



Original Article

Automation of hyperfine interaction parameters analysis using machine learning techniques

Cristian Willian Campos da Silva^{*}, Artur Wilson Carbonari^{ID}, Larissa Otubo^{ID}

Instituto de Pesquisas Energéticas e Nucleares, Universidade de São Paulo (IPEN-USP), Av. Prof. Lineu Prestes, 2242, 05508-000, São Paulo, SP, Brazil

ARTICLE INFO

Keywords:

Hyperfine interactions
Machine learning
Software
Analysis

ABSTRACT

The time-differential perturbed angular correlation (TDPAC) spectra often exhibit complex interactions between multiple hyperfine fields, making it difficult to isolate individual parameter contributions and accurately model the system. This study introduces a Python-based software designed to efficiently process files containing hyperfine parameters, which are essential for analyzing results obtained from fitting TDPAC spectra. The primary objective is to investigate hyperfine interactions in a variety of materials. The software organizes hyperfine parameters measured at different temperatures according to their respective sites and thermal regimes, enabling the automatic separation of behaviors associated with heating and cooling cycles. During the exploratory data analysis, unsupervised machine learning techniques, such as cluster analysis and Principal Component Analysis (PCA), are employed. These techniques facilitate the identification of cluster formation as a function of temperature, enabling the examination of correlations among parameters and the resulting groupings. In addition, the software implements a rule-based decision tree to automatically assign the most appropriate correlation coefficient to each pair of variables, taking into account the statistical characteristics of the dataset. This methodology represents a comprehensive and adaptable approach for the statistical analysis of TDPAC data.

1. Introduction

Time Differential Perturbed Angular Correlation (TDPAC) spectroscopy is a highly sensitive nuclear technique widely used for characterizing the local structural and magnetic properties of materials [1,2]. Its high resolution makes it especially valuable for detecting subtle variations in these properties. A key advantage of TDPAC is its ability to investigate temperature-dependent behaviors due to its non-resonant nature, offering flexibility for in situ studies under different conditions [3,4]. Typically, this method incorporates a radioactive probe, which decays via a gamma cascade such as ^{111}In (^{111}Cd), ^{181}Hf (^{181}Ta) or ^{140}La (^{140}Ce) into the material to be characterized [5–8], using diffusion or implantation processes [9,10] for the investigation of hyperfine interactions between the probe nucleus and the charge distribution in its surroundings. The coupling between nuclear moments and local electromagnetic fields from the electronic vicinity gives rise to these interactions. These fields can either be due to the electric field gradient (EFG) or magnetic in nature, indicated by the magnetic hyperfine field (B_{hf}) [11]. The electric interaction is described by the electric quadrupole Hamiltonian:

$$H_Q = \frac{eQV_{zz}}{4I(2I-1)} \left(3I_z^2 - I^2 + \eta(I_x^2 - I_y^2) \right) \quad (1)$$

where eQ represents the nuclear electric quadrupole moment, V_{zz} is the largest component of the diagonalized electric field gradient tensor (V_{kk} , with $k = x, y, z$), η is the asymmetry parameter, and I is the nuclear angular momentum. On the other hand, the magnetic interaction is described by the hyperfine Zeeman Hamiltonian, as shown in Eq. (2):

$$H_M = -\vec{\mu}_I \cdot \vec{B}_{\text{hf}} = -g_I \mu_N \vec{I} \cdot \vec{B}_{\text{hf}} \quad (2)$$

where $\vec{\mu}_I$ is the nuclear magnetic moment, $\vec{B}_{\text{hf}}$ corresponds to the hyperfine magnetic field vector, g_I is the nuclear gyromagnetic factor, and μ_N represents the nuclear magneton.

Time Differential Perturbed Angular Correlation (TDPAC) spectroscopy enables the quantification of hyperfine interactions by analyzing the angular correlation between two gamma rays emitted in cascade by a radioactive probe nucleus and measured in coincidence by distinct detectors in the same plane. A fundamental characteristic of this technique is the alignment of spin moments in the intermediate state, after the measurement of the first gamma ray (γ_1) by one detector, resulting in an anisotropic emission pattern by the second gamma ray measured by another detector. In this context, if a hyperfine interaction occurs between the charge and/or spin distribution surrounding the probe nucleus in the intermediate state, characterized by the nuclear

^{*} Corresponding author.

E-mail address: cristianwill@usp.br (C.W.C.d. Silva).

angular momentum I , the emission pattern of the second gamma ray (γ_2) will be perturbed. This perturbation in the emission pattern results in a time modulation of the emission probability function, which can provide valuable information about the local properties of the material [3,12]. The function that describes this probability is given:

$$W(\theta, t) = \sum_k A_{kk} G_{kk}(t) P_k(\cos \theta), \quad (3)$$

where θ is the angle between the two emitted gamma rays, A_{kk} are the angular anisotropy coefficients, dependent on the characteristics of the nucleus, $G_{kk}(t)$ is the time-dependent perturbation function, which describes the hyperfine interactions, and $P_k(\cos \theta)$ are the associated Legendre polynomials.

In the experimental context, samples intended for perturbed angular correlation spectroscopy are analyzed in a gamma-ray spectrometer equipped with scintillator crystal detectors coupled to high-voltage powered photomultipliers. These detectors capture energy and time information from the γ_1 and γ_2 radiation, whose signals are processed and stored for subsequent analysis.

In usual TDPAC measurements performed in a spectrometer with four- or six-detectors, twelve or thirty coincidence spectra between the γ_1 and γ_2 energies are generated, respectively, corresponding to the possible combinations between the detectors. To determine the associated parameters, a TDPAC software is used, which employs non-linear regression algorithms to sum the histograms obtained from the angular correlation measurements and generate the perturbation spectrum [10, 12].

The formalism employed in this process is described as:

$$C(\theta, t)_{\text{exp}} = n_0 \exp\left(-\frac{t}{\tau_n}\right) W(\theta, t) + Bg \quad (4)$$

where Bg represents the accidental coincidence background, τ_n is the lifetime of the probe in the intermediate state, and $W(\theta, t)$ is the angular correlation function, defined by Eq. (3). The true coincidences exhibit an exponential behavior, while the accidental coincidences form a linear baseline. To isolate the true coincidences, the background Bg is subtracted:

$$C(\theta, t) = C(\theta, t)_{\text{exp}} - Bg \quad (5)$$

The value of Bg is obtained as the average of the counts outside the relevant interval. The experimental time-dependent anisotropy ratio function is expressed as:

$$R(t) = A_{22} \cdot G_{22} = 2 \frac{\overline{W(180^\circ, t)} - \overline{W(90^\circ, t)}}{\overline{W(180^\circ, t)} + 2\overline{W(90^\circ, t)}} \quad (6)$$

where $\overline{W(180^\circ, t)}$ and $\overline{W(90^\circ, t)}$ represent the geometric means of the true coincidences between pairs of detectors at angles of 180° and 90° , calculated using the following equations:

$$\overline{W(180^\circ, t)} = \sqrt[4]{\prod_{i=1}^4 C(180^\circ, t)} \quad (7)$$

$$\overline{W(90^\circ, t)} = \sqrt[8]{\frac{1}{8} \prod_{i=1}^8 C(90^\circ, t)} \quad (8)$$

After obtaining the perturbation spectrum in the TDPAC software, the fit is performed using the *PACFIT* software, which employs Fourier transforms to model the probability density of γ_2 emission [13]. For the fit, a model that considers the fractional population (f_i) of probe nuclei occupying a certain site i is used [5,14]. This model is expressed by the following equation:

$$R(t) = A_{22} G_{22}(t) = A_{22} \sum_i f_i G_{22}^i(t) \quad (9)$$

where A_{22} is the unperturbed angular correlation coefficient, and the perturbation factor $G_{22}(t)$ provides specific information about the hyperfine interactions. The information on magnetic hyperfine interactions, obtained based on the perturbation factor, can be represented

by:

$$G_{22}(t) = 0.2 + 0.4 \cos(\omega_L t) + 0.4 \cos(2\omega_L t) \quad (10)$$

where $\omega_L = \mu_N g B_{\text{hf}} / \hbar$ is the Larmor frequency ($\nu_m = \omega_L / 2\pi$), μ_N is the nuclear magneton, and g is the nuclear g -factor. Therefore, the hyperfine magnetic field (B_{hf}) can be determined from the Larmor frequency extracted from the $G_{22}(t)$ measurements [5]. Meanwhile, electric quadrupolar interactions are modeled according to the following equation:

$$G_{22}(t) = S_{20} + \sum_{n=1}^3 S_{2n}(\eta) \cos(\omega_n t) \quad (11)$$

The transition frequencies ω_n are related to the quadrupolar frequency $\nu_Q = (eQV_{zz})/\hbar$ by the relation $\omega_n = g_n(\eta)\nu_Q$. The parameters $S_{2n}(\eta)$ and $g_n(\eta)$, as well as the transition frequencies ω_n , are expressed in terms of the asymmetry parameter η , which is defined as

$$\eta = \frac{V_{xx} - V_{yy}}{V_{zz}} \quad (12)$$

This parameter depends on the diagonal components of the EFG tensor (V_{kk}), considering that $|V_{xx}| \leq |V_{yy}| \leq |V_{zz}|$. For static interactions, both electric quadrupolar and magnetic dipolar, the width of the frequency distribution is given by a Lorentzian broadening, parameterized by the exponential factor $\delta\omega$ [11].

Analyzing TDPAC spectra involves fitting theoretical models to measured data. This process can be challenging due to the presence of multiple interacting hyperfine fields, resulting in ambiguities that hinder the accurate extraction of parameters from the data. These ambiguities can complicate the interpretation of the results. These parameters include: the quadrupolar frequency (ν_Q), the Larmor frequency (ν_M), the asymmetry parameter (η), the frequency distribution (δ), and the fraction of occupied sites (f) [5,14]. The fit performed with *PACFIT* generates a file in the .par format, containing information corresponding to the hyperfine parameters for each considered site. The characterization of materials as a function of temperature, especially with the aim of investigating phase transitions (whether structural or magnetic), requires the analysis of the correlations between these hyperfine parameters. In general, this approach allows exploring the influence of these parameters on the observed changes. However, the grouping and organization of multiple files containing hyperfine information, carried out in the conventional way, demand significant time and hinder the feasibility of this type of analysis [15,16].

In this scenario, recent advances in machine learning techniques emerge as a promising alternative, enabling the rapid, automated, and efficient processing of large collections of hyperfine parameter files. These approaches support concise and systematic exploratory analyses, allowing the identification of trends, correlations, and regime changes that may not be readily apparent through traditional methods [17,18]. As a result, the integration of spectral analysis with machine learning has emerged as a powerful framework for extracting physically meaningful information from complex spectroscopic datasets, surpassing the limitations of purely visual or manually guided analyses. For example, Han et al. applied machine learning techniques combined with laser-induced breakdown spectroscopy (LIBS) for the automated analysis of soils, employing clustering methods (K-means) and principal component analysis (PCA) to identify spectral patterns and subtle variations in elemental distribution, demonstrating higher sensitivity to regime change detection compared to conventional spectroscopic approaches [19]. Similarly, Wang et al. investigated metallic alloys using spectroscopic techniques combined with machine learning algorithms, showing that multivariate exploratory analysis approaches (such as principal component analysis (PCA) and classification methods) enable the identification of latent patterns and distinct regimes in spectroscopic data dependent on multiple parameters, thereby facilitating a more robust physical interpretation of the results [20]. These studies highlight that the integration of spectroscopy and machine learning

constitutes a powerful strategy for transforming large experimental datasets into physically interpretable information.

