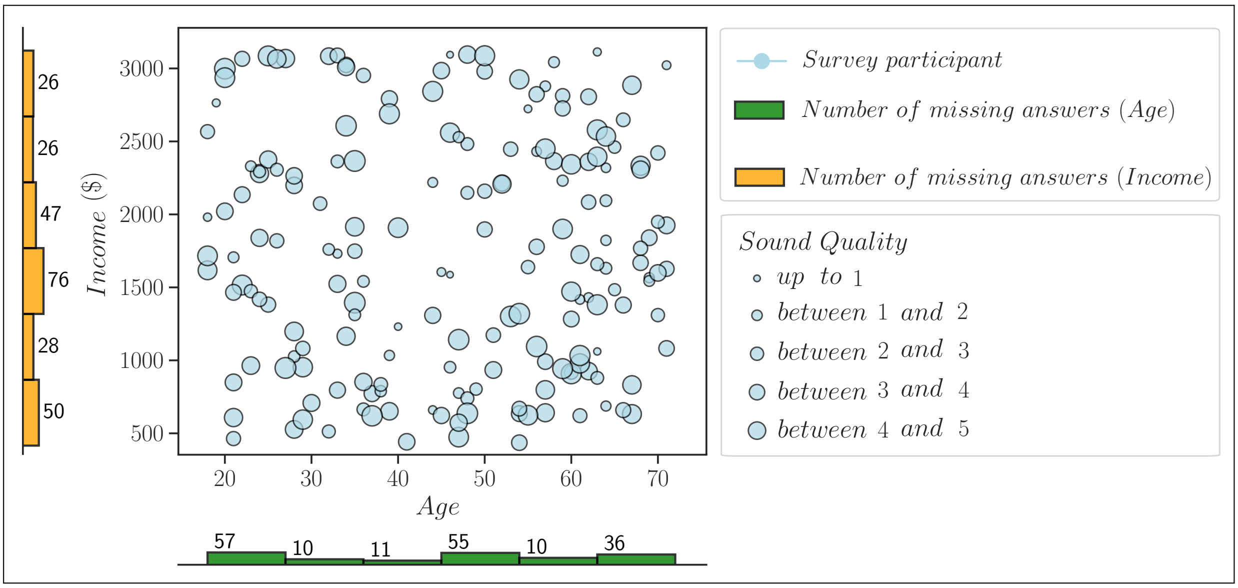


Figure 1 – Hypothetical situation illustrating the missing data mechanism MCAR.
Source: elaborated by the author.

On the other hand, the situation of MAR is depicted in Figure 2. Suppose now that owners over 50 years of age simply do not frequently use the sound system due to personal reasons, which could range from difficulty concentrating while driving and listening to music to a lack of ability to operate a touch screen with multiple options.

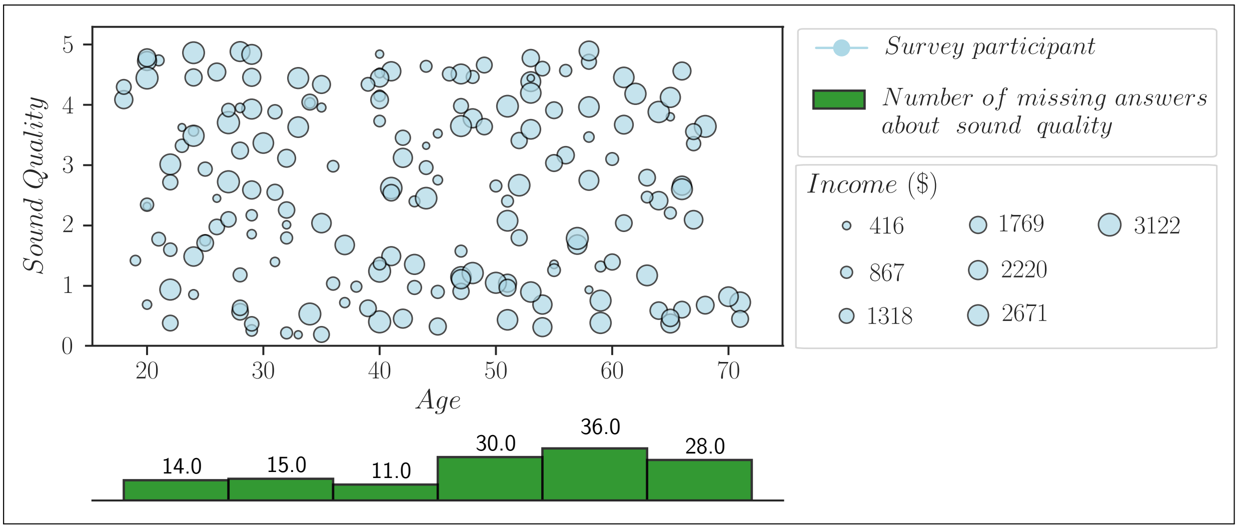


Figure 2 – Hypothetical situation illustrating the missing data mechanism MAR.
Source: elaborated by the author.

Analyzing the above figure, it is possible to identify that participants over 50 years old are less likely to respond about the quality of the sound system. Through the distribution that relates the quantity of missing responses about the sound system's quality, it becomes clear that the lack of responses is directly associated with the collected information about the participant's age. In this case, the missing responses are related to the age variable, but not directly to the quality of the sound system.

Finally, the situation of MNAR is presented in Figure 3. For instance, suppose that participants with higher income are less likely to respond to the question about their monthly income. This occurs because they may consider personal income information and prefer not to share it, regardless of age or sound system quality.

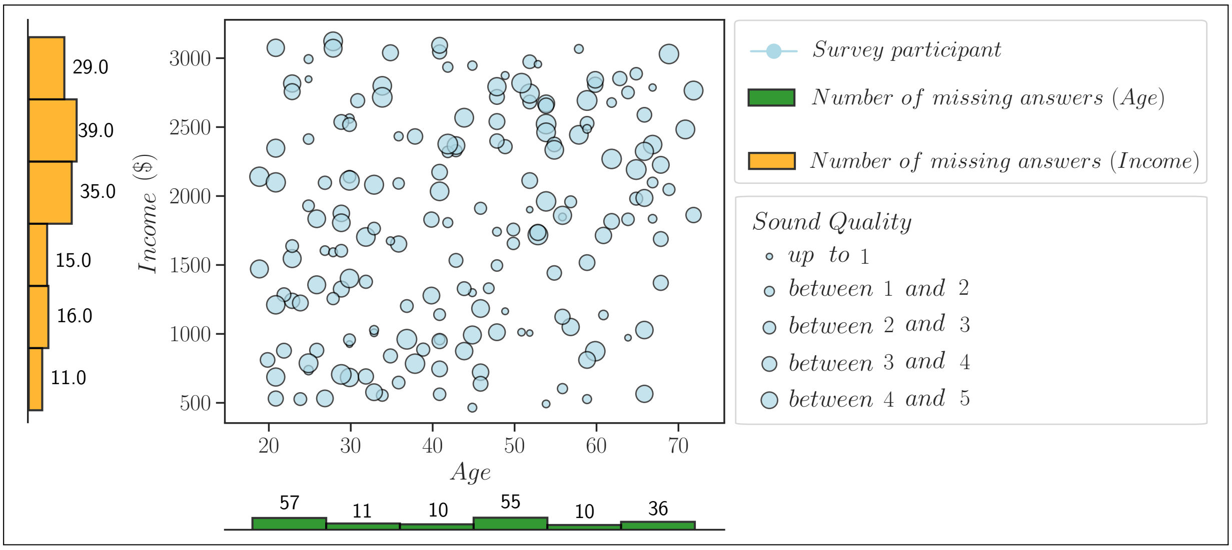


Figure 3 – Hypothetical situation illustrating the missing data mechanism MNAR.
Source: elaborated by the author.

When analyzing Figure 3, the distribution of the quantity of missing responses about income clearly shows the absence of responses for monthly incomes exceeding $2.000. In other words, the missing responses about monthly income are directly related to the actual value of income.

The characterization of the MCAR mechanism is very restrictive, to the point that its occurrence is unlikely. In this case, the reasons for absence can be neglected or treated as a special case of MAR mechanism. On the other hand, in the MNAR mechanism, there is no general method to handle this situation (GARCÍA-LAENCINA *et al.*, 2010), and unbiased estimates of missing values can be obtained by constructing an absence model (YU *et al.*, 2013). As presented in Little and Rubin (2019), MAR is a sufficient condition for likelihood-based and Bayesian inferences to be valid without the need to model the mechanism of missing data. Throughout this work, missing data is considered according to the MAR mechanism.

2.2 Modeling the Data with a Gaussian Mixture Model (GMM)

One powerful and flexible approach for modeling complex data distributions is the use of Gaussian Mixture Models (GMMs). Originally proposed by Dempster *et al.* (1977), GMMs are widely employed due to their ability to capture features and patterns present in the data, even when they originate from multimodal distributions or when there is overlap between different groups of data. A Gaussian Mixture Model (GMM) is composed of a set of Gaussian components, each representing a subpopulation or group within the observed data. Each component is described by its mean, covariance matrix, and mixture weight. The weighted combination of these components allows for modeling the joint probability distribution of the observed data, even capturing the underlying latent structure in the data.

The probability distribution of the mixture model for a vector $\boldsymbol{X}_n \in \mathcal{X}$ with N samples, where $\boldsymbol{X}_{n,obs}$ are the observed values and $\boldsymbol{X}_{n,mis}$ are the missing values, where obs and mis are the sets of indices of the observed and missing components, respectively, is given by:

$$p(\boldsymbol{X}_n|\boldsymbol{\Theta}) = \sum_{c=1}^{C} \pi^{(c)} \mathcal{N}(\boldsymbol{X}_n|\boldsymbol{\mu}^{(c)}, \boldsymbol{\Sigma}^{(c)}), \tag{2.1}$$

where $\boldsymbol{\Theta} = \{\boldsymbol{\pi}^{(c)}, \boldsymbol{\mu}^{(c)}, \boldsymbol{\Sigma}^{(c)}\}_{c=1}^{C}$ represents the set of model parameters, $\boldsymbol{\pi}^{(c)}$ represents the mixture weight associated with the cth component, $\boldsymbol{\mu}^{(c)}$ is the mean vector of the cth component, $\boldsymbol{\Sigma}^{(c)}$ is the covariance matrix of the cth component and C is the number of Gaussian components.

In the absence of information about the distributions of the random variables representing the missing components, it is necessary to make inferences based on the available observations. When deriving statistical estimates, the set of observed data is used to calculate estimates of the unknown parameters. Thus, by partitioning the mean vector and covariance matrix according to the indices of the observed and missing components, it follows

$$\boldsymbol{\mu}^{(c)} = \begin{bmatrix} \boldsymbol{\mu}^{(c)}_{obs} \\ \boldsymbol{\mu}^{(c)}_{mis} \end{bmatrix}, \quad \boldsymbol{\Sigma}^{(c)} = \begin{bmatrix} \boldsymbol{\Sigma}^{(c)}_{obs/obs} & \boldsymbol{\Sigma}^{(c)}_{obs/mis} \\ \boldsymbol{\Sigma}^{(c)}_{mis/obs} & \boldsymbol{\Sigma}^{(c)}_{mis/mis} \end{bmatrix}, \tag{2.2}$$

the conditional distribution of the missing values given the observed values is normally distributed with mean $\tilde{\boldsymbol{\mu}}_n^{(c)} = \mathbb{E}^{(c)}[\boldsymbol{X}_{n,mis}|\boldsymbol{X}_{n,obs}]$ and covariance $\tilde{\boldsymbol{\Sigma}}_n^{(c)} = \mathrm{Var}^{(c)}[\boldsymbol{X}_{n,mis}|\boldsymbol{X}_{n,obs}]$. This can be expressed as:

$$\tilde{\boldsymbol{\mu}}_n^{(c)} = \boldsymbol{\mu}_{mis}^{(c)} + \boldsymbol{\Sigma}_{mis/obs}^{(c)}(\boldsymbol{\Sigma}_{obs/obs}^{(c)})^{-1}(\boldsymbol{X}_{n,obs} - \boldsymbol{\mu}_{obs}^{(c)}), \tag{2.3}$$

$$\tilde{\boldsymbol{\Sigma}}_n^{(c)} = \boldsymbol{\Sigma}_{mis/mis}^{(c)} - \boldsymbol{\Sigma}_{mis/obs}^{(c)}(\boldsymbol{\Sigma}_{obs/obs}^{(c)})^{-1}\boldsymbol{\Sigma}_{obs/mis}^{(c)}, \tag{2.4}$$

and the moments of a specific component $x_{n,d}$ of $\boldsymbol{X}_{n,mis}$ are given by

$$\mathbb{E}[x_{n,d}] = \sum_{c=1}^{C} \pi^{(c)} \tilde{\boldsymbol{\mu}}_{n,d}^{(c)}, \tag{2.5}$$

$$\mathbb{E}[x_{n,d}^2] = \sum_{c=1}^{C} \pi^{(c)} \left([\tilde{\boldsymbol{\mu}}_{n,d}^{(c)}]^2 + \tilde{\boldsymbol{\Sigma}}_{n,d}^{(c)} \right), \tag{2.6}$$

where $\tilde{\boldsymbol{\mu}}_{n,d}^{(c)}$ represents the dth element of the vector $\tilde{\boldsymbol{\mu}}_n^{(c)}$ and $\tilde{\boldsymbol{\Sigma}}_{n,d}^{(c)}$ is the dth element of the diagonal of the matrix $\tilde{\boldsymbol{\Sigma}}_n^{(c)}$.

Treating the missing values as latent variables, the objective is to find the parameters $\boldsymbol{\Theta}$ of the model that maximize the likelihood function conditioned on the observed data. This can be expressed as:

$$\mathcal{L}(\boldsymbol{\Theta}|\boldsymbol{X}_{obs}) = \prod_{n=1}^{N} \left(\sum_{c=1}^{C} \pi^{(c)} \mathcal{N}(\boldsymbol{X}_{obs}|\boldsymbol{\mu}_{obs}^{(c)}, \boldsymbol{\Sigma}_{obs}^{(c)}) \right). \tag{2.7}$$

Therefore, the Expectation-Maximization (EM) algorithm in its extended form for handling missing data (HUNT; JORGENSEN, 2003) is used.

2.3 Expectation-Maximization (EM) with missing data

EM is an iterative technique that maximizes the likelihood of the data. EM alternates between the Expectation step (E-step), where the probabilities of data points belonging to each component are estimated, and the Maximization step (M-step), where the GMM parameters are updated based on the estimated probabilities. When using a GMM to model missing data, the EM algorithm estimates the model parameters and simultaneously imputes the missing values iteratively. This allows the model to be fitted to the observed data, taking into account the uncertainty of the missing values and the relationships between variables.

After initializing the number of components C in the mixture model and the parameters $\boldsymbol{\Theta}$, the initial value of the data likelihood in the form presented in Equation

2.7 should be computed. It is worth noting that an initial imputation strategy for $\boldsymbol{X}_{mis}$ should be used.

Next, in the E-step, for each sample n, the membership probability τ for each component c must be computed

$$\tau_{n,c} = \frac{\pi^{(c)}\mathcal{N}(\boldsymbol{X}_{n,obs}|\boldsymbol{\mu}^{(c)}_{obs}, \boldsymbol{\Sigma}^{(c)}_{obs/obs})}{\sum_l \pi^{(l)}\mathcal{N}(\boldsymbol{X}_{n,obs}|\boldsymbol{\mu}^{(l)}_{obs}, \boldsymbol{\Sigma}^{(l)}_{obs/obs})}, \tag{2.8}$$

and calculate the conditional mean and covariance according to Equations 2.3 and 2.4.

In the M-step, the parameters $\boldsymbol{\Theta}$ are updated using the probabilities from the previous step as follows:

$$\boldsymbol{\pi}^{(c)} = \frac{\sum_{n=1}^{N} \tau_{n,c}}{N}, \tag{2.9}$$

$$\boldsymbol{\mu}^{(c)} = \frac{\sum_{n=1}^{N} \tau_{n,c}\tilde{\boldsymbol{X}}^{(c)}_n}{\sum_{n=1}^{N} \tau_{n,c}}, \tag{2.10}$$

$$\boldsymbol{\Sigma}^{(c)} = \frac{\sum_{n=1}^{N} \tau_{n,c}(\tilde{\boldsymbol{X}}^{(c)}_n - \boldsymbol{\mu}^{(c)})(\tilde{\boldsymbol{X}}^{(c)}_n - \boldsymbol{\mu}^{(c)})^T + \sum_{n=1}^{N} \tau_{n,c}\boldsymbol{\Sigma}^{(c)}_n}{\sum_{n=1}^{N} \tau_{n,c}}, \tag{2.11}$$

where

$$\tilde{\boldsymbol{X}}^{(c)}_n = \begin{bmatrix} \boldsymbol{X}_{n,obs} \\ \\ \tilde{\boldsymbol{\mu}}^{(c)}_n \end{bmatrix}, \quad \boldsymbol{\Sigma}^{(c)}_n = \begin{bmatrix} \boldsymbol{0}_{obs/obs} & \boldsymbol{0}_{obs/mis} \\ \\ \boldsymbol{0}_{mis/obs} & \tilde{\boldsymbol{\Sigma}}^{(c)}_n \end{bmatrix}. \tag{2.12}$$

In simpler terms, $\tilde{\boldsymbol{X}}^{(c)}_n$ refers to the vector $\boldsymbol{X}_n$ after filling in its missing entries with the values from $\tilde{\boldsymbol{\mu}}^{(c)}_n$, while $\boldsymbol{\Sigma}^{(c)}_n$ represents the conditional covariance matrix $\tilde{\boldsymbol{\Sigma}}^{(c)}_n$ with zero-padding.

The E-step and M-step are repeated until the parameters of the GMM converge. Common approaches for convergence include analyzing the difference between the parameters based on predefined thresholds or reaching a maximum number of iterations.

2.4 Approximation Methods

Approximation methods play a crucial role in various situations, e.g., when estimating unknown parameters, adjusting theoretical models to observed data, or dealing with problems where exact solutions are difficult or impossible to obtain. Next, the theoretical foundations and the necessary mathematical formulation for using the Monte Carlo Method and the Unscented Transform are presented. These techniques allow for obtaining arbitrarily more precise estimates in complex statistical models, even considering problems where uncertainty plays a significant role.

2.4.1 Monte Carlo (MC)

A wide variety of numerical problems in fields such as science (M. Al Luhayb, 2023; LUENGO *et al.*, 2020), engineering (MORALES-RODRIGUEZ *et al.*, 2011; CUI; HASHEMI, 2023) and finance (LAI; SPANIER, 2000; MCLEISH, 2011) are handled using Monte Carlo (MC) methods. These methods belong to a class of statistical and computational techniques widely used to solve complex numerical problems, based on the idea of conducting random experiments on a computer to obtain approximate estimates of quantities of interest.

As mentioned earlier, calculating the distribution of a function of a random variable $f(\boldsymbol{x})$ for certain models can be quite challenging. A powerful yet simple alternative is to generate a set of samples $\mathcal{X} = \{x_1, x_2, ..., x_D\}$ from the distribution and approximate $f(\boldsymbol{x})$ using the empirical distribution $f(x_i)_{i=1}^{D}$. This approximation is called the Monte Carlo method and can be used to approximate the expected value of any function of a random variable (MURPHY, 2012), which can be expressed as follows:

$$\mathbb{E}[f(\boldsymbol{x})] = \int_{\mathcal{X}} f(\boldsymbol{x})p(\boldsymbol{x})d_{\boldsymbol{x}}, \tag{2.13}$$

equivalent to the empirical mean

$$\overline{h}_e = \frac{1}{D}\sum_{i=1}^{D} f(x_i). \tag{2.14}$$

When $f(\boldsymbol{x})$ has finite variance, according to the Central Limit Theorem, the error can be expressed as:

$$\frac{\overline{h}_e - \mathbb{E}[f(\boldsymbol{x})]}{\sqrt{DVar[f(\boldsymbol{x})]}} \sim \mathcal{N}(0,1), \tag{2.15}$$

and the term $Var[f(\boldsymbol{x})]$ can be approximated by the sample variance

$$\frac{1}{D-1}\sum_{i=1}^{D}\left(f(x_i) - \overline{h}_e\right)^2. \tag{2.16}$$

Among the families of MC methods, the following methods can be highlighted: Importance Sampling (IS); Rejection Sampling (RS) and Markov Chain Monte Carlo (MCMC).

2.4.1.1 Importance Sampling (IS)

It can be divided according to the number of proposed densities to extract the samples and the temporal evolution (in which the parameters are adaptive or stationary) (LUENGO *et al.*, 2020), this approach accepts all the samples and assigns weights to each of them, based on their quality in approximating the desired distribution, which can be expressed as:

$$\mathbb{E}[f(\boldsymbol{x})] = \frac{1}{D}\sum_{i=1}^{D}\omega_i f(x_i), \tag{2.17}$$

where ω_i is the associated weight.

From Equation 2.13, the function $p(\boldsymbol{x})$ is not necessarily the best Probability Density Function (PDF) to use even though it appears in the integrand. Therefore, a different PDF $g(\boldsymbol{x})$ can be introduced as follows:

$$\mathbb{E}[f(\boldsymbol{x})] = \int_{\mathcal{X}} f(\boldsymbol{x})\frac{p(\boldsymbol{x})}{g(\boldsymbol{x})}g(\boldsymbol{x})d_{\boldsymbol{x}}, \tag{2.18}$$

where $g(\boldsymbol{x}) \geq 0$, $\int g(\boldsymbol{x})d_{\boldsymbol{x}} = 1$ and $f(\boldsymbol{x})p(\boldsymbol{x})/g(\boldsymbol{x}) < \infty$. This suggests that samples from probability density functions other than $p(\boldsymbol{x})$ can also be used in the approximation, where for each $i = 1, ..., D$, x_i is assigned a weight $\omega_i = p(\boldsymbol{x})/g(\boldsymbol{x})$.

The reduction in variance occurs because it is possible to choose a density that is more similar to the function $f(\boldsymbol{x})$ being integrated. In other words, it allows determining the most important regions for integration, resulting in a reduction in the variance of the estimator.

2.4.1.2 Rejection Sampling (RS)

Another classical approach of the Monte Carlo Method is to accept or reject a sample based on a suitable test of the densities $f(\boldsymbol{x})$ and $p(\boldsymbol{x})$ (KALOS; WHITLOCK, 2008), and it can be proven that the accepted samples are indeed distributed according to the target density $f(\boldsymbol{x})$ (DEVROYE, 1986). Originally, a single uniform PDF is considered, but the sampling can be performed from any density function for which sampling is direct (LUENGO *et al.*, 2020).

Samples x_i' are obtained from $p(\boldsymbol{x})$ and accepted with probability

$$\frac{f(x_i')}{p(x_i')w_f} \leq 1, \tag{2.19}$$

where w_f is a constant such that $p(\boldsymbol{x})w_f$ represents an enveloping function for $f(\boldsymbol{x})$, meaning that, $p(\boldsymbol{x})w_f > f(\boldsymbol{x})$ for all $x_i \in \mathcal{X}$. In summary: the RS is an iterative method where, in the tth iteration, the samples $\boldsymbol{x}^{(t)} \sim p(\boldsymbol{x})$ and $u \sim \mathcal{U}(0,1)$; If

$$u \leq \frac{f(\boldsymbol{x}^{(t)})}{Kp(\boldsymbol{x}^{(t)})}, \tag{2.20}$$

$\boldsymbol{x}^{(t)}$ will be accepted, otherwise, it will be rejected.

2.4.1.3 Markov Chain Monte Carlo (MCMC)

Basically, the MCMC technique extracts samples from a proposed PDF by constructing a Markov chain, whose stationary distribution is the desired distribution $f(\boldsymbol{x})$. It accepts or rejects these candidate samples as the new state of the chain. There are various MCMC approaches, and details of algorithms such as *Metropolis-Hastings (MH)*, *Gibbs sampler*, *MH-within-Gibbs*, and other classical techniques are thoroughly explored in (LUENGO *et al.*, 2020), as well as in other references such as (MURPHY, 2012; KROESE *et al.*, 2011; HASTIE *et al.*, 2009; KALOS; WHITLOCK, 2008; ROBERT; CASELLA, 2005).

2.4.2 Unscented Transform (UT)

The approach known as Unscented Transform (UT) (JULIER; UHLMANN, 1996), was developed to handle uncertainty propagation in nonlinear systems in search of a more precision approximation. Its conception is based on the assumption that, given a fixed number of points (called sigma points), it is easier to approximate a Gaussian distribution than it is to approximate a nonlinear function or an arbitrary transformation.

According to the theory presented in Subsection 2.4.1, it can be concluded that MC techniques require several orders of magnitude more sample points to obtain a satisfactory estimation result (especially when considering non-Gaussian models), increasing the computational cost in specific cases. In this sense, the UT has advantages in terms of its usage. The number of sigma points required by the UT can be divided into two categories (EBEIGBE *et al.*, 2021): in the first category, only $(2D + 1)$ sigma points are used, while in the second category, more than $(2D + 1)$ sigma points can be used. Additionally, with the UT, it is possible to consider information about higher-order moments, such as skewness and kurtosis.

Let $\boldsymbol{x}$ be a multivariate random vector uniformly distributed with mean $\mathbb{E}[\boldsymbol{x}]$ and covariance matrix $\boldsymbol{\Sigma}_x$. The sample mean $\mathbb{E}[\varphi]$ and sample covariance $\boldsymbol{\Sigma}_\varphi$ of the nonlinear transformation $\boldsymbol{\delta} = h(\boldsymbol{x})$ can be calculated as follows.

1. Calculate the $(2D+1)$ sigma points $\boldsymbol{\gamma}_{[i]}$:

$$\boldsymbol{\gamma}_{[0]} = \mathbb{E}[\boldsymbol{x}], \tag{2.21}$$

$$\boldsymbol{\gamma}_{[i]} = \boldsymbol{\gamma}_{[0]} + \left(\sqrt{(D+\lambda)\boldsymbol{\Sigma}_x}\right)_{[i]}, \forall i \in \{1, ..., D\}, \tag{2.22}$$

$$\boldsymbol{\gamma}_{[i]} = \boldsymbol{\gamma}_{[0]} - \left(\sqrt{(D+\lambda)\boldsymbol{\Sigma}_x}\right)_{[i-D]}, \forall i \in \{D+1, ..., 2D\}, \tag{2.23}$$

where $\left(\sqrt{(D+\lambda)\boldsymbol{\Sigma}_x}\right)_{[i]}$ denotes the ith column of the square root matrix of $(D+\lambda)\boldsymbol{\Sigma}_x$ and λ is a scaling parameter, such that

$$\lambda = \hat{\alpha}^2(D+\kappa) - D, \tag{2.24}$$

where $\hat{\alpha}$ determines the spread of the sigma points around $\mathbb{E}[\boldsymbol{x}]$ (typically set to a small positive value), and κ is a secondary scaling parameter (usually $\kappa = D - 3$) (WAN; MERWE, 2001). The first, second, and third-order moments of the distribution of $\boldsymbol{\gamma}_{[i]}$ remain unchanged for any value of κ. However, both the kurtosis (fourth-order moment) and higher-order moments are scaled by $(D+\kappa)$ and geometrically with $(D+\kappa)$, respectively (JULIER *et al.*, 1995). Thus, κ can be adjusted based on the higher-order moments of the prior distribution in order to reduce errors in these terms.

2. Calculate the weight associated with the ith-$\boldsymbol{\gamma}_{[i]}$:

$$\boldsymbol{\omega}_{[0]}^{(m)} = \frac{\lambda}{D+\lambda}, \tag{2.25}$$

$$\boldsymbol{\omega}_{[0]}^{(c)} = \frac{\lambda}{D+\lambda} + (1 - \hat{\alpha}^2 + \hat{\beta}), \tag{2.26}$$

$$\boldsymbol{\omega}_{[i]}^{(m)} = \boldsymbol{\omega}_{[i]}^{(c)} = \frac{1}{2(D+\lambda)}, \forall i \in \{1, ..., 2D\}, \tag{2.27}$$

where $\hat{\beta}$ is used to incorporate prior knowledge of the distribution of $\boldsymbol{x}$ (such that $\beta = 2$ is considered ideal for Gaussian distributions (WAN; MERWE, 2001)).

3. Apply the known nonlinear function to the $\boldsymbol{\gamma}_{[i]}$ points, obtaining the transformed points $\boldsymbol{\delta}_{[i]}$:

$$\boldsymbol{\delta}_{[i]} = h(\boldsymbol{\gamma}_{[i]}). \tag{2.28}$$

4. Evaluate the sample mean from $\boldsymbol{\delta}_{[i]}$:

$$\mathbb{E}[\varphi] \approx \sum_{i=0}^{2D} \boldsymbol{\omega}_{[i]}^{(m)} \boldsymbol{\delta}_{[i]}. \tag{2.29}$$

5. Evaluate the sample covariance from $\boldsymbol{\delta}_{[i]}$:

$$\boldsymbol{\Sigma}_{\varphi} \approx \sum_{i=0}^{2D} \boldsymbol{\omega}_{[i]}^{(c)} (\boldsymbol{\delta}_{[i]} - \mathbb{E}[\varphi])(\boldsymbol{\delta}_{[i]} - \mathbb{E}[\varphi])^T. \tag{2.30}$$

2.5 Conclusion

This chapter provides the reader with a thorough understanding of fundamental knowledge required for the development of the strategies presented in the following chapters. Cases of missing data, MCAR, MAR, and MNAR, are discussed through practical examples that clearly illustrate each type of mechanism. The theory and mathematical formulation to fit a Gaussian Mixture Model to incomplete datasets are presented in an objective manner. Since the general approach to estimate the model parameters from the data is maximum likelihood estimation, the iterative nature of the EM algorithm, involving the Expectation (E-step) and Maximization (M-step) stages (enabling robust estimation in datasets with missing values), is explored. Next, the approximation techniques, Monte Carlo Method, and Unscented Transform, are introduced. These methods provide acceptable approximate solutions, balancing precision and computational efficiency.

3 EXPECTED MATÉRN KERNEL

Many machine learning algorithms are based on the relative differences between samples rather than their specific values. In this sense, a useful approach is to estimate distances directly between pairs of samples across the dataset. This allows capturing the proximity relationships between samples, which is beneficial in various applications, assisting in analysis, classification, clustering, and comparison of information in different domains.

In this chapter, some techniques that compute the expectation of the distance between instances of possibly incomplete data are presented. From the distance estimation, it is possible to obtain the estimation of the Matérn Kernel function, the subject of study in this work; therefore, the methods presented are used for comparison purposes with the proposed approaches. Next, the mathematical formulation of the problem of estimating the Matérn Kernel is presented. Finally, two new methods to estimate the Matérn Kernel when subjected to incomplete datasets directly are proposed. Both solutions are based on the assumption that the Euclidean distances of the Matérn Kernel function follow a Nakagami probability distribution.

