



Investigating the effect of non-resonant background variation on the CARS data analysis of bacteria samples and classification using machine learning

RAJENDHAR JUNJURI,^{1,2}  TOBIAS MEYER-ZEDLER,^{1,2}
JÜRGEN POPP,^{1,2}  AND THOMAS BOCKLITZ^{1,2,*} 

¹Leibniz Institute of Photonic Technology, Member of Leibniz Health Technologies, Member of the Leibniz Centre for Photonics in Infection Research (LPI), Albert-Einstein-Strasse 9, 07745 Jena, Germany

²Institute of Physical Chemistry (IPC) and Abbe Center of Photonics (ACP), Friedrich Schiller University Jena, Member of the Leibniz Centre for Photonics in Infection Research (LPI), Helmholtzweg 4, 07743 Jena, Germany

*Thomas.bocklitz@uni-jena.de

Abstract: Non-resonant background (NRB) plays a significant role in coherent anti-Stokes Raman scattering (CARS) spectroscopic applications. All the recent works primarily focused on removing the NRB using different deep learning methods, and only one study explored the effect of NRB. Hence, in this work, we systematically investigated the impact of NRB variation on Raman signal retrieval. The NRB is simulated as a linear function with different strengths relative to the resonant Raman signal, and the variance also changes for each NRB strength. The resonant part of nonlinear susceptibility is extracted from real experimental Raman data; hence, the simulated CARS data better approximate the experimental CARS spectra. Then, the corresponding Raman signal is retrieved by four different methods: maximum entropy method (MEM), Kramers-Kronig (KK), convolutional neural network (CNN), and long short-term memory (LSTM) network. Pearson correlation measurements and principal component analysis combined with linear discriminant analysis modeling revealed that MEM and KK methods have an edge over LSTM and CNN for higher NRB strengths. It is also demonstrated that normalizing the input data favors LSTM and CNN predictions. In contrast, background removal from the predictions significantly influenced Pearson correlation but not the classification accuracies for MEM and KK. Further, the LSTM performance is found to be limited and can only be applied for low NRB strengths. This comprehensive study has the potential to impact the CARS spectroscopy and microscopy applications in different areas.

© 2024 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

Coherent Anti-Stokes Raman scattering (CARS) is a third-order nonlinear spectroscopic technique that measures the vibrational response of the molecules [1,2] and provides molecular fingerprint information [3,4]. Its advantages include high-resolution, high-speed (~ 10 ms/pixel) [5,6], and label-free imaging [7], making it a powerful tool for various scientific investigations. In a typical CARS experiment, three laser beams (pump, Stokes, and probe) are focused onto the target molecule, and a coherent anti-Stokes signal is generated with orders of magnitude enhancement in signal intensity compared to spontaneous Raman scattering [8]. The enhanced signal strength and other attractive features promoted CARS as a powerful spectroscopic tool for probing molecular structures with great precision. Moreover, the recent advancements in laser technology, optics, and data analysis have propelled CARS into the forefront of modern scientific investigations such as biological tissue mapping [9–11], lipid droplet dynamics [12], polymer blend analysis [13], gas phase analysis [14,15], combustion diagnosis [16], etc [17].

Despite numeric advantages, the application of CARS remains non-trivial due to the inherent contribution of the non-resonant background (NRB) in the measured CARS signal, which significantly distorts the spectral line shapes. This can be understood from the intensity of the CARS spectral line (I_C), which is given by:

$$I_C(\nu) \propto |\chi^{(3)}|^2 \cdot I_S \cdot I_p^2 \quad (1)$$

where, $\chi^{(3)}$ is the third order nonlinear susceptibility, I_S and I_p are the intensity of the Stokes and pump laser beams, respectively.

Also, the nonlinear susceptibility $\chi^{(3)}$ can be decomposed into:

$$\chi^{(3)} = \chi_R^{(3)} + \chi_{NR}^{(3)} = \text{Re}[\chi_R^{(3)}] + \text{Im}[\chi_R^{(3)}] + \chi_{NR}^{(3)} \quad (2)$$

Here, $\chi_R^{(3)}$ represents the resonant Raman signal, which corresponds to the vibrational resonances of the molecules. This term is a superposition of several complex Lorentzian responses, each associated with different vibrational transitions, providing specific molecular fingerprint information [18]. In contrast, $\chi_{NR}^{(3)}$ describes the non-resonant contribution. This term, generated by both the molecules and the surrounding medium, contributes to the NRB in the CARS signal. The NRB primarily arises from the four-wave mixing process and involves non-resonant electronic states that are not in resonance with the vibrational modes of interest. Consequently, this NRB complicates the interpretation of the resonant signal.

Several optical approaches have been demonstrated to reduce the effect of NRB but at the cost of increasing experimental complexity [19–22]. All these alternatives not only reduce the NRB but also the intensity of CARS signal. Hence, other numerical phase retrieval approaches, such as the maximum entropy method (MEM) [23], Kramers-Kronig (KK) [24], and wavelet prism analysis [25], are introduced to extract the resonant signal from the CARS spectra [26]. New correction approaches are also proposed to improve their performance [27–29]. However, these techniques require parameter optimization and reference NRB spectrum, which would become tedious for imaging applications.

A promising way to overcome the abovementioned complications is by utilizing deep learning (DL) algorithms, which have been recently explored in the CARS spectroscopy domain. SpecNet is the first work reported using DL models based on convolutional neural network (CNN) for NRB removal [30]. It has shown promising potential for removing NRB and paved the pathways for researchers. Besides, autoencoders (named as VECTOR) [31], Bi-LSTM networks [32], and very recently generative adversarial networks (GANs) [33] have been explored by several research groups, while LSTM networks [34] have been primarily studied in our work to tackle the issue of NRB. All these works have demonstrated that there is a growing interest in harnessing the capabilities of DL algorithms to overcome the hurdle of NRB.

Recently, Ryan et al. proposed that a physics-informed neural network improves their previous model efficiency based on autoencoders (named as VECTOR2) [35]. The authors have simulated the training dataset by considering the effect of laser pulse characteristics on the spectral shape, which more accurately mimics the spectrum close to the actual scenario. Earlier, we demonstrated that training the same SpecNet CNN model with semi-synthetic data [36] and fine-tuning/transfer learning [37] significantly improves its performance. These works show that DL model efficiency is greatly affected by the training dataset, which in turn depends on resonant signal and NRB, as mentioned in Eqs. (1) and (2). In our recent work, we have also shown how different types of NRBs affect the DL model's performance [38].