Within the scope of time-differential perturbed angular correlation (TDPAC) spectroscopy, however, studies reporting the systematic use of machine learning algorithms for data processing and interpretation remain scarce. Existing approaches are generally fragmented and primarily focus on supervised learning models applied to individual spectra or on the use of machine learning techniques as support for density functional theory (DFT) calculations, aiming to assist in the prediction and interpretation of hyperfine parameters. Only a limited number of studies address the joint analysis of experimental hyperfine parameter series obtained from multiple spectra [21,22].

These limitations motivate the development of automated analysis methodologies specifically designed for the organization, correlation, and interpretation of TDPAC results, both as a means of optimizing data treatment and as a strategy to deepen the physical interpretation of the results. In particular, unsupervised machine learning techniques, which are sensitive to subtle parameter variations (such as clustering methods), are especially suitable for processing TDPAC data under different thermal regimes, where phase transitions may induce fluctuations in hyperfine parameters [23]. Since these methods are non-predictive and do not rely on a priori models, they allow the automatic identification of regimes and distinct features directly from the analyzed dataset.

In this work, we present the development of a Python software designed for the automated processing and analysis of multiple files containing temperature-dependent hyperfine parameters. The program integrates data wrangling methodologies for organizing and separating parameters according to the hyperfine site and thermal regime (heating and cooling), combined with advanced data analysis techniques such as decision trees, cluster analysis, principal component analysis (PCA), and gradient boosting. This approach enables an efficient and systematic investigation of correlations among hyperfine parameters across different thermal regimes, as well as the identification of structural and magnetic transitions through clustering. As a case study, the software is applied to TDPAC spectroscopy data of the ^{111}Cd probe in $\text{Ni/Ni}_3\text{C}$ nanoparticles as a function of temperature, demonstrating the capability of the methodology to automatically detect structural and magnetic transitions.

2. Methodology

For the development of our research project, `.par` files were used as input data. These files are obtained from the fitting of experimental TDPAC spectra, performed according to the formalism of the experimental anisotropy function $R(t)$ (Eq. (9)), which is written as a weighted sum of contributions associated with distinct hyperfine sites. The fitted hyperfine parameters, encompassing electric quadrupolar and magnetic interactions (e.g., v_Q , v_M , η , δ , f), are stored in the corresponding `.par` files, whereas the graphical representation of the fit is provided in `.fit` files.

The `.par` file structure is organized into blocks, each beginning with a key enclosed in square brackets. During the fitting procedure, hyperfine parameters are recorded within these blocks in a key-value format, where the key identifies the parameter and the value stores the associated quantitative variable. The data-processing software was designed around this file format because its organization is conceptually equivalent to a dictionary-like structure (see Table 1), enabling straightforward parsing, automated extraction, and systematic handling of the dataset associated with hyperfine interactions.

Considering the analysis of multiple `.par` files acquired at different temperatures, it is important to emphasize that the structure of this file type does not contain information regarding the experimental conditions associated with each spectrum, such as the acquisition temperature or measurement atmosphere. Given the relevance of this information for identifying structural and magnetic transitions associated with distinct thermal regimes (heating or cooling), the strategy

Table 1

Block structure of the `.par` file.

Key	Value
[General]	Measurement parameters and number of sites
[Site.n]	v_Q , v_M , η , δ , f

adopted in this work consisted of assigning the thermal conditions directly from the file names.

From this perspective, the files were grouped into a single operating system directory and named according to the measurement temperature (for example, 25.PAR for a spectrum acquired at 25 °C), which facilitates traceability, identification, and the systematic analysis of the temperature dependence of the hyperfine parameters. In addition, the direction of the thermal cycle is inferred from the presence of a suffix of the type “V” (for example, 25V), which characterizes measurements performed under the cooling regime (*cooling*); in the absence of this marker, the file is automatically classified as belonging to the heating regime (*heating*). Based on these characteristics, the software was designed to import and process these files using data analysis techniques aimed at the organization, investigation, and interpretation of the parameters as a function of temperature. Among the employed approaches are data wrangling, gradient boosting, decision trees, correlation analyses using the Pearson, Spearman, and Kendall coefficients, as well as clustering techniques and principal component analysis (PCA). These aspects are described in greater detail in the subsequent sections. Fig. 1 shows a schematic representation of the software workflow.

2.1. Data wrangling

The data wrangling process involves transforming a raw database from its original structure into a format that enables the appropriate extraction of information. In the context of obtaining hyperfine parameters from multiple sites, the files were initially imported into the Python (Spyder environment) using the `tkinter` library. Subsequently, a function was defined to utilize the keys in the `.par` file to extract the values associated with each site and, subsequently, the corresponding hyperfine parameters. In this way, each input file was structured as a dataframe using the `pandas` package, organized by site and their respective hyperfine parameters.

A classification variable was introduced to determine the validity of a given site based on the value of the frequency distribution (δ). In the context of material characterization using Time-Differential Perturbed Angular Correlation (TDPAC) spectroscopy, hyperfine interactions serve as unique identifiers of the crystal symmetries surrounding the probe. Consequently, the associated frequencies must be well defined to ensure accurate characterization. However, the diffusion or implantation of the probe into the crystalline environment under analysis is influenced by several parameters, including temperature, heating rate, and interstitial radius [24]. This process often leads to the formation of sites exhibiting distributed or poorly defined frequencies, which are typically associated with probes positioned in low-symmetry regions, resulting in a high value of δ .

The parameters associated with sites exhibiting high δ values have minimal influence on the determination of local properties. Consequently, in this analysis, only sites with a frequency distribution (δ) below 20% were considered. As a result, for each input file, all valid sites are identified and selected for further evaluation. In addition, as an additional criterion, the magnetic hyperfine field is calculated from the Larmor frequency corresponding to each site. The obtained values are subsequently incorporated into the dataset according to the formulation presented in Eq. (2).

After this step, the multiple dataframes are first grouped and separated by site, resulting in the creation of site-specific dataframes. Subsequently, these dataframes are organized according to the thermal

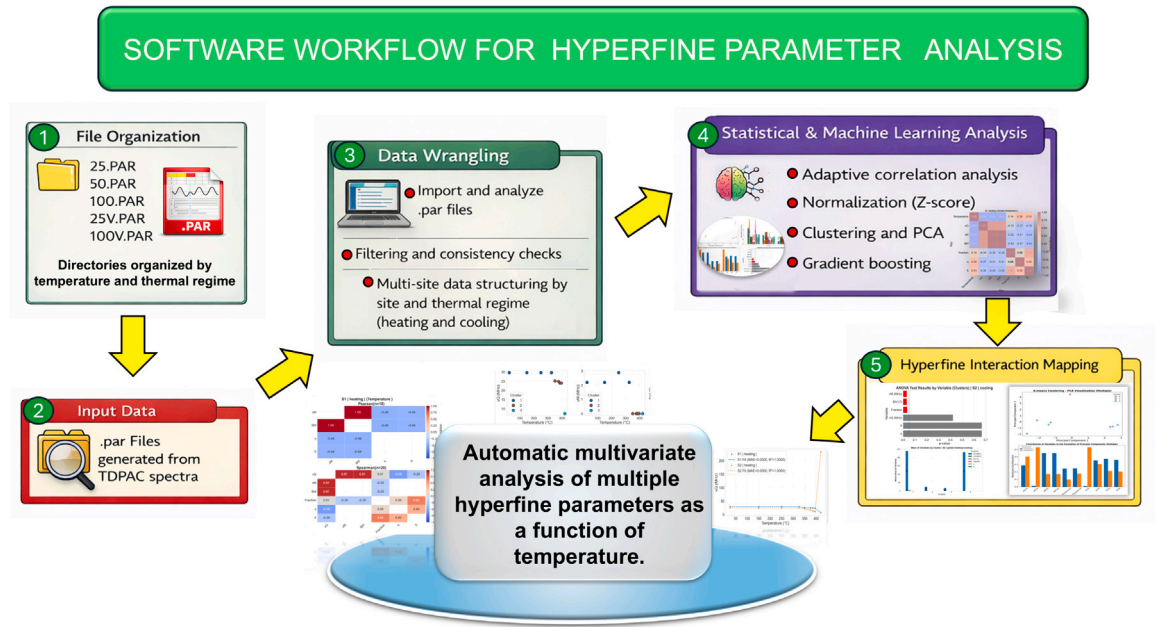


Fig. 1. Software workflow for automated multivariate analysis of hyperfine parameters as a function of temperature, including file organization, data wrangling, and statistical and machine-learning-based analysis.

Table 2

Configuration of the site-specific DataFrames generated by the software, including temperature and thermal regime (heating or cooling). The index n identifies the hyperfine site and ranges from 1 to 5.

Temperature ($^{\circ}\text{C}$)	Thermal regime	Hyperfine parameter columns indexed by Site _{n} (n from 1 to 5)
T_1	heating or cooling	$\nu_{Qn}, \nu_{Mn}, B_{hfn}, \eta_n, \delta_n, f_n$
T_2	heating or cooling	$\nu_{Qn}, \nu_{Mn}, B_{hfn}, \eta_n, \delta_n, f_n$
T_3	heating or cooling	$\nu_{Qn}, \nu_{Mn}, B_{hfn}, \eta_n, \delta_n, f_n$
T_n	heating or cooling	$\nu_{Qn}, \nu_{Mn}, B_{hfn}, \eta_n, \delta_n, f_n$

regime (heating or cooling), as inferred from the file nomenclature. In this perspective, the software was designed to import files containing up to five sites. The temperature column is generated based on the name of the original .par file. Thus, it becomes possible to automatically process multiple files and describe each site individually as a function of temperature and thermal regime. Table 2 presents an example of the configuration of these dataframes.

2.2. Adaptive correlation analysis of hyperfine parameters

Based on the generated DataFrames, the software evaluates the correlation of each hyperfine parameter as a function of temperature as an exploratory measure of the dataset, considering each thermal regime (heating and cooling) separately. In this approach, temperature is explicitly treated as a statistical variable in the correlation calculation, allowing the quantification of how each hyperfine parameter evolves with thermal variation. Therefore, the resulting correlations describe the individual thermal behavior of the hyperfine parameters, providing direct insight into their temperature-dependent evolution.

Considering the possibility of different datasets as input, no fixed correlation type (such as Pearson, Spearman, or Kendall) was imposed *a priori*. Instead, an adaptive approach was adopted, since relationships between variables may exhibit linear or nonlinear behavior. In this context, the correlation type is determined through a rule-based decision tree for adaptive selection of the correlation coefficient. This type of algorithm is particularly relevant because it enables the automatic choice of the most appropriate method according to the statistical characteristics of each dataset [25,26].