To better elucidate the methods to follow, consider a dataset $\mathcal{X} = \{\boldsymbol{X}_n\}_{n=1}^{N}$ consisting of N samples, such that $\boldsymbol{X}_{n,obs}$ represents the observed values and $\boldsymbol{X}_{n,mis}$ represents the missing values, where *obs* and *mis* are sets of indices of the observed and missing components, respectively. Also, consider $\boldsymbol{X}_i = (\mathrm{x}_{i,1}, ..., \mathrm{x}_{i,D})^T$ and $\boldsymbol{X}_j = (\mathrm{x}_{j,1}, ..., \mathrm{x}_{j,D})^T$, D-dimensional vectors in which one or more of their components are not observed, following the MAR mechanism and being independent. Thus, it is possible to model $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ as random variables.

3.1 Methods for Estimating Similarity Measures

3.1.1 Conditional Mean Imputation (CMI)

The method described by Hunt and Jorgensen (2003), called Conditional Mean Imputation (CMI), fills in the missing entries in $\boldsymbol{X}_n$ based on the mean value of a probability distribution estimated for the missing entries. The usual approach to obtain this distribution consists of first obtaining a model for the distribution from which the feature vectors of the dataset were generated, and then conditioning the observed values of each incomplete vector. This can be expressed as:

$$\int_{-\infty}^{\infty} \varPhi p(\varPhi | x_{n,obs}) d_{\varPhi}, \tag{3.1}$$

where $p(\cdot|x_{n,obs})$ is the PDF p conditioned on $\boldsymbol{X}_{n,obs}$.

Once a Gaussian Mixture Model is used to model the data, the probability density function in its general form is given by Equation 2.1. The parameters of $p(\cdot|x_{n,obs})$ are the mixture weight $\boldsymbol{\pi}^{(c)}$, the mean vector $\boldsymbol{\mu}^{(c)}$, and the covariance matrix $\boldsymbol{\Sigma}^{(c)}$, such that $c = 1 \ \ldots \ C$ is the number of components. Thus, Equation 3.1 then becomes a weighted sum of the conditioned components of the GMM. This process can be obtained as described in Section 2.2.

3.1.2 Expected Square Distance (ESD)

The Expected Square Distance (ESD) method (EIROLA *et al.*, 2013) calculates the expected value of the squared distance between vectors $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ under the assumption that they are normally distributed. The proposed solution consists of finding the expectation, conditioned on the observed values, and the variance of each missing component separately, without the need to determine the total probability density.

Under the assumption of independence between the missing entries, the expected squared distance $\mathbb{E}[||\boldsymbol{X}_i - \boldsymbol{X}_j||^2]$ can be expressed in terms of four parts, depending on which attributes are missing for the two samples:

$$\begin{aligned}\mathbb{E}\left[||\boldsymbol{X}_i - \boldsymbol{X}_j||^2\right] = &\sum_{d \notin mis_i \cup mis_j} (x_{i,d} - x_{j,d})^2 + \sum_{d \in mis_j \setminus mis_i} \mathbb{E}\left[(x_{i,d} - x_{j,d})^2\right] \\ &+ \sum_{d \in mis_i \setminus mis_j} \mathbb{E}\left[(x_{i,d} - x_{j,d})^2\right] + \sum_{d \in mis_i \cap mis_j} \mathbb{E}\left[(x_{i,d} - x_{j,d})^2\right],\end{aligned} \tag{3.2}$$

where mis_i, $mis_j \subseteq \{1 \ \cdots \ D\}$ denote the sets of indices of missing components in $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$, respectively. The fact is that all terms of Equation 3.2 can be expanded. For illustration, the expansion of the term $(d \in mis_i \cap mis_j)$ is shown below:

$$\begin{aligned}\mathbb{E}\left[(x_{i,d} - x_{j,d})^2\right] &= \mathbb{E}\left[x_{i,d}^2 - 2x_{i,d}x_{j,d} + x_{j,d}^2\right] \\ &= \mathbb{E}\left[x_{i,d}^2\right] - 2\mathbb{E}\left[x_{i,d}\right]\mathbb{E}\left[x_{j,d}\right] + \mathbb{E}\left[x_{j,d}^2\right] \\ &= \mathbb{E}\left[x_{i,d}^2\right] - 2\mathbb{E}\left[x_{i,d}\right]\mathbb{E}\left[x_{j,d}\right] + \mathbb{E}\left[x_{j,d}^2\right] - \mathbb{E}\left[x_{i,d}\right]^2 + \mathbb{E}\left[x_{i,d}\right]^2 - \mathbb{E}\left[x_{j,d}\right]^2 + \mathbb{E}\left[x_{j,d}\right]^2 \\ &= \mathbb{E}\left[x_{i,d}\right]^2 - 2\mathbb{E}\left[x_{i,d}\right]\mathbb{E}\left[x_{j,d}\right] + \mathbb{E}\left[x_{j,d}\right]^2 + \mathbb{E}\left[x_{i,d}^2\right] - \mathbb{E}\left[x_{i,d}\right]^2 + \mathbb{E}\left[x_{j,d}^2\right] - \mathbb{E}\left[x_{j,d}\right]^2\end{aligned}$$

$$= \left(\mathbb{E}\left[x_{i,d}\right] - \mathbb{E}\left[x_{j,d}\right]\right)^2 + \mathrm{Var}\left[x_{i,d}\right] + \mathrm{Var}\left[x_{j,d}\right]. \tag{3.3}$$

In a similar way, the expression $\left(x_{i,d} - \mathbb{E}\left[x_{j,d}\right]\right)^2 + \mathrm{Var}\left[x_{j,d}\right]$ is obtained for the term $(d \in mis_j \setminus mis_i)$, while the expression $\left(\mathbb{E}\left[x_{i,d}\right] - x_{j,d}\right)^2 + \mathrm{Var}\left[x_{i,d}\right]$ is obtained for the term $(d \in mis_i \setminus mis_j)$. The term $(d \notin mis_i \cup mis_j)$ is a sum over the known components in both samples, so the expectation involves calculating the distance directly.

Let $\tilde{x}_{i,d}$ be the imputed version of $x_{i,d}$, where each missing value is replaced by its conditional mean

$$\tilde{x}_{i,d} = \begin{cases} \mathbb{E}[x_{i,d}|x_{i,obs}] & \text{if } d \in mis_i, \\ x_{i,d} & \text{otherwise,} \end{cases} \tag{3.4}$$

and σ^2 corresponds to the conditional variance

$$\sigma^2_{i,d} = \begin{cases} \mathrm{Var}[x_{i,d}|x_{i,obs}] & \text{if } d \in mis_i, \\ 0 & \text{otherwise,} \end{cases} \tag{3.5}$$

the expected squared distance can be written in its compact form as follows:

$$\mathbb{E}\left[||\boldsymbol{X}_i - \boldsymbol{X}_j||^2\right] = \sum_{d=1}^{D} \left(\mathbb{E}\left[x_{i,d}\right] - \mathbb{E}\left[x_{j,d}\right]\right)^2 + \mathrm{Var}\left[x_{i,d}\right] + \mathrm{Var}\left[x_{j,d}\right]. \tag{3.6}$$

As presented by Eirola *et al.* (2013), the Equation 3.6 emphasizes how the uncertainty of missing values is taken into account.

The means and conditional covariances are obtained in a similar manner to what was presented in Section 2.2. Given that $\boldsymbol{X} = [\boldsymbol{X}_{mis}, \boldsymbol{X}_{obs}]$, the mean and covariance matrix can be partitioned as follows:

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_{mis} \\ \boldsymbol{\mu}_{obs} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{mis/mis} & \boldsymbol{\Sigma}_{mis/obs} \\ \boldsymbol{\Sigma}_{obs/mis} & \boldsymbol{\Sigma}_{obs/obs} \end{bmatrix}. \tag{3.7}$$

Therefore, the conditional distribution of the missing values given the observed values is normally distributed with mean

$$\tilde{\boldsymbol{\mu}}_{mis} = \boldsymbol{\mu}_{mis} + \boldsymbol{\Sigma}_{mis/obs}(\boldsymbol{\Sigma}_{obs/obs})^{-1}(\boldsymbol{X}_{n,obs} - \boldsymbol{\mu}_{obs}) \tag{3.8}$$

and covariance

$$\tilde{\boldsymbol{\Sigma}}_{mis} = \boldsymbol{\Sigma}_{mis/mis} - \boldsymbol{\Sigma}_{mis/obs}(\boldsymbol{\Sigma}_{obs/obs})^{-1}\boldsymbol{\Sigma}_{obs/mis}, \tag{3.9}$$

where the mean and variance of each missing value are found by extracting the corresponding element from $\tilde{\boldsymbol{\mu}}_{mis}$ or from the main diagonal of $\tilde{\boldsymbol{\Sigma}}_{mis}$. The initial estimate of the mean and covariance matrix is obtained using the Expectation Conditional Maximization (ECM) (MENG; RUBIN, 1993).

Later on, an improved version of the ESD method is presented by modeling the dataset with a Gaussian mixture model (EIROLA *et al.*, 2014).

3.1.3 Expected Euclidean Distance (EED)

The Expected Euclidean Distance (EED) method (MESQUITA *et al.*, 2017), is based on the assumption that the Euclidean distance between two potentially incomplete vectors follows a Nakagami distribution. Therefore, the expected distance can be obtained in a closed-form manner, depending only on the parameters of the distribution.

Let η be the Euclidean distance between $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$, such that

$$\eta = z^{1/2} = \sqrt{\sum_{d=1}^{D}(x_{i,d} - x_{j,d})^2}, \tag{3.10}$$

where $z = \| \boldsymbol{X}_i - \boldsymbol{X}_j \|^2 = \sum_{d=1}^{D}(\mathrm{x}_{i,d} - \mathrm{x}_{j,d})^2$. As η is the result of a non-negative transformation of $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$, it can also be considered as a random variable. Therefore, computing its expected value involves solving the following equation:

$$\mathbb{E}[\eta] = \int_0^\infty p(\eta)\eta \; d_\eta. \tag{3.11}$$

In order to define a statistical model for $p(\eta)$, some assumptions can be made based on z. By definition, z can be characterized as the sum of squares of random variables. In Roberts and Geisser (1966), sufficient conditions are presented for a squared random variable to follow a Gamma distribution. Furthermore, (COVO; ELALOUF, 2014) provides proofs that, under moderate conditions of independence, the distribution of a sum of several Gamma-distributed variables can be approximated by a Gamma distribution itself. Therefore, it is reasonable to assume that $z \sim$ Gamma $(\alpha,\, \beta)$.

As further presented in (NAKAGAMI, 1960), a Nakagami random variable can be obtained by taking the square root of a Gamma-distributed random variable. Hence, it is appropriate to choose a Nakagami distribution for η. Thus, under the assumption that $\eta \sim$ Nakagami (m, Ω), the expected value of η is directly obtained by:

$$\mathbb{E}[\eta] = \frac{\Gamma\left(m + \frac{1}{2}\right)}{\Gamma(m)}\left(\frac{\Omega}{m}\right)^{1/2}, \tag{3.12}$$

while its variance is obtained by

$$\mathrm{Var}[\eta] = \Omega\left(1 - \frac{1}{m}\left(\frac{\Gamma\left(m+\frac{1}{2}\right)}{\Gamma(m)}\right)^2\right), \tag{3.13}$$

where m and Ω represent, respectively, the shape and scale parameters of the Nakagami distribution. These parameters, in turn, can be expressed in terms of the expected squared distance between the data vectors. Thus,

$$m = \frac{(\mathbb{E}[z])^2}{\mathrm{Var}[z]}, \quad \Omega = \mathbb{E}[z]. \tag{3.14}$$

The estimation of $\mathbb{E}[z]$ can be explicitly addressed based on the exposition in Subsection 3.1.2. As also presented in Mesquita *et al.* (2017), the estimate of the expected value and the variance of z can be expressed in terms of non-central moments of $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$. Consequently, the variance, in its simplified form, can be obtained using Equation 3.15.

$$\mathrm{Var}[z] = \sum_{d=1}^{D} 4(\mathbb{E}[x_{i,d}] - \mathbb{E}[x_{j,d}])^2 + (\mathrm{Var}[x_{i,d}] + \mathrm{Var}[x_{j,d}]) + 2(\mathrm{Var}[x_{i,d}] + \mathrm{Var}[x_{j,d}])^2. \tag{3.15}$$

As in the ESD method, the dataset in EED is modeled according to a Gaussian mixture model.

3.2 Expected Matérn Kernel

The Matérn kernel function, in its general definition, is given by:

$$\mathcal{C}_\nu(\eta) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}\,\eta}{\rho}\right)^\nu K_\nu\left(\frac{\sqrt{2\nu}\,\eta}{\rho}\right), \tag{3.16}$$

where $\eta = ||\boldsymbol{X}_i - \boldsymbol{X}_j||$ it is the Euclidean distance between $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$, $\Gamma(\cdot)$ is a Gamma function, σ^2 and ρ are kernel parameters (where $\rho > 0$) and K_ν is a modified Bessel function of the second kind with order ν (where $\nu > 0$).

It is interesting to note that the smoothness of the function can be controlled through ν. The Matérn kernel reduces to the Exponential kernel for $\nu = 0.5$ and to the Gaussian kernel as $\nu \to \infty$, making it very useful for applications due to its flexibility.

It is also worth noting that for values of $\nu \in \left\{\frac{1}{2}, \frac{3}{2}, \frac{5}{2}\right\}$, the kernel is simplified respectively to:

$$\mathcal{C}_{1/2}(\eta) = \sigma^2 \; exp\left\{-\frac{\eta}{\rho}\right\}, \tag{3.17}$$

$$\mathcal{C}_{3/2}(\eta) = \sigma^2 \; \left(1 + \frac{\sqrt{3}\eta}{\rho}\right) exp\left\{-\frac{\sqrt{3}\eta}{\rho}\right\}, \tag{3.18}$$

$$\mathcal{C}_{5/2}(\eta) = \sigma^2 \; \left(1 + \frac{\sqrt{5}\eta}{\rho} + \frac{5\eta^2}{3\rho^2}\right) exp\left\{-\frac{\sqrt{5}\eta}{\rho}\right\}. \tag{3.19}$$

3.2.1 Formulation of the Estimation Problem

Let $k_{\mathcal{C}_\nu(\eta)}(\boldsymbol{X}_i, \boldsymbol{X}_j)$ be the symmetric function of the Matérn kernel $\mathcal{C}_\nu(\eta)$ that represents the similarity between $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$, given the vector $\boldsymbol{X} \in \mathcal{X}$ without repetition, the elements resulting from $k_{\mathcal{C}_\nu(\eta)}(\boldsymbol{X}_i, \boldsymbol{X}_j)$ are positive semi-definite. As mentioned earlier, $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ are characterized as random variables, and similarly $\eta_{ij \neq i} > 0$ is a random variable. Therefore, the goal is to estimate the expected Matérn Kernel $\mathcal{C}_\nu(\eta)$. This can be mathematically expressed in a general form as follows:

$$\mathbb{E}[k_{\mathcal{C}_\nu(\eta)}(\boldsymbol{X}_i, \boldsymbol{X}_j)] = \mathbb{E}[\mathcal{C}_\nu(\eta)]. \tag{3.20}$$

Considering that the function $k_{\mathcal{C}_\nu(\eta)}(\boldsymbol{X}_i, \boldsymbol{X}_j) = f(\eta)$ is a measurable function, the estimation problem is equivalent to estimating the expected value of the function $f(\eta)$ under a probability distribution $p(\eta)$,

$$\mathrm{E}[f(\eta)] = \int_{-\infty}^{\infty} p(\eta) f(\eta) d_\eta. \tag{3.21}$$

To solve the above problem, once again, a statistical model for $p(\eta)$ is required. Based on the theory presented in Subsection 3.1.3, it is possible to assume that $\eta \sim$ Nakagami(m, Ω). Therefore, the PDF for η is given by:

$$p(\eta | m, \Omega) = \frac{2m^m}{\Gamma(m)\Omega^m} \eta^{2m-1} exp\left\{-\frac{m}{\Omega}\eta^2\right\}, \tag{3.22}$$

thus, the expected value of the $\mathcal{C}_{1/2}(\eta)$ can be obtained by solving the following equation:

$$\int_0^{\infty} \sigma^2 \; exp\left\{-\frac{\eta}{\rho}\right\} \frac{2m^m}{\Gamma(m)\Omega^m} \eta^{2m-1} exp\left\{-\frac{m}{\Omega}\eta^2\right\} d_\eta. \tag{3.23}$$

Similarly, the expected value of $\mathcal{C}_{3/2}(\eta)$ can be obtained by solving

$$\int_0^\infty \sigma^2 \left(1 + \frac{\sqrt{3}\ \eta}{\rho}\right) exp\left\{-\frac{\sqrt{3}\ \eta}{\rho}\right\} \frac{2m^m}{\Gamma(m)\Omega^m}\eta^{2m-1} exp\left\{-\frac{m}{\Omega}\eta^2\right\} d_\eta, \tag{3.24}$$

and finally, the expected value of the Matérn kernel $\mathcal{C}_{5/2}(\eta)$ can be obtained by solving

$$\int_0^\infty \sigma^2 \left(1 + \frac{\sqrt{5}\eta}{\rho} + \frac{5\eta^2}{3\rho^2}\right) exp\left\{-\frac{\sqrt{5}\eta}{\rho}\right\} \frac{2m^m}{\Gamma(m)\Omega^m}\eta^{2m-1} exp\left\{-\frac{m}{\Omega}\eta^2\right\} d_\eta. \tag{3.25}$$

So far, no analytical or closed-form solutions using moment generating functions are known for solving Equations 3.23-3.25. In situations where the complexity of the model makes it very challenging to obtain an explicit representation, such as in latent variable model structures (ROBERT; CASELLA, 2005), it is entirely appropriate to resort to approximation methods.

In this work, the approximation techniques Monte Carlo Method and Unscented Transform (UT) are used in the direct estimation process of the Matérn kernel $\mathcal{C}_\nu(\eta)$, without any preprocessing step. The basic idea is to obtain the estimation of Equation 3.20 by approximating the distribution of the Euclidean distance between $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ after the imputation process of missing values. In other words, given that $\eta \sim$ Nakagami(m, Ω), with known parameters m and Ω, it is possible to estimate $\mathbb{E}[k_{\mathcal{C}_\nu(\eta)}(\boldsymbol{X}_i, \boldsymbol{X}_j)]$ using the mentioned approximation methods.

3.2.2 Expected Matérn Kernel via Monte Carlo Method (EMK-MC) - Proposal 1

After obtaining the values of m and Ω, it is possible to generate samples via MC simulation from the distribution $\psi \sim$ Nakagami(m, Ω), which represents the distribution of the Euclidean distance between $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ after the process of imputing missing values (see Section 2.2). Then, based on a sampling of the generated ψ distribution, the estimation of the Matérn Kernel $\mathcal{C}_\nu(\eta)$ can be directly obtained through $\mathbb{E}[\mathcal{C}_\nu(\psi)]$. This estimation process constitutes the proposed method called Expected Matérn Kernel via Monte Carlo Method (EMK-MC). This approach can be implemented based on the procedure described in Figure 4, namely:

1. Estimate the parameters of the Gaussian mixture distribution in $\mathcal{X}$: The method starts by estimating the parameters of the mixture model distribution (Equation

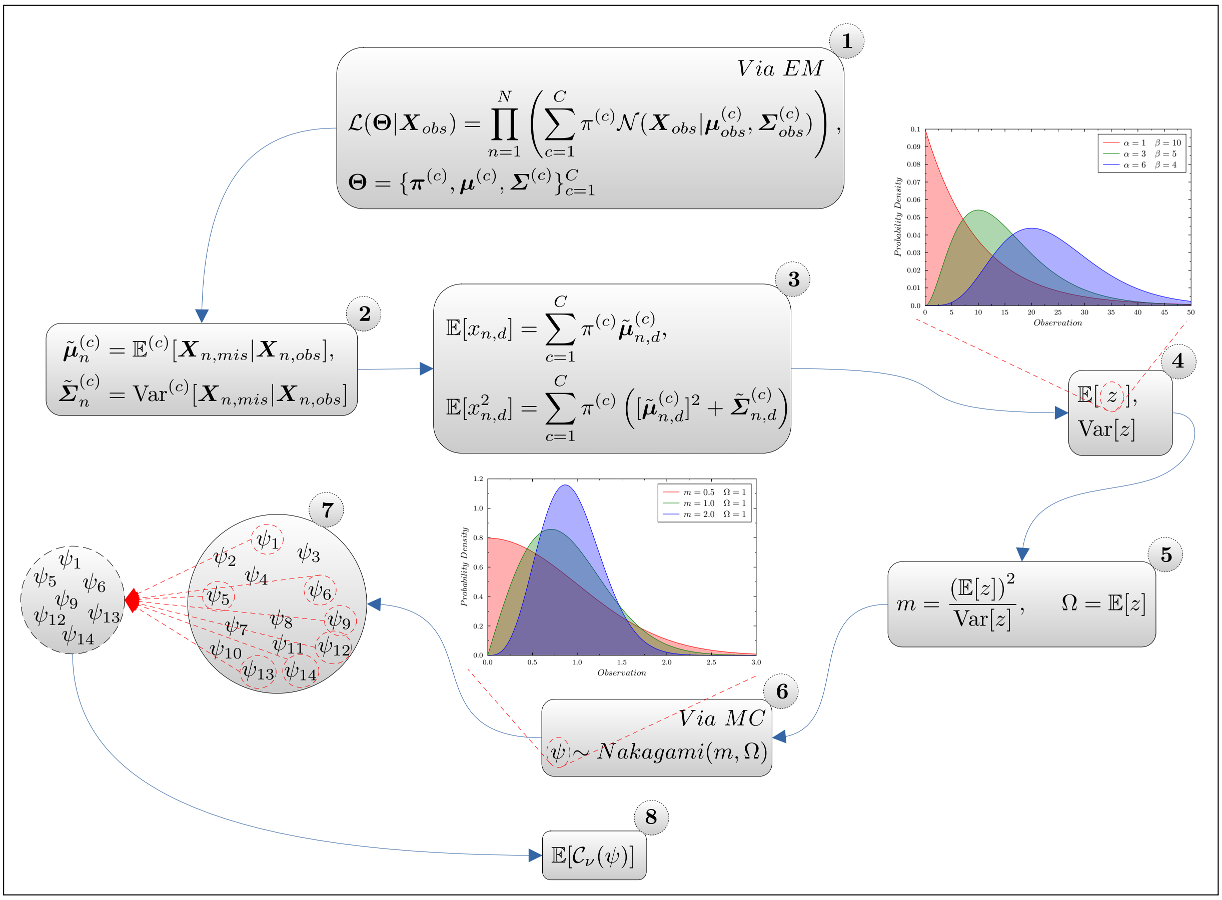


Figure 4 – Estimation process of the Matérn Kernel performed through the EMK-MC method: the implementation process of the method is detailed through the steps to be executed, containing all equations involved in the estimation.
Source: elaborated by the author.

2.7) using the maximum likelihood method. The likelihood function is maximized using the EM algorithm extended for incomplete data described in Section 2.3.

2. Calculate the mean and conditional covariance of each missing component: For each missing component $\boldsymbol{X}_{i,mis}$ and $\boldsymbol{X}_{j,mis}$, their mean and conditional covariances are computed for each model component c (Equations 2.3 and 2.4).
3. Compute the first and second moments of each missing component: The first and second moments of each missing component $\boldsymbol{X}_{i,mis}$ and $\boldsymbol{X}_{j,mis}$ are computed using Equations 2.5 and 2.6.
4. Compute the mean and variance of the squared distance: The mean $\mathbb{E}[z]$ and variance $\mathrm{Var}[z]$ of the squared distance z between the components $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ are calculated (Equations 3.6 and 3.15).
5. Obtain the parameters m and Ω of the η distribution: The parameters m and Ω of the η distribution, which is a Nakagami distribution, are obtained (Equation 3.14).

6. Generate samples via MC simulation from the ψ distribution: Samples are generated using Monte Carlo simulation from the ψ distribution, which follows a Nakagami distribution with the parameters m and Ω obtained earlier.
7. Perform a sampling of the generated ψ distribution: A sampling of the generated ψ distribution from the previous step is done. This involves selecting random values from this distribution.
8. Compute the expected value of the Matérn Kernel: Finally, the expected Matérn Kernel, denoted by $\mathcal{C}_\nu(\eta)$, is calculated through $\mathbb{E}[\mathcal{C}_\nu(\psi)]$. This calculation is performed to obtain an estimation of the expected value of the Matérn kernel using the distribution that represents η after the imputation process.

3.2.3 Expected Matérn Kernel via Unscented Transform (EMK-UT) - Proposal 2

On the other hand, through the UT, it is possible to select a deterministic set of samples from the original distribution and obtain estimates of the statistical moments of a probability distribution. In this case, the proposed method called Expected Matérn Kernel via Unscented Transform (EMK-UT), is capable of directly estimating the Matérn Kernel function $\mathcal{C}_\nu(\eta)$ through the sample mean $\mathbb{E}[\mathcal{C}_\nu(\varphi)]$, which results from summing a set of transformed points weighted by their respective weights.

Similar to what was presented in Subsection 3.2.2, this approach can be implemented based on the information provided in Figure 5, namely:

1. Estimate the parameters of the Gaussian mixture distribution in $\mathcal{X}$: The method starts by estimating the parameters of the mixture model distribution (Equation 2.7) using the maximum likelihood method. The likelihood function is maximized using the EM algorithm extended for incomplete data described in Section 2.3.
2. Calculate the mean and conditional covariance of each missing component: For each missing component $\boldsymbol{X}_{i,mis}$ and $\boldsymbol{X}_{j,mis}$, their mean and conditional covariances are computed for each model component c (Equations 2.3 and 2.4).
3. Compute the first and second moments of each missing component: The first and second moments of each missing component $\boldsymbol{X}_{i,mis}$ and $\boldsymbol{X}_{j,mis}$ are computed using Equations 2.5 and 2.6.
4. Compute the mean and variance of the squared distance: The mean $\mathbb{E}[z]$ and variance

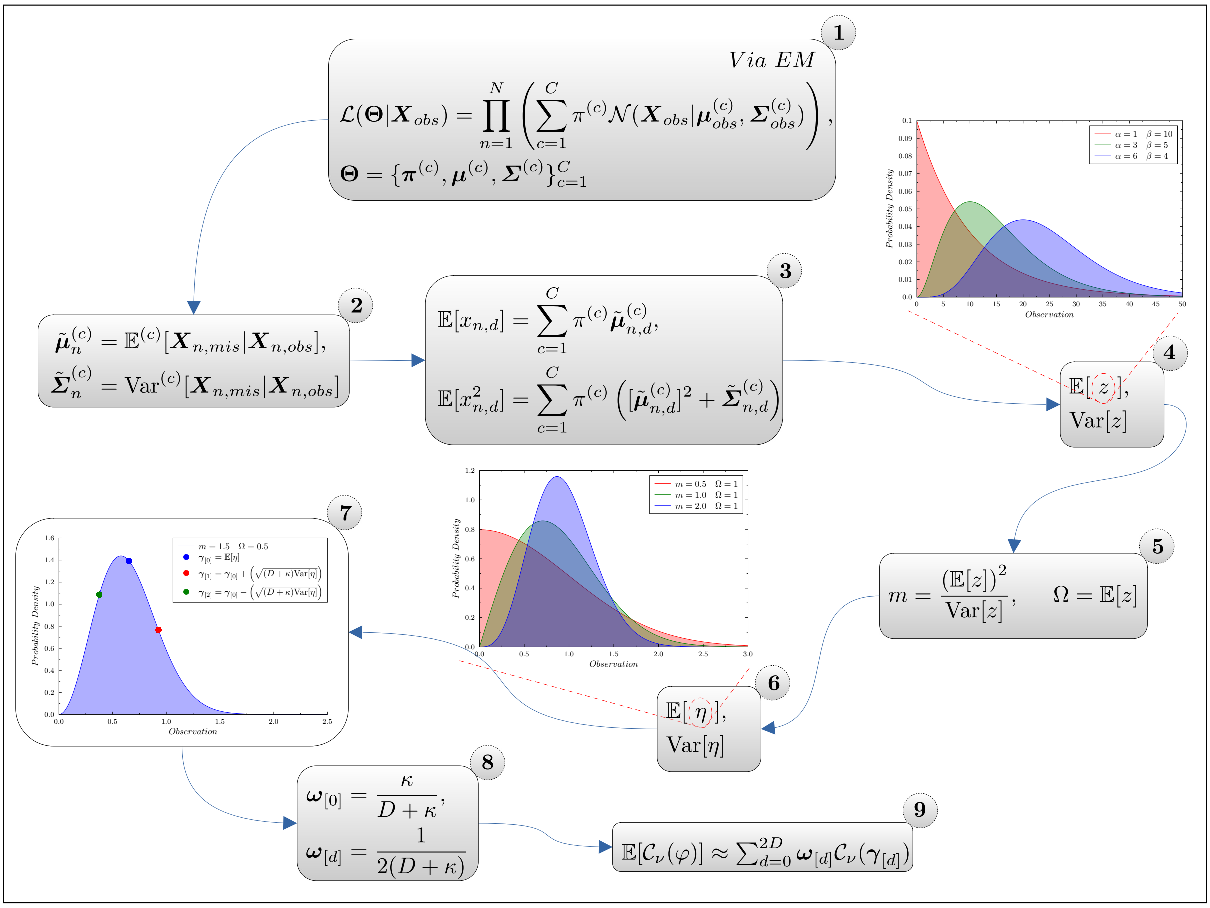


Figure 5 – Estimation process of the Matérn Kernel performed through the EMK-UT method: the implementation process of the method is detailed through the steps to be executed, containing all equations involved in the estimation.
Source: elaborated by the author.

Var$[z]$ of the squared distance z between the components $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ are calculated (Equations 3.6 and 3.15).

5. Obtain the parameters m and Ω of the η distribution: The parameters m and Ω of the η distribution, which is a Nakagami distribution, are obtained (Equation 3.14).
6. Compute the mean $\mathbb{E}[\eta]$ and variance $\mathrm{Var}[\eta]$ of the Euclidean distance η between $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ (Equations 3.12 and 3.13).
7. Select a set of $(2D+1)$ sigma points $\boldsymbol{\gamma}_{[d]}$ from the distribution $\eta \sim \mathrm{Nakagami}(m,\Omega)$, such that:

$$\boldsymbol{\gamma}_{[0]} = \mathbb{E}[\eta], \tag{3.26}$$

$$\boldsymbol{\gamma}_{[d]} = \boldsymbol{\gamma}_{[0]} + \left(\sqrt{(D+\kappa)\mathrm{Var}[\eta]}\right)_{[d]}, \forall d \in \{1, ..., D\}, \tag{3.27}$$

$$\boldsymbol{\gamma}_{[d]} = \boldsymbol{\gamma}_{[0]} - \left(\sqrt{(D+\kappa)\mathrm{Var}[\eta]}\right)_{[d-D]}, \forall d \in \{D+1, ..., 2D\}, \tag{3.28}$$

where κ is a scaling parameter.