In summary, while all recent studies have mainly focused on applying different DL algorithms for NRB removal, no study has systematically investigated the influence of NRB variation on CARS data analysis. Therefore, the present work aims to address this gap by investigating the effect of NRB variation and its consequences on the corresponding Raman signal retrieval in more detail.

NRB in CARS is influenced by multiple factors, including electronic resonances, the sample environment, laser beam profiles, and detector response. However, due to its broad and featureless nature, primarily resulting from electronic transitions, NRB is often approximated as real and frequency-independent in CARS studies [18]. This approximation influences the selection of NRB models in DL studies for Raman signal retrieval. Consequently, recent studies have employed various functional forms to model NRB, such as constant [34], sigmoid [38], product of two sigmoid [30,31], Gaussian [34], polynomial [32] and multiplicative functions [39]. The choice of these functions often reflects the assumption of frequency independence, tailored to different DL model architectures. While these functional forms have been effective for simulating CARS data with varying NRB shapes, they may rely on arbitrary assumptions that do not fully align with the underlying physical principles of NRB.

In our study, we opted to model NRB as a linear function. This choice is justified by general theoretical modelling (see above) and its effectiveness in scenarios where the resonant signal is dominant, such as in the CH band region of biological samples, where a linear approximation of NRB is often sufficient [18,33]. Additionally, modelling NRB as a linear function allows for straightforward and systematic variations in strength and variance, facilitating controlled simulations and consistent visualizations.

For our analysis, CARS data was generated with varying NRB strengths and variances, with strength referring to the relative intensity of the NRB compared to the resonant Raman signal. Also, the resonant part of nonlinear susceptibility is extracted from real experimental Raman data [36,37]; hence, the simulated CARS data better approximate the experimental CARS spectra. Then, the corresponding Raman signal is extracted by four methods/models in which the first two are numerical phase retrieval approaches, MEM [23] and KK [24]. The remaining two are recently explored DL learning algorithms, SpecNet CNN [30] and LSTM [34]. Later, we present a comparative study of the imaginary part prediction capabilities of these four models for different NRB strengths. Finally, we have addressed the limitations of this approach and outlined future work to explore more complex NRB models.

2. Materials and methods

This section presents a detailed overview of the CARS data simulation procedure. First, a brief overview of the experimental Raman data is presented in section 2.1. Details of the phase retrieval methods are presented in section 2.2. The NRB simulation details are discussed in section 2.3. Final CARS data simulation details are given in section 2.4.

2.1. Experimental Raman data

The Raman spectra of the six different bacteria samples were considered, and the details are given in Supplement 1 Table S1. The sample preparation procedure and experimental details are described here [40]. The six bacterial species are *Escherichia coli* DSM 423, *Klebsiella terrigena* DSM 2687, *Pseudomonas stutzeri* DSM 5190, *Listeria innocua* DSM 20649, *Staphylococcus warneri* DSM 20316, and *Staphylococcus cohnii* DSM 20261, from Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ). All these samples were cultivated in nine independent biological replicates. The suspension (1 μ L) was placed on nickel foil and air-dried at room temperature prior to acquiring the Raman spectra. Also, all these measurements were done immediately after the sample cultivation. Raman microscope (Bio Particle Explorer, rap.ID Particle Systems GmbH) comprised of Nd:YAG laser (LCM-s-111-NNP25, Laser-export Company Ltd.) operating at 532 nm is utilized in the experiment. A 100x objective lens (MPLFLN-BD, Olympus) was used to focus the laser beam onto the nickel foil. It offers a lateral resolution of less than 1 μ m with a maximum laser output of \sim 3 mW. The backscattered light from the sample was focused onto a single-stage monochromator (HE 532, Horiba). It is equipped with a grating of 920 lines/mm and coupled to a thermoelectrically cooled CCD

detector (DV401A-BV, Andor). The resulting spectral resolution was around $\sim 8 \text{ cm}^{-1}$. Finally, a total of 5420 spectra were acquired in these conditions, and the raw Raman spectra are presented in [Supplement 1 Fig. S1](#).

2.2. Phase retrieval methods

In this study, four-phase retrieval methods, MEM, KK, CNN, and LSTM are employed to remove the NRB and extract resonant Raman information from CARS spectra. The details of traditional phase retrieval methods (MEM and KK) are well documented in the literature [23,24,26] and also presented in our recent work [34]; thus, they are not described here. In-house written Python codes were used for KK and MEM analysis. In the case of DL methods, CNN (SpecNet) [30] and LSTM [34] networks were utilized. For the analysis, we directly adapted the pre-trained LSTM model weights from our previous work [34], while for CNN, the model weights were obtained from their GitHub repository [41]. The architecture and hyperparameters of both models are provided in the [Supplement 1](#). Further, the pre-trained weights of recently proposed DL models, such as VECTOR2 [35] and GAN [33], are not readily available in the literature for direct comparison. Thus, they were not considered in this study.

2.3. NRB simulation details

The NRB is simulated as a linear function with different strengths. In this context, ‘Strength’ represents the intensity of the NRB relative to the resonant Raman signal, where its maximum value is set to 1. Therefore, the NRB strength is systematically varied relative to the Raman signal by considering the following values: 0.01, 0.1, 1, 10, 100, and 1000, with each step increasing the strength by an order of magnitude. Additionally, for each strength (μ), the NRB is generated within a specific variance (σ) value, assuming a Gaussian distribution. For example, a typical schematic of NRBs simulated with $\mu = 100$ and $\sigma = 1$ is shown in Fig. 1(a). The dotted sky-blue line at 100 represents the NRB strength, and the dotted red lines approximately correspond to the 3σ variance of the Gaussian distribution. The NRB is generated as random lines with different slopes and intercept values within this variance region. The wavenumber scale is flipped after each iteration to obtain the negative slope for the NRB, ensuring that all possible variations within the selected variance are captured, as depicted in Fig. 1(a).