For each pair of variables, an initial filtering step is performed to exclude pairs of observations containing missing (null) values, after which the number of valid paired observations, defined as N , is quantified. Correlations are computed only when the number of valid pairs is greater than or equal to four ($N \geq 4$). For small samples ($4 \leq N \leq 5$), Kendall's coefficient (τ), a rank-based measure, is employed due to its robustness for limited datasets [27]. For numbers of pairs greater than five ($N > 5$), outliers are evaluated using the Z-score; when $|Z_i| \geq 3$, Spearman's correlation (ρ) is selected because of its reduced sensitivity to extreme values and its suitability for monotonic relationships. In the absence of outliers, both Pearson's correlation coefficient (r), which quantifies linear associations, and Spearman's coefficient (ρ) are calculated. When both coefficients exhibit the same sign and the difference between their absolute values satisfies $\|r| - |\rho|\| \leq 0.15$, the relationship is considered approximately linear and Pearson's correlation is adopted. Otherwise, Spearman's correlation is selected, indicating monotonic or nonlinear behavior. The threshold $\|r| - |\rho|\| \leq 0.15$ was empirically adopted as a heuristic criterion to indicate approximate linearity. This strategy enables the automatic selection of the most appropriate correlation coefficient for each pair of variables, ensuring a statistically consistent analysis of the hyperfine parameters [27,28].

Based on the correlation type selected for each site and thermal regime, three full correlation heatmaps (Pearson, Spearman, and Kendall) are generated, in which all computable correlation coefficients are shown after pairwise filtering to remove missing or non-finite values. The coefficients selected by the rule-based decision tree are highlighted with a green outline, enabling the direct identification of the variable pairs for which a given correlation metric was deemed the most appropriate (the mathematical expressions of the correlation coefficients are provided in the Supplementary Material).

In addition to these full matrices, the software also produces compact heatmaps containing only the correlations selected by the decision tree. In this representation, the correlation matrix is masked such that non-selected entries are omitted, and the plot is automatically cropped to retain only the rows and columns associated with at least one selected pair. This approach provides a concise and focused visualization of the statistically retained relationships, thereby facilitating comparisons across different thermal regimes and between sites. Fig. 2 presents an illustrative example of this procedure.

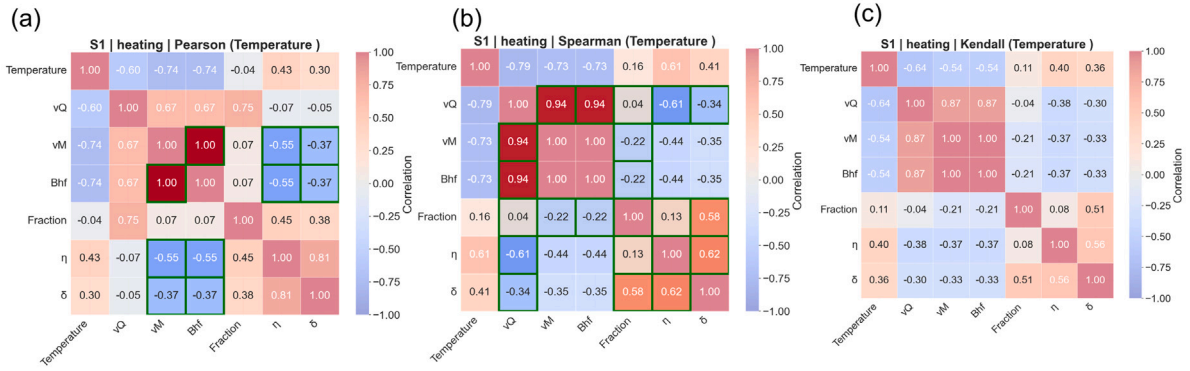


Fig. 2. Illustrative example of adaptive correlation heatmaps for a generic site (S1) under the heating regime: (a) Pearson correlation matrix, (b) Spearman correlation matrix, and (c) Kendall correlation matrix. Green contours indicate the correlation coefficients selected by the rule-based decision tree as the most appropriate for each pair of variables.

2.3. Gradient boosting, clusterization, and PCA.

To model and analyze the relationship between the hyperfine parameters and temperature, the software employs the Gradient-BoostingRegressor algorithm from the `scikit-learn` library. In order to ensure a systematic evaluation, six related physical parameters are considered: quadrupole frequency (ν_Q), magnetic frequency (ν_M), magnetic hyperfine field (B_{hf}), fraction (f), asymmetry parameter (η), and frequency distribution (δ). Temperature is defined as the independent variable (x), since it directly influences the physical phenomena under investigation.

The modeling procedure is performed independently for each hyperfine parameter and for each site identified in the previously consolidated *dataframes*. Furthermore, when more than one thermal regime is present, the data are separated according to the thermal regime (heating or cooling). For each distinct combination of site, hyperfine parameter, and thermal regime, the dependent variables (y) are extracted from the dataset and associated with specific labels, which are used to identify the corresponding physical quantities in the generated plots.

The Gradient Boosting model is based on an ensemble of decision trees trained sequentially, in which each new estimator is fitted to the residuals of the previous ensemble in order to minimize the overall prediction error. This approach is widely employed for the nonlinear modeling of experimental data [29]. After training, the fitted values are predicted only at the experimentally measured temperatures and compared with the observed values. The quality of the fit is evaluated by calculating the mean absolute error (MAE) and the coefficient of determination (R^2), which quantify the discrepancy between experimental and model-estimated values. These metrics provide a measure of the reliability of the fit for each analyzed variable, site, and thermal regime [17,18].

In the investigation of possible structural and magnetic changes across different thermal regimes, the software also incorporates unsupervised machine learning techniques, specifically cluster analysis and principal component analysis (PCA), in order to explore latent patterns and internal structures within multivariate datasets [23].

To ensure comparability among variables with different numerical scales, the data are previously standardized using the Z-score method implemented through the `StandardScaler` function of the `scikit-learn` package. This procedure transforms each variable to have zero mean and unit standard deviation prior to the clustering and PCA steps. The Z-score standardization of an observation X_{ji} is defined as

$$Z_{X_{ji}} = \frac{X_{ji} - \bar{X}_j}{s_j}, \quad (13)$$

where $Z_{X_{ji}}$ denotes the standardized value of the observation X_{ji} , X_{ji} is the value of the j th variable in the i th observation, $\bar{X}_j$ represents the

mean of the j th variable, and s_j is its standard deviation. This transformation ensures that all variables contribute in a balanced manner to scale-sensitive multivariate analyses.

After standardization, the software performs cluster analysis in order to organize the data into homogeneous groups (clusters), such that observations belonging to the same cluster exhibit higher similarity to each other than to those assigned to different clusters [17,23,30].

Within the implementation framework, the data are analyzed using two complementary clustering approaches. First, Agglomerative Hierarchical Clustering (HAC) is applied using Euclidean distance and the Ward linkage criterion (minimum variance method), which minimizes the increase in intra-cluster variance at each merging step, thereby favoring the formation of compact and physically interpretable clusters. In parallel, the partition-based *K-means* method is employed in an exploratory manner, mainly as a supporting tool for assessing the appropriate number of clusters [23].

The selection of the optimal number of clusters is supported by internal validation metrics, including the elbow method, the silhouette index, the Calinski-Harabasz index, and the Davies-Bouldin index, which provide quantitative criteria to evaluate partition quality and guide the choice of more consistent configurations [31,32]. Each metric plays a complementary role: the elbow method evaluates the reduction of intra-cluster variance as a function of k ; the silhouette index simultaneously quantifies cohesion and separation; the Calinski-Harabasz index favors solutions with high inter-cluster separation and low internal dispersion; whereas the Davies-Bouldin index penalizes configurations characterized by overlapping clusters. Taken together, these criteria provide a robust quantitative basis for selecting the most appropriate number of clusters.

In addition, cluster stability is evaluated using the Adjusted Rand Index under a *leave-one-out* scheme (ARI_{LOO}), which quantifies the agreement between the clustering obtained from the complete dataset and the solutions resulting from the successive removal of individual data points. High ARI_{LOO} values indicate that the identified clusters are robust against small perturbations of the dataset, thereby reinforcing the reliability of the extracted structures.

Beyond cluster identification, the software provides additional quantitative information that enables the interpretation of the factors responsible for cluster separation. For this purpose, the mean values and standard deviations of the variables within each cluster are computed, allowing direct comparison among different groups. Furthermore, a one-way analysis of variance (ANOVA) is performed for each variable to evaluate whether the differences observed between clusters are statistically significant. Variables presenting p -values below the adopted significance level (5%) are considered relevant contributors to cluster discrimination. Complementarily, the software generates graphical representations of the variable distributions as a function of

cluster membership and temperature, enabling the simultaneous assessment of group separation and the thermal evolution of the hyperfine parameters [33].

This set of procedures not only allows the identification of the existence of clusters but also provides insight into which physical parameters are determinant for cluster formation and how these parameters evolve across different thermal regimes, thereby strengthening the physicochemical interpretation of the results.

Finally, in order to reduce data dimensionality and facilitate the visualization of clustering results, the software applies Principal Component Analysis (PCA). PCA projects the data onto a lower-dimensional space defined by orthogonal linear combinations of the original variables, while preserving the largest possible fraction of the total variance. This step enables the interpretation of clusters in terms of the main axes of variation of the dataset [17,18].

2.4. Case study

As a case study, .par files obtained from the fitting of TDPAC spectra using the PACFIT program were analyzed. These files originate from measurements employing ^{111}In (^{111}Cd) probe nuclei incorporated into hierarchical Ni/Ni₃C nanoparticles, consisting of a metallic Ni core coated by a nickel carbide shell (Ni₃C), investigated as a function of temperature under a nitrogen atmosphere and two distinct thermal regimes: heating and cooling.

For the heating regime, spectra acquired at 25, 150, 250, 325, 350, 375, 385, 395, and 420 °C were considered. For the cooling regime, .par files obtained from the fitting of spectra measured at 25, -73, -196, and -228 °C were analyzed, corresponding to measurements performed after completion of the heating cycle. All these files were jointly imported into the developed software, which automatically organized and structured the data into *dataframes*, according to the methodology described previously. The experimental methodology employed for data acquisition was developed at the Hyperfine Interactions Laboratory (LHI), located at the Research Reactor Center of IPEN-CNEN/SP.

In addition, aspects related to structural phase transitions in these systems have been previously discussed in the literature [10]. In this context, the present work focuses on the processing, treatment, and statistical analysis of the extracted hyperfine parameters, with emphasis on the application of the developed software, aiming to demonstrate that the statistical analysis of multiple hyperfine parameters constitutes a robust tool for the interpretation of local physical changes, going beyond simple data visualization. Fig. 3 schematically illustrates the data acquisition and processing workflow for the generation and analysis of TDPAC files.