8. Compute the weights $\boldsymbol{\omega}_{[d]}$ associated with the sigma points $\boldsymbol{\gamma}_{[d]}$:

$$\boldsymbol{\omega}_{[0]} = \frac{\kappa}{D+\kappa}, \tag{3.29}$$

$$\boldsymbol{\omega}_{[d]} = \frac{1}{2(D+\kappa)}, \forall d \in \{1, ..., 2D\}. \tag{3.30}$$

9. Compute the expected value of the Matérn Kernel $\mathcal{C}_\nu(\eta)$ through $\mathbb{E}[\mathcal{C}_\nu(\varphi)]$: Finally, the estimation of the Matérn Kernel $\mathcal{C}_\nu(\eta)$ via UT can be obtained directly, resulting from the sum of the transformed points $\mathcal{C}_\nu(\boldsymbol{\gamma}_{[d]})$ weighted by their respective weights $\boldsymbol{\omega}_{[d]}$:

$$\mathbb{E}[\mathcal{C}_\nu(\varphi)] \approx \sum_{d=0}^{2D} \boldsymbol{\omega}_{[d]} \mathcal{C}_\nu(\boldsymbol{\gamma}_{[d]}). \tag{3.31}$$

3.2.4 Specificities of the employed approaches

It is important to highlight a common characteristic among the methods used in the comparison: the calculation of the kernel is performed a posteriori, based on the estimated Euclidean distance. Given two arbitrary instances:

a) The CMI method fills in the missing entries with the conditional expected value given the observed entries, then calculates the squared distance based on the imputed values, and finally, the kernel is obtained from the square root of the calculated distance:

$$k_{\mathcal{C}_\nu(\eta^{CMI})}(\boldsymbol{X}_i, \boldsymbol{X}_j) = \mathcal{C}_\nu\left(\sqrt{\parallel \mathbb{E}[\boldsymbol{X}_i] - \mathbb{E}[\boldsymbol{X}_j] \mid \boldsymbol{X}_{i,obs}, \boldsymbol{X}_{j,obs} \parallel^2}\right) \tag{3.32}$$

b) The ESD method first estimates the expected squared distance given the observed entries. Taking the square root of the estimated value allows obtaining the Euclidean distance between the data vectors, which is then used to obtain the kernel:

$$k_{\mathcal{C}_\nu(\eta^{ESD})}(\boldsymbol{X}_i, \boldsymbol{X}_j) = \mathcal{C}_\nu\left(\sqrt{\mathbb{E}[\ \parallel \boldsymbol{X}_i - \boldsymbol{X}_j \parallel^2\ \mid \boldsymbol{X}_{i,obs}, \boldsymbol{X}_{j,obs}]}\right) \tag{3.33}$$

c) Similarly, the EED method estimates the Euclidean distance between possibly incomplete data vectors, and then uses the estimate to obtain the kernel:

$$k_{\mathcal{C}_\nu(\eta^{EED})}(\boldsymbol{X}_i, \boldsymbol{X}_j) = \mathcal{C}_\nu\left(\mathbb{E}\left[\sqrt{\parallel \boldsymbol{X}_i - \boldsymbol{X}_j \parallel^2}\ \middle|\ \boldsymbol{X}_{i,obs}, \boldsymbol{X}_{j,obs}\right]\right) \tag{3.34}$$

However, the estimation of the Matérn Kernel function by the EMK-MC and EMK-UT method is obtained directly from the points obtained in ψ and φ, respectively.

This can be expressed as:

$$k_{\mathcal{C}_\nu(\eta^{EMK-MC})}(\boldsymbol{X}_i, \boldsymbol{X}_j) = \mathbb{E}[\mathcal{C}_\nu(\psi)] \tag{3.35}$$

$$k_{\mathcal{C}_\nu(\eta^{EMK-UT})}(\boldsymbol{X}_i, \boldsymbol{X}_j) = \mathbb{E}[\mathcal{C}_\nu(\varphi)]. \tag{3.36}$$

3.3 Conclusion

Kernel methods are widely used in machine learning to perform tasks such as classification, regression, and clustering. These methods rely on calculating similarity measures between pairs of samples using a kernel function. One example of a kernel function popularly used in Gaussian processes is the Matérn Kernel. Characterized by a covariance function based on the Euclidean distance between samples, the Matérn Kernel assumes that the data is fully observed and thus cannot naturally handle incomplete data.

In this chapter, robust techniques for estimating similarity measures were presented. Once the Euclidean distance between two data instances with missing or unobserved attributes is estimated, the Matérn Kernel can be successfully computed. However, the computation of the kernel is performed a posteriori. Two new approaches to deal with this situation were introduced. The EMK-MC and EMK-UT methods estimate the Matérn Kernel function directly, depending only on the parameters of the distribution that represents the Euclidean distance between possibly incomplete data vectors. These parameters are obtained from the estimation of the squared distance between the incomplete data vectors. It is also worth noting that the proposed EMK-UT method requires $O(1)$ samples, precisely three, regardless of the size of the dataset or the number of missing attributes involved.

4 EXPERIMENTS AND RESULTS

In this chapter, the entire methodology used in the experiments is presented. The characteristics of the selected databases to evaluate the performance of the methods proposed in this thesis, as well as the results obtained from the computational simulations for the other methods used, are presented throughout this chapter.

4.1 Methodology

To analyze the efficiency of the proposed approaches, experiments were conducted on synthetic ($\mathcal{X}_{sint}$) and real ($\mathcal{X}_{real}$) datasets for estimating the Matérn Kernel $\mathcal{C}_{1/2}(\eta)$, $\mathcal{C}_{3/2}(\eta)$ and $\mathcal{C}_{5/2}(\eta)$. All datasets used are completely observed, so an initial step involves assigning missing data to the original dataset. Initially, the data is normalized with zero mean and unit standard deviation, and then missing data is generated according to the *Missing At Random* mechanism to simulate missing attributes, such that the Rate of Missing Instances ($rMiss$) varies from 10% to 80%, with each instance containing 70% of its attributes missing.

For each $rMiss$, all methods (CMI, ESD, EED, EMK-MC e EMK-UT) perform the estimation of the Matérn Kernel $\mathcal{C}_\nu(\cdot)$, where $\nu \in \left\{\frac{1}{2}; \frac{3}{2}; \frac{5}{2}\right\}$. The performance analysis is verified based on the Average Root Mean Square Error (ARMSE) for 30 independent runs. The error is obtained by computing the difference between the estimated value of the kernel $\mathbb{E}[\mathcal{C}_\nu(\cdot)]$ from the imputed versions of $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$, and the actual value of $\mathcal{C}_\nu(\cdot)$ considering the vectors $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ completely observed.

In Section 4.2, $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ are drawn from multivariate normal distributions with known parameters. Both $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ may contain missing entries. The aim of this experiment is to analyze the performance of the methods without the influence of the dataset's distribution estimation. In Section 4.3, the results obtained for real datasets are presented. The Section 4.4 presents a case study in which the estimated Matérn Kernel is integrated into a Machine Learning algorithm.

4.2 Multivariate normal data with known parameters

Three sets of synthetic data were created, each containing 100 samples and 3 attributes (data type: real). Each dataset was generated from a multivariate normal distribution $\{\mathcal{N}^{(c)}\}_{c=1}^{3}$, where c represents the number of normal components used to generate the data. The parameters $\boldsymbol{\mu}^{(c)}$ and $\boldsymbol{\Sigma}^{(c)}$ for each distribution, as well as the probability (*prob*) of a generated data belonging to the respective component c, are presented in Tables 1, 2 and 3, respectively, for distributions $\mathcal{N}^{(1)}$, $\mathcal{N}^{(2)}$ and $\mathcal{N}^{(3)}$.

Table 1 – Distribution Parameters $\mathcal{N}^{(1)}$.

$\boldsymbol{\mu}$	$\boldsymbol{\Sigma}$
$\begin{bmatrix}-0.30\\0.10\\2\end{bmatrix}$	$\begin{bmatrix}0.40 & 0.15 & 0.25\\0.15 & 0.25 & 0.10\\0.25 & 0.10 & 0.30\end{bmatrix}$

Source: elaborated by the author.

Table 2 – Distribution Parameters $\mathcal{N}^{(2)}$.

#Component 1 ($prob = 0.6$)		**#Component 2** ($prob = 0.4$)	
$\boldsymbol{\mu}^{(1)}$	$\boldsymbol{\Sigma}^{(1)}$	$\boldsymbol{\mu}^{(2)}$	$\boldsymbol{\Sigma}^{(2)}$
$\begin{bmatrix}-0.30\\0.10\\0.20\end{bmatrix}$	$\begin{bmatrix}0.40 & 0.15 & 0.25\\0.15 & 0.25 & 0.10\\0.25 & 0.10 & 0.30\end{bmatrix}$	$\begin{bmatrix}-0.45\\1.00\\0.40\end{bmatrix}$	$\begin{bmatrix}0.79 & 0.58 & 0.69\\0.58 & 0.53 & 0.40\\0.69 & 0.40 & 0.85\end{bmatrix}$

Source: elaborated by the author.

Table 3 – Distribution Parameters $\mathcal{N}^{(3)}$.

#Component 1 ($prob = 0.4$)		**#Component 2** ($prob = 0.3$)		**#Component 3** ($prob = 0.3$)	
$\boldsymbol{\mu}^{(1)}$	$\boldsymbol{\Sigma}^{(1)}$	$\boldsymbol{\mu}^{(2)}$	$\boldsymbol{\Sigma}^{(2)}$	$\boldsymbol{\mu}^{(3)}$	$\boldsymbol{\Sigma}^{(3)}$
$\begin{bmatrix}-0.10\\0.50\\0.35\end{bmatrix}$	$\begin{bmatrix}0.79 & 0.59 & 0.71\\0.59 & 0.64 & 0.47\\0.71 & 0.47 & 0.72\end{bmatrix}$	$\begin{bmatrix}-0.30\\0.10\\0.20\end{bmatrix}$	$\begin{bmatrix}0.40 & 0.15 & 0.25\\0.15 & 0.25 & 0.10\\0.25 & 0.10 & 0.30\end{bmatrix}$	$\begin{bmatrix}-0.45\\1.00\\0.40\end{bmatrix}$	$\begin{bmatrix}0.79 & 0.58 & 0.69\\0.58 & 0.53 & 0.40\\0.69 & 0.40 & 0.85\end{bmatrix}$

Source: elaborated by the author.

As mentioned earlier, the estimation of the dataset's distribution has no impact on the performance of the evaluated methods in this experiment, as the components used to generate the datasets are known. Thus, it is possible to make a percentage-based comparison to highlight the improvement of a specific method compared to others. Let M_1 and M_2 be any two methods. The Percentage of Improvement (p^{IMP}) of method M_1 compared to method M_2 can be obtained through the following calculation:

$$p^{\text{IMP}} = \frac{ARMSE^{(M_2)} - ARMSE^{(M_1)}}{ARMSE^{(M_2)}} \times 100. \tag{4.1}$$

The Table 4 presents the ARMSE's obtained along with the associated standard deviation for each type of estimated kernel, considering the dataset $\mathcal{X}_{sint} \sim \mathcal{N}^{(1)}$.

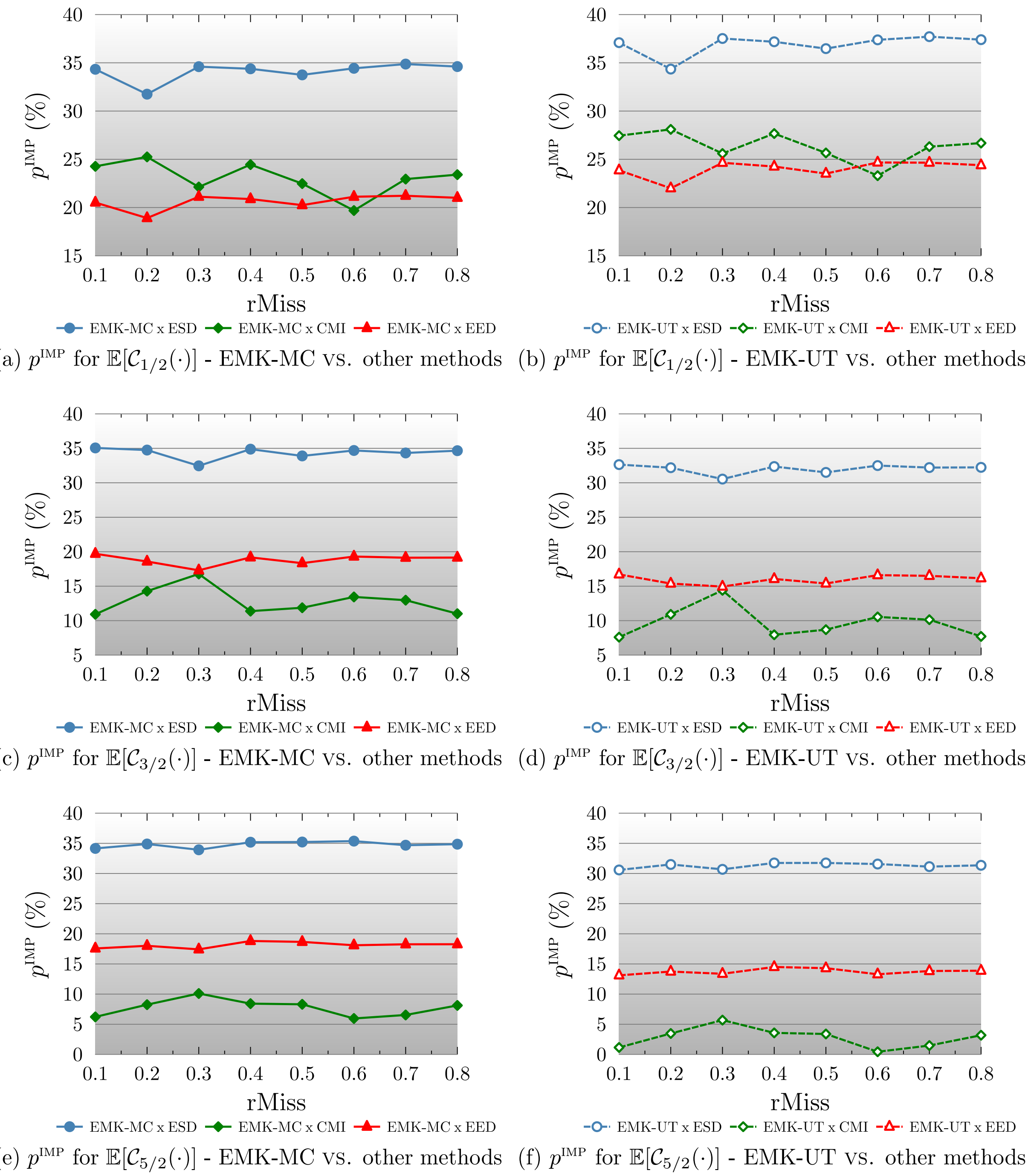
By analyzing the obtained data, it can be concluded that the EMK-UT method shows the best performance in estimating the Matérn kernel $\mathcal{C}_{1/2}(\cdot)$, followed by the EMK-MC, EED, CMI and ESD methods. Regarding the estimation of the Matérn kernel $\mathcal{C}_{3/2}(\cdot)$ and $\mathcal{C}_{5/2}(\cdot)$, the proposed EMK-MC approach presents the best results, followed by the EMK-UT, CMI, EED and ESD methods.

In addition to the previous results, Figure 6 presents the percentage of improvement of the proposed approaches compared to the other methods based on the results presented in Table 4.

Table 4 – Synthetic dataset - ARMSE for $\mathcal{X}_{sint} \sim \mathcal{N}^{(1)}$.

Kernel	**rMiss**	**CMI**	**ESD**	**EED**	**EMK-MC**	**EMK-UT**
$\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$	0.10	0.2395 ±0.04	0.2762 ±0.05	0.2282 ±0.04	0.1814 ±0.03	**0.1738 ±0.03**
	0.20	0.2561 ±0.03	0.2805 ±0.03	0.2361 ±0.03	0.1914 ±0.02	**0.1841 ±0.02**
	0.30	0.2482 ±0.03	0.2955 ±0.04	0.2450 ±0.03	0.1933 ±0.03	**0.1846 ±0.02**
	0.40	0.2480 ±0.02	0.2856 ±0.02	0.2368 ±0.02	0.1874 ±0.02	**0.1794 ±0.02**
	0.50	0.2482 ±0.02	0.2904 ±0.03	0.2413 ±0.03	0.1924 ±0.03	**0.1845 ±0.03**
	0.60	0.2401 ±0.03	0.2941 ±0.02	0.2445 ±0.02	0.1928 ±0.02	**0.1842 ±0.02**
	0.70	0.2429 ±0.03	0.2874 ±0.02	0.2376 ±0.02	0.1872 ±0.02	**0.1790 ±0.02**
	0.80	0.2439 ±0.02	0.2857 ±0.02	0.2365 ±0.02	0.1868 ±0.02	**0.1788 ±0.02**
$\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$	0.10	0.2729 ±0.05	0.3743 ±0.05	0.3027 ±0.05	**0.2431 ±0.04**	0.2522 ±0.04
	0.20	0.2735 ±0.04	0.3593 ±0.04	0.2880 ±0.04	**0.2345 ±0.03**	0.2437 ±0.03
	0.30	0.2890 ±0.04	0.3562 ±0.04	0.2909 ±0.04	**0.2406 ±0.03**	0.2474 ±0.03
	0.40	0.2734 ±0.03	0.3720 ±0.03	0.2997 ±0.03	**0.2423 ±0.02**	0.2516 ±0.02
	0.50	0.2724 ±0.03	0.3632 ±0.03	0.2940 ±0.03	**0.2401 ±0.02**	0.2488 ±0.02
	0.60	0.2788 ±0.03	0.3695 ±0.03	0.2990 ±0.03	**0.2413 ±0.02**	0.2494 ±0.02
	0.70	0.2754 ±0.04	0.3650 ±0.03	0.2964 ±0.03	**0.2397 ±0.03**	0.2475 ±0.03
	0.80	0.2737 ±0.03	0.3729 ±0.03	0.3013 ±0.03	**0.2436 ±0.03**	0.2526 ±0.03
$\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$	0.10	0.2710 ±0.03	0.3859 ±0.06	0.3083 ±0.05	**0.2541 ±0.03**	0.2679 ±0.04
	0.20	0.2722 ±0.03	0.3836 ±0.04	0.3046 ±0.03	**0.2498 ±0.02**	0.2628 ±0.02
	0.30	0.2779 ±0.03	0.3781 ±0.05	0.3025 ±0.04	**0.2498 ±0.03**	0.2621 ±0.03
	0.40	0.2787 ±0.03	0.3937 ±0.03	0.3143 ±0.03	**0.2552 ±0.02**	0.2687 ±0.02
	0.50	0.2754 ±0.03	0.3899 ±0.04	0.3105 ±0.03	**0.2525 ±0.02**	0.2661 ±0.02
	0.60	0.2657 ±0.03	0.3865 ±0.03	0.3050 ±0.03	**0.2498 ±0.02**	0.2645 ±0.02
	0.70	0.2777 ±0.03	0.3974 ±0.03	0.3175 ±0.03	**0.2595 ±0.02**	0.2736 ±0.02
	0.80	0.2746 ±0.03	0.3874 ±0.02	0.3087 ±0.03	**0.2523 ±0.02**	0.2659 ±0.02

Source: elaborated by the author.



(a) p^{IMP} for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$ - EMK-MC vs. other methods (b) p^{IMP} for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$ - EMK-UT vs. other methods

(c) p^{IMP} for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$ - EMK-MC vs. other methods (d) p^{IMP} for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$ - EMK-UT vs. other methods

(e) p^{IMP} for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$ - EMK-MC vs. other methods (f) p^{IMP} for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$ - EMK-UT vs. other methods

Figure 6 – p^{IMP} between the evaluated methods on $\mathcal{X}_{sint} \sim \mathcal{N}^{(1)}$.

Source: elaborated by the author.

Considering the estimation of the Matérn Kernel $\mathcal{C}_{1/2}(\cdot)$ (Figures 6a and 6b), the proposed approach EMK-UT shows the best results. On average, this method presents an improvement of 24% compared to EED, about 26% compared to CMI and almost 37% compared to ESD method.

Analyzing the results obtained for the estimation of the Matérn Kernel $\mathcal{C}_{3/2}(\cdot)$ (Figures 6c and 6d), the EMK-MC method proves to be superior. On average, the proposed approach presents an improvement of almost 13% compared to CMI, about 19% compared to EED and over 34% compared to ESD method.

Finally, considering the estimation of the Matérn Kernel $\mathcal{C}_{5/2}(\cdot)$ (Figures 6e and 6f), once again the proposed EMK-MC approach shows the best results. On average, this approach presents an improvement of 7.74% compared to CMI, approximately 18% compared to EED and almost 35% compared to ESD method.

The Table 5 presents the results related to the methods CMI, ESD, EED, EMK-MC and EMK-UT when applied to the dataset $\mathcal{X}_{sint} \sim \mathcal{N}^{(2)}$. The Figure 7, shows the percentage of improvement of the proposed approaches compared to the other methods based on the results presented in Table 5.

Once again, the EMK-UT method presents the lowest ARMSE's in estimating the kernel $\mathcal{C}_{1/2}(\cdot)$, followed by the EMK-MC, EED, ESD and CMI methods. Regarding the estimation of the Matérn Kernel $\mathcal{C}_{3/2}(\cdot)$ and $\mathcal{C}_{5/2}(\cdot)$, the proposed EMK-MC approach shows the best results, followed by the EMK-UT, EED, CMI, and ESD methods.

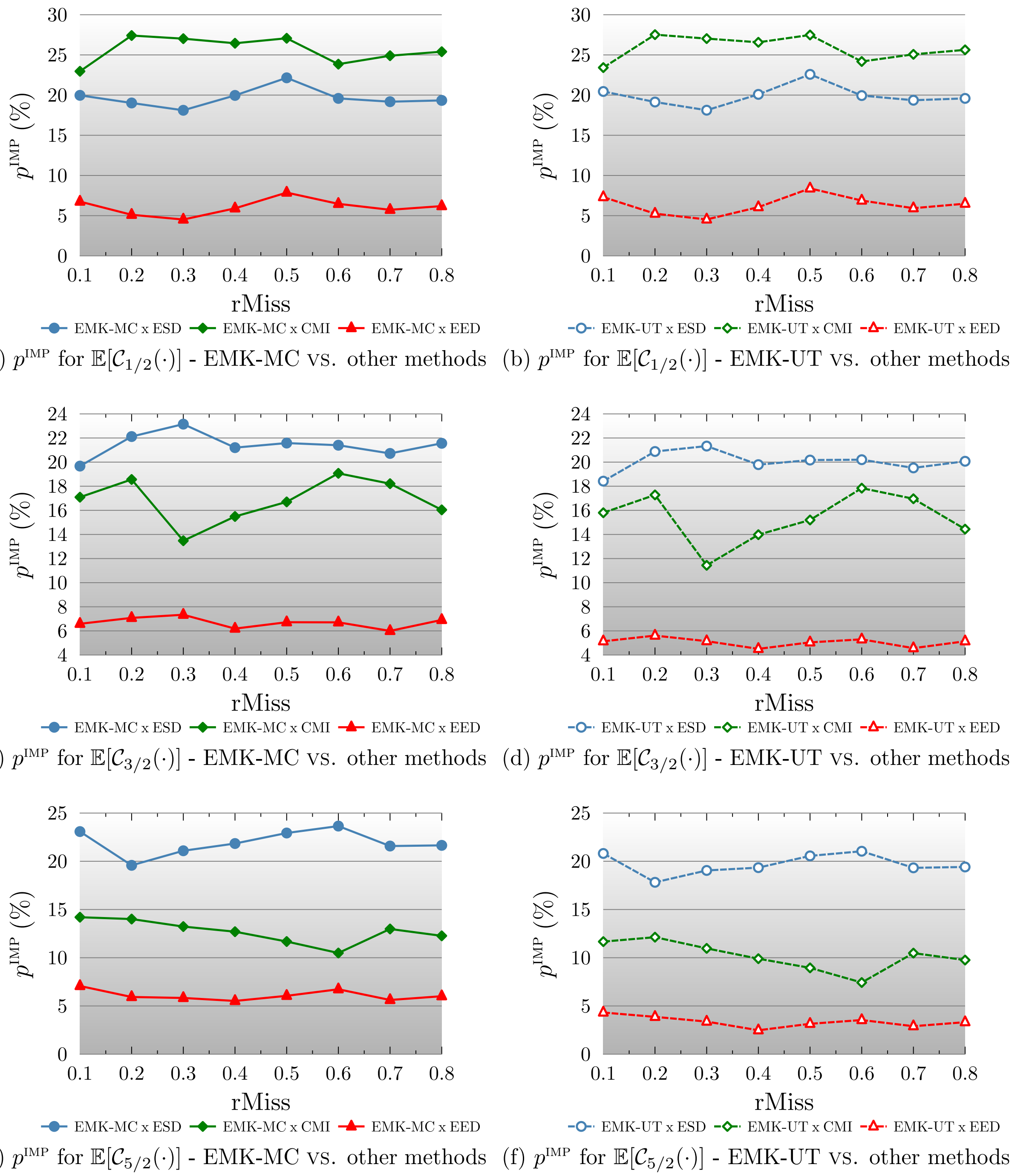
Regarding the estimation of the Matérn Kernel $\mathcal{C}_{1/2}(\cdot)$ (Figures 7a and 7b), the proposed approaches EMK-UT and EMK-MC can be considered equivalent. Taking into account the slight advantage of the EMK-UT method, on average, this approach presents an improvement of 6.35% compared to EED, almost 20% compared to ESD, and over 25% compared to the CMI method.

For the estimation of the Matérn Kernel $\mathcal{C}_{3/2}(\cdot)$ (Figures 7c and 7d), once again, the proposed approaches are equivalent. Analogously, due to the slight advantage of the EMK-MC method, on average, this approach presents an improvement of almost 7% compared to EED, about 17% compared to CMI, and 21.4% compared to the ESD method.

Table 5 – Synthetic dataset - ARMSE for $\mathcal{X}_{sint} \sim \mathcal{N}^{(2)}$.

Kernel	**rMiss**	**CMI**	**ESD**	**EED**	**EMK-MC**	**EMK-UT**
$\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$	0.10	0.1986 ±0.04	0.1912 ±0.03	0.1641 ±0.03	0.1530 ±0.02	**0.1521 ±0.02**
	0.20	0.1988 ±0.03	0.1782 ±0.02	0.1521 ±0.02	0.1443 ±0.02	**0.1441 ±0.02**
	0.30	0.1988 ±0.02	0.1771 ±0.02	0.1519 ±0.02	0.1451 ±0.02	**0.1450 ±0.02**
	0.40	0.1926 ±0.03	0.1770 ±0.03	0.1505 ±0.02	0.1417 ±0.01	**0.1414 ±0.01**
	0.50	0.2008 ±0.03	0.1881 ±0.02	0.1590 ±0.02	0.1465 ±0.02	**0.1456 ±0.02**
	0.60	0.1962 ±0.02	0.1858 ±0.02	0.1597 ±0.01	0.1494 ±0.01	**0.1488 ±0.01**
	0.70	0.1966 ±0.02	0.1827 ±0.02	0.1566 ±0.01	0.1476 ±0.01	**0.1473 ±0.01**
	0.80	0.2002 ±0.03	0.1851 ±0.02	0.1592 ±0.02	0.1493 ±0.02	**0.1489 ±0.02**
$\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$	0.10	0.2085 ±0.04	0.2152 ±0.05	0.1851 ±0.03	**0.1729 ±0.03**	0.1756 ±0.02
	0.20	0.2025 ±0.04	0.2117 ±0.04	0.1775 ±0.03	**0.1649 ±0.03**	0.1675 ±0.03
	0.30	0.1946 ±0.02	0.2191 ±0.02	0.1817 ±0.01	**0.1684 ±0.01**	0.1723 ±0.01
	0.40	0.1989 ±0.03	0.2133 ±0.03	0.1792 ±0.02	**0.1681 ±0.02**	0.1711 ±0.02
	0.50	0.2014 ±0.02	0.2140 ±0.02	0.1799 ±0.02	**0.1678 ±0.02**	0.1708 ±0.02
	0.60	0.2029 ±0.02	0.2089 ±0.02	0.1760 ±0.02	**0.1642 ±0.01**	0.1667 ±0.01
	0.70	0.2025 ±0.02	0.2089 ±0.02	0.1762 ±0.02	**0.1656 ±0.02**	0.1682 ±0.02
	0.80	0.2051 ±0.02	0.2196 ±0.02	0.1850 ±0.02	**0.1722 ±0.02**	0.1755 ±0.02
$\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$	0.10	0.1935 ±0.04	0.2158 ±0.04	0.1786 ±0.03	**0.1660 ±0.02**	0.1709 ±0.02
	0.20	0.2013 ±0.03	0.2152 ±0.04	0.1840 ±0.03	**0.1731 ±0.02**	0.1769 ±0.02
	0.30	0.1968 ±0.03	0.2164 ±0.03	0.1814 ±0.03	**0.1708 ±0.02**	0.1752 ±0.02
	0.40	0.1889 ±0.03	0.2110 ±0.03	0.1746 ±0.02	**0.1649 ±0.02**	0.1702 ±0.02
	0.50	0.1920 ±0.03	0.2200 ±0.02	0.1805 ±0.02	**0.1695 ±0.02**	0.1748 ±0.02
	0.60	0.1855 ±0.03	0.2175 ±0.02	0.1780 ±0.02	**0.1660 ±0.02**	0.1717 ±0.02
	0.70	0.1915 ±0.03	0.2125 ±0.02	0.1766 ±0.02	**0.1667 ±0.02**	0.1715 ±0.02
	0.80	0.1937 ±0.02	0.2169 ±0.02	0.1808 ±0.02	**0.1699 ±0.02**	0.1748 ±0.02

Source: elaborated by the author.



(a) p^{IMP} for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$ - EMK-MC vs. other methods
(b) p^{IMP} for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$ - EMK-UT vs. other methods
(c) p^{IMP} for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$ - EMK-MC vs. other methods
(d) p^{IMP} for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$ - EMK-UT vs. other methods
(e) p^{IMP} for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$ - EMK-MC vs. other methods
(f) p^{IMP} for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$ - EMK-UT vs. other methods

Figure 7 – p^{IMP} between the evaluated methods on $\mathcal{X}_{sint} \sim \mathcal{N}^{(2)}$.

Source: elaborated by the author.