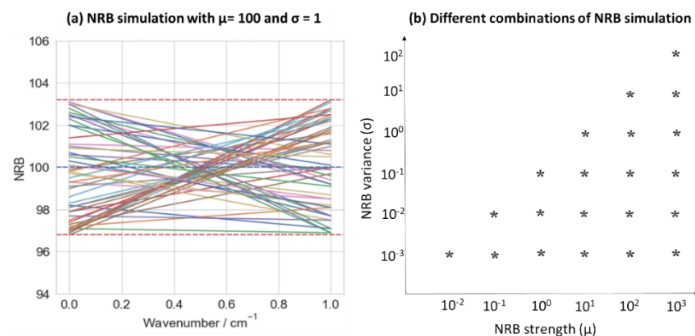


Fig. 1. a) Typical schematic of the NRBs simulation with strength of 100 and variance of 1. The dotted sky-blue line at 100 represents the NRB strength, and the dotted red lines correspond to the 3σ variance approximately. The NRB is generated within this range with different slopes and intercept values. b) Different combinations (*) of strengths and variances are considered for NRB simulation for the present study.

Furthermore, the variance values significantly influence the level of fluctuations present in the simulated NRB, thereby impacting the overall shape and intensity of the final CARS signal.

Hence, the variance is varied in the range of 0.001 to 100 for each NRB strength. However, it's important to note that when the strength is less than or equal to the variance ($\mu \leq \sigma$), it may result in negative NRB values in the Gaussian distribution, which are physically meaningless in CARS spectroscopy. Consequently, such combinations (e.g., $\mu = 0.1$ and $\sigma = 0.1, 1, 10, 100$) are omitted from consideration, resulting in a total of 21 possible combinations, as depicted in Fig. 1(b) with an asterisk (*). Finally, the NRBs are generated for all these 21 combinations by following the same procedure and subsequently utilized for the CARS data simulation.

2.4. CARS data simulation details

The simulation of the CARS signal involves both resonant and non-resonant components, as described in Eqs. (1) and (2). The NRB simulation process has already been detailed in the previous section. In this section, the resonant signal is simulated from the experimental Raman spectra acquired from six different bacteria across nine independent batches. These spectra exhibit considerable batch-to-batch variations commonly encountered in biomedical applications. Consequently, the simulated CARS spectra can better approximate the experimental CARS spectra, and the data can be considered semisynthetic. First, the real part of the resonant signal is extracted from the raw Raman data using the KK relation [36]. However, it exhibited an erroneous shape at the ends. Hence, the input Raman data is background corrected using the statistics-sensitive nonlinear iterative peak-clipping (SNIP) algorithm, followed by min-max normalization. The processed Raman spectrum and its corresponding real part are depicted in the first two rows of Fig. 2(a), respectively. The real part now has proper line shapes. Finally, CARS data is simulated using specific NRB strength and variance combinations. For instance, the CARS spectrum simulated with NRB $\mu = 100$ and $\sigma = 1$ is shown in Fig. 2(a) with a blue line. This procedure is repeated for all 21 possible combinations, as illustrated in Fig. 1(b).

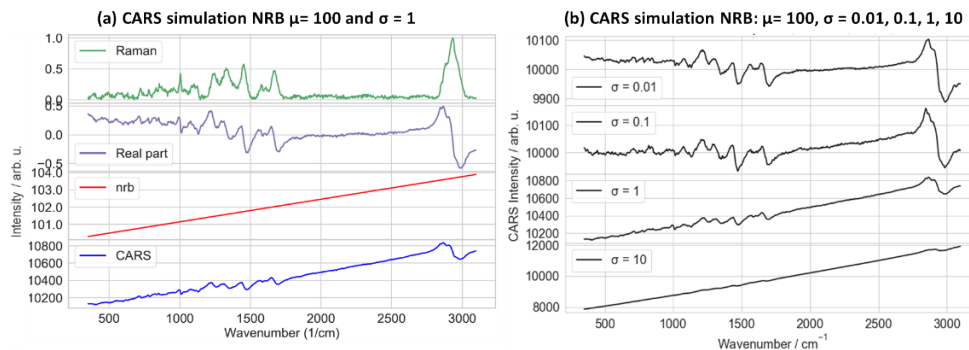


Fig. 2. (a) Sequential steps involved in the CARS data simulation. Raman spectrum background is corrected and normalized (green). The corresponding real part was extracted using the KK algorithm (purple). NRB spectrum simulated with a strength of 100 and variance of 1 (red). Final simulated CARS spectrum (blue). (b) CARS data simulated with NRB strength 100 and variances of 0.01 (Top), 0.1, 1, and 10 (bottom).

Further, the level of variance significantly influences the resulting CARS spectra for each NRB strength. For instance, Fig. 2(b) illustrates the CARS spectra simulated at NRB $\mu = 100$ with $\sigma = 0.01$ (top), 0.1, 1, and 100 (bottom). The positive slopes are considered solely for better visualization purposes. A systematic variation in the spectra is observed with increasing variance, as illustrated in the figure. Higher variances tend to hinder or even obscure spectral features, as evidenced by the last row of Fig. 2(b). In contrast, spectra simulated with lower variances exhibit only subtle differences, as depicted in the first two rows of Fig. 2(b). A similar trend is noted for CARS spectra simulated with NRB strengths of $\mu = 100$ and 1000 and variances of $\sigma = 0.001$,

0.01, and 0.1. These observations suggest that the NRB simulated at very low variance and strength approximates it as a constant value compared to the resonant signal as assumed in the literature [34].

3. Results and discussion

This section discusses the results obtained from the four different methods: MEM, KK, CNN, and LSTM. Initially, these methods are employed to retrieve the Raman signal from the simulated CARS spectra. Subsequently, the predictive performance is evaluated using Pearson correlation analysis, which compares the predicted output with the ground truth. Furthermore, PCA-LDA analysis is conducted to estimate the classification accuracy. In the text, the following terminology is utilized to represent the NRB strength and variance: “NRB of 10 ± 1 ” indicates simulation with a strength of 10 and variance of 1.