3. Results

Fig. 4 presents the plots obtained from the analysis of the .par files for the heating regime, which originate from the fitting of experimental TDPAC spectra. The fits were performed assuming a three-site model based on Eq. (9), and the corresponding Fast Fourier Transforms (FFT) were subsequently obtained. These fits are in agreement with experimental results (see Supplementary Material).

Site 1 (S1) is characterized by the coexistence of quadrupolar (ν_{Q1}), with a frequency of approximately 29 MHz, and magnetic (ν_{M1}) interactions, with a frequency around 2.25 MHz. This site is attributed to probe located in carbon-deficient Ni₃C_{1-x} environments, likely in regions close to the metallic Ni core. Site 2 (S2) exhibits exclusively a quadrupolar interaction (ν_{Q2}), with an approximate quadrupolar frequency of 24 MHz, and is assigned to probes incorporated into stoichiometric Ni₃C regions. Site 3 is broadly distributed, presenting a large distribution parameter ($\delta > 20\%$), and is attributed to probes situated in locally low-symmetry environments, typical of structurally disordered regions frequently observed in PAC spectroscopy [10,34].

It should be emphasized that the values discussed correspond to the parameters determined at the initial temperature of 25 °C. The remaining spectra systematically reveal the thermal evolution and progressive modifications of these parameters throughout the heating regime, as shown in Fig. 4(a-f). In Fig. 4(g-i), the gradient boosting results are presented, describing the evolution of the hyperfine parameters as a function of temperature under the heating regime, considering sites 1 (S1) and 2 (S2). The results indicate that the hyperfine parameters begin to exhibit significant changes at approximately 325 °C. In particular, the magnetic frequency associated with S1 disappears above ~350 °C, consistent with the loss of magnetic ordering in Ni upon reaching its Curie temperature. Beyond this thermal regime, more pronounced variations in the quadrupolar frequencies are observed, suggesting a local structural reorganization. This behavior is likely associated with the progressive migration of carbon atoms toward surface regions, ultimately leading to the formation of a new crystalline environment attributed to the onset of graphitization [10].

In this context, a systematic investigation of the correlations between temperature and the hyperfine parameters enables a global analysis of these transformations for each site. Fig. 5 presents the heatmaps corresponding to the correlation matrices for S1 and S2. The correlation analyses clearly demonstrate that the two sites correspond to physically distinct hyperfine environments, exhibiting significantly different thermal responses and parameter-coupling patterns. For S1, strong negative correlations are observed between temperature and the principal quantities associated with the hyperfine interactions (ν_Q , ν_M , and B_{hf}), accompanied by a moderate positive correlation with the asymmetry parameter η (Fig. 5a). This behavior indicates a systematic weakening of the magnetic hyperfine interaction upon heating, concomitant with an increase in local distortion. Furthermore, ν_Q and ν_M , exhibit very strong positive correlations among themselves (which is expected), revealing a tightly coupled block of common physical origin, whereas δ and Fraction show positive correlation, suggesting an association between local disorder and site population (Fig. 5b).

In contrast, for S2, temperature does not correlate directly and consistently with the quadrupolar frequency (Fig. 5c); instead, stronger correlations are observed with Fraction, η , and δ , along with a strong coupling between ν_Q and Fraction (Fig. 5d). These results indicate that the primary thermal response of this site is associated with increasing structural disorder and population redistribution rather than with a direct variation in the strength of the hyperfine interaction. Additionally, the differences observed between Pearson and Spearman coefficients for S2 point to predominantly non-linear relationships, whereas in S1 both coefficients agree in sign and magnitude, characterizing a nearly monotonic linear behavior. Overall, these findings reinforce that S1 corresponds to an environment more sensitive to thermally induced magnetic collapse, while the thermal evolution of S2 is predominantly governed by occupation effects and local structural disorder.

The changes in the hyperfine environments become evident from the hierarchical clustering analysis (Ward method) performed for the heating regime, which reveals contrasting behaviors for sites S1 and S2. For S1, the clustering quality assessment ($k = 3$, $N = 9$) indicates an overall consistent and stable partition (Fig. 6a). The silhouette coefficient (0.467) suggests a moderate separation among groups, indicating that the clusters are reasonably well defined, although not fully isolated. The Calinski-Harabasz index (8.962) points to a favorable ratio between inter-cluster dispersion and intra-cluster compactness, whereas the Davies-Bouldin index (0.557), lower than 1, indicates adequate compactness and separability. In addition, the high stability obtained from leave-one-out validation ($\text{ARI}_{\text{LOO}} = 0.975$) demonstrates that the solution is highly robust against small variations in the dataset. Therefore, $k = 3$ is justified as the smallest number of clusters that simultaneously satisfies the stability criterion ($\text{ARI}_{\text{LOO}} \geq 0.7$) and yields a silhouette value close to the best observed result, leading to a statistically consistent and physically interpretable segmentation for S1 [35,36].

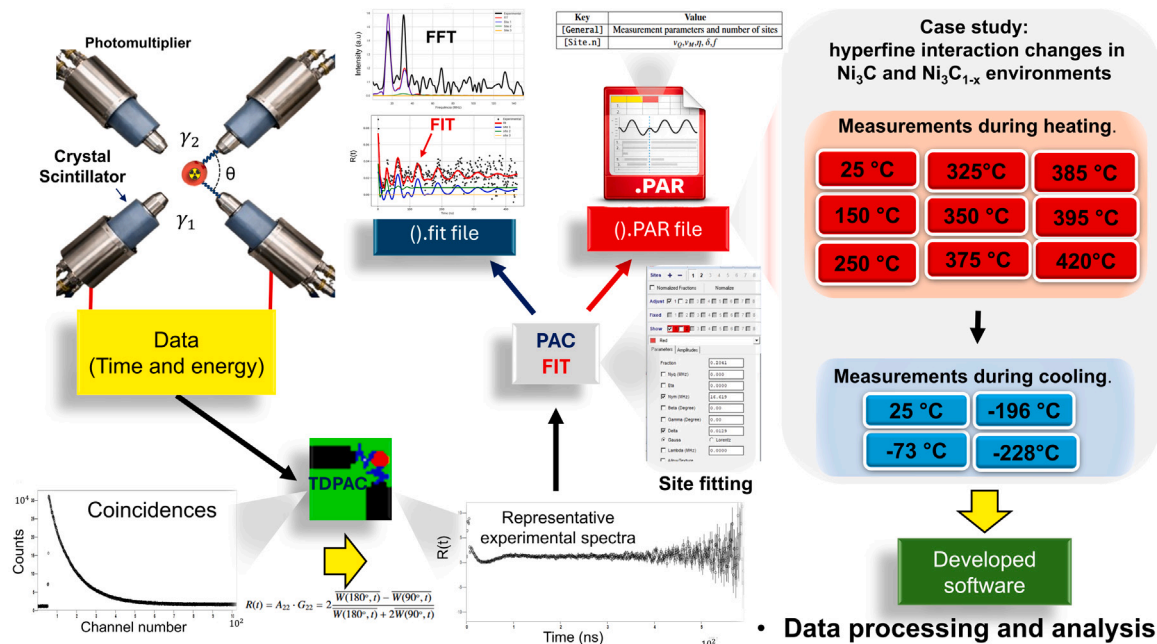


Fig. 3. Schematic representation of the data acquisition and TDPAC spectrum processing, together with the representation of the case study. Coincidence events recorded by the detector array are processed using the TDPAC software to generate time-differential perturbation spectra, which are subsequently fitted using the PACFIT software to obtain the hyperfine interaction parameters and generate the .fit and .par files. The .par files related to the case study, involving hyperfine interaction changes in Ni_3C and Ni_3C_{1-x} environments under heating and cooling regimes, were then analyzed using the developed software.

The ANOVA test (Fig. 6b) together with the mean values of the variables by cluster (Fig. 6c) indicate which parameters contribute significantly to cluster separation and how their cluster-averaged values evolve with temperature. The results show that the quadrupolar frequency (ν_Q), which presents the lowest p -value ($p \ll 0.001$) and the largest F -statistic, decreases systematically from cluster 1 to cluster 3, highlighting its dominant role in distinguishing the groups. Similarly, the magnetic frequency (ν_M) and the hyperfine magnetic field (B_{hf}) also exhibit statistical significance and display high mean values exclusively in the low-temperature cluster, vanishing in the clusters associated with higher temperatures. These analyses confirm that the clustering clearly discriminates between magnetic and non-magnetic regimes, being primarily governed by three statistically significant parameters: ν_Q , ν_M , and B_{hf} . The fractional contribution also decreases with increasing cluster index, indicating that population redistribution contributes to the differentiation during the heating regime. In contrast, the asymmetry parameter (η) and the frequency distribution (δ) exhibit very similar mean values across clusters, consistent with their non-significant p -values in the ANOVA. Altogether, these results demonstrate that S1 clustering is primarily driven by variations in the electric and magnetic frequencies, as well as by population changes, yielding a physically consistent interpretation of the thermal evolution of the system.

In contrast, for S2 the hierarchical clustering under the heating regime reveals a more abrupt separation between thermal regimes (Fig. 6d). The clustering quality assessment ($k = 3$, $N = 9$) indicates a highly consistent and stable partition, with a silhouette coefficient of 0.574, a Calinski–Harabasz index of 15.480, and a Davies–Bouldin index of 0.384, all pointing to well-separated and compact clusters. The high stability obtained from leave-one-out validation ($ARI_{LOO} = 0.975$) further demonstrates the robustness of the solution. These quality metrics are systematically better than those obtained for S1, indicating a more sharply defined clustering structure for S2 [35,36].

The ANOVA results for S2 identify the quadrupolar frequency (ν_Q) as the dominant variable controlling cluster formation, presenting an extremely low p -value ($p \ll 0.001$) and a very large F -statistic. The fractional contribution (Fraction) is also highly significant, followed

by the frequency distribution parameter (δ), whereas the asymmetry parameter (η) does not show statistical significance (Fig. 6e). Consistently, the mean values by cluster reveal pronounced variations in quadrupolar frequency and Fraction across the clusters, while the remaining parameters display comparatively smaller changes (Fig. 6f). These results indicate that the clustering of S2 is primarily driven by strong modifications of the local electric field gradient and by population redistribution.

These results are illustrated in Fig. 7, where the PCA maps for sites S1 and S2 are presented together with the distribution of variables within the clusters as a function of temperature. The projection of the data onto the space defined by the first two principal components further supports this interpretation, as the experimental points organize into three well-defined groups associated with distinct thermal regimes and fully consistent with the clusters obtained through the hierarchical Ward method (Fig. 7b and Fig. 7d).

Overall, based on the multivariate analysis, the results reveal three distinct configurations for both sites throughout the thermal evolution, as represented in Fig. 7e: (i) the initial configuration of the nanoparticles; (ii) the magnetic transition in S1 concomitant with the onset of carbon migration in S2, beginning at approximately 350 °C; and (iii) a high-temperature stage around 420 °C. In this latter stage, a decrease in the quadrupolar frequency is observed for S1, whereas a pronounced increase in the quadrupolar frequency occurs for S2 (~230 MHz). This behavior is consistent with the structural phase transition previously reported for Ni_3C nanostructures, culminating in the formation of metallic Ni surrounded by a graphitic shell [10,34].