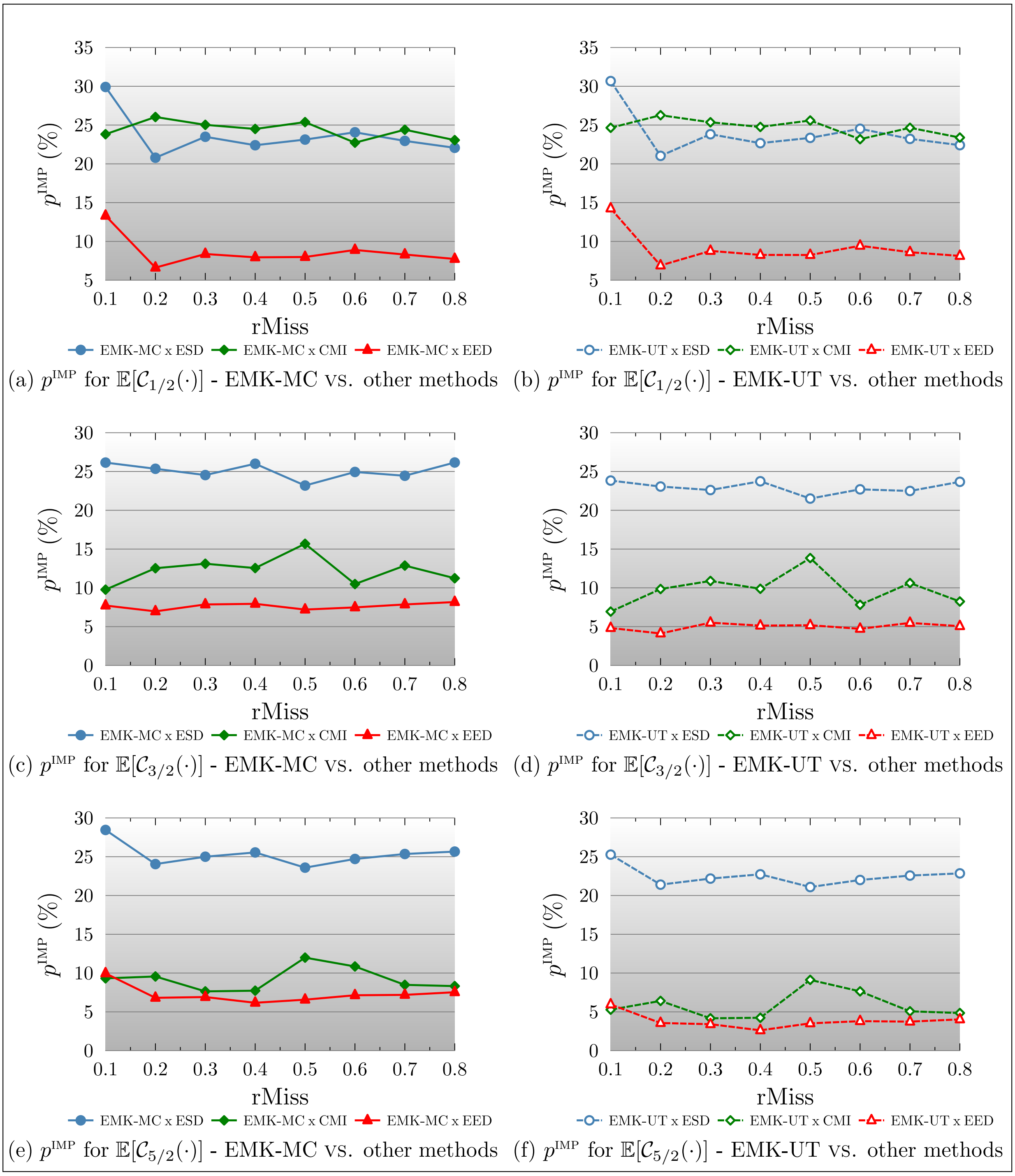
Finally, for the estimation of the Matérn Kernel $\mathcal{C}_{5/2}(\cdot)$ (Figures 7e and 7f), the proposed EMK-MC approach shows the best results. On average, this approach presents an improvement of 6.09% compared to EED, about 12% compared to CMI, and almost 22% compared to the ESD method.

The ARMSE's obtained considering $\mathcal{X}_{sint} \sim \mathcal{N}^{(3)}$ are shown in Table 6. The Figure 8 illustrates the percentage of improvement of the proposed approaches compared to the other methods based on the results presented in the following table.

Table 6 – Synthetic dataset - ARMSE for $\mathcal{X}_{sint} \sim \mathcal{N}^{(3)}$.

Kernel	**rMiss**	**CMI**	**ESD**	**EED**	**EMK-MC**	**EMK-UT**
$\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$	0.10	0.1800 ±0.03	0.1957 ±0.03	0.1582 ±0.02	0.1371 ±0.02	**0.1356 ±0.02**
	0.20	0.1948 ±0.03	0.1819 ±0.03	0.1543 ±0.02	0.1440 ±0.02	**0.1436 ±0.02**
	0.30	0.1893 ±0.03	0.1855 ±0.03	0.1549 ±0.02	0.1419 ±0.02	**0.1413 ±0.02**
	0.40	0.1896 ±0.02	0.1845 ±0.02	0.1555 ±0.02	0.1432 ±0.02	**0.1427 ±0.02**
	0.50	0.1894 ±0.02	0.1838 ±0.03	0.1536 ±0.02	0.1413 ±0.02	**0.1409 ±0.02**
	0.60	0.1871 ±0.02	0.1904 ±0.02	0.1587 ±0.02	0.1446 ±0.02	**0.1437 ±0.02**
	0.70	0.1915 ±0.02	0.1879 ±0.02	0.1579 ±0.02	0.1448 ±0.01	**0.1443 ±0.01**
	0.80	0.1896 ±0.02	0.1872 ±0.02	0.1581 ±0.02	0.1459 ±0.02	**0.1453 ±0.02**
$\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$	0.10	0.1802 ±0.04	0.2202 ±0.05	0.1762 ±0.04	**0.1626 ±0.03**	0.1677 ±0.03
	0.20	0.1784 ±0.02	0.2090 ±0.03	0.1677 ±0.02	**0.1561 ±0.02**	0.1608 ±0.02
	0.30	0.1904 ±0.03	0.2192 ±0.03	0.1795 ±0.02	**0.1654 ±0.02**	0.1697 ±0.02
	0.40	0.1816 ±0.03	0.2146 ±0.03	0.1725 ±0.03	**0.1588 ±0.03**	0.1636 ±0.03
	0.50	0.1968 ±0.03	0.2160 ±0.02	0.1788 ±0.02	**0.1659 ±0.02**	0.1696 ±0.02
	0.60	0.1858 ±0.02	0.2216 ±0.03	0.1798 ±0.02	**0.1663 ±0.02**	0.1713 ±0.02
	0.70	0.1879 ±0.02	0.2167 ±0.02	0.1777 ±0.01	**0.1637 ±0.01**	0.1679 ±0.01
	0.80	0.1818 ±0.02	0.2185 ±0.02	0.1757 ±0.02	**0.1613 ±0.01**	0.1668 ±0.01
$\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$	0.10	0.1745 ±0.04	0.2211 ±0.05	0.1757 ±0.04	**0.1582 ±0.03**	0.1652 ±0.03
	0.20	0.1821 ±0.03	0.2168 ±0.03	0.1767 ±0.03	**0.1646 ±0.02**	0.1704 ±0.02
	0.30	0.1793 ±0.02	0.2208 ±0.02	0.1779 ±0.02	**0.1656 ±0.02**	0.1719 ±0.02
	0.40	0.1786 ±0.02	0.2214 ±0.03	0.1756 ±0.02	**0.1648 ±0.02**	0.1710 ±0.02
	0.50	0.1886 ±0.02	0.2172 ±0.03	0.1777 ±0.02	**0.1660 ±0.02**	0.1714 ±0.02
	0.60	0.1882 ±0.02	0.2228 ±0.03	0.1806 ±0.02	**0.1677 ±0.02**	0.1738 ±0.02
	0.70	0.1789 ±0.02	0.2194 ±0.02	0.1765 ±0.02	**0.1638 ±0.01**	0.1699 ±0.01
	0.80	0.1823 ±0.02	0.2248 ±0.02	0.1807 ±0.02	**0.1671 ±0.02**	0.1735 ±0.02

Source: elaborated by the author.



(a) p^{IMP} for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$ - EMK-MC vs. other methods (b) p^{IMP} for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$ - EMK-UT vs. other methods

(c) p^{IMP} for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$ - EMK-MC vs. other methods (d) p^{IMP} for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$ - EMK-UT vs. other methods

(e) p^{IMP} for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$ - EMK-MC vs. other methods (f) p^{IMP} for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$ - EMK-UT vs. other methods

Figure 8 – p^{IMP} between the evaluated methods on $\mathcal{X}_{sint} \sim \mathcal{N}^{(3)}$.

Source: elaborated by the author.

From Table 6, it can be observed once again that the EMK-UT method presents the best performance in estimating the Matérn Kernel $\mathcal{C}_{1/2}(\cdot)$, followed by the methods EMK-MC, EED, ESD, and CMI. As for the estimation of the Matérn Kernel $\mathcal{C}_{3/2}(\cdot)$ and $\mathcal{C}_{5/2}(\cdot)$, the proposed EMK-MC approach shows the best results, followed by the methods EMK-UT, EED, CMI, and ESD.

Regarding Figures 8a and 8b, similar to the results observed for the estimation of the Matérn Kernel $\mathcal{C}_{1/2}(\cdot)$ from $\mathcal{X}_{sint} \sim \mathcal{N}^{(2)}$, the proposed methods EMK-MC and EMK-UT are again equivalent in terms of percentage of improvement. Considering the slight advantage of the EMK-UT method, on average, this approach shows an improvement of 9.08% compared to the EED method, nearly 24% compared to the ESD method, and about 25% compared to the CMI method.

A larger margin of difference between the proposed approaches for the estimation of the Matérn Kernel $\mathcal{C}_{3/2}(\cdot)$ (Figures 8c and 8d) gives the EMK-MC method an advantage. Thus, on average, the proposed approach presents an improvement of 7.65% compared to the EED, about 12% compared to the CMI, and 25.1% compared to the ESD method.

Finally, for the estimation of the Matérn Kernel $\mathcal{C}_{5/2}(\cdot)$ (Figures 8e and 8f), the proposed EMK-MC approach shows the best results. On average, this approach presents an improvement of over 7% compared to the EED, about 9% compared to the CMI, and slightly over 25% compared to the ESD method.

In summary, concerning the synthetic data, the superiority of the proposed EMK-MC and EMK-UT methods is evident, as they achieve the best performance in all evaluated situations. From the results presented in Tables 4-6, it can be noticed that the largest performance differences are obtained in estimating the Matérn Kernel $\mathcal{C}_{1/2}(\eta)$. However, contrary to expectations, the ARMSE does not follow a trend of increasing with the number of samples with missing data for any evaluated method, regardless of the type of estimated kernel.

4.3 Experiments on real-world data

A total of 22 datasets from the UCI Machine Learning repository (KELLY *et al.*, 2013) were evaluated. The specific characteristics of each dataset are presented in Table 7.

Table 7 – Datasets description.

Dataset	**Size**	**Features**	**Task**	**Attribute Type**
Iris	150	4	Classification	Real
Servo	167	4	Regression	Categorical, Integer
Concrete Slump	103	7	Regression	Real
Haberman	306	3	Classification	Integer
Breast tissue	106	9	Classification	Real
Coluna	310	6	Classification	Real
Glass	214	9	Classification	Real
Wine	178	13	Classification	Integer, Real
Ecoli	336	7	Classification	Real
Auto MPG	392	7	Regression	Categorical, Real
Monk 1	556	6	Classification	Categorical
Monk 2	601	6	Classification	Categorical
Monk 3	554	6	Classification	Categorical
Pima Indians Diabetes	768	8	Classification	Integer, Real
Energy	768	8	Classification, Regression	Integer, Real
Forest Fires	517	12	Regression	Real
Boston Housing	506	13	Regression	Integer, Real
Concrete Compression	1030	8	Regression	Real
Liver Disorders BUPA Medical Research	345	5	Classification	Categorical, Integer, Real
Computer Hardware	209	6	Regression	Integer
Statlog (vehicle silhouettes)	846	18	Classification	Integer
Heart Disease (Cleveland)	303	13	Classification	Categorical, Integer, Real

Source: elaborated by the author.

To estimate the probabilistic model for the data, a GMM with up to 10 components is used. The best model is selected after 10 runs based on the Bayesian Information Criterion (BIC). Due to the various situations when estimating the Matérn Kernel $\mathcal{C}_{1/2}(\eta)$, $\mathcal{C}_{3/2}(\eta)$, and $\mathcal{C}_{5/2}(\eta)$ with $rMiss \in \{0.1 ... 0.8\}$ for 22 datasets, the Relative Success Rate (RSR) is used as a performance measure for the methods. This rate measures how many times a method outperforms the others considering the number of evaluated configurations, defined as:

$$RSR = \frac{\#\textsc{sms}}{\#\textsc{tcs}}, \tag{4.2}$$

where $\#\textsc{sms}$ is the number of times a method outperformed the others considering the ARMSE for the estimated kernel value and $\#\textsc{tcs}$ is the total number of evaluated configurations. It is worth mentioning that for this metric, for the purpose of fair comparison, the EMK-MC and EMK-UT methods are compared separately to the other methods considered.

The results in terms of RSR are divided into 3 groups according to the percentage of instance removal, namely:

1. Overall removal rate ($\mathcal{G}_1$): $rMiss \in \{0.1 \ldots 0.8\}$;
2. Low/Medium removal rate ($\mathcal{G}_2$): $rMiss \in \{0.1 \ldots 0.4\}$;
3. Medium/High removal rate ($\mathcal{G}_3$): $rMiss \in \{0.5 \ldots 0.8\}$.

Example 2 *Considering the complete scenario of the experiment, there are 528 evaluated configurations (*3 *kernel types* $\times$ 8 *instance removal percentages* $\times$ 22 *datasets). If a certain method obtained the lowest ARMSE's values in 250 cases, it means its relative success rate is approximately* 0.47.

Example 3 *Considering a scenario where only the estimation of the Matérn Kernel* $\mathcal{C}_{1/2}(\cdot)$ *is performed based on* $\mathcal{G}_1$*, there are 88 evaluated configurations (*1 *kernel type* $\times$ 4 *instance removal percentages* $\times$ 22 *datasets). If a certain method obtained the lowest ARMSE's values in 34 cases, it means its relative success rate is approximately* 0.38.

It is worth noting that all results obtained in terms of ARMSE (along with the associated standard deviation) comparing the methods CMI, ESD, EED, EMK-MC, and EMK-UT when applied to each dataset from the UCI Machine Learning repository, considering each type of estimated kernel, are provided additionally in Appendix B.

Additionally, for each group, the RSR is computed for the estimation of all kernels together and considering each individual kernel. The results obtained in terms of RSR are presented in Tables 8 and 9, respectively, for the EMK-MC and EMK-UT methods.

When considering the Tables 8 and 9, it can be inferred that the RSR of the CMI method is consistently the lowest. This indicates that the CMI method had inferior performance compared to the other methods, regardless of the kernel used. Among the ESD, EED, and CMI methods, there is no clear differentiation in performance based on the estimated kernel type, considering only the RSR.

Similar to the synthetic data, the ARMSE for the proposed methods does not increase with the amount of missing data. This can be inferred from the relative success rates in both the overall context ($\mathcal{G}_1$) and the specific cases represented by groups $\mathcal{G}_2$ and $\mathcal{G}_3$. However, once again, the proposed EMK-MC and EMK-UT methods consistently outperform the other methods in most evaluated scenarios, providing evidence that the assumptions about their formulations do not degrade their performance for real-world datasets. Additionally, in Appendix A, Figures 9, 10, 11, and 12 directly illustrate the comparison between the number of times a particular method outperformed the others in each group $\mathcal{G}_n$.

Table 8 – UCI dataset - RSR considering the EMK-MC method.

Group	Kernel	CMI	ESD	EED	EMK-MC
$\mathcal{G}_1$	$\mathbb{E}[\mathcal{C}_\nu(\cdot)]$	0.0947	0.1667	0.3030	**0.4356**
	$\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$	0.0852	0.1648	0.3409	**0.4091**
	$\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$	0.0966	0.1647	0.2898	**0.4489**
	$\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$	0.1023	0.1704	0.2784	**0.4489**
$\mathcal{G}_2$	$\mathbb{E}[\mathcal{C}_\nu(\cdot)]$	0.1023	0.1780	0.2803	**0.4394**
	$\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$	0.1023	0.1704	0.3523	**0.3750**
	$\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$	0.1023	0.1818	0.2386	**0.4773**
	$\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$	0.1023	0.1818	0.2500	**0.4659**
$\mathcal{G}_3$	$\mathbb{E}[\mathcal{C}_\nu(\cdot)]$	0.0871	0.1553	0.3258	**0.4318**
	$\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$	0.0682	0.1591	0.3295	**0.4432**
	$\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$	0.0909	0.1477	0.3409	**0.4205**
	$\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$	0.1023	0.1591	0.3068	**0.4318**

Source: elaborated by the author.

Table 9 – UCI dataset - RSR considering the EMK-UT method.

Group	Kernel	CMI	ESD	EED	EMK-UT
$\mathcal{G}_1$	$\mathbb{E}[\mathcal{C}_\nu(\cdot)]$	0.1230	0.1667	0.3239	**0.3864**
	$\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$	0.0909	0.1648	**0.4375**	0.3068
	$\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$	0.1250	0.1648	0.2898	**0.4204**
	$\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$	0.1534	0.1705	0.2443	**0.4318**
$\mathcal{G}_2$	$\mathbb{E}[\mathcal{C}_\nu(\cdot)]$	0.1250	0.1780	0.3106	**0.3864**
	$\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$	0.0909	0.1705	**0.4659**	0.2727
	$\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$	0.1364	0.1818	0.2500	**0.4318**
	$\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$	0.1477	0.1818	0.2160	**0.4545**
$\mathcal{G}_3$	$\mathbb{E}[\mathcal{C}_\nu(\cdot)]$	0.1212	0.1553	0.3371	**0.3864**
	$\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$	0.0909	0.1591	**0.4091**	0.3409
	$\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$	0.1136	0.1477	0.3296	**0.4091**
	$\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$	0.1591	0.1591	0.2727	**0.4091**

Source: elaborated by the author.

Considering the performance in terms of ARMSE for both small and large amounts of missing samples, the proposed approaches outperform the other methods in most evaluated scenarios. As expected, the performance differences between EMK and EED are smaller than the differences when any other method is considered. It is worth noting that the EED method was designed to estimate the Euclidean distance between incomplete data vectors. Since the Matérn Kernel specifies the covariance between measurements based on the Euclidean distance between points of interest, the estimation error of the kernel should be reduced if the distance estimation is made with arbitrarily better precision.

However, it is important to highlight the difference in the level at which the estimation problem is addressed by the approaches developed in this thesis. The EMK-MC and EMK-UT methods have the capacity to directly address the kernel estimation problem, estimating the transformation of interest rather than embedding it in a lower-level structure. From the results obtained in MESQUITA (2017), a statement can be made by applying Jensen's inequality directly. Formally, Jensen's inequality states that for a convex function f and a random variable X with probability density function $p(x)$, it holds that:

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)], \tag{4.3}$$

in other words, for convex functions, the function of the mean is less than or equal to the mean of the function.

Given that η is a random variable and follows a Nakagami distribution, the following statement can be obtained:

$$k_{\mathcal{C}_\nu(\eta^{EED})}(\boldsymbol{X}_i, \boldsymbol{X}_j) \leq k_{\mathcal{C}_\nu(\eta^{EMK-MC})}(\boldsymbol{X}_i, \boldsymbol{X}_j), \tag{4.4}$$

in the same way that

$$k_{\mathcal{C}_\nu(\eta^{EED})}(\boldsymbol{X}_i, \boldsymbol{X}_j) \leq k_{\mathcal{C}_\nu(\eta^{EMK-UT})}(\boldsymbol{X}_i, \boldsymbol{X}_j). \tag{4.5}$$

Therefore, from the inequalities presented earlier, it is possible to conclude that estimating the Euclidean distance between incomplete data vectors to only then obtain the kernel tends to underestimate the expected value of the desired kernel. On the other hand, in the proposed approaches, the kernel function estimates are obtained directly from the approximated functions represented in ψ and φ.

4.4 Case-study: LSSVR with Matérn Kernel for Incomplete Data

A more comprehensive evaluation of the proposed approaches involves applying the estimated Matérn Kernel to a Machine Learning algorithm, specifically, the Least Squares Support Vector Regression (LSSVR) model (SUYKENS; VANDEWALLE, 1999; SUYKENS *et al.*, 2001). This way, the proposed EMK-MC and EMK-UT approaches are integrated into a state-of-the-art kernel machine designed for regression tasks, rather than being limited exclusively to isolated estimation techniques.

4.4.1 Least Squares Support Vector Regression (LSSVR)

Two modifications were made to create the LSSVR model from the Support Vector Regression (SVR) (VAPNIK, 1995). These modifications are present in the primal optimization formulation. The first modification involves using equality constraints instead of inequalities. Consequently, the solution is obtained by solving a set of linear equations instead of a quadratic problem. The second change is in the cost function, achieved by incorporating the sum of squared approximation errors, weighted by a regularization term.

Let $\mathcal{X} = \{(\boldsymbol{X}_1, Y_1),\ \dots\ ,(\boldsymbol{X}_N, Y_N)\}$ be a dataset, where $\boldsymbol{X} \in \mathbb{R}^D$ and $Y \in \mathbb{R}$. The goal is to estimate the parameters of the nonlinear regression model

$$f(\boldsymbol{X}) = \langle \boldsymbol{\omega}, \varrho(\boldsymbol{X}) \rangle + b, \tag{4.6}$$

where $\langle \cdot, \cdot \rangle$ denotes the dot-product, $\boldsymbol{\omega} \in \mathbb{R}^D$ is the weight vector, $b \in \mathbb{R}$ is the bias, and $\varrho(\boldsymbol{X})$ represents the nonlinear mapping from the input feature space to a higher-dimensional space. The goal is to minimize the following cost function:

$$min\ J_P(\boldsymbol{\omega}, \boldsymbol{e}) = \frac{1}{2}\boldsymbol{\omega}^T\boldsymbol{\omega} + \zeta\frac{1}{2}\sum_{i=1}^{N} e_i^2, \tag{4.7}$$

$$\text{subject to} \quad Y_i = \langle \boldsymbol{\omega}^T, \varrho(\boldsymbol{X}_i) \rangle + b + e_i, \quad i = 1, \dots, N,$$

where $e_i = Y_i - f(\boldsymbol{X}_i)$ is the error in the ith observation, and $\zeta > 0$ is the regularization parameter. By proceeding with the solution using the method of *Lagrange multipliers*, the problem given in Equation 4.7 can be rewritten in its dual form as follows:

$$L(\boldsymbol{\omega}, b, \boldsymbol{e}, \boldsymbol{\iota}) = \frac{1}{2}\boldsymbol{\omega}^T\boldsymbol{\omega} + \zeta\frac{1}{2}\sum_{i=1}^{N} e_i^2 - \sum_{i=1}^{N} \iota_i\{\boldsymbol{\omega}^T\varrho(\boldsymbol{X}_i) + b + e_i - Y_i\}, \tag{4.8}$$

where $\{\iota_i\}_{i=1}^{N} \in \mathbb{R}$ are the *Lagrange multipliers.*

Given the optimality conditions, the solution can then be directly found by solving the following linear system:

$$\mathbf{Az} = \mathbf{l}, \tag{4.9}$$

such that:

$$\mathbf{A} = \left[\begin{array}{c|c} 0 & \mathbf{1}^T \\ \hline \mathbf{1} & \boldsymbol{\Psi} + \zeta^{-1}\mathbf{I} \end{array}\right], \quad \mathbf{z} = \left[\begin{array}{c} b \\ \hline \boldsymbol{\iota} \end{array}\right], \quad \mathbf{l} = \left[\begin{array}{c} 0 \\ \hline \boldsymbol{Y} \end{array}\right], \tag{4.10}$$

where $\boldsymbol{\Psi} \in \mathbb{R}^{N \times N}$ is the kernel matrix whose entries are given by $\Psi_{i,j} = \langle \varrho(\boldsymbol{X}_i), \varrho(\boldsymbol{X}_j) \rangle$, $i, j = 1, \ldots, N$. Additionally, $\boldsymbol{Y} = [Y_1 \ \cdots \ Y_N]^T$ and the symbol $\mathbf{1}$ represents a vector of 1's with dimension N.

Applying the *kernel trick* to $\Psi_{i,j}$, it is obtained:

$$\begin{aligned} \Psi_{i,j} &= \langle \varrho(\boldsymbol{X}_i), \varrho(\boldsymbol{X}_j) \rangle \\ &= k(\boldsymbol{X}_i, \boldsymbol{X}_j), \quad i, j = 1, \ldots, N, \end{aligned} \tag{4.11}$$

where $k(\cdot, \cdot)$ is the kernel function.

Effectively, the LSSVR model is defined in terms of the coefficients $\boldsymbol{\iota}$, b and the kernel function $k(\cdot, \cdot)$. Therefore, the resulting model for nonlinear regression is given by:

$$f(\mathbf{x}) = \sum_{i=1}^{N} \iota_i k(\boldsymbol{X}_i, \boldsymbol{X}_j) + b, \tag{4.12}$$

where ι_i and b are the solution of the linear system presented in Equation 4.9. For further details, refer to Suykens *et al.* (2002).

4.4.2 Experiments and Results

Considering the absence of attributes in the training observations, the direct estimation of the kernel matrix $\boldsymbol{\Psi}$ becomes a challenge. In the following experiments, the CMI, ESD, and EED methods were used to complete the data matrix, which was then subjected to the LSSVR model for the regression task. However, it should be noted that the proposed methods EMK-MC and EMK-UT were applied to directly estimate the kernel matrices $\Psi_{i,j}$.

The *Concrete Slump* dataset was considered for conducting the experiments, so that the performance analysis is evaluated in terms of ARMSE for 30 independent runs when estimating each of the three output variables of the dataset (Output$_{(1)}$: *Slump (cm)*; Output$_{(2)}$: *Flow (cm)*; Output$_{(3)}$: *28-day Compressive Strength (Mpa)*). For simplification, the regularization parameter ζ was fixed at 0.05. The results obtained in terms of ARMSE along with their standard deviation for each missing instance rate while also considering the application of the Matérn kernels $\mathcal{C}_{1/2}$, $\mathcal{C}_{3/2}$, and $\mathcal{C}_{5/2}$, are detailed in Appendix C. The following tables relate the results obtained in terms of the Relative Success Rate, considering the estimation of the three kernels for group $\mathcal{G}_1$.

Table 10 – Case-Study - RSR for *Concrete Slump* dataset considering the EMK-MC method.

	Output$_{(1)}$	Output$_{(2)}$	Output$_{(3)}$
CMI	0.0417	0.0000	0.1667
ESD	0.4583	0.5000	0.1250
EED	0.0417	0.0000	0.0000
EMK-MC	**0.4583**	**0.5000**	**0.7083**

Source: elaborated by the author.

Table 11 – Case-Study - RSR for *Concrete Slump* dataset considering the EMK-UT method.

	Output$_{(1)}$	Output$_{(2)}$	Output$_{(3)}$
CMI	0.0417	0.0000	0.2083
ESD	0.3750	0.3333	0.1250
EED	0.0000	0.0000	0.0000
EMK-UT	**0.5833**	**0.6667**	**0.6667**

Source: elaborated by the author.

From the results presented in Tables 10 and 11, it can be observed that the proposed methods achieve the best results for all three output types of the evaluated dataset. Furthermore, a substantial difference margin is notable between the frequency at which the EMK-UT approach attains the lowest mean squared error in comparison to the other methods.

4.5 Conclusion

In this chapter, the methodology used to evaluate the proposed methods, as well as the characteristics of the datasets used, were described. A total of 22 datasets from the UCI Machine Learning repository, along with 3 synthetic datasets, were included. The performance metrics used to compare the proposals in this thesis with three other methods when applied to the datasets used in the simulations were also presented.

Comparative tables that highlight the performance of the methods under different scenarios are provided, as well as figures illustrating the improvements of the proposed approaches over other methods. The tests were performed under various conditions, in addition to applying the estimated Matérn Kernel in the LSSVR method, to demonstrate the robustness of the proposed methods and provide consistent evidence for the conclusions presented in Chapter 5.

5 CONCLUSIONS AND FUTURE WORKS

The purpose of this chapter is to provide the conclusions and final considerations regarding the proposed methods developed in this thesis. Additionally, final remarks about the achievement of the objectives described in the introduction will be discussed, as well as suggestions for future research in this area.

5.1 Conclusions

In this work, two new methods for estimating the Matérn Kernel in the presence of incomplete datasets are presented. Typically, these methods assume that the dataset is fully observed; however, incomplete data is a common occurrence in various domains.

For modeling the data considering the existence of missing attributes, a Gaussian Mixture Model is used. As presented in Chapter 2, the parameters of the model that maximize the likelihood function, conditioning on the observed data, are estimated using the Expectation-Maximization algorithm in its extended form to deal with incomplete data. At the same time, the EM algorithm iteratively estimates the missing values.

In Chapter 3, the estimation problem is formulated under the assumption that the Euclidean distances of the kernel function follow a Nakagami probability distribution. Once the distribution parameters are known, approximation techniques are used as an alternative to the infeasibility of solving the problem analytically.

The method Expected Matérn Kernel via Monte Carlo (EMK-MC), generates samples through MC simulation that represent the distribution of the Euclidean distance between two vectors; then, the kernel estimation is directly obtained through the expected value of the generated distribution. On the other hand, the method Expected Matérn Kernel via Unscented Transform (EMK-UT), can estimate the Matérn Kernel function by summing a set of transformed points, weighted by their respective weights.

The performances of the proposed methods are compared to three other methods on 22 real-world datasets, as well as 3 synthetic datasets, as discussed in Chapter 4. According to the results obtained, it is possible to infer that the EMK-MC and EMK-UT methods are valid alternatives for estimating the Matérn Kernel function in incomplete datasets, as the consistent performance is evident in the majority of the various scenarios evaluated for real-world datasets and in all scenarios evaluated on synthetic datasets. The proposed approaches provide the possibility of working with the $\mathcal{C}_{1/2}$, $\mathcal{C}_{3/2}$, and $\mathcal{C}_{5/2}$ kernels even in the presence of a large number of instances with missing attributes, where it would be impossible to use the original formulation of the Matérn Kernel.

It is also worth noting that the EMK-UT approach is based on the UT. This means that the asymptotic complexity requires only $O(1)$ samples, or more precisely, three samples. Another point to highlight is that both proposed methodologies inherently consider the uncertain nature of missing entries. The use of direct imputation techniques would result in the loss of this information, as discussed in Chapter 2.

As discussed at the end of Section 4.3, it was expected that the proposed methods and EED would yield the two best results. Considering the general case for real-world problems (where 528 test scenarios are evaluated in $\mathcal{G}_1$), the proposed method EMK-MC outperformed in 230 scenarios, while the second-best model (EED) succeeded in 160 scenarios. On the other hand, the EMK-UT method was superior in 204 scenarios, while the EED method had an advantage in 171 scenarios.

When analyzing only the estimation of the Matérn Kernel $\mathcal{C}_{1/2}(\cdot)$ (totaling 176 test scenarios in $\mathcal{G}_1$), the difference between the two best methods was more subtle. However, again, the EMK-MC method proved superior in 72 scenarios, while the EED method achieved the best results in 60 scenarios. Regarding the EMK-UT approach, it succeeded in 54 scenarios, while the EED method showed superiority in 77 evaluated scenarios.