3.1. Extraction of the imaginary part

Figure 3(a-d) represents the imaginary part predicted by the MEM, KK, CNN, and LSTM models, respectively. The CARS spectrum simulated with the NRB of 100 ± 1 is used as input to the models. This spectrum is arbitrarily considered as an example from the entire data set only for visualizing the efficacy of the four methods. In each plot, the input CARS spectrum is presented at the top with a blue line. The predicted output and ground truth are shown in the middle, with red and green lines. It is worth noting that the different intensity scales are considered for ground truth and prediction, which helps readers to interpret the data more efficiently. Their absolute difference is estimated and presented with the black line at the bottom. This error plot serves as a visualization tool for evaluating the predictive performance of the four methods. All the methods except for LSTM have successfully retrieved the total spectral lines from the CARS spectra, as shown in Fig. 3(a-d). It is observed that MEM and KK predictions have been dominated by a significant background and retrieved spectral features are hindered by it. In the case of CNN, the background is minimal, albeit the output is noisy compared to others. The LSTM model consistently failed to extract the spectral features, resulting in a straight line with spurious lines at random places, as shown in Fig. 3(d). A similar behavior was demonstrated for LSTM when dealing with the real experimental CARS spectra [32]. This could be due to the simple architecture of the LSTM model.

Despite differences in the absolute intensity values, the predicted line shapes, including the background, are almost similar for MEM and KK methods. Also, the presence of a broad background contributes to a notable increase in error, as evidenced by the results shown in the last rows of Fig. 3(a), (b). Hence, the background from the extracted output is removed by applying the SNIP algorithm to minimize this error, and the results of the four models are presented in Fig. 3(e-h), respectively. After removing the background, the predicted output closely resembles the ground truth with negligible variations in intensities. All the spectral lines, even with minute intensities, are correctly retrieved by the MEM and KK methods, as shown in Fig. 3(e), (f). The background removal improved the CNN performance, but the intensities of the few retrieved lines deviated from the actual one. In addition, the noisy output limits its performance. This limited efficiency of the DL models could be due to the original models being trained with the 0-1 normalized CARS data as input [30], but raw CARS data is given as input here. Hence, the input raw CARS data is normalized between 0-1 to tackle the issue, and the retrieved imaginary parts are presented in Supplement 1 Fig. S2(c). The normalization improved the efficiency of the CNN model, which removed the noisy pattern, and no background was observed.

In the case of LSTM, some peaks in the fingerprint region are properly identified, as shown in Supplement 1 Fig. S2(d). However, MEM and KK predictions still have the background, as shown in Supplement 1 Fig. S2(a), (b). Hence, the background is removed, and the results are presented in Supplement 1 Fig. S2(e), (f). Nevertheless, the extracted line intensities deviated

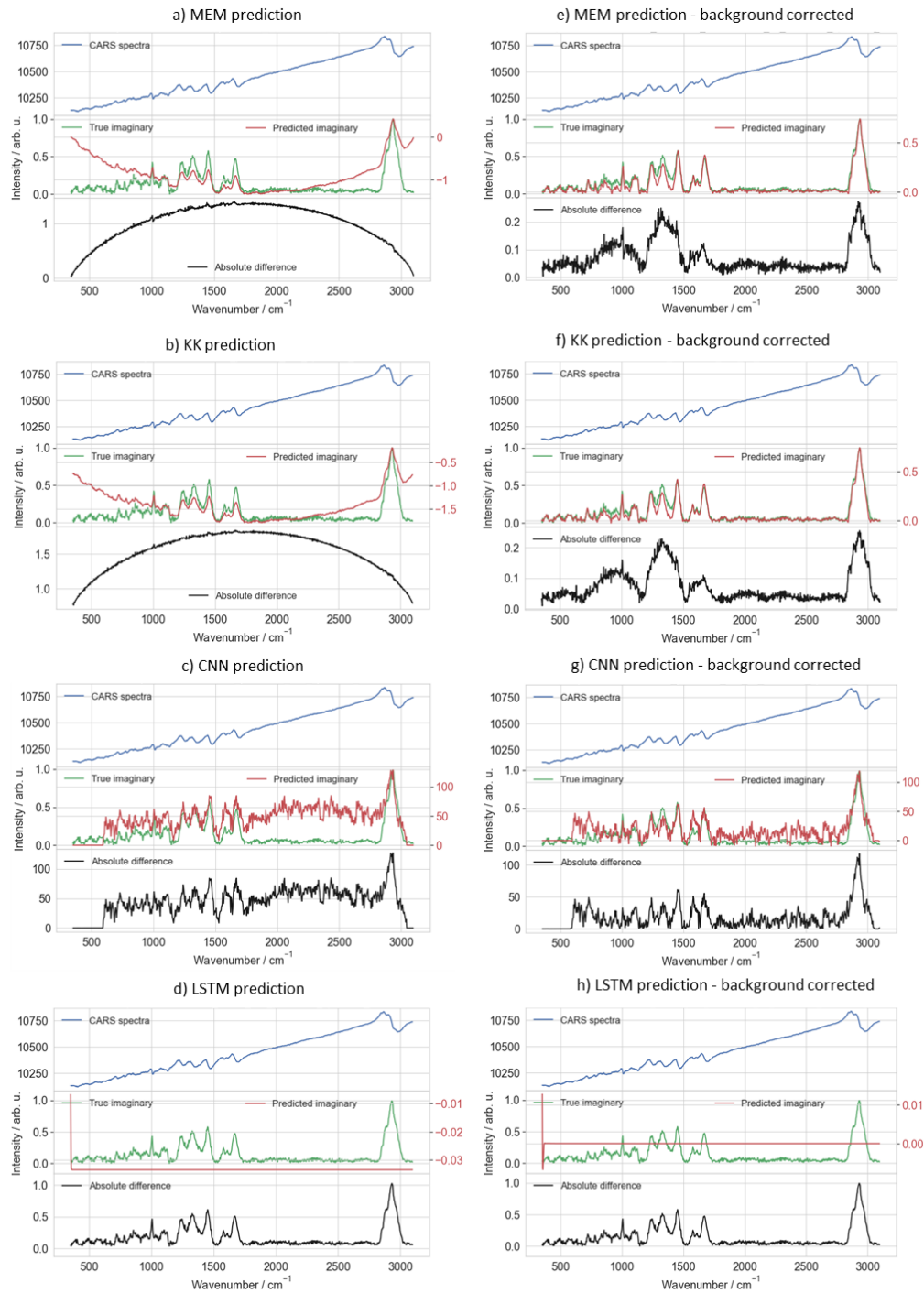


Fig. 3. (a–d) Represents the predictions of the MEM, KK, CNN, and LSTM models, respectively. The input to models is the raw CARS data (blue line) simulated with NRB of 100 ± 1 . The middle row in each figure represents the predicted imaginary part (red line) and ground truth (green line). The last row visualizes the absolute difference between the predicted output and ground truth. The prediction was found to be superior for MEM and KK, followed by CNN and LSTM models. (e–f): Represents background corrected imaginary part for the four models, respectively. The SNIP algorithm is used to remove the background.

from ground truth compared to raw data as input (see Fig. 3(e), (f)). It suggests that raw data as an input is best for MEM and KK, whereas normalized data gives optimum results for CNN and LSTM.