After this stage, the thermally induced modifications were assessed by analyzing the spectra acquired in the cooling regime. Measurements performed at 30, -73, -196, and -228 °C were considered, corresponding to the experiments carried out after completion of the heating cycle. The results obtained at 30 °C after thermal treatment confirm the transitions previously identified by the clustering analysis in the heating regime. The spectral fit indicates the presence of two hyperfine sites. Site 1 is characterized by a purely magnetic interaction, with a magnetic frequency of approximately 16 MHz and a magnetic hyperfine field of 6.85 T, which are typical values for the ^{111}Cd probe in a

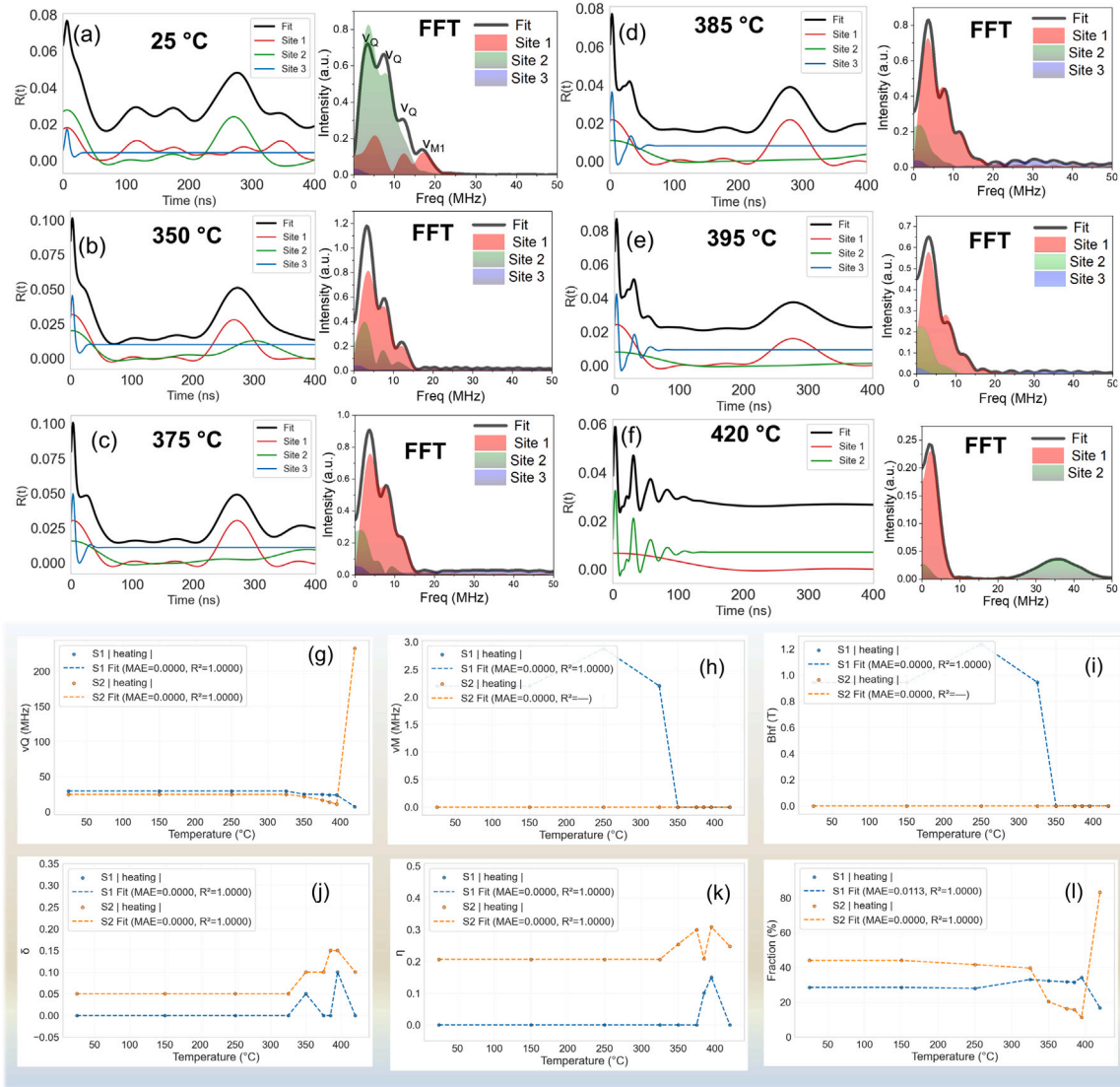


Fig. 4. Evolution of hyperfine parameters as a function of temperature for the heating regime. Panels (a–f) show the plots obtained from the .par files and their corresponding Fast Fourier Transforms (FFT) for Sites 1 (S1), 2 (S2), and 3 (S3). Panels (g–i) present the temperature evolution of hyperfine parameters extracted from the fits for crystalline sites ($\delta < 20\%$), processed by the gradient boosting model, describing the temperature dependence of selected hyperfine configurations for S1 and S2.

substitutional Ni-fcc environment. Site 2 exhibits a high quadrupole frequency on the order of 238 MHz, combined with a magnetic frequency of about 3.2 MHz, and is assigned to probes located in the interfacial region between Ni-fcc and graphite, as shown in Fig. 8a and Fig. 8b.

The results in Fig. 8(c–h) show the evolution of the hyperfine parameters associated with sites 1 and 2 throughout the cooling regime. The most pronounced variations are observed in the relative fractions of the two sites, a trend that becomes even clearer when considering the Kendall correlation matrices shown in Fig. 8(i–l). Note that the Kendall coefficient was selected according to the characteristics of the dataset, following the criteria established by the rule-based decision tree.

For site 1, the ν_M and B_{hf} plots (Fig. 8(d,e)) reveal a progressive increase of these parameters as the temperature decreases, reflecting the strengthening of magnetic order. The correlation matrices further indicate a positive correlation between ν_M and B_{hf} , as well as negative correlations of these parameters with temperature, confirming that the enhancement of magnetism at low temperatures governs their coupled evolution (Fig. 8i). Consequently, the magnetic modulation of the perturbation spectrum becomes more pronounced, increasing the spectral weight of this interaction in the fit. Consistently, the relative fraction

of site 1 (Fig. 8f) shows a systematic increase at low temperatures, which should be interpreted as a larger spectral dominance of the magnetically ordered component.

For site 2, the correlation matrix reveals a strong anticorrelation between temperature and the quadrupole frequency ν_Q , along with positive correlations among ν_Q , ν_M , and B_{hf} . The fraction also exhibits a positive correlation with temperature (Fig. 8j), indicating that the spectral contributions from both environments respond in a complementary manner to the strengthening of the magnetic interaction at low temperatures. The agreement between the statistical correlations and the fits obtained via Gradient Boosting supports that the observed changes are associated with the intensification of magnetic order and the progressive stabilization of the Ni-fcc phase and the Ni-fcc/graphite interface at low temperatures.

The hierarchical clustering results obtained using Ward's method for sites 1 and 2 in the cooling regime are presented in Fig. 9(a,b), and reveal the formation of two main clusters ($k = 2$), although with distinct quality levels. For site 1, the choice of $k = 2$ is supported by a silhouette coefficient of 0.540, indicating moderate separation between groups, together with a Calinski–Harabasz index of 5.93 and

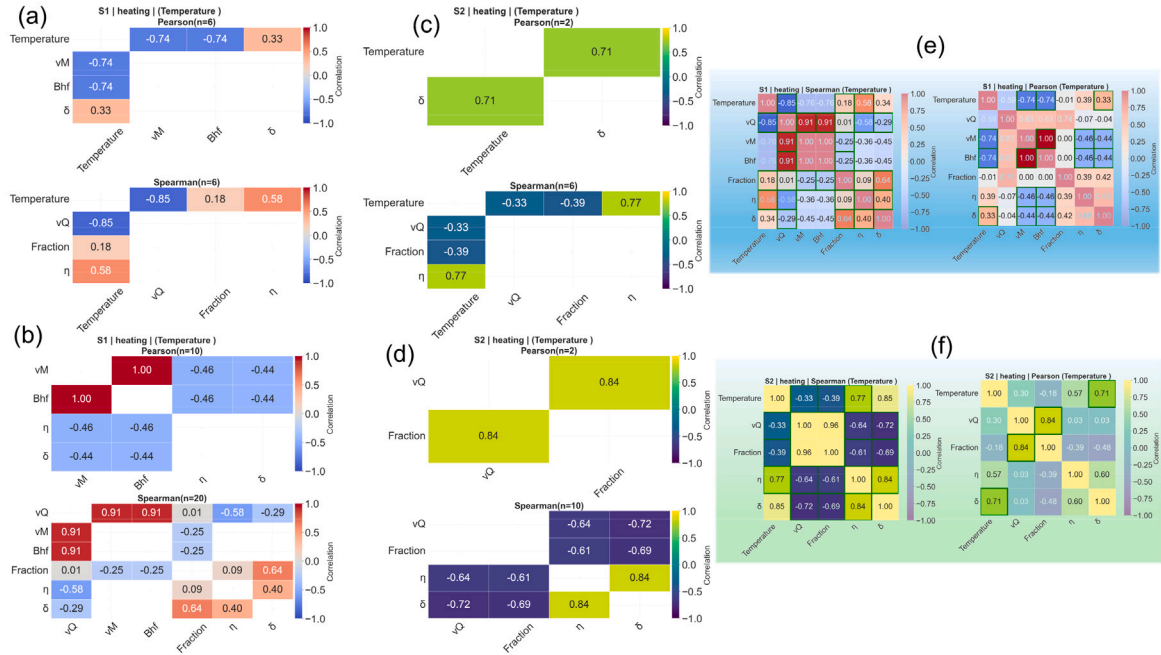


Fig. 5. Heatmaps of correlation matrices for Sites 1 (S1) and 2 (S2) under the heating regime. Panels (a,b) show the correlations between temperature and hyperfine parameters for S1 obtained using Pearson and Spearman coefficients, respectively, while panels (c,d) present the corresponding results for S2. Panels (e,f) display the full correlation matrices for S1 and S2, respectively, comparing Pearson and Spearman coefficients. In all heatmaps, n denotes the number of masks applied for a given specific correlation.