Regarding the estimation of the Matérn Kernel $\mathcal{C}_{3/2}(\cdot)$ and $\mathcal{C}_{5/2}(\cdot)$ (totaling 176 test scenarios in $\mathcal{G}_1$ for each kernel), the difference in performance between the two methods increases again. In the case of $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$, the EMK-MC method achieves the best results in 79 scenarios, and the EED method in 51 scenarios. On the other hand, the EMK-UT method achieves the best results in 74 scenarios, compared to 51 for the EED method.

Finally, in $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$, the EMK-MC approach has an advantage in 79 evaluated scenarios, while the EED method achieves the best results in 49 scenarios. Meanwhile, the proposed EMK-UT approach achieves the best results in 76 evaluated scenarios, compared to 43 scenarios for the EED method.

When considering the results obtained regarding the application of the estimated Matérn Kernel in a Machine Learning algorithm (presented in Appendix C), it is possible to observe that the best results for the methods developed in this Thesis are achieved when applying the Matérn Kernels $\mathcal{C}_{3/2}$ and $\mathcal{C}_{5/2}$, where the smallest errors in the evaluated regression task are evident.

5.2 Future Works

As future work, the use of parametric and semi-parametric densities for modeling the dataset is intended to be evaluated. The developments presented do not depend on the specific assumption that the data distribution can be represented by a Gaussian mixture model, therefore, it can be easily replaced.

Another suggested work is to model the squared distance present in the Matérn Kernel $\mathcal{C}_{5/2}$ based on a specific distribution, instead of computing the distance from the resulting distributions of the approximation processes. From the discussions conducted, this measure can be considered a random variable following a Gamma distribution. Thus, its modeling will depend only on the parameters of that distribution.

Additionally, it is also suggested to generalize the presented results to other kernels constructed based on the standard Matérn Kernel $\mathcal{C}_{1/2}(\eta)$, $\mathcal{C}_{3/2}(\eta)$, and $\mathcal{C}_{5/2}(\eta)$. Suggestions for extensions are briefly presented in Appendix D.

Lastly, but not least, a clear extension of this study consists of the direct application of the methods discussed in the previous chapters to Machine Learning algorithms. An initial application was carried out as discussed in Section 4.4. Therefore, the intention is to replicate the application to more regression problems with the aim of confirming the initial conclusions obtained. The next step, will be to organize the results in order to synthesize the study resulting from this thesis, with the aim of submitting articles to scientific journals. It is worth noting that at least one article is already in an advanced stage of writing and will be submitted for publication in the coming months.

BIBLIOGRAPHY

BELANCHE, L. A.; KOBAYASHI, V.; ALUJA, T. Handling missing values in kernel methods with application to microbiology data. **Neurocomputing**, v. 141, p. 110–116, 2014. ISSN 0925-2312.

COVO, S.; ELALOUF, A. A novel single-gamma approximation to the sum of independent gamma variables, and a generalization to infinitely divisible distributions. **Electronic Journal of Statistics**, Institute of Mathematical Statistics and Bernoulli Society, v. 8, n. 1, p. 894 – 926, 2014.

CUI, C.; HASHEMI, S. Deep neural network aided monte carlo simulation in solder joint failure probability analysis. **Materials Letters**, v. 347, p. 134663, 2023. ISSN 0167-577X. Disponível em: https://www.sciencedirect.com/science/article/pii/S0167577X23008480.

De Souto, M. C.; JASKOWIAK, P. A.; COSTA, I. G. Impact of missing data imputation methods on gene expression clustering and classification. **BMC bioinformatics**, BioMed Central, v. 16, n. 1, p. 1–9, 2015.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the em algorithm. **Journal of the Royal Statistical Society. Series B (Methodological)**, [Royal Statistical Society, Wiley], v. 39, n. 1, p. 1–38, 1977. ISSN 00359246.

DEVROYE, L. **Non-Uniform Random Variate Generation**. New York, NY, USA: Springer-Verlag, 1986.

DU, J.; CHEN, H.; ZHANG, W. A deep learning method for data recovery in sensor networks using effective spatio-temporal correlation data. **Sensor Review**, Emerald Publishing Limited, v. 39, n. 2, p. 208–217, 2019.

EBEIGBE, D.; BERRY, T.; NORTON, M. M.; WHALEN, A. J.; SIMON, D.; SAUER, T.; SCHIFF, S. J. **A Generalized Unscented Transformation for Probability Distributions**. 2021.

EIROLA, E.; DOQUIRE, G.; VERLEYSEN, M.; LENDASSE, A. Distance estimation in numerical data sets with missing values. **Information Sciences**, v. 240, p. 115–128, 2013. ISSN 0020-0255.

EIROLA, E.; LENDASSE, A.; VANDEWALLE, V.; BIERNACKI, C. Mixture of gaussians for distance estimation with missing data. **Neurocomputing**, v. 131, p. 32–42, 2014. ISSN 0925-2312.

FARHANGFAR, A.; KURGAN, L. A.; PEDRYCZ, W. A novel framework for imputation of missing values in databases. **IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans**, v. 37, n. 5, p. 692–709, 2007.

FERNSTAD, S. J.; GLEN, R. C. Visual analysis of missing data — to see what isn't there. In: **2014 IEEE Conference on Visual Analytics Science and Technology (VAST)**, 2014. p. 249–250.

GARCÍA-LAENCINA, P. J.; SANCHO-GÓMEZ, J.-L.; FIGUEIRAS-VIDAL, A. R. Pattern classification with missing data: a review. **Neural Computing and Applications**, v. 19, n. 2, p. 263–282, 2010. ISSN 1433-3058.

GULER, S.; EREM, B.; COHEN, A. L.; AFACAN, O.; WARFIELD, S. Dynamic missing-data completion reduces leakage of motion artifact caused by temporal filtering that remains after scrubbing. In: **2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)**, 2020. p. 1–4.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. Springer, 2009. (Springer series in statistics). ISBN 9780387848846. Disponível em: https://books.google.com.br/books?id=eBSgoAEACAAJ.

HUNT, L.; JORGENSEN, M. Mixture model clustering for mixed data with missing information. **Computational Statistics & Data Analysis**, v. 41, n. 3, p. 429–440, 2003. ISSN 0167-9473. Recent Developments in Mixture Model.

JAFRASTEH, B.; HERNÁNDEZ-LOBATO, D.; LUBIÁN-LÓPEZ, S. P.; BENAVENTE-FERNÁNDEZ, I. Gaussian processes for missing value imputation. **Knowledge-Based Systems**, v. 273, p. 110603, 2023. ISSN 0950-7051.

JULIER, S.; UHLMANN, J.; DURRANT-WHYTE, H. A new approach for filtering nonlinear systems. In: **Proceedings of 1995 American Control Conference - ACC'95**, 1995. v. 3, p. 1628–1632 vol.3.

JULIER, S. J.; UHLMANN, J. K. A general method for approximating nonlinear transformations of probability distributions. In: , 1996.

KALOS, M. H.; WHITLOCK, P. A. **Monte Carlo methods**. Weinheim: WILEY-VCH, 2008. XII, 203 S. p. ISBN 3-527-40760-X, 978-3-527-40760-6.

KANG, H. The prevention and handling of the missing data. **Korean J. Anesthesiol.**, The Korean Society of Anesthesiologists, v. 64, n. 5, p. 402–406, maio 2013.

KELLY, M.; LONGJOHN, R.; NOTTINGHAM, K. **The UCI Machine Learning Repository**. 2013. Disponível em: https://archive.ics.uci.edu.

KROESE, D.; TAIMRE, T.; BOTEV, Z. **Handbook of Monte Carlo Methods**. Wiley, 2011. (Wiley Series in Probability and Statistics). ISBN 9780470177938. Disponível em: https://books.google.com.br/books?id=-j3bmyGXKUIC.

LAI, Y.; SPANIER, J. Applications of monte carlo/quasi-monte carlo methods in finance: Option pricing. In: NIEDERREITER, H.; SPANIER, J. (Ed.). **Monte-Carlo and Quasi-Monte Carlo Methods 1998**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000. p. 284–295. ISBN 978-3-642-59657-5.

LITTLE, R. J. A.; RUBIN, D. B. **Statistical Analysis with Missing Data**. Second. New Jersey: Wiley, 2002. 397 p.

LITTLE, R. J. A.; RUBIN, D. B. **Statistical Analysis with Missing Data**. Third. New Jersey: Wiley, 2019. 462 p.

LIU, C.; MA, Y.; ZHAO, J.; NUSSINOV, R.; ZHANG, Y.-C.; CHENG, F.; ZHANG, Z.-K. Computational network biology: Data, models, and applications. **Physics Reports**, v. 846, p. 1–66, 2020. ISSN 0370-1573. Computational network biology: Data, models, and applications.

LUENGO, D.; MARTINO, L.; BUGALLO, M.; ELVIRA, V.; SÄRKKÄ, S. A survey of monte carlo methods for parameter estimation. **EURASIP Journal on Advances in Signal Processing**, v. 2020, n. 1, p. 25, 2020. ISSN 1687-6180.

LUO, H.; LI, M.; YANG, M.; WU, F.-X.; LI, Y.; WANG, J. Biomedical data and computational models for drug repositioning: a comprehensive review. **Briefings in Bioinformatics**, v. 22, n. 2, p. 1604–1619, 02 2020. ISSN 1477-4054.

M. Al Luhayb, A. S. The bootstrap method for monte carlo integration inference. **Journal of King Saud University - Science**, p. 102768, 2023. ISSN 1018-3647. Disponível em: https://www.sciencedirect.com/science/article/pii/S1018364723002306.

MADHU, G.; BHARADWAJ, B.; NAGACHANDRIKA, G.; VARDHAN, K. A novel algorithm for missing data imputation on machine learning. In: **2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)**, 2019. p. 173–177.

MCLEISH, D. L. **Monte Carlo simulation and finance**: John Wiley & Sons, 2011. v. 276.

MENG, X.-L.; RUBIN, D. B. Maximum likelihood estimation via the ecm algorithm: A general framework. **Biometrika**, [Oxford University Press, Biometrika Trust], v. 80, n. 2, p. 267–278, 1993. ISSN 00063444. Disponível em: http://www.jstor.org/stable/2337198.

MESQUITA, D. P.; GOMES, J. P.; CORONA, F.; SOUZA, A. H.; NOBRE, J. S. Gaussian kernels for incomplete data. **Applied Soft Computing**, v. 77, p. 356–365, 2019. ISSN 1568-4946.

MESQUITA, D. P.; GOMES, J. P.; Souza Junior, A. H.; NOBRE, J. S. Euclidean distance estimation in incomplete datasets. **Neurocomputing**, v. 248, p. 11–18, 2017. ISSN 0925-2312. Neural Networks : Learning Algorithms and Classification Systems.

MESQUITA, D. P. P. **MACHINE LEARNING FOR INCOMPLETE DATA**. 56 p. Dissertação (Mestrado), 2017.

MOLENBERGHS, G.; FITZMAURICE, G.; KENWARD, M.; TSIATIS, A.; VERBEKE, G. **Handbook of Missing Data Methodology**. Hoboken, NJ: Chapman and Hall/CRC, 2014. Disponível em: https://doi.org/10.1201/b17622.

MORALES-RODRIGUEZ, R.; MEYER, A. S.; GERNAEY, K. V.; SIN, G. A framework for model-based optimization of bioprocesses under uncertainty: Identifying critical parameters and operating variables. In: PISTIKOPOULOS, E.; GEORGIADIS, M.; KOKOSSIS, A. (Ed.). **21st European Symposium on Computer Aided Process Engineering**. Elsevier, 2011, (Computer Aided Chemical Engineering, v. 29). p. 1455–1459. Disponível em: https://www.sciencedirect.com/science/article/pii/B9780444542984500702.

MURPHY, K. P. **Machine Learning: A Probabilistic Perspective**: The MIT Press, 2012. ISBN 0262018020.

NAKAGAMI, M. The m-distribution—a general formula of intensity distribution of rapid fading. In: HOFFMAN, W. (Ed.). **Statistical Methods in Radio Wave Propagation**: Pergamon, 1960. p. 3–36. ISBN 978-0-08-009306-2.

ROBERT, C. P.; CASELLA, G. **Monte Carlo Statistical Methods (Springer Texts in Statistics)**. Berlin, Heidelberg: Springer-Verlag, 2005. ISBN 0387212396.

ROBERTS, C.; GEISSER, S. A necessary and sufficient condition for the square of a random variable to be gamma. **Biometrika**, [Oxford University Press, Biometrika Trust], v. 53, n. 1/2, p. 275–278, 1966. ISSN 00063444.

SCHÖLKOPF, B.; SMOLA, A. J.; BACH, F. *et al.* **Learning with kernels: support vector machines, regularization, optimization, and beyond**: MIT press, 2002.

SHAWE-TAYLOR, J.; CRISTIANINI, N. **Kernel Methods for Pattern Analysis**: Cambridge University Press, 2004.

SILVA-RAMÍREZ, E.-L.; PINO-MEJÍAS, R.; LÓPEZ-COELLO, M. Single imputation with multilayer perceptron and multiple imputation combining multilayer perceptron and k-nearest neighbours for monotone patterns. **Applied Soft Computing**, Elsevier, v. 29, p. 65–74, 2015.

SUYKENS, J.; GESTEL, T.; BRABANTER, J.; VANDEWALLE, J. **Least Squares Support Vector Machines**. Singapore: World Scientific, 2002.

SUYKENS, J.; VANDEWALLE, J. Least squares support vector machine classifiers. **Neural Processing Letters**, v. 9, n. 3, p. 293–300, 1999. ISSN 1573-773X.

SUYKENS, J.; VANDEWALLE, J.; De Moor, B. Optimal control by least squares support vector machines. **Neural Networks**, v. 14, n. 1, p. 23–35, 2001. ISSN 0893-6080. Disponível em: https://www.sciencedirect.com/science/article/pii/S0893608000000770.

TANG, X.; YAO, H.; SUN, Y.; AGGARWAL, C.; MITRA, P.; WANG, S. Joint modeling of local and global temporal dynamics for multivariate time series forecasting with missing values. In: **Proceedings of the AAAI Conference on Artificial Intelligence**, 2020. v. 34, n. 04, p. 5956–5963.

VAPNIK, V. N. **The nature of statistical learning theory**: Springer-Verlag New York, Inc., 1995. ISBN 0-387-94559-8.

WAN, E. A.; MERWE, R. van der. The unscented kalman filter. In: _____. **Kalman Filtering and Neural Networks**. John Wiley & Sons, Ltd, 2001. cap. 7, p. 221–280. ISBN 9780471221548. Disponível em: https://onlinelibrary.wiley.com/doi/abs/10.1002/0471221546.ch7.

YU, Q.; MICHE, Y.; EIROLA, E.; van Heeswijk, M.; SÉVERIN, E.; LENDASSE, A. Regularized extreme learning machine for regression with missing data. **Neurocomputing**, v. 102, p. 45–51, 2013. ISSN 0925-2312. Advances in Extreme Learning Machines (ELM 2011).

ZHANG, S.; JIN, Z.; ZHU, X.; ZHANG, J. Missing data analysis: A kernel-based multi-imputation approach. In: _____. **Transactions on Computational Science III**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. p. 122–142. ISBN 978-3-642-00212-0.

ZHANG, S.; QIN, Y.; ZHU, X.; ZHANG, J.; ZHANG, C. Kernel-based multi-imputation for missing data. In: **Proceedings of the 2006 Conference on Advances in Intelligent IT: Active Media Technology 2006**. NLD: IOS Press, 2006. p. 106–111. ISBN 1586036157.

ZHANG, Y.-F.; THORBURN, P. J.; XIANG, W.; FITCH, P. Ssim—a deep learning approach for recovering missing time series sensor data. **IEEE Internet of Things Journal**, v. 6, n. 4, p. 6618–6628, 2019.

ZHOU, L.; LAI, K. K. Adaboost models for corporate bankruptcy prediction with missing data. **Computational Economics**, v. 50, n. 1, p. 69–94, 2017. ISSN 1572-9974.

APPENDIX A – #SMS FOR EACH $\mathcal{G}$ GROUP (UCI DATASET)

Overall Analysis ($\mathbb{E}[\mathcal{C}_\nu(\cdot)]$)

The following figure shows the number of times each method was better than the others. It considers the estimation of the 3 types of kernels $\left(\mathcal{C}_\nu(\cdot)\ \forall\ \nu \in \{\frac{1}{2}; \frac{3}{2}; \frac{5}{2}\}\right)$ evaluated on 22 datasets from UCI. Thus, 528 scenarios were evaluated in $\mathcal{G}_1$. Considering groups $\mathcal{G}_2$ and $\mathcal{G}_3$, 264 scenarios were evaluated in each group.

Scenarios evaluated in $\mathcal{G}_1$:

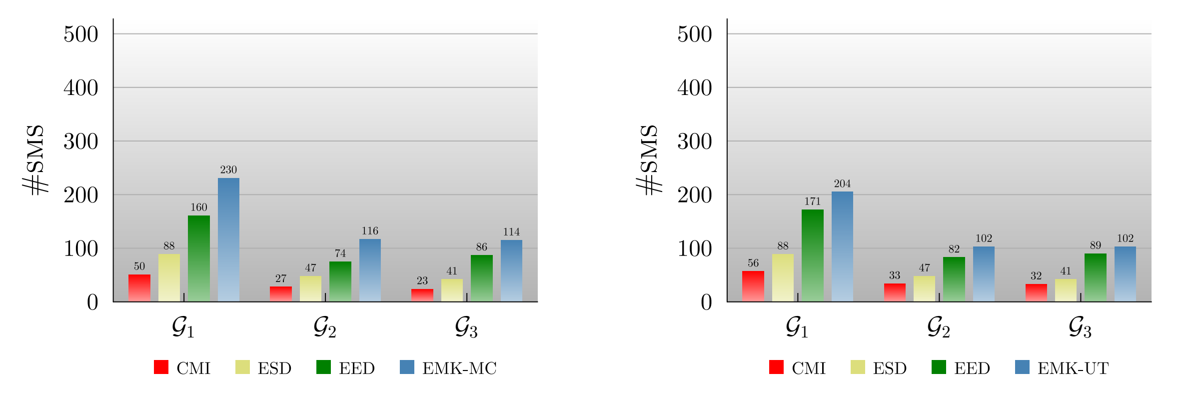
$$(3\ types\ of\ kernel) \times (rMiss \in \{0.1\ \cdots\ 0.8\}) \times (22\ datasets) \tag{A.1}$$

Scenarios evaluated in $\mathcal{G}_2$

$$(3\ types\ of\ kernel) \times (rMiss \in \{0.1\ \cdots\ 0.4\}) \times (22\ datasets) \tag{A.2}$$

Scenarios evaluated in $\mathcal{G}_3$

$$(3\ types\ of\ kernel) \times (rMiss \in \{0.5\ \cdots\ 0.8\}) \times (22\ datasets) \tag{A.3}$$



(a) EMK-MC x other methods (b) EMK-UT x other methods

Figure 9 – UCI dataset - #SMS for $\mathbb{E}[\mathcal{C}_\nu(\cdot)]$.

Source: elaborated by the author.

Specific analysis for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$

The following figure shows the number of times each method was better than the others. It considers only the estimation of the Matern kernel $\mathcal{C}_{1/2}(\cdot)$ evaluated on 22 datasets from UCI. Thus, 176 scenarios were evaluated in $\mathcal{G}_1$. Considering groups $\mathcal{G}_2$ and $\mathcal{G}_3$, 88 scenarios were evaluated in each group.

Scenarios evaluated in $\mathcal{G}_1$:

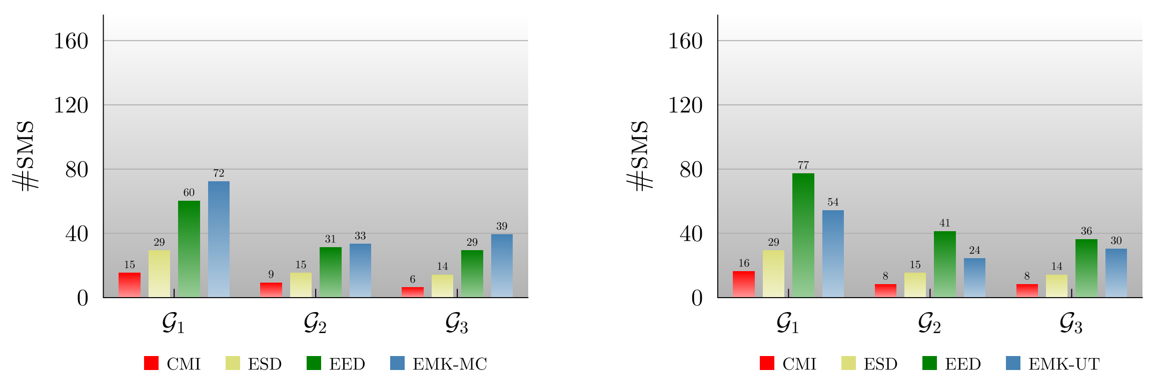
$$(1\ type\ of\ kernel) \times (rMiss \in \{0.1\ \cdots\ 0.8\}) \times (22\ datasets) \tag{A.4}$$

Scenarios evaluated in $\mathcal{G}_2$

$$(1\ type\ of\ kernel) \times (rMiss \in \{0.1\ \cdots\ 0.4\}) \times (22\ datasets) \tag{A.5}$$

Scenarios evaluated in $\mathcal{G}_3$

$$(1\ type\ of\ kernel) \times (rMiss \in \{0.5\ \cdots\ 0.8\}) \times (22\ datasets) \tag{A.6}$$



(a) EMK-MC x other methods (b) EMK-UT x other methods

Figure 10 – UCI dataset - #SMS for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$.

Source: elaborated by the author.

Specific analysis for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$

The following figure shows the number of times each method was better than the others. It considers only the estimation of the Matern kernel $\mathcal{C}_{3/2}(\cdot)$ evaluated on 22 datasets from UCI. Thus, 176 scenarios were evaluated in $\mathcal{G}_1$. Considering groups $\mathcal{G}_2$ and $\mathcal{G}_3$, 88 scenarios were evaluated in each group.

Scenarios evaluated in $\mathcal{G}_1$:

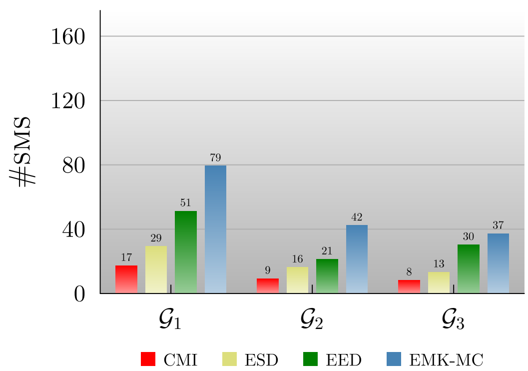
$$(1\ type\ of\ kernel) \times (rMiss \in \{0.1\ \cdots\ 0.8\}) \times (22\ datasets) \tag{A.7}$$

Scenarios evaluated in $\mathcal{G}_2$

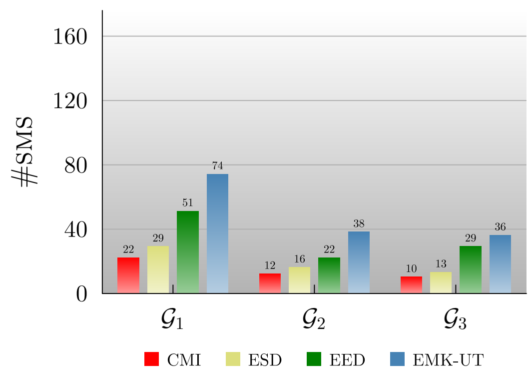
$$(1\ type\ of\ kernel) \times (rMiss \in \{0.1\ \cdots\ 0.4\}) \times (22\ datasets) \tag{A.8}$$

Scenarios evaluated in $\mathcal{G}_3$

$$(1\ type\ of\ kernel) \times (rMiss \in \{0.5\ \cdots\ 0.8\}) \times (22\ datasets) \tag{A.9}$$



(a) EMK-MC x other methods



(b) EMK-UT x other methods

Figure 11 – UCI dataset - #SMS for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$.

Source: elaborated by the author.

Specific analysis for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$

The following figure shows the number of times each method was better than the others. It considers only the estimation of the Matern kernel $\mathcal{C}_{5/2}(\cdot)$ evaluated on 22 datasets from UCI. Thus, 176 scenarios were evaluated in $\mathcal{G}_1$. Considering groups $\mathcal{G}_2$ and $\mathcal{G}_3$, 88 scenarios were evaluated in each group.

Scenarios evaluated in $\mathcal{G}_1$:

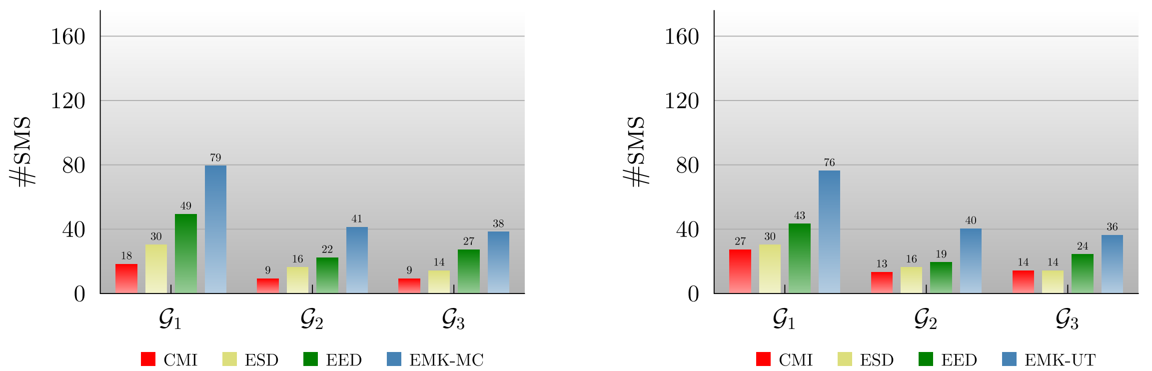
$$(1\ type\ of\ kernel) \times (rMiss \in \{0.1\ \cdots\ 0.8\}) \times (22\ datasets) \tag{A.10}$$

Scenarios evaluated in $\mathcal{G}_2$

$$(1\ type\ of\ kernel) \times (rMiss \in \{0.1\ \cdots\ 0.4\}) \times (22\ datasets) \tag{A.11}$$

Scenarios evaluated in $\mathcal{G}_3$

$$(1\ type\ of\ kernel) \times (rMiss \in \{0.5\ \cdots\ 0.8\}) \times (22\ datasets) \tag{A.12}$$



(a) EMK-MC x other methods (b) EMK-UT x other methods

Figure 12 – UCI dataset - #SMS for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$.
Source: elaborated by the author.

APPENDIX B – ARMSE FOR UCI DATASETS

Breast tissue

Table 12 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$ (Breast tissue).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	**0.1399 ±0.03**	0.2796 ±0.05	0.2341 ±0.04	0.1672 ±0.03	0.1597 ±0.03
0.20	**0.1430 ±0.02**	0.2737 ±0.04	0.2299 ±0.04	0.1649 ±0.03	0.1577 ±0.02
0.30	**0.1389 ±0.02**	0.2944 ±0.04	0.2497 ±0.04	0.1751 ±0.03	0.1617 ±0.02
0.40	**0.1504 ±0.03**	0.2975 ±0.04	0.2548 ±0.05	0.1786 ±0.03	0.1680 ±0.03
0.50	**0.1598 ±0.04**	0.3054 ±0.05	0.2627 ±0.05	0.1879 ±0.04	0.1702 ±0.02
0.60	**0.1508 ±0.03**	0.2784 ±0.05	0.2347 ±0.05	0.1689 ±0.04	0.1579 ±0.02
0.70	**0.1469 ±0.02**	0.2751 ±0.04	0.2319 ±0.04	0.1674 ±0.03	0.1566 ±0.02
0.80	**0.1615 ±0.03**	0.2743 ±0.05	0.2319 ±0.05	0.1741 ±0.03	0.1635 ±0.03

Source: elaborated by the author.

Table 13 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$ (Breast tissue).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	**0.1525 ±0.04**	0.3572 ±0.06	0.2864 ±0.06	0.2076 ±0.04	0.2273 ±0.04
0.20	**0.1857 ±0.05**	0.3663 ±0.05	0.3033 ±0.05	0.2226 ±0.04	0.2441 ±0.06
0.30	**0.1760 ±0.03**	0.3852 ±0.05	0.3206 ±0.05	0.2292 ±0.03	0.2430 ±0.03
0.40	**0.1855 ±0.04**	0.3938 ±0.07	0.3324 ±0.07	0.2384 ±0.05	0.2509 ±0.05
0.50	**0.1986 ±0.03**	0.4061 ±0.07	0.3480 ±0.07	0.2494 ±0.05	0.2592 ±0.04
0.60	**0.1917 ±0.03**	0.3712 ±0.06	0.3112 ±0.07	0.2296 ±0.05	0.2394 ±0.04
0.70	**0.1901 ±0.04**	0.3737 ±0.07	0.3114 ±0.07	0.2252 ±0.05	0.2360 ±0.04
0.80	**0.1916 ±0.04**	0.3504 ±0.07	0.2904 ±0.07	0.2137 ±0.04	0.2227 ±0.04

Source: elaborated by the author.

Table 14 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$ (Breast tissue).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	**0.1709 ±0.04**	0.3695 ±0.05	0.2990 ±0.05	0.2128 ±0.03	0.2291 ±0.04
0.20	**0.1890 ±0.04**	0.3948 ±0.05	0.3260 ±0.05	0.2344 ±0.04	0.2521 ±0.04
0.30	**0.1948 ±0.04**	0.4096 ±0.06	0.3425 ±0.06	0.2422 ±0.04	0.2588 ±0.04
0.40	**0.1905 ±0.05**	0.4139 ±0.08	0.3438 ±0.09	0.2532 ±0.06	0.2683 ±0.05
0.50	**0.2057 ±0.04**	0.4109 ±0.07	0.3455 ±0.07	0.2492 ±0.05	0.2623 ±0.05
0.60	**0.1925 ±0.03**	0.3868 ±0.06	0.3187 ±0.06	0.2306 ±0.04	0.2458 ±0.04
0.70	**0.2092 ±0.04**	0.3652 ±0.07	0.3033 ±0.07	0.2265 ±0.05	0.2377 ±0.04
0.80	**0.2110 ±0.05**	0.3743 ±0.07	0.3122 ±0.07	0.2294 ±0.05	0.2407 ±0.05

Source: elaborated by the author.