Further correlation analysis is performed in the next section to estimate the efficacy of the four models.

3.2. Pearson correlation analysis

Pearson correlation is a statistical technique that measures the strength of the linear relationship between two sets of continuous variables and provides a single numerical value, i.e., Pearson correlation coefficient (PCC). Here, it represents the similarity percentage between actual and predicted imaginary parts. The measured PCC values can vary between 100 and -100, representing positive and negative linear correlations, respectively. Overall, PCC value 100 represents the best prediction; the predicted imaginary part is identical to the ground truth, whereas 0 corresponds to no similarity. The PCC value is computed for each test spectrum for all four models, which gives a unique value for the predictions. Therefore, it can be considered as a performance metric for evaluating the predictive performance of all the models. First, the PCC values are computed for all the predictions of one dataset (ex: CARS spectra simulated with NRB of 100 ± 1) without any background correction, and their average is calculated. The same procedure is repeated for all the combinations, and results are presented as a heat map. For example, the results of the MEM predictions without background correction are visualized in Fig. 4(a). The x and y axes correspond to NRB strength and variance, respectively, and the heat scale (-25 to 100) represents the average PCC value obtained for each combination. For example, the first box on the bottom left corresponds to the average PCC value estimated for NRB of 0.01 ± 0.001 . A similar analysis is done for the remaining three models, KK, CNN, and LSTM, and the results are shown in Fig. 4(b-d), respectively.

It is noticed that the PCCs of MEM predictions are decreasing with increasing variance for each NRB strength, as shown in Fig. 4(a). It directly correlates that the predictive performance decreases with increasing variance. However, the change is drastic for the variance of 1, irrespective of the NRB strength, where PCC values are reduced to half of its previous variance. For example, the PCC value is 93.49 at NRB of 100 ± 0.1 but suddenly dropped to 42.14 for NRB of 100 ± 1 . It suggests that NRB simulated with a variance of 1 significantly affects predictive performance. The significant impact observed at a variance of 1 can be explained by the relative magnitude of the NRB compared to the resonant Raman signal. In simulations, the maximum value of the resonant Raman signal is set to 1. When the variance is 1, the minimum NRB strength considered is 10, meaning the NRB is an order of magnitude stronger than the Raman signal. This substantial difference causes the NRB to dominate the CARS signal, leading to noticeable distortions that adversely affect the accuracy of predictions.

The efficiency further deteriorated with its increase, as shown in Fig. 4(a). It is because the NRB becomes even more dominant, further exaggerating its influence on the CARS signal and causing another significant drop in predictive accuracy. In addition, a different scenario is observed for lower NRB strength (0.01) where the negative PCCs are achieved. It is merely due to the inverted shape after 2000 cm^{-1} in the predicted imaginary part, as shown in Supplement 1 Fig. S3(a). The CH band in the $2800\text{--}3100 \text{ cm}^{-1}$ region mainly contributed to a negative correlation. A similar trend is observed for the KK predictions, and the measured PCCs differ by less than 10 percent compared to MEM, as shown in Fig. 4(b). However, a different scenario is observed for CNN, where the PCCs of lower NRB strength have shown a positive correlation with more than 80% similarity.

The extracted line shapes are presented in Supplement 1 Fig. S3(b). It is evident from the figure that the CNN prediction works well with the low NRB strengths. However, the performance greatly decreased for higher NRB strengths, as shown in Fig. 4(c). Further, the LSTM model