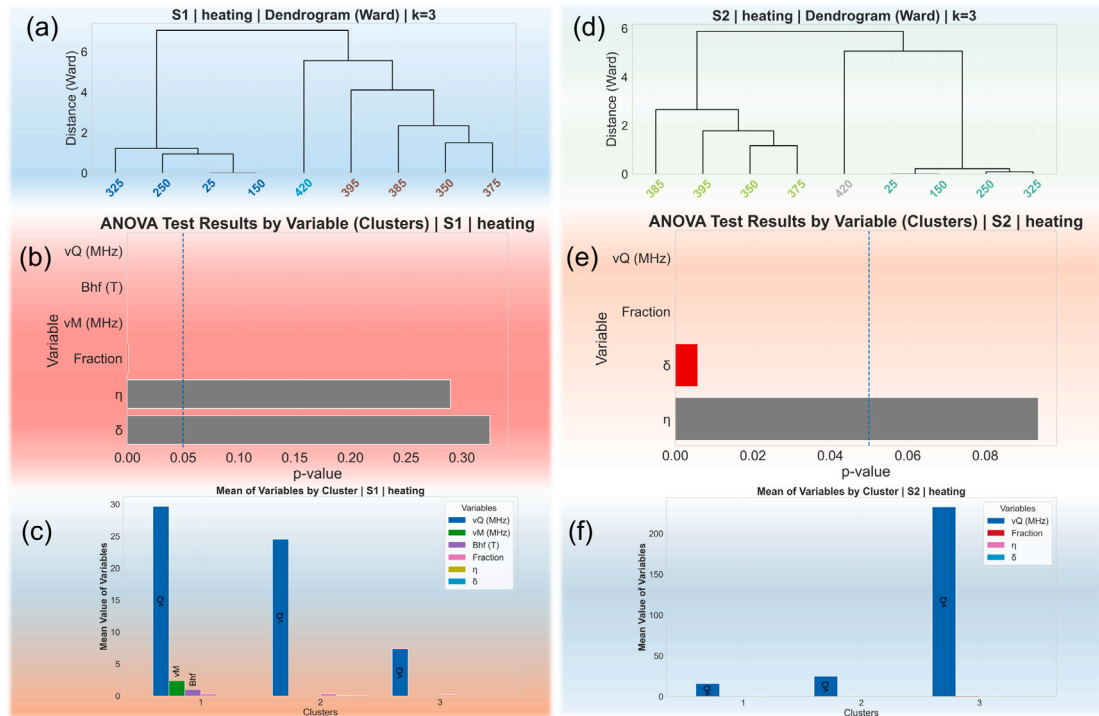


Fig. 6. Hierarchical clustering and statistical analysis of hyperfine parameters for sites S1 and S2 under the heating regime. (a,d) Dendrograms obtained using Ward's method with $k = 3$ clusters for S1 and S2, respectively. (b,e) One-way ANOVA test results by variable for each site, showing the corresponding p -values and the statistical significance of the hyperfine parameters in cluster separation. The dashed vertical line indicates the 5% significance level. (c,f) Mean values of the hyperfine parameters for each cluster in S1 and S2, respectively, illustrating how cluster-averaged quantities evolve throughout the heating regime.

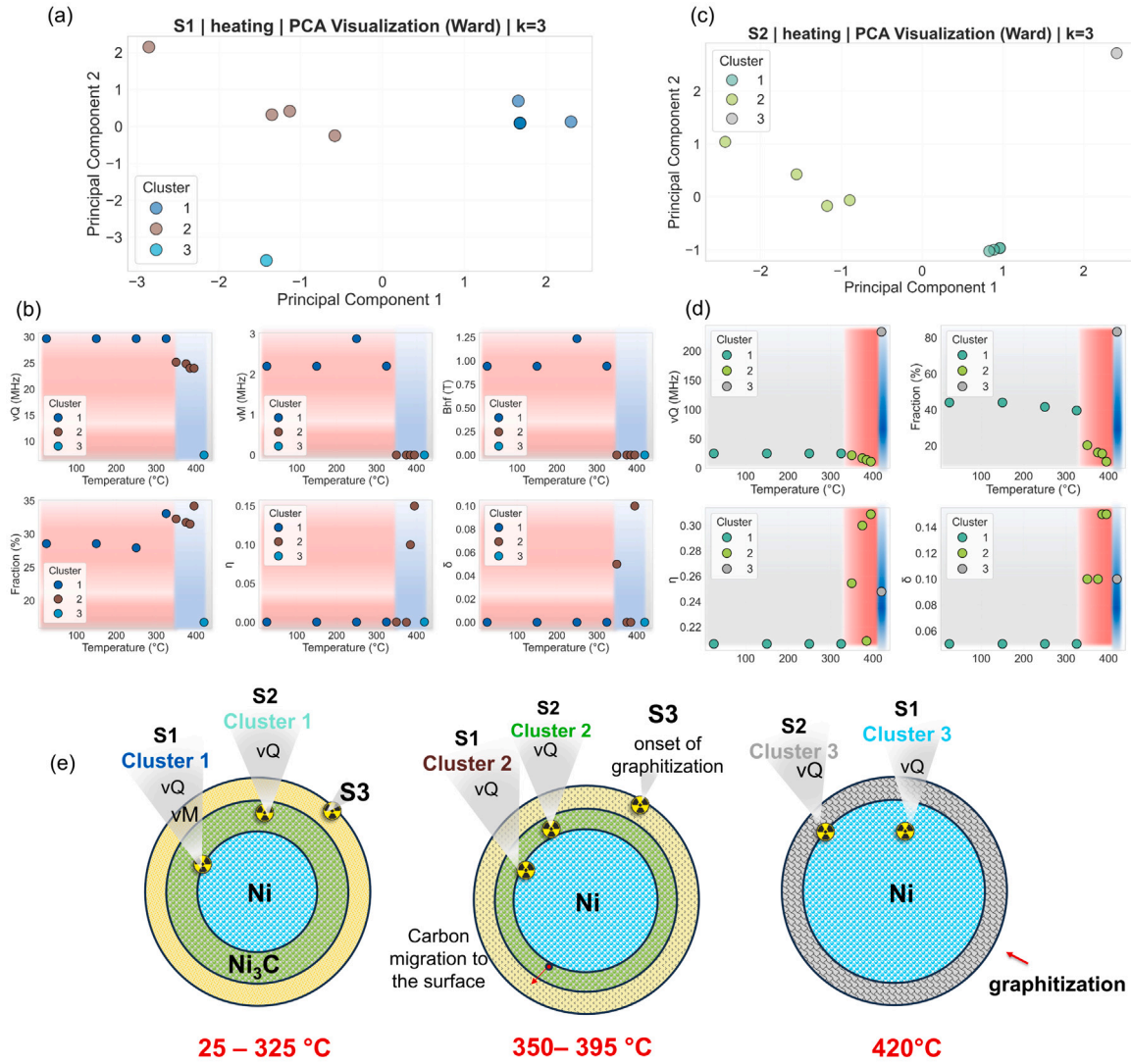


Fig. 7. PCA visualization and physical interpretation of clustering results for S1 and S2 under the heating regime. (a,c) PCA maps obtained using the first two principal components, with points colored according to Ward clusters ($k = 3$). (b,d) Hyperfine parameters as a function of temperature, colored by cluster. (e) Schematic representation of the three thermally induced configurations.

a Davies–Bouldin index of 0.44, consistent with relatively compact and well-defined clusters. Owing to the limited number of experimental points ($N = 4$), stability via leave-one-out cross-validation (ARI_{LOO}) could not be assessed; nevertheless, the selection of k was based on the criterion of choosing the smallest value whose silhouette remained within 0.03 of the best result, ensuring a parsimonious and statistically consistent partition.

For site 2, although $k = 2$ was also selected, the quality indicators point to weaker separation between groups. The silhouette coefficient is 0.197, while the Calinski–Harabasz (2.55) and Davies–Bouldin (0.43) indices indicate less compact clusters with greater overlap. Similarly to site 1, ARI_{LOO} stability was not evaluated, and the value of k was defined using the same silhouette-based criterion.

Structural analysis of the dendrograms shows that, for site 1, there is a clear separation between measurements performed at lower temperatures (−196 and −228 °C) and those obtained at intermediate temperatures (30 and −73 °C), suggesting a transition associated with the progressive strengthening of the magnetic interaction (Fig. 9a). Analysis of variance (ANOVA) indicates that the variable most sensitive to cluster formation is the relative fraction, followed by the magnetic

parameters v_M and B_{hf} , whereas the frequency distribution parameter δ exhibits low discriminative power (Fig. 9c). Cluster-wise mean values confirm a systematic increase in v_M and B_{hf} within the group associated with lower temperatures, accompanied by an increase in the fraction of site 1, indicating greater spectral dominance of the magnetically ordered component as temperature decreases (Fig. 9c).

For site 2, the dendrogram reveals two main clusters with a more gradual hierarchical structure, reflecting a continuous evolution of the hyperfine parameters (Fig. 9d). ANOVA shows that the quadrupole frequency v_Q is the variable with the highest discriminative power between clusters, whereas v_M , B_{hf} , and the fraction provide smaller relative contributions. The parameters η and δ exhibit low sensitivity to group separation (Fig. 9e). Cluster-wise mean values indicate that v_Q remains high in both regimes, displaying only subtle variations, while v_M and B_{hf} show a slight increase in the cluster associated with lower temperatures, suggesting coupling between electric and magnetic interactions in the Ni-fcc/graphite interfacial environment (Fig. 9f).

These results are directly confirmed by the Principal Component Analysis (PCA) projections and by the distribution of clusters as a function of temperature, presented in Fig. 10. For site 1 in the cooling