Liver Disorders BUPA Medical Research

Table 15 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$ (Liver Disorders BUPA Medical Research).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2214 ±0.03	0.2229 ±0.09	0.1962 ±0.08	0.1718 ±0.04	**0.1582 ±0.02**
0.20	0.2281 ±0.03	0.1868 ±0.06	0.1633 ±0.06	0.1633 ±0.03	**0.1603 ±0.02**
0.30	0.2233 ±0.03	0.1630 ±0.05	**0.1390 ±0.04**	0.1497 ±0.02	0.1537 ±0.02
0.40	0.2359 ±0.03	0.1505 ±0.02	**0.1327 ±0.01**	0.1536 ±0.02	0.1610 ±0.02
0.50	0.2271 ±0.03	0.1561 ±0.02	**0.1329 ±0.02**	0.1493 ±0.02	0.1566 ±0.02
0.60	0.2224 ±0.03	0.1554 ±0.02	**0.1298 ±0.02**	0.1444 ±0.02	0.1512 ±0.02
0.70	0.2263 ±0.02	0.1833 ±0.05	**0.1559 ±0.05**	0.1581 ±0.02	0.1599 ±0.02
0.80	0.2204 ±0.03	0.1838 ±0.04	0.1530 ±0.03	**0.1529 ±0.01**	0.1574 ±0.01

Source: elaborated by the author.

Table 16 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$ (Liver Disorders BUPA Medical Research).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2588 ±0.05	0.3605 ±0.12	0.3251 ±0.12	0.2560 ±0.06	**0.2525 ±0.06**
0.20	0.2719 ±0.04	0.2708 ±0.11	0.2403 ±0.10	0.2138 ±0.05	**0.2109 ±0.05**
0.30	0.2778 ±0.03	0.2136 ±0.08	**0.1887 ±0.07**	0.1937 ±0.04	0.1897 ±0.04
0.40	0.2711 ±0.03	0.2035 ±0.03	**0.1765 ±0.03**	0.1857 ±0.02	0.1831 ±0.02
0.50	0.2658 ±0.03	0.2094 ±0.05	**0.1777 ±0.05**	0.1843 ±0.03	0.1823 ±0.03
0.60	0.2641 ±0.03	0.2101 ±0.04	**0.1766 ±0.03**	0.1834 ±0.02	0.1827 ±0.02
0.70	0.2777 ±0.03	0.2128 ±0.06	**0.1865 ±0.05**	0.1944 ±0.03	0.1915 ±0.03
0.80	0.2742 ±0.03	0.2468 ±0.05	0.2050 ±0.04	**0.1945 ±0.02**	0.1947 ±0.01

Source: elaborated by the author.

Table 17 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$ (Liver Disorders BUPA Medical Research).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2836 ±0.05	0.3202 ±0.13	0.2918 ±0.12	0.2534 ±0.06	**0.2501 ±0.06**
0.20	0.2673 ±0.04	0.2739 ±0.12	0.2447 ±0.10	0.2185 ±0.05	**0.2155 ±0.05**
0.30	0.2585 ±0.02	0.2517 ±0.09	0.2108 ±0.08	**0.2003 ±0.04**	0.2014 ±0.04
0.40	0.2772 ±0.03	0.2405 ±0.09	0.2153 ±0.08	0.2119 ±0.05	**0.2088 ±0.04**
0.50	0.2732 ±0.03	0.2147 ±0.04	**0.1880 ±0.04**	0.1948 ±0.02	0.1926 ±0.02
0.60	0.2785 ±0.04	0.2066 ±0.02	**0.1837 ±0.02**	0.1941 ±0.03	0.1913 ±0.02
0.70	0.2716 ±0.03	0.2369 ±0.06	0.1993 ±0.05	**0.1969 ±0.03**	0.1979 ±0.03
0.80	0.2702 ±0.03	0.2496 ±0.07	0.2088 ±0.07	**0.1991 ±0.04**	0.1997 ±0.03

Source: elaborated by the author.

Heart Disease Cleveland

Table 18 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$ (Heart Disease Cleveland).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2125 ±0.03	**0.0748 ±0.01**	0.0811 ±0.01	0.1656 ±0.01	0.1955 ±0.01
0.20	0.2071 ±0.02	**0.0798 ±0.01**	0.0809 ±0.01	0.1614 ±0.01	0.1909 ±0.01
0.30	0.2105 ±0.02	**0.0768 ±0.01**	0.0806 ±0.01	0.1633 ±0.01	0.1929 ±0.01
0.40	0.2005 ±0.02	0.0799 ±0.02	**0.0788 ±0.01**	0.1582 ±0.01	0.1882 ±0.01
0.50	0.2029 ±0.02	**0.0731 ±0.01**	0.0781 ±0.01	0.1623 ±0.01	0.1923 ±0.01
0.60	0.2065 ±0.02	0.0844 ±0.02	**0.0834 ±0.01**	0.1615 ±0.01	0.1912 ±0.01
0.70	0.2077 ±0.02	**0.0752 ±0.01**	0.0791 ±0.01	0.1615 ±0.01	0.1916 ±0.01
0.80	0.2014 ±0.01	0.0802 ±0.01	**0.0770 ±0.01**	0.1557 ±0.01	0.1863 ±0.01

Source: elaborated by the author.

Table 19 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$ (Heart Disease Cleveland).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.3046 ±0.04	**0.1095 ±0.03**	0.1219 ±0.02	0.2174 ±0.02	0.2017 ±0.02
0.20	0.3022 ±0.03	**0.1123 ±0.02**	0.1181 ±0.01	0.2109 ±0.01	0.1953 ±0.01
0.30	0.3028 ±0.02	**0.1136 ±0.02**	0.1188 ±0.01	0.2115 ±0.01	0.1957 ±0.01
0.40	0.3017 ±0.02	**0.1129 ±0.02**	0.1207 ±0.01	0.2138 ±0.01	0.1987 ±0.01
0.50	0.2908 ±0.02	0.1144 ±0.02	**0.1144 ±0.01**	0.2048 ±0.01	0.1901 ±0.01
0.60	0.2909 ±0.02	**0.1071 ±0.02**	0.1114 ±0.01	0.2062 ±0.01	0.1917 ±0.01
0.70	0.2840 ±0.02	0.1182 ±0.02	**0.1159 ±0.01**	0.2053 ±0.01	0.1919 ±0.01
0.80	0.2801 ±0.02	0.1205 ±0.02	**0.1135 ±0.02**	0.2019 ±0.01	0.1885 ±0.01

Source: elaborated by the author.

Table 20 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$ (Heart Disease Cleveland).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.3195 ±0.03	**0.1258 ±0.02**	0.1272 ±0.02	0.2180 ±0.02	0.1947 ±0.02
0.20	0.3221 ±0.03	**0.1284 ±0.02**	0.1303 ±0.01	0.2192 ±0.01	0.1961 ±0.01
0.30	0.3245 ±0.03	**0.1225 ±0.02**	0.1298 ±0.01	0.2227 ±0.02	0.1989 ±0.02
0.40	0.3220 ±0.03	**0.1291 ±0.02**	0.1324 ±0.01	0.2224 ±0.01	0.1996 ±0.01
0.50	0.3236 ±0.02	**0.1286 ±0.02**	0.1315 ±0.02	0.2208 ±0.02	0.1985 ±0.02
0.60	0.3181 ±0.02	0.1320 ±0.02	**0.1319 ±0.02**	0.2188 ±0.01	0.1966 ±0.01
0.70	0.3258 ±0.03	**0.1251 ±0.02**	0.1329 ±0.02	0.2238 ±0.02	0.2003 ±0.02
0.80	0.3082 ±0.02	0.1333 ±0.01	**0.1263 ±0.01**	0.2116 ±0.01	0.1901 ±0.01

Source: elaborated by the author.

Coluna

Table 21 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$ (Coluna).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.1632 ±0.03	0.1625 ±0.02	**0.1280 ±0.02**	0.1292 ±0.02	0.1341 ±0.02
0.20	0.1746 ±0.03	0.1580 ±0.02	**0.1287 ±0.02**	0.1342 ±0.02	0.1389 ±0.02
0.30	0.1659 ±0.03	0.1800 ±0.05	0.1466 ±0.05	0.1398 ±0.03	**0.1374 ±0.02**
0.40	0.1708 ±0.02	0.1530 ±0.02	**0.1223 ±0.02**	0.1293 ±0.01	0.1352 ±0.01
0.50	0.1751 ±0.02	0.1706 ±0.03	**0.1377 ±0.02**	0.1380 ±0.02	0.1429 ±0.01
0.60	0.1727 ±0.02	0.1865 ±0.04	0.1529 ±0.03	**0.1436 ±0.02**	0.1468 ±0.02
0.70	0.1706 ±0.02	0.1839 ±0.05	0.1530 ±0.05	**0.1462 ±0.03**	0.1593 ±0.08
0.80	0.1658 ±0.02	0.1700 ±0.04	0.1381 ±0.04	**0.1339 ±0.02**	0.1364 ±0.02

Source: elaborated by the author.

Table 22 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$ (Coluna).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.1980 ±0.04	0.2195 ±0.05	0.1719 ±0.05	**0.1691 ±0.03**	0.1741 ±0.04
0.20	0.2088 ±0.03	0.2108 ±0.04	0.1665 ±0.03	**0.1644 ±0.02**	0.1669 ±0.02
0.30	0.2042 ±0.02	0.2257 ±0.06	0.1834 ±0.05	**0.1768 ±0.03**	0.1800 ±0.03
0.40	0.2014 ±0.03	0.2329 ±0.06	0.1850 ±0.05	**0.1720 ±0.03**	0.1756 ±0.03
0.50	0.2148 ±0.03	0.2155 ±0.05	0.1771 ±0.04	**0.1768 ±0.03**	0.1793 ±0.03
0.60	0.2154 ±0.03	0.2276 ±0.05	0.1834 ±0.03	**0.1807 ±0.02**	0.1829 ±0.02
0.70	0.1984 ±0.02	0.2477 ±0.06	0.1983 ±0.06	**0.1768 ±0.03**	0.1811 ±0.03
0.80	0.2145 ±0.03	0.2187 ±0.04	0.1766 ±0.03	**0.1748 ±0.02**	0.1764 ±0.02

Source: elaborated by the author.

Table 23 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$ (Coluna).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2130 ±0.03	0.2224 ±0.05	**0.1796 ±0.05**	0.1798 ±0.03	0.1809 ±0.03
0.20	0.2015 ±0.03	0.2330 ±0.05	0.1806 ±0.05	**0.1750 ±0.03**	0.1770 ±0.02
0.30	0.2120 ±0.03	0.2631 ±0.06	0.2063 ±0.06	**0.1899 ±0.04**	0.1942 ±0.03
0.40	0.2124 ±0.04	0.2274 ±0.05	0.1810 ±0.04	**0.1767 ±0.02**	0.1802 ±0.02
0.50	0.2191 ±0.02	0.2315 ±0.04	0.1832 ±0.03	**0.1806 ±0.02**	0.1837 ±0.02
0.60	0.2145 ±0.03	0.2792 ±0.07	0.2257 ±0.08	**0.1998 ±0.05**	0.2014 ±0.04
0.70	0.2159 ±0.03	0.2448 ±0.05	0.1904 ±0.04	**0.1791 ±0.02**	0.1833 ±0.02
0.80	0.2152 ±0.03	0.2650 ±0.09	0.2181 ±0.08	**0.1966 ±0.04**	0.1980 ±0.04

Source: elaborated by the author.

Computer Hardware

Table 24 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$ (Computer Hardware).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2053 ±0.05	0.3034 ±0.05	0.2553 ±0.04	0.2012 ±0.03	**0.1918 ±0.03**
0.20	0.1986 ±0.03	0.3151 ±0.05	0.2646 ±0.04	0.2042 ±0.03	**0.1927 ±0.03**
0.30	0.2085 ±0.03	0.2870 ±0.05	0.2403 ±0.04	0.1962 ±0.03	**0.1903 ±0.02**
0.40	0.2111 ±0.03	0.2777 ±0.05	0.2314 ±0.04	0.1906 ±0.03	**0.1857 ±0.02**
0.50	0.2069 ±0.03	0.2793 ±0.03	0.2342 ±0.03	0.1911 ±0.02	**0.1867 ±0.02**
0.60	0.2039 ±0.03	0.2630 ±0.03	0.2162 ±0.03	0.1800 ±0.02	**0.1778 ±0.02**
0.70	0.2079 ±0.03	0.2484 ±0.03	0.2061 ±0.02	0.1778 ±0.02	**0.1778 ±0.01**
0.80	0.2110 ±0.03	0.2477 ±0.04	0.2073 ±0.04	0.1793 ±0.02	**0.1771 ±0.02**

Source: elaborated by the author.

Table 25 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$ (Computer Hardware).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2535 ±0.07	0.3801 ±0.07	0.3127 ±0.06	**0.2491 ±0.04**	0.2604 ±0.04
0.20	**0.2298 ±0.04**	0.3795 ±0.05	0.3072 ±0.04	0.2439 ±0.03	0.2597 ±0.03
0.30	**0.2413 ±0.04**	0.3584 ±0.07	0.2939 ±0.06	0.2417 ±0.04	0.2557 ±0.04
0.40	**0.2428 ±0.04**	0.3668 ±0.06	0.3011 ±0.05	0.2443 ±0.03	0.2568 ±0.03
0.50	0.2385 ±0.04	0.3242 ±0.06	0.2652 ±0.05	**0.2216 ±0.03**	0.2327 ±0.03
0.60	0.2377 ±0.03	0.3270 ±0.04	0.2654 ±0.03	**0.2247 ±0.02**	0.2362 ±0.02
0.70	0.2469 ±0.03	0.3204 ±0.06	0.2646 ±0.05	**0.2297 ±0.03**	0.2397 ±0.03
0.80	0.2363 ±0.04	0.2962 ±0.03	0.2403 ±0.03	**0.2149 ±0.02**	0.2246 ±0.02

Source: elaborated by the author.

Table 26 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$ (Computer Hardware).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	**0.2396 ±0.08**	0.3947 ±0.07	0.3176 ±0.06	0.2584 ±0.05	0.2776 ±0.05
0.20	**0.2386 ±0.05**	0.3825 ±0.05	0.3079 ±0.04	0.2503 ±0.03	0.2677 ±0.03
0.30	**0.2503 ±0.04**	0.3916 ±0.07	0.3198 ±0.06	0.2571 ±0.03	0.2737 ±0.04
0.40	**0.2413 ±0.05**	0.3787 ±0.07	0.3062 ±0.06	0.2489 ±0.04	0.2658 ±0.04
0.50	0.2444 ±0.05	0.3414 ±0.06	0.2796 ±0.05	**0.2393 ±0.03**	0.2528 ±0.03
0.60	0.2562 ±0.04	0.3384 ±0.07	0.2816 ±0.06	**0.2443 ±0.03**	0.2563 ±0.04
0.70	0.2460 ±0.04	0.3098 ±0.06	0.2537 ±0.05	**0.2274 ±0.03**	0.2379 ±0.03
0.80	0.2300 ±0.04	0.3066 ±0.04	0.2437 ±0.04	**0.2183 ±0.03**	0.2313 ±0.02

Source: elaborated by the author.

Concrete Compression

Table 27 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$ (Concrete Compression).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2418 ±0.02	0.1833 ±0.03	0.1488 ±0.03	**0.1436 ±0.02**	0.1571 ±0.01
0.20	0.2397 ±0.02	0.1818 ±0.03	0.1482 ±0.03	**0.1430 ±0.02**	0.1583 ±0.01
0.30	0.2458 ±0.02	0.1841 ±0.03	0.1523 ±0.03	**0.1465 ±0.01**	0.1612 ±0.01
0.40	0.2394 ±0.02	0.1832 ±0.03	0.1530 ±0.03	**0.1475 ±0.01**	0.1609 ±0.01
0.50	0.2385 ±0.02	0.1784 ±0.03	0.1452 ±0.03	**0.1411 ±0.02**	0.1563 ±0.01
0.60	0.2401 ±0.02	0.1662 ±0.03	**0.1371 ±0.02**	0.1456 ±0.01	0.1637 ±0.01
0.70	0.2427 ±0.02	0.1775 ±0.03	**0.1484 ±0.03**	0.1510 ±0.02	0.1658 ±0.02
0.80	0.2480 ±0.02	0.1581 ±0.03	**0.1304 ±0.03**	0.1488 ±0.02	0.1661 ±0.02

Source: elaborated by the author.

Table 28 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$ (Concrete Compression).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.3166 ±0.02	0.2565 ±0.03	0.2087 ±0.03	0.1889 ±0.02	**0.1824 ±0.02**
0.20	0.3164 ±0.02	0.2659 ±0.04	0.2200 ±0.04	0.1906 ±0.02	**0.1835 ±0.02**
0.30	0.3194 ±0.02	0.2521 ±0.04	0.2076 ±0.04	0.1876 ±0.02	**0.1795 ±0.02**
0.40	0.3094 ±0.03	0.2557 ±0.04	0.2055 ±0.03	0.1785 ±0.02	**0.1753 ±0.02**
0.50	0.3104 ±0.02	0.2557 ±0.03	0.2078 ±0.03	0.1873 ±0.02	**0.1809 ±0.02**
0.60	0.3134 ±0.03	0.2614 ±0.04	0.2124 ±0.03	0.1872 ±0.01	**0.1813 ±0.02**
0.70	0.3159 ±0.02	0.2317 ±0.03	0.1905 ±0.03	0.1877 ±0.01	**0.1787 ±0.01**
0.80	0.3204 ±0.03	0.2285 ±0.04	0.1891 ±0.04	0.1908 ±0.02	**0.1818 ±0.02**

Source: elaborated by the author.

Table 29 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$ (Concrete Compression).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.3273 ±0.04	0.2721 ±0.05	0.2208 ±0.05	0.1948 ±0.03	**0.1865 ±0.03**
0.20	0.3295 ±0.02	0.2680 ±0.03	0.2156 ±0.03	0.1926 ±0.02	**0.1861 ±0.02**
0.30	0.3325 ±0.03	0.2819 ±0.05	0.2299 ±0.05	0.2012 ±0.02	**0.1947 ±0.02**
0.40	0.3302 ±0.03	0.2813 ±0.03	0.2260 ±0.03	0.1976 ±0.02	**0.1924 ±0.02**
0.50	0.3421 ±0.03	0.2717 ±0.04	0.2263 ±0.04	0.2000 ±0.02	**0.1899 ±0.02**
0.60	0.3255 ±0.02	0.2733 ±0.03	0.2209 ±0.03	0.1972 ±0.02	**0.1906 ±0.02**
0.70	0.3342 ±0.03	0.2698 ±0.05	0.2217 ±0.04	0.1940 ±0.02	**0.1857 ±0.02**
0.80	0.3370 ±0.03	0.2456 ±0.04	0.2041 ±0.03	0.2035 ±0.02	**0.1940 ±0.02**

Source: elaborated by the author.

Pima Indians Diabetes

Table 30 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$ (Pima Indians Diabetes).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2708 ±0.03	0.1327 ±0.01	**0.1181 ±0.01**	0.1576 ±0.01	0.1733 ±0.01
0.20	0.2737 ±0.03	0.1253 ±0.02	**0.1160 ±0.02**	0.1613 ±0.01	0.1782 ±0.01
0.30	0.2687 ±0.02	0.1301 ±0.01	**0.1167 ±0.01**	0.1587 ±0.01	0.1754 ±0.01
0.40	0.2681 ±0.02	0.1322 ±0.02	**0.1178 ±0.01**	0.1575 ±0.02	0.1739 ±0.02
0.50	0.2663 ±0.03	0.1324 ±0.01	**0.1186 ±0.01**	0.1582 ±0.02	0.1747 ±0.02
0.60	0.2662 ±0.02	0.1339 ±0.02	**0.1173 ±0.01**	0.1548 ±0.01	0.1708 ±0.01
0.70	0.2723 ±0.02	0.1333 ±0.01	**0.1221 ±0.01**	0.1631 ±0.01	0.1794 ±0.01
0.80	0.2607 ±0.02	0.1380 ±0.01	**0.1210 ±0.01**	0.1551 ±0.01	0.1710 ±0.01

Source: elaborated by the author.

Table 31 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$ (Pima Indians Diabetes).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.3533 ±0.03	0.1855 ±0.03	**0.1705 ±0.03**	0.2033 ±0.02	0.1891 ±0.02
0.20	0.3496 ±0.03	0.1827 ±0.02	**0.1683 ±0.02**	0.2005 ±0.02	0.1880 ±0.02
0.30	0.3382 ±0.03	0.1865 ±0.03	**0.1684 ±0.02**	0.2005 ±0.02	0.1917 ±0.03
0.40	0.3448 ±0.02	0.1799 ±0.02	**0.1649 ±0.01**	0.2002 ±0.02	0.1866 ±0.01
0.50	0.3442 ±0.02	0.1791 ±0.02	**0.1620 ±0.01**	0.1960 ±0.02	0.1827 ±0.01
0.60	0.3394 ±0.02	0.1838 ±0.03	**0.1638 ±0.02**	0.1945 ±0.02	0.1817 ±0.02
0.70	0.3316 ±0.02	0.1812 ±0.02	**0.1605 ±0.02**	0.1934 ±0.02	0.1808 ±0.02
0.80	0.3440 ±0.03	0.1857 ±0.02	**0.1710 ±0.02**	0.2036 ±0.02	0.1902 ±0.02

Source: elaborated by the author.

Table 32 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$ (Pima Indians Diabetes).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.3620 ±0.03	0.1949 ±0.02	**0.1771 ±0.02**	0.2073 ±0.02	0.1923 ±0.02
0.20	0.3552 ±0.03	0.2006 ±0.03	**0.1818 ±0.03**	0.2074 ±0.02	0.1921 ±0.02
0.30	0.3696 ±0.03	0.1925 ±0.02	**0.1821 ±0.02**	0.2135 ±0.02	0.1970 ±0.02
0.40	0.3579 ±0.03	0.1980 ±0.03	**0.1777 ±0.02**	0.2032 ±0.02	0.1870 ±0.02
0.50	0.3603 ±0.03	0.1911 ±0.02	**0.1803 ±0.02**	0.2128 ±0.02	0.1968 ±0.02
0.60	0.3655 ±0.02	0.1951 ±0.02	**0.1797 ±0.02**	0.2090 ±0.02	0.1934 ±0.02
0.70	0.3492 ±0.03	0.1995 ±0.02	**0.1774 ±0.02**	0.2023 ±0.02	0.1888 ±0.02
0.80	0.3540 ±0.04	0.2007 ±0.03	**0.1831 ±0.02**	0.2049 ±0.02	0.1905 ±0.01

Source: elaborated by the author.

Ecoli

Table 33 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$ (Ecoli).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2484 ±0.04	0.1925 ±0.03	**0.1601 ±0.02**	0.1607 ±0.02	0.1675 ±0.02
0.20	0.2565 ±0.03	0.2138 ±0.04	0.1820 ±0.04	**0.1717 ±0.02**	0.1731 ±0.02
0.30	0.2611 ±0.03	0.2057 ±0.04	0.1729 ±0.03	**0.1634 ±0.02**	0.1663 ±0.02
0.40	0.2502 ±0.03	0.2120 ±0.04	0.1790 ±0.04	**0.1659 ±0.02**	0.1668 ±0.02
0.50	0.2452 ±0.03	0.2154 ±0.04	0.1817 ±0.04	0.1646 ±0.02	**0.1637 ±0.01**
0.60	0.2467 ±0.03	0.2243 ±0.05	0.1929 ±0.05	0.1731 ±0.02	**0.1700 ±0.02**
0.70	0.2410 ±0.03	0.2370 ±0.05	0.2024 ±0.05	0.1726 ±0.02	**0.1656 ±0.01**
0.80	0.2419 ±0.03	0.2293 ±0.05	0.1937 ±0.05	0.1693 ±0.02	**0.1669 ±0.02**

Source: elaborated by the author.

Table 34 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$ (Ecoli).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.3072 ±0.04	0.2736 ±0.05	0.2272 ±0.04	0.2093 ±0.03	**0.2092 ±0.03**
0.20	0.3084 ±0.05	0.2589 ±0.05	0.2152 ±0.04	0.2059 ±0.03	**0.2048 ±0.03**
0.30	0.2991 ±0.04	0.2759 ±0.05	0.2315 ±0.05	0.2106 ±0.02	**0.2095 ±0.02**
0.40	0.3023 ±0.03	0.2859 ±0.05	0.2358 ±0.05	**0.2106 ±0.03**	0.2114 ±0.03
0.50	0.2924 ±0.03	0.2928 ±0.06	0.2462 ±0.06	**0.2146 ±0.03**	0.2150 ±0.03
0.60	0.2985 ±0.03	0.2567 ±0.04	0.2145 ±0.04	0.2043 ±0.02	**0.2025 ±0.02**
0.70	0.2954 ±0.04	0.2998 ±0.07	0.2578 ±0.06	**0.2233 ±0.03**	0.2234 ±0.03
0.80	0.2922 ±0.03	0.3313 ±0.08	0.2832 ±0.08	**0.2331 ±0.05**	0.2342 ±0.05

Source: elaborated by the author.

Table 35 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$ (Ecoli).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.3056 ±0.05	0.2692 ±0.04	0.2171 ±0.03	**0.2084 ±0.03**	0.2092 ±0.03
0.20	0.3060 ±0.04	0.2706 ±0.04	0.2203 ±0.03	**0.2096 ±0.02**	0.2107 ±0.02
0.30	0.3029 ±0.04	0.3047 ±0.06	0.2463 ±0.05	**0.2201 ±0.03**	0.2244 ±0.03
0.40	0.3058 ±0.04	0.2935 ±0.05	0.2426 ±0.05	**0.2182 ±0.03**	0.2195 ±0.02
0.50	0.3093 ±0.03	0.3272 ±0.07	0.2781 ±0.07	**0.2324 ±0.03**	0.2324 ±0.03
0.60	0.2920 ±0.02	0.3147 ±0.08	0.2636 ±0.07	**0.2258 ±0.04**	0.2276 ±0.04
0.70	0.3019 ±0.04	0.3400 ±0.07	0.2865 ±0.06	**0.2348 ±0.03**	0.2376 ±0.03
0.80	0.3004 ±0.04	0.3281 ±0.08	0.2810 ±0.08	**0.2394 ±0.04**	0.2395 ±0.04

Source: elaborated by the author.

Energy

Table 36 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$ (Energy).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2075 ±0.02	0.1280 ±0.02	**0.1091 ±0.01**	0.1428 ±0.01	0.1569 ±0.01
0.20	0.1995 ±0.02	0.1312 ±0.02	**0.1115 ±0.02**	0.1431 ±0.01	0.1568 ±0.01
0.30	0.2037 ±0.01	0.1320 ±0.02	**0.1121 ±0.01**	0.1439 ±0.01	0.1576 ±0.01
0.40	0.2057 ±0.02	0.1305 ±0.02	**0.1116 ±0.02**	0.1443 ±0.01	0.1581 ±0.01
0.50	0.2107 ±0.01	0.1312 ±0.02	**0.1127 ±0.01**	0.1458 ±0.01	0.1596 ±0.01
0.60	0.2097 ±0.02	0.1388 ±0.02	**0.1156 ±0.02**	0.1420 ±0.01	0.1549 ±0.01
0.70	0.2014 ±0.02	0.1338 ±0.02	**0.1134 ±0.02**	0.1439 ±0.01	0.1572 ±0.01
0.80	0.2013 ±0.03	0.1298 ±0.02	**0.1108 ±0.01**	0.1434 ±0.01	0.1573 ±0.01

Source: elaborated by the author.

Table 37 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$ (Energy).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2619 ±0.02	0.1865 ±0.03	**0.1598 ±0.02**	0.1883 ±0.01	0.1827 ±0.01
0.20	0.2627 ±0.02	0.1850 ±0.02	**0.1561 ±0.02**	0.1834 ±0.01	0.1768 ±0.01
0.30	0.2538 ±0.02	0.1822 ±0.02	**0.1511 ±0.02**	0.1783 ±0.02	0.1724 ±0.02
0.40	0.2548 ±0.02	0.1822 ±0.02	**0.1521 ±0.02**	0.1798 ±0.02	0.1735 ±0.02
0.50	0.2606 ±0.02	0.1889 ±0.02	**0.1599 ±0.02**	0.1850 ±0.02	0.1793 ±0.02
0.60	0.2552 ±0.03	0.1879 ±0.02	**0.1566 ±0.02**	0.1823 ±0.01	0.1765 ±0.01
0.70	0.2644 ±0.02	0.1792 ±0.02	**0.1522 ±0.02**	0.1812 ±0.01	0.1743 ±0.01
0.80	0.2591 ±0.02	0.1907 ±0.02	**0.1591 ±0.02**	0.1851 ±0.02	0.1798 ±0.02

Source: elaborated by the author.

Table 38 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$ (Energy).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2659 ±0.03	0.1849 ±0.02	**0.1562 ±0.02**	0.1866 ±0.02	0.1779 ±0.02
0.20	0.2726 ±0.03	0.1999 ±0.02	**0.1667 ±0.02**	0.1885 ±0.02	0.1811 ±0.01
0.30	0.2792 ±0.02	0.1933 ±0.03	**0.1670 ±0.02**	0.1935 ±0.01	0.1838 ±0.01
0.40	0.2629 ±0.02	0.1996 ±0.03	**0.1650 ±0.02**	0.1889 ±0.02	0.1819 ±0.02
0.50	0.2777 ±0.03	0.1935 ±0.02	**0.1667 ±0.02**	0.1917 ±0.01	0.1826 ±0.01
0.60	0.2708 ±0.03	0.1902 ±0.02	**0.1618 ±0.02**	0.1895 ±0.02	0.1806 ±0.01
0.70	0.2767 ±0.02	0.1950 ±0.03	**0.1663 ±0.02**	0.1917 ±0.01	0.1833 ±0.01
0.80	0.2765 ±0.03	0.1913 ±0.02	**0.1645 ±0.02**	0.1925 ±0.01	0.1831 ±0.01

Source: elaborated by the author.

Forest Fires

Table 39 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$ (Forest Fires).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.1869 ±0.03	0.1159 ±0.02	**0.0932 ±0.02**	0.1373 ±0.02	0.1601 ±0.02
0.20	0.1917 ±0.02	0.1198 ±0.02	**0.0999 ±0.01**	0.1418 ±0.02	0.1646 ±0.02
0.30	0.1860 ±0.02	0.1198 ±0.02	**0.0964 ±0.02**	0.1362 ±0.01	0.1588 ±0.01
0.40	0.1828 ±0.02	0.1208 ±0.02	**0.0961 ±0.02**	0.1339 ±0.01	0.1577 ±0.01
0.50	0.1904 ±0.02	0.1134 ±0.02	**0.0924 ±0.01**	0.1382 ±0.01	0.1613 ±0.01
0.60	0.1845 ±0.02	0.1273 ±0.02	**0.1023 ±0.02**	0.1357 ±0.01	0.1590 ±0.01
0.70	0.1854 ±0.02	0.1255 ±0.04	**0.1024 ±0.03**	0.1348 ±0.02	0.1570 ±0.01
0.80	0.1799 ±0.02	0.1181 ±0.03	**0.0960 ±0.02**	0.1351 ±0.01	0.1571 ±0.01

Source: elaborated by the author.