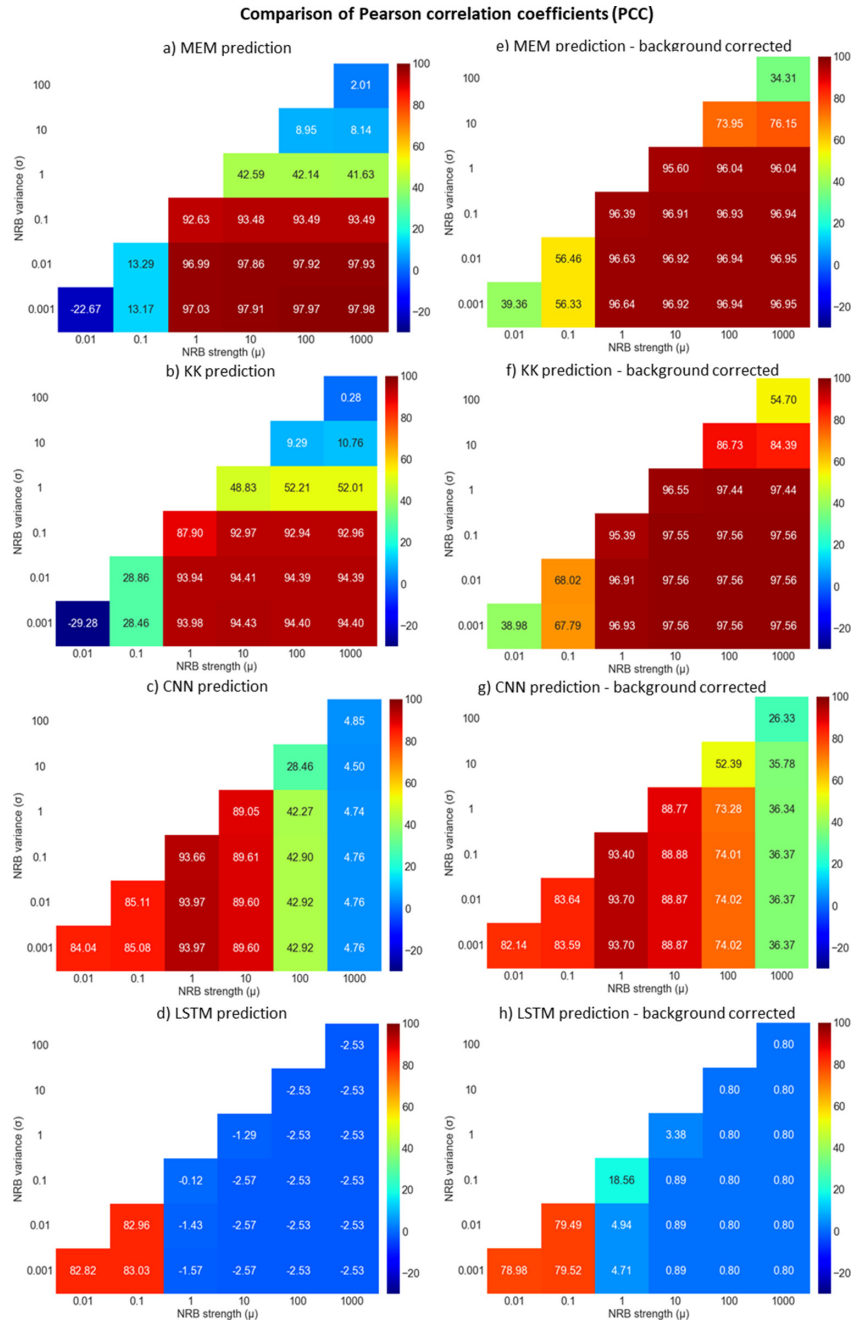


Fig. 4. (a–d) Average of Pearson correlation coefficients (PCC) obtained from the MEM, KK, CNN, and LSTM predictions, respectively. Raw CARS data is used as input to the models. Higher PCC values were achieved for MEM and KK, followed by CNN and LSTM. Also, a negative correlation was observed for MEM and KK prediction for $NRB\ 0.01 \pm 0.001$ due to the inverted shape after 2000cm^{-1} in the predicted imaginary part. (e–h) PCC values obtained after removing the background from the predictions. The correlation increased after background removal but did not affect the LSTM predictions.

prediction was only good for the first 2 NRB strengths with more than 80% positive correlation, whereas it is less than 10% for the remaining.

The PCCs are also estimated after removing the background from the predicted imaginary part, and the results are presented in Fig. 4(e-h). The correlation increased for all the models except for the LSTM. Also, it primarily impacted the predictions at higher NRB strengths. For example, the PCC is only 8.95% for MEM prediction at $\text{NRB } 100 \pm 10$ and increased to 73.95% after background removal. A similar trend is noticed for the KK prediction, with a relatively higher correlation of more than 10% compared to MEM. In the case of CNN, the correlation is increased around six times for the NRB strength of 1000. However, the final PCCs are not even half of the MEM and KK predictions. Also, the LSTM predictions are not affected by the background correction; hence, the PCCs are not changed, as shown in Fig. 4(h).

As mentioned in the previous section, the CNN and LSTM predictions improved after normalizing the input CARS data. Hence, the PCCs are also estimated for the predictions with normalized CARS data as an input to the models, and the results are presented in [Supplement 1 Fig. S4](#). The correlation is significantly increased for CNN and LSTM at higher NRB strengths as compared to raw data as an input. It is primarily attributed to the reduced noise in the predicted imaginary part along with no background, as shown in [Supplement 1 Fig. S2\(c\), \(d\)](#). In the case of MEM and KK, the correlation decreased due to background and noise in the predictions. Nevertheless, the correlation is later improved after background correction, as shown in [Supplement 1 Fig. S4\(e\), \(f\)](#). However, it is still less than the raw data as input to models, as shown in Fig. 4(e), (f). These observations manifest that the raw CARS data as input would be best for MEM and KK. In contrast, normalized data demonstrates optimum results for CNN and LSTM. Also, the predictive efficiency significantly improves after background correction for MEM and KK, and it would not be required for the other two methods.

Even though the PCC value can be considered as the best metric for comparing the predicted output with the ground truth, it only measures the strength of linear associations between pairs of continuous variables. Hence, a linear classification model is utilized in the next section, and its accuracy is used as a performance metric for the evaluation.

3.3. *Principal-component analysis (PCA) – linear-discriminant analysis (LDA)*

Principal-component analysis (PCA) followed by linear-discriminant analysis (LDA) is applied as a classification model to separate six bacteria species based on the predictions made on basis of the spectra generated by the four methods (MEM, KK CNN, and LSTM). PCA, being an unsupervised technique, reduces high dimensional input data into a lower dimensional space by retaining most of the information. It primarily projects datasets in the direction representing the largest possible variance, called principal components (PCs). All the PCs are arranged such that maximum variance is retained by the first PC, i.e., PC1, and followed by subsequent PCs.

In contrast, LDA is a supervised method applied in the PCA space, which provides a classification model and is evaluated using classification accuracy [40,42]. Here, the Batchout cross-validation method is used for the validation, where one batch is utilized as a validation set out of nine independent batches, and the classification model is trained based on the remaining batches [43]. The same procedure is repeated for all batches as a validation set, and finally, average accuracy is estimated. This validation approach ensures that the independent datasets are being used for training and validation while evaluating the reliable and unbiased estimate of model performance. The entire analysis is carried out using in-house written functions in Python 3.9.

First, the PCA is applied to the model predictions and retrieved the PCs. Later, the LDA model is built by considering the first 10 PCs as input. Figure 5(a-c) represents the PCA loadings plot obtained from the MEM, KK, and CNN predictions, respectively. The three rows in each figure correspond to the loadings of the first three PCs. As seen from the figure, a huge background

contributes to the first PC for all the models, whereas little background is observed in the remaining PCs. In addition, the loadings look noisy for the CNN model. It is worth noting that all the LSTM model predictions were found to be the same for NRB strength of more than 10.