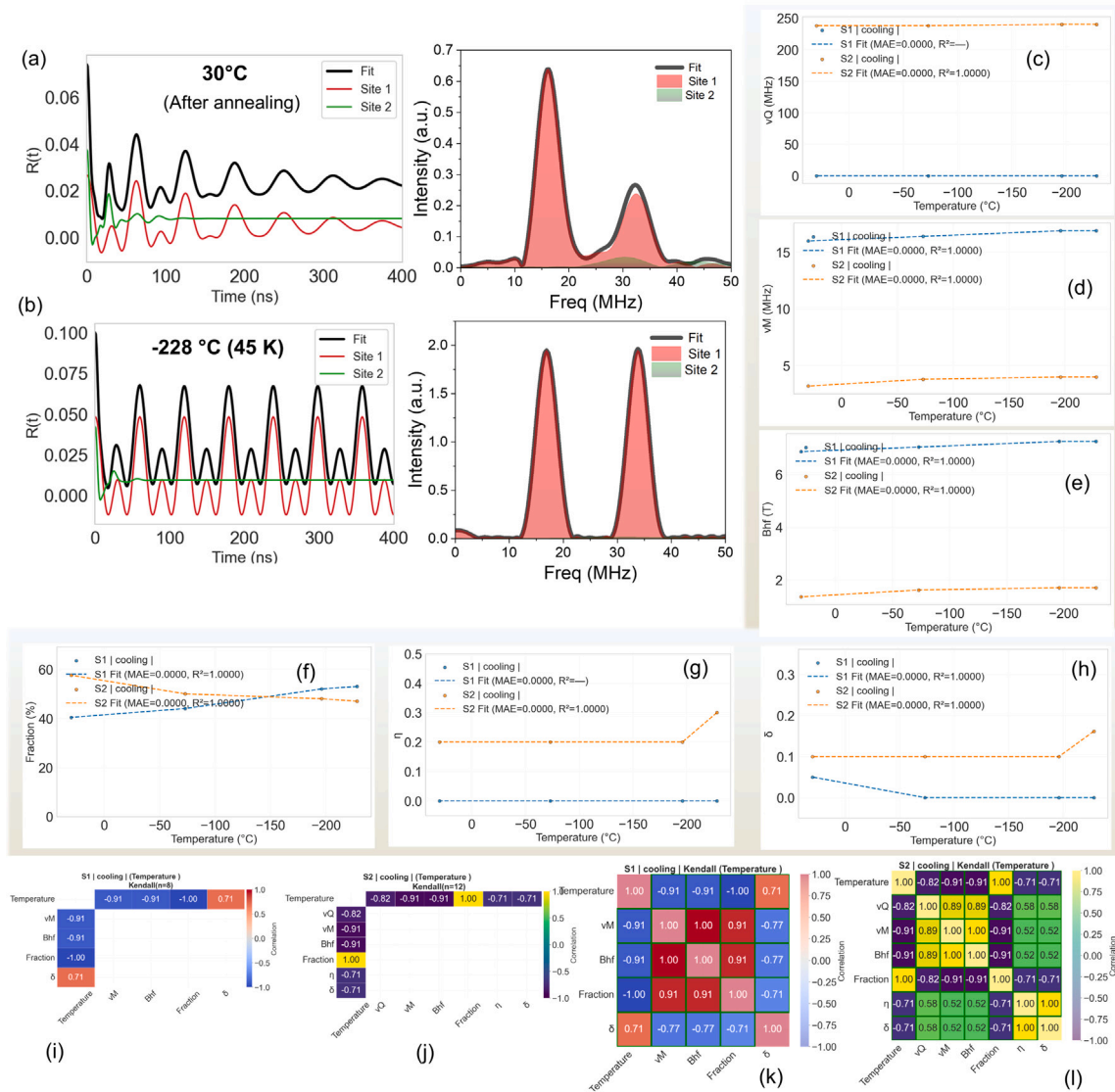


Fig. 8. Cooling-regime evolution of hyperfine parameters for sites S1 and S2. Panels (a,b) show perturbation functions and corresponding FFT spectra at 30 °C and -228 °C. Panels (c-h) present the temperature dependence of selected hyperfine parameters obtained from the fits and processed by the Gradient Boosting model. Panels (i-l) display the Kendall correlation matrices, showing correlations between temperature and hyperfine parameters during cooling.

regime (Fig. 10(a,b)), the experimental points are organized into two well-defined groups in the space of the first two principal components, in excellent agreement with the Ward partition ($k = 2$). For site 2 (Fig. 10(c,d)), PCA also reveals the formation of two clusters, although with greater proximity between points, reflecting the lower separation quality indicated by the clustering metrics. The distribution of clusters as a function of temperature further illustrates these trends and reinforces their physical interpretation.

Overall, the consistency among PCA, hierarchical clustering, correlation analyses, and Gradient Boosting fits demonstrates that the identified clusters effectively reflect distinct physical regimes in the cooling stage. Specifically, the evolution of site 1 is predominantly governed by the strengthening of magnetic order at low temperatures, whereas site 2 is mainly controlled by the stability of the electric field gradient, with secondary magnetic contributions. These results fully corroborate the previous descriptions and reinforce the robustness of the proposed model for describing the hyperfine environments present in the system.

3.1. Conclusion

This study described the development of Python-based software designed to process and analyze multiple hyperfine parameter files as a function of temperature. By integrating both supervised and unsupervised machine learning techniques, the program enabled the identification of correlations among parameters and the recognition of potential clusters, contributing to a deeper understanding of the temperature dependence of these parameters in different thermal regimes (heating and cooling). As a case study, hyperfine interactions in Ni/Ni₃C nanoparticles were analyzed. The results demonstrated that applied machine learning techniques were effective in identifying significant variations, such as phase transitions and fluctuations in the hyperfine parameters. Thus, the proposed methodology is versatile and can be applied to any set of .par files containing models with up to five sites. This approach enhances the analysis of hyperfine parameters, offering a more comprehensive interpretation of their relationship with local structural and electronic variations, as observed through Time-Differential Perturbed Angular Correlation (TDPAC) spectroscopy.

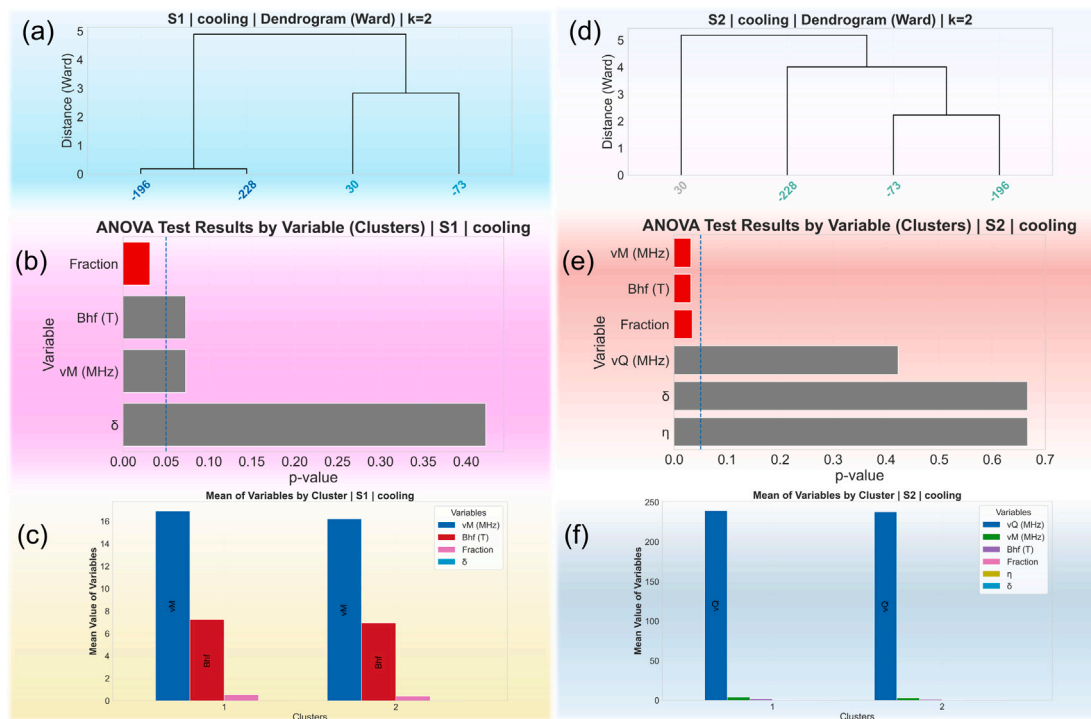


Fig. 9. Hierarchical clustering of hyperfine parameters for sites S1 and S2 under the cooling regime. (a,d) Ward dendrograms with $k = 2$ clusters. (b,e) ANOVA results by variable. (c,f) Mean hyperfine parameters for each cluster, illustrating their evolution with decreasing temperature.

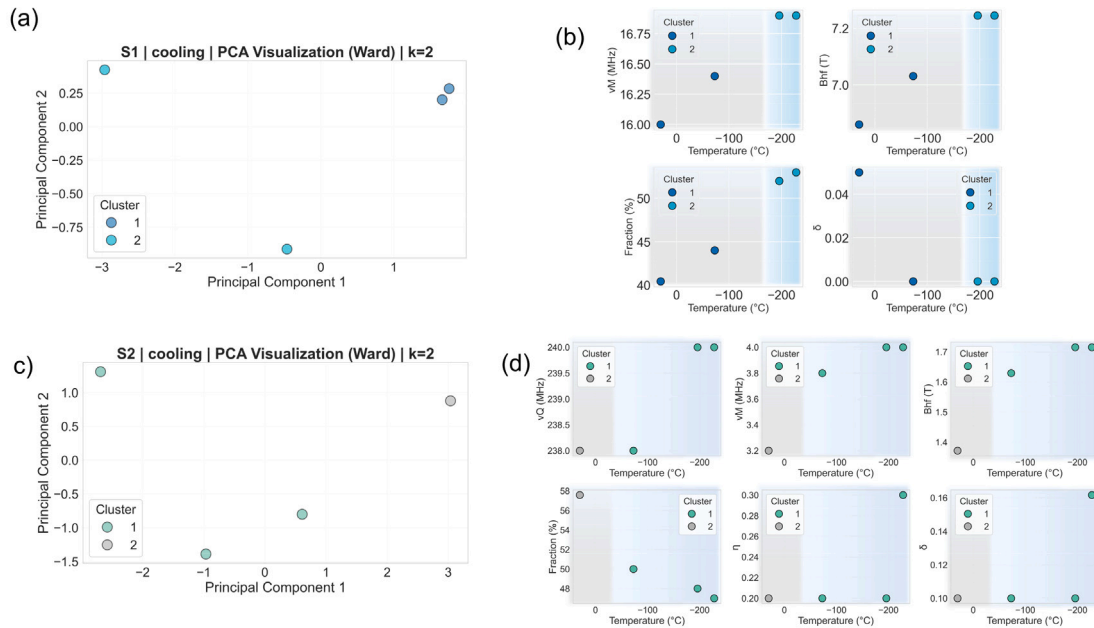


Fig. 10. PCA visualization and clustering for sites S1 and S2 under the cooling regime. (a) PCA projection for S1 obtained using the first two principal components. (b) Distribution of hyperfine parameters of S1 as a function of temperature. (c) Corresponding PCA projection for S2. (d) Distribution of hyperfine parameters of S2 as a function of temperature.

CRedit authorship contribution statement

Crystian Willian Campos da Silva: Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Artur Wilson Carbonari:** Writing – review & editing, Validation, Supervision. **Larissa Otubo:** Writing – original draft, Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Artur Wilson Carbonari reports financial support was provided by Brazilian Nuclear Energy Commission. Artur Wilson Carbonari reports a relationship with National Council for Scientific and Technological Development that includes: funding grants. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was financed in part by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil under grant # 140981/2023-3 and by the Fundação de Apoio à Pesquisa do Estado de São Paulo (FAPESP), Brazil under grant # 2021/06712-9. A.W. Carbonari acknowledges the financial support received from CNPq, Brazil under grant # 307322/2021-1.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.net.2026.104233>.