Table 40 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$ (Forest Fires).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2547 ±0.03	0.1672 ±0.02	**0.1344 ±0.02**	0.1762 ±0.02	0.1684 ±0.02
0.20	0.2652 ±0.03	0.1585 ±0.02	**0.1304 ±0.02**	0.1799 ±0.02	0.1708 ±0.01
0.30	0.2572 ±0.02	0.1687 ±0.02	**0.1320 ±0.02**	0.1721 ±0.02	0.1645 ±0.02
0.40	0.2573 ±0.02	0.1676 ±0.02	**0.1389 ±0.01**	0.1792 ±0.02	0.1720 ±0.02
0.50	0.2557 ±0.03	0.1638 ±0.03	**0.1333 ±0.02**	0.1765 ±0.02	0.1691 ±0.02
0.60	0.2429 ±0.03	0.1633 ±0.03	**0.1328 ±0.03**	0.1735 ±0.02	0.1670 ±0.02
0.70	0.2460 ±0.03	0.1810 ±0.05	**0.1471 ±0.04**	0.1735 ±0.02	0.1668 ±0.01
0.80	0.2400 ±0.03	0.1691 ±0.03	**0.1363 ±0.02**	0.1728 ±0.02	0.1667 ±0.01

Source: elaborated by the author.

Table 41 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$ (Forest Fires).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2654 ±0.03	0.1796 ±0.03	**0.1427 ±0.02**	0.1848 ±0.02	0.1737 ±0.02
0.20	0.2758 ±0.03	0.1850 ±0.02	**0.1491 ±0.02**	0.1858 ±0.02	0.1750 ±0.02
0.30	0.2700 ±0.03	0.1837 ±0.03	**0.1440 ±0.02**	0.1828 ±0.02	0.1724 ±0.02
0.40	0.2731 ±0.02	0.1789 ±0.02	**0.1477 ±0.02**	0.1921 ±0.02	0.1808 ±0.02
0.50	0.2664 ±0.02	0.1879 ±0.02	**0.1469 ±0.02**	0.1810 ±0.01	0.1712 ±0.01
0.60	0.2719 ±0.03	0.1895 ±0.03	**0.1495 ±0.02**	0.1794 ±0.02	0.1686 ±0.02
0.70	0.2601 ±0.03	0.1921 ±0.05	**0.1586 ±0.05**	0.1853 ±0.02	0.1760 ±0.02
0.80	0.2624 ±0.03	0.1790 ±0.03	**0.1481 ±0.02**	0.1876 ±0.02	0.1773 ±0.01

Source: elaborated by the author.

Glass

Table 42 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$ (Glass).

rMiss	**CMI**	**ESD**	**EED**	**EMK-MC**	**EMK-UT**
0.10	0.1807 ±0.04	0.2167 ±0.04	0.1777 ±0.04	**0.1570 ±0.03**	0.1632 ±0.02
0.20	0.1745 ±0.03	0.2089 ±0.04	0.1699 ±0.03	**0.1537 ±0.02**	0.1607 ±0.02
0.30	0.1676 ±0.02	0.2228 ±0.03	0.1808 ±0.02	**0.1544 ±0.01**	0.1557 ±0.02
0.40	0.1813 ±0.03	0.2255 ±0.04	0.1872 ±0.04	**0.1646 ±0.03**	0.1661 ±0.02
0.50	0.1839 ±0.03	0.2239 ±0.03	0.1868 ±0.04	0.1647 ±0.02	**0.1630 ±0.01**
0.60	0.1786 ±0.03	0.2307 ±0.03	0.1896 ±0.03	**0.1623 ±0.02**	0.1633 ±0.02
0.70	0.1750 ±0.03	0.2233 ±0.04	0.1839 ±0.04	**0.1597 ±0.02**	0.1608 ±0.02
0.80	0.1719 ±0.03	0.2253 ±0.04	0.1870 ±0.03	0.1615 ±0.02	**0.1606 ±0.02**

Source: elaborated by the author.

Table 43 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$ (Glass).

rMiss	**CMI**	**ESD**	**EED**	**EMK-MC**	**EMK-UT**
0.10	0.2192 ±0.05	0.2901 ±0.05	0.2295 ±0.05	**0.2000 ±0.03**	0.2061 ±0.03
0.20	0.2249 ±0.04	0.2897 ±0.06	0.2342 ±0.05	**0.2074 ±0.03**	0.2111 ±0.03
0.30	0.2208 ±0.03	0.2917 ±0.05	0.2341 ±0.04	**0.2073 ±0.02**	0.2147 ±0.03
0.40	0.2204 ±0.03	0.2857 ±0.06	0.2274 ±0.06	**0.2042 ±0.03**	0.2107 ±0.03
0.50	0.2321 ±0.03	0.3023 ±0.05	0.2471 ±0.05	**0.2146 ±0.03**	0.2189 ±0.03
0.60	0.2159 ±0.04	0.2771 ±0.04	0.2211 ±0.03	**0.1977 ±0.02**	0.2034 ±0.02
0.70	**0.2220 ±0.03**	0.3352 ±0.08	0.2792 ±0.08	0.2296 ±0.04	0.2383 ±0.04
0.80	0.2185 ±0.03	0.2720 ±0.04	0.2159 ±0.04	**0.1981 ±0.02**	0.2044 ±0.02

Source: elaborated by the author.

Table 44 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$ (Glass).

rMiss	**CMI**	**ESD**	**EED**	**EMK-MC**	**EMK-UT**
0.10	0.2335 ±0.06	0.2787 ±0.05	0.2184 ±0.05	**0.1994 ±0.04**	0.2043 ±0.03
0.20	0.2407 ±0.04	0.2908 ±0.05	0.2323 ±0.06	**0.2131 ±0.04**	0.2146 ±0.03
0.30	0.2360 ±0.04	0.2801 ±0.04	0.2195 ±0.03	**0.2080 ±0.02**	0.2131 ±0.02
0.40	0.2404 ±0.04	0.2993 ±0.07	0.2420 ±0.06	**0.2198 ±0.04**	0.2237 ±0.03
0.50	0.2366 ±0.04	0.3099 ±0.06	0.2443 ±0.05	**0.2175 ±0.03**	0.2364 ±0.08
0.60	0.2261 ±0.04	0.3007 ±0.06	0.2379 ±0.06	**0.2114 ±0.03**	0.2180 ±0.03
0.70	0.2312 ±0.03	0.3311 ±0.08	0.2664 ±0.08	**0.2244 ±0.04**	0.2343 ±0.04
0.80	0.2323 ±0.03	0.2885 ±0.05	0.2312 ±0.05	**0.2103 ±0.03**	0.2157 ±0.03

Source: elaborated by the author.

Haberman

Table 45 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$ (Haberman).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2422 ±0.04	0.1713 ±0.02	**0.1524 ±0.02**	0.1546 ±0.02	0.1566 ±0.02
0.20	0.2446 ±0.03	0.1699 ±0.02	**0.1501 ±0.01**	0.1520 ±0.02	0.1538 ±0.02
0.30	0.2444 ±0.03	0.1849 ±0.02	0.1616 ±0.02	**0.1588 ±0.02**	0.1596 ±0.02
0.40	0.2463 ±0.03	0.1761 ±0.03	**0.1574 ±0.02**	0.1575 ±0.01	0.1581 ±0.01
0.50	0.2470 ±0.03	0.1871 ±0.03	0.1662 ±0.02	0.1617 ±0.02	**0.1614 ±0.02**
0.60	0.2482 ±0.03	0.1909 ±0.03	0.1687 ±0.03	0.1626 ±0.02	**0.1621 ±0.02**
0.70	0.2531 ±0.03	0.1962 ±0.03	0.1756 ±0.02	0.1680 ±0.02	**0.1662 ±0.01**
0.80	0.2531 ±0.03	0.2088 ±0.04	0.1878 ±0.04	0.1749 ±0.02	**0.1694 ±0.02**

Source: elaborated by the author.

Table 46 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$ (Haberman).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2456 ±0.04	0.2037 ±0.03	0.1795 ±0.03	0.1714 ±0.02	**0.1711 ±0.02**
0.20	0.2573 ±0.03	0.1982 ±0.03	0.1801 ±0.02	0.1730 ±0.02	**0.1709 ±0.02**
0.30	0.2565 ±0.04	0.2123 ±0.03	0.1901 ±0.03	0.1793 ±0.03	**0.1788 ±0.02**
0.40	0.2534 ±0.03	0.2097 ±0.04	0.1914 ±0.04	0.1794 ±0.02	**0.1776 ±0.02**
0.50	0.2608 ±0.03	0.2143 ±0.04	0.1963 ±0.03	0.1849 ±0.03	**0.1836 ±0.03**
0.60	0.2564 ±0.03	0.2332 ±0.03	0.2087 ±0.03	0.1878 ±0.02	**0.1866 ±0.02**
0.70	0.2684 ±0.03	0.2306 ±0.04	0.2108 ±0.04	0.1941 ±0.03	**0.1914 ±0.02**
0.80	0.2643 ±0.03	0.2430 ±0.04	0.2201 ±0.04	0.1968 ±0.03	**0.1945 ±0.03**

Source: elaborated by the author.

Table 47 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$ (Haberman).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2532 ±0.06	0.2036 ±0.03	0.1893 ±0.05	0.1806 ±0.05	**0.1799 ±0.04**
0.20	0.2453 ±0.03	0.2143 ±0.04	0.1930 ±0.03	**0.1785 ±0.02**	0.1793 ±0.02
0.30	0.2418 ±0.03	0.2072 ±0.03	0.1854 ±0.03	**0.1740 ±0.02**	0.1751 ±0.02
0.40	0.2557 ±0.03	0.2192 ±0.04	0.2027 ±0.03	0.1866 ±0.02	**0.1861 ±0.02**
0.50	0.2519 ±0.04	0.2406 ±0.07	0.2168 ±0.06	**0.1927 ±0.04**	0.1941 ±0.04
0.60	0.2538 ±0.03	0.2291 ±0.04	0.2092 ±0.03	**0.1886 ±0.03**	0.1889 ±0.03
0.70	0.2425 ±0.03	0.2311 ±0.05	0.2024 ±0.04	**0.1814 ±0.03**	0.1829 ±0.03
0.80	0.2564 ±0.04	0.2579 ±0.06	0.2285 ±0.05	0.1965 ±0.03	**0.1962 ±0.03**

Source: elaborated by the author.

Housing

Table 48 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$ (Housing).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.1740 ±0.02	0.1919 ±0.03	0.1619 ±0.03	**0.1330 ±0.02**	0.1802 ±0.08
0.20	0.1778 ±0.02	0.1984 ±0.03	0.1692 ±0.03	**0.1432 ±0.02**	0.1954 ±0.07
0.30	0.1749 ±0.02	0.1924 ±0.03	0.1654 ±0.03	**0.1423 ±0.02**	0.1775 ±0.04
0.40	0.1776 ±0.02	0.1993 ±0.03	0.1709 ±0.03	**0.1451 ±0.01**	0.1817 ±0.03
0.50	0.1755 ±0.02	0.1845 ±0.03	0.1551 ±0.03	**0.1346 ±0.01**	0.1729 ±0.03
0.60	0.1843 ±0.02	0.1829 ±0.02	0.1549 ±0.02	**0.1422 ±0.02**	0.1802 ±0.03
0.70	0.1820 ±0.02	0.1931 ±0.03	0.1658 ±0.03	**0.1433 ±0.01**	0.1736 ±0.03
0.80	0.1788 ±0.02	0.1978 ±0.03	0.1701 ±0.03	**0.1428 ±0.01**	0.1660 ±0.01

Source: elaborated by the author.

Table 49 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$ (Housing).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2564 ±0.04	0.2383 ±0.03	0.1950 ±0.03	**0.1906 ±0.02**	0.1926 ±0.02
0.20	0.2430 ±0.03	0.2615 ±0.04	0.2213 ±0.04	**0.1877 ±0.02**	0.2079 ±0.07
0.30	0.2465 ±0.03	0.2628 ±0.03	0.2244 ±0.03	**0.1954 ±0.02**	0.2141 ±0.05
0.40	0.2661 ±0.03	0.2196 ±0.02	**0.1801 ±0.02**	0.1950 ±0.02	0.1940 ±0.02
0.50	0.2601 ±0.03	0.2262 ±0.04	**0.1840 ±0.03**	0.1944 ±0.01	0.1943 ±0.01
0.60	0.2463 ±0.03	0.2683 ±0.04	0.2306 ±0.04	**0.1958 ±0.02**	0.2031 ±0.03
0.70	0.2554 ±0.02	0.2233 ±0.03	**0.1828 ±0.03**	0.1972 ±0.01	0.1978 ±0.01
0.80	0.2587 ±0.03	0.2161 ±0.03	**0.1754 ±0.03**	0.1955 ±0.02	0.1958 ±0.02

Source: elaborated by the author.

Table 50 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$ (Housing).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2754 ±0.04	0.2746 ±0.04	0.2257 ±0.03	**0.2153 ±0.02**	0.2154 ±0.02
0.20	0.2773 ±0.04	0.2473 ±0.03	**0.1982 ±0.03**	0.2035 ±0.02	0.2013 ±0.02
0.30	0.2818 ±0.03	0.2372 ±0.04	**0.1975 ±0.03**	0.2133 ±0.02	0.2087 ±0.02
0.40	0.2732 ±0.03	0.2520 ±0.03	**0.2051 ±0.02**	0.2126 ±0.01	0.2113 ±0.01
0.50	0.2771 ±0.03	0.2471 ±0.03	**0.1992 ±0.03**	0.2122 ±0.02	0.2102 ±0.02
0.60	0.2681 ±0.03	0.2536 ±0.03	**0.2018 ±0.03**	0.2057 ±0.02	0.2056 ±0.02
0.70	0.2794 ±0.02	0.2372 ±0.03	**0.1909 ±0.02**	0.2066 ±0.02	0.2027 ±0.02
0.80	0.2781 ±0.03	0.2653 ±0.03	**0.2155 ±0.03**	0.2174 ±0.02	0.2170 ±0.02

Source: elaborated by the author.

Iris

Table 51 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$ (Iris).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2044 ±0.04	0.2736 ±0.04	0.2239 ±0.03	0.1808 ±0.03	**0.1755 ±0.03**
0.20	0.2288 ±0.03	0.2763 ±0.03	0.2292 ±0.03	0.1921 ±0.02	**0.1876 ±0.02**
0.30	0.2300 ±0.03	0.2914 ±0.05	0.2434 ±0.04	0.1990 ±0.03	**0.1916 ±0.02**
0.40	0.2308 ±0.03	0.2924 ±0.04	0.2455 ±0.04	0.2001 ±0.02	**0.1922 ±0.02**
0.50	0.2355 ±0.03	0.2949 ±0.04	0.2482 ±0.04	0.2020 ±0.03	**0.1929 ±0.02**
0.60	0.2301 ±0.03	0.3029 ±0.05	0.2585 ±0.04	0.2065 ±0.03	**0.1934 ±0.02**
0.70	0.2391 ±0.04	0.2924 ±0.06	0.2490 ±0.06	0.2029 ±0.04	**0.1909 ±0.03**
0.80	0.2435 ±0.04	0.3212 ±0.05	0.2783 ±0.05	0.2215 ±0.03	**0.2032 ±0.02**

Source: elaborated by the author.

Table 52 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$ (Iris).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2526 ±0.04	0.3329 ±0.05	0.2636 ±0.04	**0.2239 ±0.03**	0.2338 ±0.03
0.20	0.2655 ±0.04	0.3471 ±0.04	0.2797 ±0.03	**0.2354 ±0.02**	0.2449 ±0.02
0.30	0.2626 ±0.04	0.3418 ±0.03	0.2765 ±0.03	**0.2342 ±0.02**	0.2433 ±0.02
0.40	0.2495 ±0.04	0.3512 ±0.04	0.2810 ±0.04	**0.2322 ±0.02**	0.2430 ±0.03
0.50	0.2520 ±0.03	0.3502 ±0.05	0.2840 ±0.05	**0.2327 ±0.03**	0.2419 ±0.03
0.60	0.2603 ±0.03	0.3651 ±0.06	0.2978 ±0.05	**0.2424 ±0.03**	0.2523 ±0.03
0.70	0.2626 ±0.04	0.3924 ±0.06	0.3268 ±0.06	**0.2579 ±0.04**	0.2666 ±0.03
0.80	0.2814 ±0.03	0.3999 ±0.07	0.3410 ±0.07	**0.2701 ±0.05**	0.2758 ±0.04

Source: elaborated by the author.

Table 53 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$ (Iris).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2672 ±0.05	0.3626 ±0.04	0.2888 ±0.03	**0.2433 ±0.02**	0.2560 ±0.02
0.20	0.2404 ±0.05	0.3606 ±0.04	0.2775 ±0.03	**0.2308 ±0.02**	0.2467 ±0.02
0.30	0.2501 ±0.04	0.3507 ±0.03	0.2743 ±0.03	**0.2313 ±0.02**	0.2452 ±0.02
0.40	0.2655 ±0.03	0.3619 ±0.06	0.2886 ±0.05	**0.2440 ±0.03**	0.2571 ±0.03
0.50	0.2555 ±0.04	0.3552 ±0.06	0.2830 ±0.05	**0.2370 ±0.03**	0.2495 ±0.03
0.60	0.2606 ±0.04	0.3861 ±0.05	0.3124 ±0.05	**0.2526 ±0.03**	0.2666 ±0.03
0.70	0.2690 ±0.04	0.3880 ±0.07	0.3204 ±0.07	**0.2581 ±0.04**	0.2690 ±0.04
0.80	**0.2743 ±0.05**	0.4205 ±0.07	0.3544 ±0.08	0.2773 ±0.05	0.2865 ±0.04

Source: elaborated by the author.

Monk 1

Table 54 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$ (Monk 1).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2242 ±0.02	**0.0905 ±0.01**	0.1060 ±0.01	0.1614 ±0.01	0.1741 ±0.01
0.20	0.2310 ±0.02	**0.0935 ±0.02**	0.1065 ±0.01	0.1601 ±0.01	0.1724 ±0.01
0.30	0.2281 ±0.02	**0.0894 ±0.02**	0.1043 ±0.01	0.1595 ±0.01	0.1722 ±0.01
0.40	0.2325 ±0.02	**0.0937 ±0.02**	0.1069 ±0.01	0.1607 ±0.01	0.1730 ±0.01
0.50	0.2253 ±0.02	**0.0921 ±0.02**	0.1044 ±0.01	0.1578 ±0.01	0.1702 ±0.01
0.60	0.2278 ±0.01	**0.0973 ±0.01**	0.1087 ±0.01	0.1611 ±0.01	0.1734 ±0.01
0.70	0.2259 ±0.02	**0.0878 ±0.01**	0.1026 ±0.01	0.1581 ±0.01	0.1707 ±0.01
0.80	0.2266 ±0.02	**0.0889 ±0.02**	0.1039 ±0.01	0.1590 ±0.01	0.1715 ±0.01

Source: elaborated by the author.

Table 55 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$ (Monk 1).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2972 ±0.02	**0.1254 ±0.02**	0.1497 ±0.01	0.1873 ±0.01	0.1698 ±0.01
0.20	0.2946 ±0.02	**0.1241 ±0.01**	0.1471 ±0.01	0.1848 ±0.02	0.1671 ±0.01
0.30	0.3000 ±0.02	**0.1218 ±0.02**	0.1518 ±0.01	0.1922 ±0.01	0.1740 ±0.01
0.40	0.2905 ±0.02	**0.1288 ±0.02**	0.1479 ±0.01	0.1847 ±0.01	0.1676 ±0.01
0.50	0.2880 ±0.01	**0.1256 ±0.02**	0.1476 ±0.01	0.1860 ±0.01	0.1690 ±0.01
0.60	0.2993 ±0.01	**0.1224 ±0.01**	0.1509 ±0.01	0.1902 ±0.01	0.1721 ±0.01
0.70	0.2959 ±0.02	**0.1263 ±0.01**	0.1511 ±0.01	0.1893 ±0.01	0.1716 ±0.01
0.80	0.2961 ±0.01	**0.1267 ±0.02**	0.1499 ±0.01	0.1873 ±0.01	0.1697 ±0.01

Source: elaborated by the author.

Table 56 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$ (Monk 1).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.3090 ±0.02	**0.1331 ±0.02**	0.1623 ±0.01	0.1894 ±0.01	0.1699 ±0.01
0.20	0.3154 ±0.02	**0.1325 ±0.02**	0.1665 ±0.01	0.1936 ±0.01	0.1733 ±0.01
0.30	0.3073 ±0.02	**0.1350 ±0.02**	0.1632 ±0.02	0.1902 ±0.02	0.1709 ±0.02
0.40	0.3067 ±0.01	**0.1318 ±0.02**	0.1599 ±0.01	0.1869 ±0.01	0.1674 ±0.01
0.50	0.3084 ±0.02	**0.1373 ±0.01**	0.1633 ±0.01	0.1891 ±0.01	0.1702 ±0.01
0.60	0.3007 ±0.02	**0.1427 ±0.02**	0.1628 ±0.01	0.1880 ±0.01	0.1700 ±0.01
0.70	0.3038 ±0.02	**0.1324 ±0.02**	0.1597 ±0.01	0.1867 ±0.01	0.1676 ±0.01
0.80	0.3030 ±0.01	**0.1412 ±0.01**	0.1629 ±0.01	0.1881 ±0.01	0.1700 ±0.01

Source: elaborated by the author.

Monk 2

Table 57 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$ (Monk 2).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2266 ±0.02	**0.0968 ±0.02**	0.1074 ±0.01	0.1593 ±0.01	0.1716 ±0.01
0.20	0.2247 ±0.02	**0.0930 ±0.01**	0.1047 ±0.01	0.1585 ±0.01	0.1710 ±0.01
0.30	0.2284 ±0.01	**0.0940 ±0.02**	0.1057 ±0.01	0.1587 ±0.01	0.1709 ±0.01
0.40	0.2326 ±0.02	**0.0953 ±0.02**	0.1090 ±0.01	0.1624 ±0.01	0.1745 ±0.01
0.50	0.2286 ±0.02	**0.0932 ±0.02**	0.1054 ±0.01	0.1587 ±0.01	0.1709 ±0.01
0.60	0.2316 ±0.02	**0.0923 ±0.02**	0.1056 ±0.01	0.1594 ±0.01	0.1716 ±0.01
0.70	0.2302 ±0.02	**0.0906 ±0.02**	0.1058 ±0.01	0.1609 ±0.01	0.1733 ±0.01
0.80	0.2265 ±0.02	**0.0945 ±0.02**	0.1074 ±0.01	0.1613 ±0.01	0.1736 ±0.01

Source: elaborated by the author.

Table 58 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$ (Monk 2).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2871 ±0.02	**0.1283 ±0.01**	0.1474 ±0.01	0.1849 ±0.02	0.1679 ±0.01
0.20	0.2975 ±0.02	**0.1256 ±0.01**	0.1501 ±0.01	0.1881 ±0.01	0.1705 ±0.01
0.30	0.2909 ±0.02	**0.1243 ±0.02**	0.1452 ±0.01	0.1825 ±0.01	0.1652 ±0.01
0.40	0.2938 ±0.02	**0.1281 ±0.02**	0.1498 ±0.01	0.1872 ±0.01	0.1703 ±0.01
0.50	0.2947 ±0.02	**0.1196 ±0.02**	0.1468 ±0.01	0.1863 ±0.01	0.1685 ±0.01
0.60	0.2984 ±0.02	**0.1272 ±0.02**	0.1529 ±0.01	0.1910 ±0.01	0.1734 ±0.01
0.70	0.2886 ±0.02	**0.1337 ±0.01**	0.1523 ±0.01	0.1885 ±0.01	0.1720 ±0.01
0.80	0.2962 ±0.02	**0.1311 ±0.02**	0.1542 ±0.01	0.1910 ±0.01	0.1739 ±0.01

Source: elaborated by the author.

Table 59 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$ (Monk 2).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.3139 ±0.02	**0.1352 ±0.02**	0.1669 ±0.01	0.1944 ±0.01	0.1747 ±0.01
0.20	0.3047 ±0.02	**0.1371 ±0.02**	0.1642 ±0.01	0.1913 ±0.01	0.1723 ±0.01
0.30	0.3051 ±0.02	**0.1336 ±0.02**	0.1604 ±0.01	0.1868 ±0.02	0.1674 ±0.01
0.40	0.3067 ±0.02	**0.1339 ±0.02**	0.1623 ±0.01	0.1896 ±0.01	0.1702 ±0.01
0.50	0.3117 ±0.01	**0.1349 ±0.02**	0.1629 ±0.01	0.1887 ±0.01	0.1692 ±0.01
0.60	0.3026 ±0.02	**0.1403 ±0.02**	0.1640 ±0.01	0.1901 ±0.01	0.1718 ±0.01
0.70	0.3074 ±0.02	**0.1343 ±0.02**	0.1629 ±0.01	0.1894 ±0.01	0.1701 ±0.01
0.80	0.3058 ±0.02	**0.1347 ±0.01**	0.1613 ±0.01	0.1881 ±0.01	0.1687 ±0.01

Source: elaborated by the author.

Monk 3

Table 60 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$ (Monk 3).

rMiss	**CMI**	**ESD**	**EED**	**EMK-MC**	**EMK-UT**
0.10	0.2260 ±0.03	**0.0899 ±0.02**	0.1027 ±0.01	0.1570 ±0.01	0.1696 ±0.01
0.20	0.2307 ±0.01	**0.0959 ±0.02**	0.1090 ±0.01	0.1621 ±0.01	0.1744 ±0.01
0.30	0.2257 ±0.02	**0.0956 ±0.02**	0.1074 ±0.01	0.1601 ±0.01	0.1723 ±0.01
0.40	0.2319 ±0.02	**0.0937 ±0.02**	0.1091 ±0.01	0.1631 ±0.01	0.1754 ±0.01
0.50	0.2288 ±0.02	**0.0883 ±0.02**	0.1014 ±0.01	0.1564 ±0.01	0.1690 ±0.01
0.60	0.2329 ±0.02	**0.0908 ±0.01**	0.1073 ±0.01	0.1626 ±0.01	0.1751 ±0.01
0.70	0.2352 ±0.02	**0.0912 ±0.01**	0.1082 ±0.01	0.1637 ±0.01	0.1762 ±0.01
0.80	0.2300 ±0.02	**0.0962 ±0.01**	0.1084 ±0.01	0.1611 ±0.01	0.1733 ±0.01

Source: elaborated by the author.

Table 61 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$ (Monk 3).

rMiss	**CMI**	**ESD**	**EED**	**EMK-MC**	**EMK-UT**
0.10	0.2885 ±0.02	**0.1276 ±0.02**	0.1501 ±0.01	0.1883 ±0.01	0.1712 ±0.01
0.20	0.2917 ±0.01	**0.1241 ±0.02**	0.1479 ±0.01	0.1867 ±0.01	0.1693 ±0.01
0.30	0.2972 ±0.02	**0.1253 ±0.01**	0.1494 ±0.01	0.1874 ±0.01	0.1698 ±0.01
0.40	0.2912 ±0.02	**0.1323 ±0.02**	0.1517 ±0.01	0.1883 ±0.01	0.1716 ±0.01
0.50	0.2945 ±0.02	**0.1201 ±0.01**	0.1478 ±0.01	0.1881 ±0.01	0.1701 ±0.01
0.60	0.2969 ±0.02	**0.1240 ±0.01**	0.1509 ±0.01	0.1898 ±0.01	0.1720 ±0.01
0.70	0.2974 ±0.02	**0.1247 ±0.02**	0.1498 ±0.01	0.1879 ±0.01	0.1704 ±0.01
0.80	0.2939 ±0.02	**0.1260 ±0.02**	0.1492 ±0.01	0.1874 ±0.01	0.1700 ±0.01

Source: elaborated by the author.

Table 62 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$ (Monk 3).

rMiss	**CMI**	**ESD**	**EED**	**EMK-MC**	**EMK-UT**
0.10	0.3072 ±0.02	**0.1394 ±0.02**	0.1644 ±0.01	0.1902 ±0.02	0.1716 ±0.01
0.20	0.3088 ±0.02	**0.1396 ±0.02**	0.1661 ±0.02	0.1923 ±0.02	0.1734 ±0.01
0.30	0.3053 ±0.01	**0.1368 ±0.02**	0.1611 ±0.01	0.1869 ±0.01	0.1681 ±0.01
0.40	0.3117 ±0.02	**0.1349 ±0.02**	0.1653 ±0.02	0.1923 ±0.02	0.1726 ±0.02
0.50	0.3060 ±0.02	**0.1365 ±0.02**	0.1629 ±0.02	0.1899 ±0.02	0.1708 ±0.01
0.60	0.3091 ±0.02	**0.1345 ±0.02**	0.1632 ±0.01	0.1899 ±0.01	0.1708 ±0.01
0.70	0.3049 ±0.02	**0.1360 ±0.02**	0.1609 ±0.01	0.1867 ±0.01	0.1675 ±0.01
0.80	0.3011 ±0.02	**0.1332 ±0.01**	0.1604 ±0.01	0.1881 ±0.01	0.1687 ±0.01

Source: elaborated by the author.

Auto MPG

Table 63 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$ (Auto MPG).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.1735 ±0.02	0.1695 ±0.04	0.1390 ±0.04	**0.1354 ±0.02**	0.1392 ±0.02
0.20	0.1784 ±0.02	0.1673 ±0.03	0.1356 ±0.02	**0.1320 ±0.01**	0.1376 ±0.01
0.30	0.1837 ±0.02	0.1717 ±0.03	0.1415 ±0.03	**0.1369 ±0.01**	0.1410 ±0.01
0.40	0.1770 ±0.02	0.1657 ±0.03	**0.1359 ±0.02**	0.1366 ±0.01	0.1423 ±0.01
0.50	0.1756 ±0.02	0.1730 ±0.03	0.1409 ±0.02	**0.1352 ±0.01**	0.1404 ±0.01
0.60	0.1812 ±0.03	0.1876 ±0.03	0.1540 ±0.03	**0.1440 ±0.02**	0.1478 ±0.02
0.70	0.1780 ±0.02	0.1710 ±0.03	0.1405 ±0.02	**0.1385 ±0.02**	0.1435 ±0.02
0.80	0.1747 ±0.02	0.1779 ±0.03	0.1448 ±0.03	**0.1356 ±0.02**	0.1392 ±0.02

Source: elaborated by the author.

Table 64 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$ (Auto MPG).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2142 ±0.03	0.2124 ±0.03	0.1714 ±0.03	**0.1692 ±0.02**	0.1713 ±0.02
0.20	0.2093 ±0.03	0.2292 ±0.04	0.1823 ±0.04	**0.1689 ±0.02**	0.1719 ±0.02
0.30	0.2111 ±0.02	0.2307 ±0.04	0.1855 ±0.04	**0.1729 ±0.02**	0.1765 ±0.02
0.40	0.2188 ±0.02	0.2344 ±0.04	0.1896 ±0.04	**0.1772 ±0.02**	0.1801 ±0.02
0.50	0.2070 ±0.03	0.2180 ±0.04	0.1717 ±0.03	**0.1642 ±0.02**	0.1681 ±0.02
0.60	0.2144 ±0.03	0.2284 ±0.04	0.1837 ±0.03	**0.1719 ±0.02**	0.1740 ±0.02
0.70	0.2110 ±0.02	0.2347 ±0.04	0.1890 ±0.04	**0.1741 ±0.02**	0.1770 ±0.02
0.80	0.2011 ±0.02	0.2221 ±0.04	0.1763 ±0.03	**0.1667 ±0.02**	0.1707 ±0.02

Source: elaborated by the author.