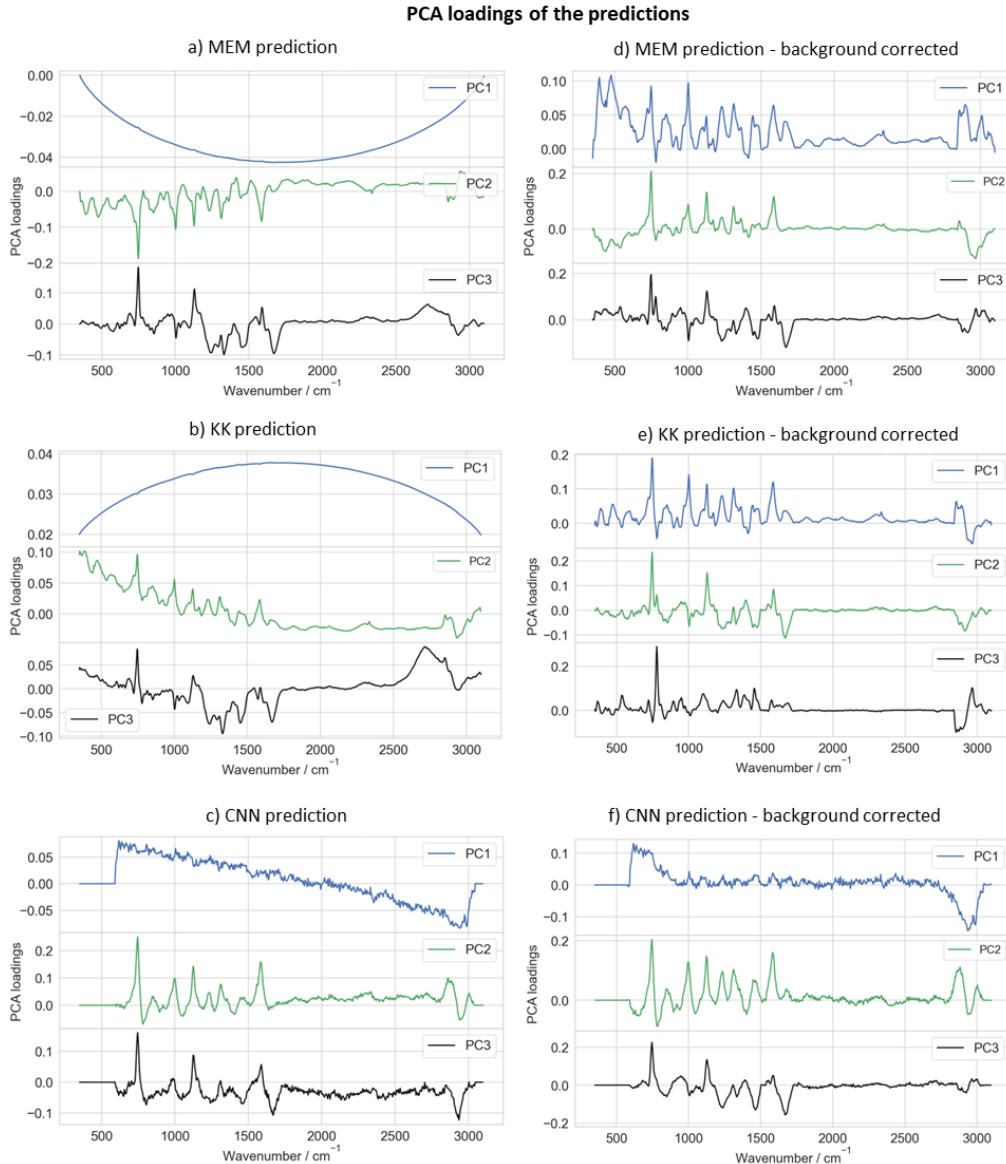


Fig. 5. (a–c) Represent the PCA loadings plot obtained from the predictions of the four models, MEM, KK, and CNN, respectively. PC1, PC2, and PC3 correspond to the first three principal components (PCs). A large background is observed in the first PC compared to the remaining PCs for all the models. The LSTM model predictions were found to be the same for NRB strength of more than 10. Hence, the PCA-LDA could not fit the data, and the loadings are not visualized here. (d–f) Represents the PCA loadings plot obtained after removing the background from the predictions of MEM, KK, and CNN models, respectively. The spectral features buried in the spectra have appeared in the first PCs after removing the background.

Thus, the PCA-LDA could not fit the data, and the loadings are not visualized in Fig. 5. The same analysis repeated after removing the background from the predictions and corresponding loadings are illustrated in Fig. 5(d–f). Now, the baseline of all the loadings almost looks flat, and spectral features buried in the background before its removal have appeared in the first PC. As aforementioned, LDA was applied to the first 10 PCs, and accuracies obtained for each model are visualized in the heat map, as shown in Fig. 6. The percentages represent the average classification accuracy for each combination of NRB strength and variance.