References

- [1] M. Zacate, H. Jaeger, Perturbed angular correlation spectroscopy – a tool for the study of defects and diffusion at the atomic scale, *Defect Diffus. Forum* 311 (2011) 3–38, <http://dx.doi.org/10.4028/www.scientific.net/DDF.311.3>.
- [2] A.W. Carbonari, J. Mestnik-Filho, R.N. Saxena, Impurities in magnetic materials studied by PAC spectroscopy, *Defect Diffus. Forum* 311 (2011) 39–61, <http://dx.doi.org/10.4028/www.scientific.net/DDF.311.39>.
- [3] A.C. Campos, S.C. Paes, B.S. Correa, G.A. Cabrera-Pasca, M.S. Costa, C.S. Costa, L. Otubo, A.W. Carbonari, Growth of long ZnO nanowires with high density on the ZnO surface for gas sensors, *ACS Appl. Nano Mater.* 3 (2020) 175–185, <http://dx.doi.org/10.1021/acsnanm.9b01888>.
- [4] R.V. Santos, G.A. Cabrera-Pasca, C.S. Costa, B. Bosch-Santos, L. Otubo, L.F.D. Pereira, B.S. Correa, F.B. Effenberger, A. Burimova, R.S. Freitas, A.W. Carbonari, Crystalline and magnetic properties of CoO nanoparticles locally investigated by using radioactive indium tracer, *Sci. Rep.* 11 (2021) 21028, <http://dx.doi.org/10.1038/s41598-021-99810-y>.
- [5] A.W. Carbonari, W. Pendl, R.N. Attili, R.N. Saxena, Magnetic hyperfine fields in the Heusler alloys Co₂YZ (Y=Sc, Ti, Hf, V, Nb; Z=Al, Ga, Si, Ge, Sn), *Hyperfine Interact.* 80 (1993) 971–976, <http://dx.doi.org/10.1007/BF00567449>.
- [6] A.W. Carbonari, J. Mestnik-Filho, R.N. Saxena, Investigation of hyperfine interactions in CeIn₃ by TDPAC, *Hyperfine Interact.* 133 (2001) 77–81, <http://dx.doi.org/10.1023/A:1012205300493>.
- [7] B. Bosch-Santos, A.W. Carbonari, G.A. Cabrera-Pasca, M.S. Costa, R.N. Saxena, Effect of ge substitution for si on the magnetic hyperfine field in LaMn₂Si₂ compound measured by perturbed angular correlation spectroscopy with ¹⁴⁰Ce as probe nuclei, *J. Appl. Phys.* 113 (2013) 17E124, <http://dx.doi.org/10.1063/1.4795419>.
- [8] L.G. Shpinkova, A.W. Carbonari, S.M. Nikitin, J. Mestnik-Filho, Influence of electron capture after-effects on the stability of ¹¹¹In(¹¹¹Cd)-complexes with organic ligands, *Chem. Phys.* 279 (2002) 255–263, [http://dx.doi.org/10.1016/S0301-0104\(02\)00492-5](http://dx.doi.org/10.1016/S0301-0104(02)00492-5).
- [9] A.M. Gerami, J. Heiniger-Schell, E.L. da Silva, M.S. Costa, C.S. Costa, J.G. Monteiro, J.J. Pires, D.R. Pereira, C. Díaz-Guerra, A.W. Carbonari, K. Lorenz, J.G. Correia, Cd implantation in α-MoO₃: An atomic scale study, *Phys. Rev. Mater.* 7 (2023) 033603, <http://dx.doi.org/10.1103/PhysRevMaterials.7.033603>.
- [10] C.W.C. Silva, G.A. Cabrera-Pasca, K.S. Souza, C.S. Costa, M.S. Costa, B.S. Correa, W.L. Ferreira, R.S. Freitas, A.W. Carbonari, L. Otubo, Investigation of Ni/Ni₃C nanoparticle synthesis for application as a catalyst in carbon nanostructure growth, *ACS Appl. Nano Mater.* 6 (2023) 20623–20638, <http://dx.doi.org/10.1021/acsnanm.3c03205>.
- [11] G. Schatz, A. Weidinger, *Nuclear Condensed Matter Physics: Nuclear Methods and Applications*, Wiley, New York, 1996.
- [12] A.V. Tsyvashchenko, L.N. Fomicheva, V.B. Brudanin, O.I. Kochetov, A.V. Salamatina, A. Velichkov, M. Wiertel, M. Budzynski, A.A. Sorokin, G.K. Rysany, B.A. Komissarova, The TDPAC study of the hyperfine interactions at ¹¹¹Cd nuclei in Al₂ compounds synthesized under high pressure, *Solid State Commun.* 142 (2007) 664–669, <http://dx.doi.org/10.1016/j.ssc.2007.02.027>.
- [13] T.T. Dang, J. Schell, A.G. Boa, D. Lewin, G. Marschick, A. Dubey, M. Escobar-Castillo, C. Noll, R. Beck, D.V. Zybkin, K. Glukhov, I.C.J. Yap, A.M. Gerami, D.C. Lupascu, Temperature dependence of the local electromagnetic field at the Fe site in multiferroic bismuth ferrite, *Phys. Rev. B* 106 (2022) 054416, <http://dx.doi.org/10.1103/PhysRevB.106.054416>.
- [14] S.V. Santos, C.S. Costa, W. Paraguassu, C.W.C. Silva, L. Otubo, K.S. Souza, B.S. Correa, A.A. Miranda-Filho, W.L. Ferreira, A.W. Carbonari, G.A. Cabrera-Pasca, Synthesis and local characterization of CoO nanoparticles in distinct phases: Unveiling polymorphic structures, *ACS Omega* 9 (2024) 42883–42894, <http://dx.doi.org/10.1021/acsomega.4c05308>.
- [15] P. Karande, B. Gallagher, T.Y.-J. Han, A strategic approach to machine learning for material science: How to tackle real-world challenges and avoid pitfalls, *Chem. Mater.* 34 (17) (2022) 7650–7665, <http://dx.doi.org/10.1021/acs.chemmater.2c01333>.
- [16] S. Lee, C. Chen, G. Garcia, A. Oliynyk, Machine learning descriptors in materials chemistry used in multiple experimentally validated studies: Oliynyk elemental property dataset, *Data Brief* 53 (2024) 110178, <http://dx.doi.org/10.1016/j.dib.2024.110178>.
- [17] L.P. Fávero, P. Belfiore, *Data Science, Analytics and Machine Learning with R*, first ed., Academic Press, London, 2022.
- [18] M. Collins, *Machine Learning: An Introduction to Supervised and Unsupervised Learning Algorithms*, CreateSpace Independent Publishing Platform, Charleston, SC, 2017.
- [19] B. Han, W. Gao, J. Feng, A. Iroshan, J. Yang, G. Chen, Y. Zhang, N. Aizezi, Y. Liu, Laser-induced breakdown spectroscopy for imaging and distribution analysis of heavy metal elements in soil, *J. Hazard. Mater.* 496 (2025) 139284, <http://dx.doi.org/10.1016/j.jhazmat.2025.139284>.
- [20] Z. Wang, N. Aizezi, Y. Ye, Y. Chen, J. Li, B. Yu, W. Sun, Y. Liu, Enhanced classification of aluminium alloys via time-frequency dual-domain acoustic feature fusion with laser-induced breakdown spectroscopy, *Talanta* 300 (2026) 129208, <http://dx.doi.org/10.1016/j.talanta.2025.129208>.
- [21] K. Choudhary, J.N. Ansari, I.I. Mazin, K.L. Sauer, Density functional theory-based electric field gradient database, *Sci. Data* 7 (2020) 362, <http://dx.doi.org/10.1038/s41597-020-00707-8>.
- [22] M.O. Zacate, W.E. Evenson, Fluctuating hyperfine interactions: an updated computational implementation, *Hyperfine Interact.* 231 (2015) 143–155, <http://dx.doi.org/10.1007/s10751-014-1083-y>.
- [23] T. Shuku, K.-K. Phoon, Y. Yokota, K. Date, I1 trend clustering: Trend estimation, clustering, and change detection for spatial-temporal ground displacement data from satellite remote sensing, *Geod. AI* 3 (2025) 100020, <http://dx.doi.org/10.1016/j.geoai.2025.100020>.
- [24] G.N. Darriba, E.L. Muñoz, A.W. Carbonari, M. Rentería, Experimental TDPAC and theoretical DFT study of structural, electronic, and hyperfine properties in (¹¹¹In → ¹¹¹Cd)-doped SnO₂ semiconductor: *Ab Initio* modeling of the electron-capture-decay after-effects phenomenon, *J. Phys. Chem. C* 122 (2018) 17423–17436, <http://dx.doi.org/10.1021/acs.jpcc.8b03724>.
- [25] Z. Hui, Z. Nie, H. Xiao, Adaptive model fusion algorithm for decision trees and association rules, in: *Proceedings of the 2023 2nd Asia Conference on Algorithms, Computing and Machine Learning*, ACM, New York, NY, USA, 2023, pp. 13–18, <http://dx.doi.org/10.1145/3590003.3590006>.
- [26] S. Lee, K. Choi, D. Yoo, Building a core rule-based decision tree to explain the causes of insolvency in small and medium-sized enterprises more easily, *Humanit. Soc. Sci. Commun.* 10 (2023) 933, <http://dx.doi.org/10.1057/s41599-023-02382-7>.
- [27] E.F. El-Hashash, R.H.A. Shiekh, A comparison of the pearson, spearman rank and Kendall tau correlation coefficients using quantitative variables, *Asian J. Probab. Stat.* (2022) 36–48, <http://dx.doi.org/10.9734/ajpas/2022/v20i3425>.
- [28] P. Schober, C. Boer, L.A. Schwarte, Correlation coefficients: Appropriate use and interpretation, *Anesth. Analg.* 126 (2018) 1763–1768, <http://dx.doi.org/10.1213/ANE.0000000000002864>.
- [29] K.T. Butler, D.W. Davies, H. Cartwright, O. Isayev, A. Walsh, Machine learning for molecular and materials science, *Nature* 559 (2018) 547–555, <http://dx.doi.org/10.1038/s41586-018-0337-2>.

- [30] H.E.G. Lopes, M.S. Gosling, Cluster analysis in practice: Dealing with outliers in managerial research, *Rev. Adm. Contemp.* 25 (2021) <http://dx.doi.org/10.1590/1982-7849rac2021200081>.
- [31] J.M. Contreras, E.M. Portillo, J.M.F. Luna, Evaluation of hierarchical clustering methodologies for identifying patterns in timeout requests in EuroLeague basketball, *Mathematics* 13 (2025) 2414, <http://dx.doi.org/10.3390/math13152414>.
- [32] N. Pakgohar, A. Lengyel, Z. Botta-Dukát, Quantitative evaluation of internal clustering validation indices using binary datasets, 2023, <http://dx.doi.org/10.1101/2023.08.09.552566>.
- [33] Z. Kanetaki, C. Stergiou, G. Bekas, C. Troussas, C. Sgouropoulou, Analysis of engineering student data in online higher education during the COVID-19 pandemic, *Int. J. Eng. Pedagog. (IJEP)* 11 (2021) 27–49, <http://dx.doi.org/10.3991/ijep.v11i6.23259>.
- [34] Y. Zhu, Y. Zhang, K. Wang, D. Li, F. Ding, D. Meng, X. Wang, Z. Zhang, One-pot synthesis of Ni/Ni₃C/Ni₃N nanocomposite: Structure, growth mechanism and magnetic properties, *Mater. Res. Bull.* 95 (2017) 79–85, <http://dx.doi.org/10.1016/j.materresbull.2017.07.001>.
- [35] L. Hubert, P. Arabie, Comparing partitions, *J. Classification* 2 (1985) 193–218, <http://dx.doi.org/10.1007/BF01908075>.
- [36] A.K. Jain, Data clustering: 50 years beyond K-means, *Pattern Recognit. Lett.* 31 (2010) 651–666, <http://dx.doi.org/10.1016/j.patrec.2009.09.011>.