Table 65 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$ (Auto MPG).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2221 ±0.03	0.2242 ±0.05	0.1807 ±0.04	**0.1774 ±0.03**	0.1788 ±0.03
0.20	0.2246 ±0.02	0.2406 ±0.05	0.1939 ±0.04	**0.1807 ±0.03**	0.1834 ±0.03
0.30	0.2227 ±0.03	0.2301 ±0.05	0.1849 ±0.04	**0.1791 ±0.02**	0.1814 ±0.02
0.40	0.2122 ±0.03	0.2383 ±0.05	0.1898 ±0.04	**0.1780 ±0.02**	0.1822 ±0.02
0.50	0.2167 ±0.02	0.2408 ±0.05	0.1924 ±0.04	**0.1771 ±0.02**	0.1802 ±0.02
0.60	0.2164 ±0.02	0.2433 ±0.04	0.1934 ±0.04	**0.1845 ±0.02**	0.1887 ±0.02
0.70	0.2108 ±0.03	0.2323 ±0.06	0.1852 ±0.05	**0.1779 ±0.03**	0.1815 ±0.03
0.80	0.2215 ±0.02	0.2308 ±0.04	0.1864 ±0.04	**0.1764 ±0.02**	0.1783 ±0.02

Source: elaborated by the author.

Servo

Table 66 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$ (Servo).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2658 ±0.04	0.1622 ±0.05	**0.1448 ±0.04**	0.1532 ±0.03	0.1541 ±0.02
0.20	0.2614 ±0.03	0.1636 ±0.05	**0.1482 ±0.04**	0.1584 ±0.02	0.1688 ±0.07
0.30	0.2610 ±0.03	0.1568 ±0.03	**0.1406 ±0.02**	0.1536 ±0.02	0.1586 ±0.02
0.40	0.2641 ±0.03	0.1438 ±0.02	**0.1341 ±0.01**	0.1552 ±0.01	0.1610 ±0.02
0.50	0.2712 ±0.03	0.1659 ±0.06	**0.1540 ±0.06**	0.1643 ±0.04	0.1740 ±0.08
0.60	0.2739 ±0.03	0.1531 ±0.05	**0.1437 ±0.04**	0.1597 ±0.03	0.1595 ±0.01
0.70	0.2651 ±0.03	0.1585 ±0.04	**0.1469 ±0.04**	0.1615 ±0.03	0.1720 ±0.07
0.80	0.2693 ±0.03	0.1397 ±0.02	**0.1310 ±0.01**	0.1537 ±0.02	0.1597 ±0.02

Source: elaborated by the author.

Table 67 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$ (Servo).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2791 ±0.05	0.2113 ±0.07	0.1936 ±0.07	**0.1891 ±0.05**	0.1900 ±0.08
0.20	0.2827 ±0.04	0.2006 ±0.06	0.1853 ±0.05	0.1861 ±0.03	**0.1803 ±0.03**
0.30	0.2895 ±0.03	0.2060 ±0.07	0.1913 ±0.06	0.1874 ±0.04	**0.1810 ±0.03**
0.40	0.2844 ±0.03	0.1807 ±0.03	**0.1722 ±0.02**	0.1806 ±0.02	0.1738 ±0.02
0.50	0.2840 ±0.03	0.1958 ±0.05	0.1806 ±0.04	0.1815 ±0.03	**0.1745 ±0.03**
0.60	0.2919 ±0.03	0.1811 ±0.02	**0.1731 ±0.02**	0.1810 ±0.02	0.1739 ±0.02
0.70	0.2870 ±0.02	0.1839 ±0.02	**0.1755 ±0.01**	0.1834 ±0.02	0.1766 ±0.02
0.80	0.2913 ±0.03	0.1820 ±0.02	**0.1739 ±0.02**	0.1820 ±0.02	0.1746 ±0.02

Source: elaborated by the author.

Table 68 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$ (Servo).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2811 ±0.03	0.1968 ±0.04	0.1856 ±0.03	0.1834 ±0.03	**0.1779 ±0.02**
0.20	0.2825 ±0.04	0.2238 ±0.09	0.2074 ±0.08	0.1929 ±0.05	**0.1868 ±0.04**
0.30	0.2959 ±0.03	0.1828 ±0.02	0.1822 ±0.02	0.1860 ±0.02	**0.1784 ±0.02**
0.40	0.2928 ±0.02	0.1787 ±0.02	0.1758 ±0.01	0.1803 ±0.01	**0.1730 ±0.01**
0.50	0.2955 ±0.05	0.2487 ±0.11	0.2293 ±0.10	0.2051 ±0.06	**0.1983 ±0.05**
0.60	0.2837 ±0.02	0.1797 ±0.02	0.1735 ±0.02	0.1777 ±0.01	**0.1709 ±0.01**
0.70	0.2833 ±0.02	0.1872 ±0.02	0.1813 ±0.02	0.1842 ±0.02	**0.1784 ±0.02**
0.80	0.2872 ±0.03	0.1861 ±0.02	0.1801 ±0.02	0.1828 ±0.02	**0.1764 ±0.02**

Source: elaborated by the author.

Concrete Slump

Table 69 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$ (Concrete Slump).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2053 ±0.04	0.2014 ±0.04	0.1716 ±0.04	0.1484 ±0.03	**0.1456 ±0.03**
0.20	0.2007 ±0.03	0.2021 ±0.03	0.1717 ±0.03	**0.1440 ±0.03**	0.1484 ±0.03
0.30	0.2034 ±0.03	0.2348 ±0.06	0.2076 ±0.06	0.1586 ±0.03	**0.1381 ±0.02**
0.40	0.2000 ±0.02	0.1947 ±0.04	0.1630 ±0.04	**0.1422 ±0.02**	0.1464 ±0.02
0.50	0.2151 ±0.02	0.2399 ±0.06	0.2126 ±0.06	0.1650 ±0.03	**0.1469 ±0.02**
0.60	0.2032 ±0.03	0.2195 ±0.05	0.1909 ±0.05	0.1497 ±0.03	**0.1389 ±0.02**
0.70	0.2138 ±0.02	0.1932 ±0.06	0.1726 ±0.05	0.1554 ±0.02	**0.1515 ±0.02**
0.80	0.2082 ±0.03	0.2153 ±0.07	0.1939 ±0.06	0.1595 ±0.02	**0.1463 ±0.02**

Source: elaborated by the author.

Table 70 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$ (Concrete Slump).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2667 ±0.04	0.2908 ±0.07	0.2559 ±0.07	0.1853 ±0.04	**0.1722 ±0.03**
0.20	0.2860 ±0.04	0.2771 ±0.05	0.2388 ±0.05	**0.1994 ±0.03**	0.2129 ±0.09
0.30	0.2627 ±0.03	0.3304 ±0.06	0.2946 ±0.07	**0.2075 ±0.04**	0.2201 ±0.09
0.40	0.2658 ±0.04	0.3238 ±0.07	0.2869 ±0.08	**0.2113 ±0.04**	0.2117 ±0.08
0.50	0.2700 ±0.02	0.2440 ±0.08	0.2136 ±0.07	0.1877 ±0.03	**0.1778 ±0.03**
0.60	0.2721 ±0.04	0.3108 ±0.09	0.2757 ±0.08	0.2073 ±0.05	**0.2029 ±0.06**
0.70	0.2839 ±0.03	0.2636 ±0.08	0.2365 ±0.07	**0.2004 ±0.03**	0.2038 ±0.06
0.80	0.2858 ±0.03	0.2785 ±0.10	0.2578 ±0.10	0.2151 ±0.04	**0.2124 ±0.09**

Source: elaborated by the author.

Table 71 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$ (Concrete Slump).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2878 ±0.05	0.2982 ±0.06	0.2542 ±0.06	0.2041 ±0.04	**0.1907 ±0.04**
0.20	0.3022 ±0.04	0.3001 ±0.05	0.2582 ±0.05	0.2114 ±0.03	**0.2039 ±0.06**
0.30	0.2981 ±0.04	0.3425 ±0.08	0.3068 ±0.08	0.2294 ±0.05	**0.2119 ±0.04**
0.40	0.2792 ±0.02	0.2922 ±0.06	0.2441 ±0.06	0.1938 ±0.03	**0.1847 ±0.03**
0.50	0.2788 ±0.04	0.3349 ±0.07	0.2943 ±0.07	0.2146 ±0.03	**0.2015 ±0.03**
0.60	0.2915 ±0.04	0.2667 ±0.07	0.2307 ±0.07	0.2015 ±0.03	**0.1905 ±0.02**
0.70	0.2777 ±0.03	0.3406 ±0.09	0.3023 ±0.09	0.2200 ±0.04	**0.2010 ±0.03**
0.80	0.3003 ±0.03	0.3453 ±0.12	0.3195 ±0.10	0.2388 ±0.04	**0.2137 ±0.03**

Source: elaborated by the author.

Statlog (vehicle silhouettes)

Table 72 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$ (Statlog (vehicle silhouettes)).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	**0.1033 ±0.02**	0.1697 ±0.02	0.1379 ±0.02	0.1056 ±0.01	0.1365 ±0.01
0.20	**0.1052 ±0.01**	0.1646 ±0.02	0.1332 ±0.02	0.1063 ±0.01	0.1365 ±0.01
0.30	**0.1034 ±0.02**	0.1576 ±0.02	0.1261 ±0.02	0.1052 ±0.01	0.1377 ±0.01
0.40	**0.1009 ±0.02**	0.1538 ±0.02	0.1205 ±0.02	0.1022 ±0.01	0.1347 ±0.01
0.50	0.1085 ±0.02	0.1568 ±0.02	0.1247 ±0.02	**0.1073 ±0.01**	0.1371 ±0.01
0.60	**0.1000 ±0.02**	0.1616 ±0.02	0.1282 ±0.02	0.1066 ±0.01	0.1326 ±0.01
0.70	**0.1026 ±0.01**	0.1523 ±0.02	0.1193 ±0.02	0.1052 ±0.01	0.1332 ±0.01
0.80	0.1037 ±0.02	0.1543 ±0.02	0.1217 ±0.02	**0.1037 ±0.01**	0.1275 ±0.01

Source: elaborated by the author.

Table 73 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$ (Statlog (vehicle silhouettes)).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	**0.1466 ±0.03**	0.2254 ±0.03	0.1831 ±0.03	0.1478 ±0.02	0.1666 ±0.02
0.20	**0.1430 ±0.02**	0.2264 ±0.03	0.1841 ±0.03	0.1511 ±0.02	0.1700 ±0.02
0.30	0.1563 ±0.02	0.2159 ±0.04	0.1750 ±0.04	**0.1543 ±0.02**	0.1716 ±0.02
0.40	0.1475 ±0.02	0.2150 ±0.03	0.1697 ±0.03	**0.1442 ±0.01**	0.1622 ±0.01
0.50	**0.1487 ±0.02**	0.2236 ±0.03	0.1796 ±0.03	0.1500 ±0.02	0.1675 ±0.01
0.60	0.1485 ±0.02	0.2102 ±0.03	0.1662 ±0.03	**0.1466 ±0.01**	0.1631 ±0.01
0.70	**0.1407 ±0.02**	0.2158 ±0.02	0.1708 ±0.02	0.1433 ±0.01	0.1602 ±0.01
0.80	**0.1423 ±0.02**	0.2080 ±0.03	0.1641 ±0.03	0.1434 ±0.01	0.1594 ±0.01

Source: elaborated by the author.

Table 74 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$ (Statlog (vehicle silhouettes)).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	**0.1612 ±0.02**	0.2487 ±0.03	0.2038 ±0.03	0.1619 ±0.01	0.1787 ±0.01
0.20	0.1611 ±0.02	0.2411 ±0.02	0.1931 ±0.02	**0.1600 ±0.01**	0.1753 ±0.01
0.30	0.1620 ±0.02	0.2413 ±0.03	0.1943 ±0.03	**0.1610 ±0.02**	0.1764 ±0.01
0.40	0.1698 ±0.02	0.2408 ±0.04	0.1943 ±0.04	**0.1630 ±0.02**	0.1773 ±0.01
0.50	**0.1569 ±0.02**	0.2458 ±0.03	0.1979 ±0.03	0.1613 ±0.01	0.1775 ±0.01
0.60	**0.1585 ±0.02**	0.2311 ±0.03	0.1833 ±0.03	0.1591 ±0.01	0.1734 ±0.01
0.70	**0.1562 ±0.02**	0.2378 ±0.02	0.1893 ±0.02	0.1598 ±0.01	0.1751 ±0.01
0.80	**0.1560 ±0.02**	0.2371 ±0.04	0.1882 ±0.03	0.1605 ±0.02	0.1758 ±0.02

Source: elaborated by the author.

Wine

Table 75 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{1/2}(\cdot)]$ (Wine).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.1752 ±0.03	0.1438 ±0.03	**0.1151 ±0.03**	0.1309 ±0.02	0.1595 ±0.02
0.20	0.1669 ±0.02	0.1390 ±0.02	**0.1114 ±0.02**	0.1329 ±0.01	0.1621 ±0.01
0.30	0.1659 ±0.02	0.1411 ±0.03	**0.1141 ±0.02**	0.1316 ±0.01	0.1589 ±0.01
0.40	0.1675 ±0.02	0.1393 ±0.03	**0.1150 ±0.03**	0.1368 ±0.01	0.1648 ±0.01
0.50	0.1682 ±0.02	0.1373 ±0.03	**0.1165 ±0.02**	0.1386 ±0.02	0.1641 ±0.01
0.60	0.1602 ±0.02	0.1249 ±0.03	**0.1018 ±0.02**	0.1314 ±0.02	0.1557 ±0.01
0.70	0.1689 ±0.02	0.1153 ±0.02	**0.0982 ±0.01**	0.1383 ±0.01	0.1620 ±0.01
0.80	0.1560 ±0.02	0.1288 ±0.04	**0.1081 ±0.03**	0.1322 ±0.02	0.1548 ±0.02

Source: elaborated by the author.

Table 76 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{3/2}(\cdot)]$ (Wine).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2460 ±0.04	0.2051 ±0.04	**0.1650 ±0.03**	0.1745 ±0.02	0.1702 ±0.02
0.20	0.2372 ±0.03	0.1986 ±0.03	**0.1602 ±0.03**	0.1774 ±0.02	0.1733 ±0.01
0.30	0.2333 ±0.03	0.2013 ±0.04	**0.1638 ±0.03**	0.1760 ±0.02	0.1732 ±0.01
0.40	0.2369 ±0.03	0.1832 ±0.04	**0.1495 ±0.03**	0.1765 ±0.02	0.1727 ±0.02
0.50	0.2242 ±0.03	0.2050 ±0.05	**0.1665 ±0.05**	0.1729 ±0.02	0.1737 ±0.02
0.60	0.2377 ±0.03	0.1935 ±0.04	**0.1633 ±0.03**	0.1827 ±0.02	0.1788 ±0.02
0.70	0.2288 ±0.03	0.1784 ±0.03	**0.1498 ±0.02**	0.1806 ±0.02	0.1775 ±0.02
0.80	0.2245 ±0.03	0.1894 ±0.05	**0.1608 ±0.05**	0.1762 ±0.02	0.1732 ±0.02

Source: elaborated by the author.

Table 77 – UCI dataset - ARMSE for $\mathbb{E}[\mathcal{C}_{5/2}(\cdot)]$ (Wine).

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	0.2675 ±0.04	0.2267 ±0.04	0.1823 ±0.04	0.1874 ±0.02	**0.1774 ±0.02**
0.20	0.2593 ±0.03	0.2195 ±0.03	**0.1770 ±0.03**	0.1904 ±0.02	0.1817 ±0.02
0.30	0.2539 ±0.03	0.2222 ±0.04	**0.1808 ±0.03**	0.1891 ±0.02	0.1823 ±0.02
0.40	0.2578 ±0.02	0.1971 ±0.04	**0.1653 ±0.02**	0.1960 ±0.02	0.1878 ±0.01
0.50	0.2482 ±0.03	0.2044 ±0.04	**0.1687 ±0.03**	0.1879 ±0.02	0.1812 ±0.02
0.60	0.2394 ±0.03	0.2053 ±0.04	**0.1658 ±0.04**	0.1849 ±0.02	0.1785 ±0.01
0.70	0.2504 ±0.03	0.2134 ±0.05	**0.1791 ±0.04**	0.1926 ±0.02	0.1852 ±0.01
0.80	0.2406 ±0.04	0.2158 ±0.06	0.1843 ±0.05	0.1874 ±0.02	**0.1812 ±0.02**

Source: elaborated by the author.

APPENDIX C – ARMSE FOR THE CASE-STUDY DATASET

Concrete Slump - Output: *Slump (cm)*

Table 78 – ARMSE in estimating the output *Slump (cm)* using $\mathcal{C}_{1/2}$ (*Concrete Slump*)

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	8.6067 ±0.92	**8.6023 ±0.91**	8.6042 ±0.91	8.6051 ±0.92	8.6048 ±0.93
0.20	8.6609 ±1.16	**8.6503 ±1.19**	8.6542 ±1.18	8.6584 ±1.17	8.6587 ±1.17
0.30	7.9508 ±1.03	7.9532 ±1.01	7.9506 ±1.02	7.9442 ±1.03	**7.9426 ±1.03**
0.40	8.1368 ±1.19	**8.1203 ±1.21**	8.1242 ±1.21	8.1263 ±1.20	8.1274 ±1.20
0.50	8.2066 ±1.19	**8.1830 ±1.21**	8.1885 ±1.21	8.1873 ±1.20	8.1867 ±1.20
0.60	8.2608 ±1.20	**8.2415 ±1.22**	8.2455 ±1.22	8.2496 ±1.22	8.2511 ±1.22
0.70	8.8063 ±1.32	8.8029 ±1.35	8.8019 ±1.34	**8.8006 ±1.34**	8.8007 ±1.34
0.80	8.9076 ±1.25	**8.8946 ±1.28**	8.8961 ±1.27	8.8966 ±1.28	8.8971 ±1.28

Source: elaborated by the author.

Table 79 – ARMSE in estimating the output *Slump (cm)* using $\mathcal{C}_{3/2}$ (*Concrete Slump*)

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	8.3979 ±1.12	**8.3924 ±1.13**	8.3936 ±1.13	8.3938 ±1.13	8.3930 ±1.13
0.20	8.3252 ±1.11	8.3192 ±1.13	8.3188 ±1.13	8.3189 ±1.13	**8.3183 ±1.13**
0.30	**8.5163 ±1.08**	8.5249 ±1.09	8.5210 ±1.09	8.5173 ±1.09	8.5178 ±1.09
0.40	8.6790 ±1.08	8.6830 ±1.10	8.6805 ±1.09	8.6717 ±1.10	**8.6708 ±1.10**
0.50	8.1670 ±1.12	8.1671 ±1.12	8.1659 ±1.12	8.1512 ±1.13	**8.1490 ±1.14**
0.60	8.3022 ±1.36	8.2889 ±1.38	8.2894 ±1.37	8.2888 ±1.38	**8.2878 ±1.38**
0.70	8.5626 ±1.20	8.5442 ±1.22	8.5474 ±1.21	8.5446 ±1.22	**8.5421 ±1.22**
0.80	8.5051 ±1.21	8.4831 ±1.23	8.4861 ±1.22	8.4847 ±1.23	**8.4825 ±1.23**

Source: elaborated by the author.

Table 80 – ARMSE in estimating the output *Slump (cm)* using $\mathcal{C}_{5/2}$ (*Concrete Slump*)

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	8.3699 ±1.33	8.3735 ±1.35	8.3704 ±1.35	8.3677 ±1.34	**8.3676 ±1.34**
0.20	8.7044 ±1.25	8.6971 ±1.27	8.7014 ±1.27	8.6962 ±1.26	**8.6940 ±1.26**
0.30	8.5514 ±1.18	**8.5458 ±1.19**	8.5500 ±1.18	8.5478 ±1.19	8.5467 ±1.19
0.40	8.4084 ±1.15	8.4039 ±1.16	8.4037 ±1.16	8.3922 ±1.17	**8.3906 ±1.17**
0.50	8.2336 ±1.54	**8.1979 ±1.55**	8.2066 ±1.54	8.2030 ±1.55	8.1983 ±1.55
0.60	8.4939 ±1.16	8.4937 ±1.17	8.4912 ±1.17	8.4787 ±1.18	**8.4779 ±1.18**
0.70	8.4017 ±1.18	8.3813 ±1.19	8.3851 ±1.19	8.3789 ±1.19	**8.3764 ±1.20**
0.80	8.6048 ±1.30	8.5874 ±1.30	8.5893 ±1.30	8.5845 ±1.31	**8.5828 ±1.31**

Source: elaborated by the author.

Concrete Slump - Output: ***Flow (cm)***

Table 81 – ARMSE in estimating the output *Flow (cm)* using $\mathcal{C}_{1/2}$ (*Concrete Slump*)

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	17.2083 ±2.01	17.2059 ±1.98	17.2071 ±2.00	17.2057 ±2.02	**17.2044 ±2.02**
0.20	16.3056 ±1.77	**16.2727 ±1.80**	16.2823 ±1.79	16.2908 ±1.78	16.2926 ±1.77
0.30	16.6831 ±1.86	**16.6573 ±1.85**	16.6652 ±1.85	16.6706 ±1.86	16.6715 ±1.87
0.40	16.4933 ±2.00	**16.4640 ±2.06**	16.4660 ±2.04	16.4713 ±2.02	16.4759 ±2.01
0.50	16.7360 ±1.60	**16.6923 ±1.58**	16.7018 ±1.58	16.7046 ±1.59	16.7068 ±1.59
0.60	16.8114 ±1.99	**16.7737 ±2.04**	16.7811 ±2.03	16.7790 ±2.01	16.7791 ±2.01
0.70	16.9304 ±2.05	**16.8918 ±2.08**	16.8982 ±2.07	16.8976 ±2.06	16.8982 ±2.06
0.80	17.3097 ±2.00	**17.2633 ±2.03**	17.2722 ±2.03	17.2772 ±2.03	17.2793 ±2.03

Source: elaborated by the author.

Table 82 – ARMSE in estimating the output *Flow (cm)* using $\mathcal{C}_{3/2}$ (*Concrete Slump*)

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	15.9765 ±1.61	15.9809 ±1.62	15.9801 ±1.62	**15.9747 ±1.61**	15.9751 ±1.61
0.20	15.6443 ±1.91	15.6408 ±1.94	15.6363 ±1.93	15.6287 ±1.92	**15.6267 ±1.93**
0.30	16.1068 ±1.41	16.0994 ±1.40	16.1009 ±1.41	16.0836 ±1.40	**16.0788 ±1.40**
0.40	16.3101 ±2.15	16.2919 ±2.18	16.2915 ±2.17	16.2799 ±2.18	**16.2753 ±2.18**
0.50	16.5550 ±1.55	16.4962 ±1.58	16.5115 ±1.57	16.4950 ±1.56	**16.4838 ±1.56**
0.60	16.1321 ±2.02	16.1096 ±2.02	16.1104 ±2.02	16.0985 ±2.03	**16.0940 ±2.03**
0.70	16.9831 ±1.72	16.9490 ±1.73	16.9537 ±1.72	16.9301 ±1.73	**16.9236 ±1.73**
0.80	16.4650 ±1.92	16.3758 ±1.99	16.3947 ±1.97	16.3848 ±1.97	**16.3725 ±1.98**

Source: elaborated by the author.

Table 83 – ARMSE in estimating the output *Flow (cm)* using $\mathcal{C}_{5/2}$ (*Concrete Slump*)

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	16.4856 ±1.94	**16.4651 ±1.93**	16.4692 ±1.94	16.4735 ±1.95	16.4710 ±1.95
0.20	16.3403 ±1.97	16.3224 ±2.01	16.3252 ±2.00	16.3233 ±1.99	**16.3205 ±1.99**
0.30	16.5209 ±1.90	16.5318 ±1.96	16.5238 ±1.94	16.5011 ±1.93	**16.4994 ±1.94**
0.40	16.5367 ±2.45	16.5145 ±2.49	16.5218 ±2.48	16.5006 ±2.47	**16.4951 ±2.47**
0.50	16.5665 ±1.42	16.5220 ±1.48	16.5349 ±1.46	16.5250 ±1.44	**16.5185 ±1.45**
0.60	16.9118 ±1.92	16.8759 ±1.97	16.8834 ±1.95	16.8703 ±1.95	**16.8647 ±1.96**
0.70	16.6877 ±1.94	16.6226 ±1.95	16.6370 ±1.95	16.6213 ±1.95	**16.6118 ±1.95**
0.80	16.5412 ±1.97	16.4434 ±1.98	16.4652 ±1.98	16.4468 ±1.98	**16.4330 ±1.98**

Source: elaborated by the author.

Concrete Slump - Output: *28-day Compressive Strength (Mpa)*

Table 84 – ARMSE in estimating the output *28-day Compressive Strength (Mpa)* using $\mathcal{C}_{1/2}$ (*Concrete Slump*)

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	7.2148 ±0.90	**7.2114 ±0.90**	7.2117 ±0.90	7.2131 ±0.90	7.2141 ±0.90
0.20	**7.2242 ±1.41**	7.2268 ±1.41	7.2245 ±1.41	7.2247 ±1.41	7.2258 ±1.41
0.30	7.8237 ±1.05	**7.8176 ±1.06**	7.8201 ±1.05	7.8215 ±1.05	7.8211 ±1.05
0.40	7.2509 ±1.34	7.2460 ±1.32	7.2472 ±1.33	7.2458 ±1.33	**7.2444 ±1.33**
0.50	7.3743 ±1.24	7.3775 ±1.24	7.3756 ±1.24	7.3697 ±1.25	**7.3678 ±1.25**
0.60	7.2970 ±1.09	7.2956 ±1.10	7.2953 ±1.10	7.2936 ±1.10	**7.2932 ±1.09**
0.70	7.3607 ±1.05	7.3709 ±1.03	7.3677 ±1.04	7.3602 ±1.04	**7.3580 ±1.04**
0.80	**7.3932 ±1.16**	7.3952 ±1.17	7.3939 ±1.17	7.3934 ±1.17	7.3937 ±1.17

Source: elaborated by the author.

Table 85 – ARMSE in estimating the output *28-day Compressive Strength (Mpa)* using $\mathcal{C}_{3/2}$ (*Concrete Slump*)

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	**7.0204 ±1.34**	7.0227 ±1.33	7.0216 ±1.33	7.0206 ±1.34	7.0208 ±1.34
0.20	7.4283 ±1.33	7.4381 ±1.33	7.4342 ±1.33	**7.4282 ±1.33**	7.4284 ±1.33
0.30	7.2814 ±1.25	7.2847 ±1.25	7.2850 ±1.25	7.2786 ±1.25	**7.2774 ±1.25**
0.40	7.3814 ±1.07	7.3830 ±1.07	7.3802 ±1.07	7.3768 ±1.06	**7.3763 ±1.06**
0.50	7.3006 ±1.10	7.2974 ±1.12	7.2963 ±1.11	7.2939 ±1.11	**7.2933 ±1.11**
0.60	7.4910 ±1.17	7.4995 ±1.18	7.4954 ±1.18	**7.4878 ±1.17**	7.4883 ±1.17
0.70	7.2183 ±1.23	7.2211 ±1.23	7.2179 ±1.23	7.2124 ±1.23	**7.2123 ±1.23**
0.80	7.3669 ±1.07	**7.3523 ±1.07**	7.3545 ±1.07	7.3604 ±1.06	7.3597 ±1.06

Source: elaborated by the author.

Table 86 – ARMSE in estimating the output *28-day Compressive Strength (Mpa)* using $\mathcal{C}_{5/2}$ (*Concrete Slump*)

rMiss	CMI	ESD	EED	EMK-MC	EMK-UT
0.10	6.6997 ±1.05	6.6991 ±1.04	6.6976 ±1.04	6.6968 ±1.05	**6.6963 ±1.05**
0.20	7.2591 ±0.88	7.2620 ±0.88	7.2593 ±0.88	7.2565 ±0.88	**7.2565 ±0.88**
0.30	**7.1011 ±1.00**	7.1288 ±1.01	7.1164 ±1.00	7.1045 ±1.00	7.1069 ±1.01
0.40	7.2818 ±1.26	7.2821 ±1.25	7.2816 ±1.25	7.2749 ±1.25	**7.2743 ±1.25**
0.50	7.3898 ±0.97	7.3800 ±1.00	7.3839 ±0.99	7.3774 ±0.98	**7.3753 ±0.99**
0.60	7.4376 ±1.07	7.4399 ±1.06	7.4381 ±1.07	**7.4360 ±1.07**	7.4362 ±1.06
0.70	7.2512 ±1.06	7.2436 ±1.07	7.2431 ±1.06	7.2372 ±1.06	**7.2364 ±1.06**
0.80	7.4903 ±1.35	7.4757 ±1.36	7.4784 ±1.36	7.4712 ±1.35	**7.4691 ±1.35**

Source: elaborated by the author.

APPENDIX D – PROPOSED EXTENSIONS FOR OTHER KERNELS

There are several common strategies for constructing kernels, one of which is to combine simpler kernels. Let $k_1(\boldsymbol{X}_i, \boldsymbol{X}_i)$ and $k_2(\boldsymbol{X}_i, \boldsymbol{X}_i)$, be two valid kernels. Then, the following kernels are also valid:

$$k(\boldsymbol{X}_i, \boldsymbol{X}_j) = ck_1(\boldsymbol{X}_i, \boldsymbol{X}_j), \quad \text{constant } c > 0 \tag{D.1}$$

$$k(\boldsymbol{X}_i, \boldsymbol{X}_j) = k_1(\boldsymbol{X}_i, \boldsymbol{X}_j) + k_2(\boldsymbol{X}_i, \boldsymbol{X}_j) \tag{D.2}$$

$$k(\boldsymbol{X}_i, \boldsymbol{X}_j) = q\left(k_1(\boldsymbol{X}_i, \boldsymbol{X}_j\right), \quad q(\cdot) \text{ it is a polynomial with non-negative coefficients} \tag{D.3}$$

$$k(\boldsymbol{X}_i, \boldsymbol{X}_j) = \frac{1}{1 + \alpha k_1(\boldsymbol{X}_i, \boldsymbol{X}_j)} \tag{D.4}$$

$$k(\boldsymbol{X}_i, \boldsymbol{X}_j) = \sum_{n=1}^{N} \beta_n k_n(\boldsymbol{X}_i, \boldsymbol{X}_j), \quad \text{constant } \beta > 0 \tag{D.5}$$