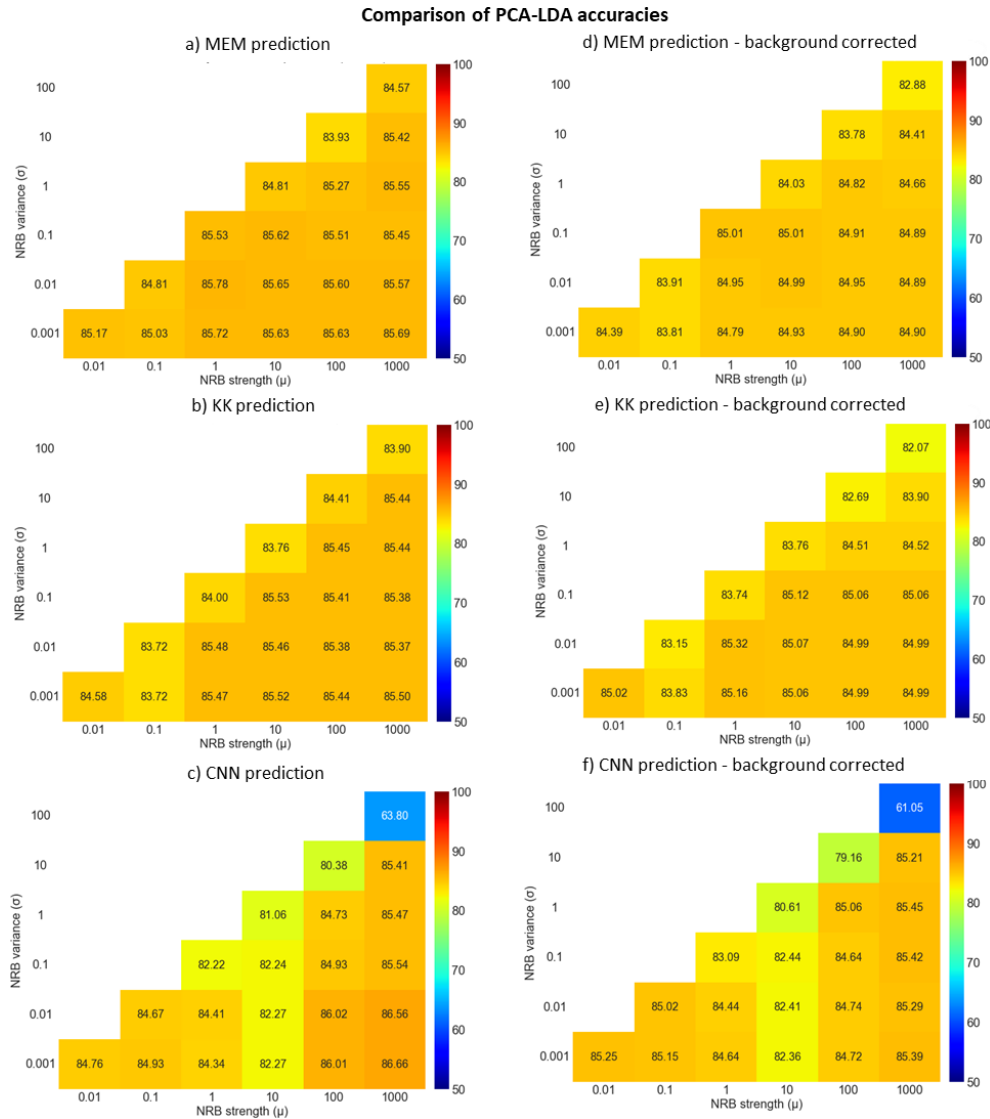


Fig. 6. (a–c) Represent the PCA-LDA classification accuracies obtained MEM, KK, and CNN predictions, respectively. Here, raw CARS data is used as an input to the models. The accuracies for a particular NRB strength decreased slowly with increasing variance. The accuracies of MEM and KK predictions are approximately the same. (e–f) Represents the classification accuracies obtained after removing the background from the predictions. Higher accuracies were achieved for MEM and KK, followed by CNN. Also, the classification accuracy is found to be the same even after background correction.

For each NRB strength, the accuracies slowly decreased with increasing variance, but the change is noticeable for the variance of more than 0.1. It is observed that the measured classification accuracies are approximately the same for the MEM and KK, as illustrated in Fig. 6(a), (b). In the case of CNN, the accuracies are slightly lower than MEM and KK except for the NRB of 1000 ± 100 . Also, no significant differences were observed after removing the background from the predictions, as shown in Fig. 6(d), (e). The change in accuracy is also found to be less than 1.5% for all the combinations except for the NRB of 1000 ± 100 . It could be due to the LDA model built based on the PCs not having any background, as shown in the coefficient plot in Supplement 1 Fig. S5. It suggests that the PCA-LDA model can classify samples even with a background whose classification accuracies are close to predictions after background correction. The same analysis is repeated for the predictions obtained for 0-1 normalized CARS data as an input to the models. The loadings plot is shown in Supplement 1 Fig. S6, and corresponding accuracies are shown in Supplement 1 Fig. S7. The accuracies decreased compared to raw data as input. The reduction is less than 4% for all the NRB strengths up to 0.1 variance. However, the reduction is 15–20% for the variances ≥ 1 . All these results suggest that the PCA-LDA analysis is effective in classifying the predictions, even in the presence of background. The classification accuracies for the MEM and KK methods range between 80–85%, which is consistent with the PCA-LDA classification accuracies obtained for the spontaneous Raman data in our previous work [40]. In contrast, the CNN and LSTM models exhibit slightly lower accuracies, ranging from 50–85%.

The overall findings of this study highlight the importance of using or adapting different phase retrieval methods across various CARS applications like bioimaging, material characterization, and chemical analysis. The superior prediction performance of the MEM and KK models, especially when applied to raw data, suggests that these methods are robust even in the presence of strong NRB, making them ideal for applications involving highly complex experimental CARS spectra such as in chemical mixtures, nanomaterials, etc. On the other hand, the CNN and LSTM models performed best when applied to normalized data, indicating that DL methods might be more suitable for cases where data preprocessing is possible, and NRB strength is less variable. Mainly, this could be beneficial in biological imaging (based on the CH band, for which the resonant signal is more intense than NRB [33]), where it can serve as a rapid spectroscopic tool for medical diagnosis and research.

Further, the linear approximation of the NRB used in this study represents a simplification. In more complex systems, particularly at higher laser intensities or in the presence of strong electronic resonances, the NRB may exhibit nonlinear behavior or frequency dependence. To address this, future studies will explore scenarios where the NRB deviates from a linear approximation. We aim to develop and implement more sophisticated models that account for potential nonlinearities in $\chi_{NR}^{(3)}$. For instance, the NRB could be simulated using a sigmoid [38], a product of two sigmoids [30,31], Gaussian [34], polynomial [32], and multiplicative functions [39], or a combination of these approaches. This approach will provide a more accurate representation of the NRB under varying experimental conditions and improve the overall precision of the CARS data analysis. Further, noise effects were not considered in this study to extract the pure dependency of the phase retrieval on the NRB strengths and variances. Thus, future work will also include investigating the impact of noise on the generated signals to understand better and improve the robustness of the phase retrieval methods. Also, the analysis would extend to other DL algorithms, such as VECTOR2 [31] and GANs [33], which have been explored very recently. In summary, the adaptability of all these conditions opens new opportunities for improving CARS data analysis for various applications.

4. Conclusions

In this work, we demonstrated the systematic study of NRB variation on CARS data analysis. The CARS data was generated with different NRB strengths by assuming it as a linear function. The NRB strengths were varied in the range of 0.01–1000 in the steps of increasing an order each time. Different variance levels were also considered for each strength. During the CARS data simulation, the real part of the resonant nonlinear susceptibility term was extracted from the experimental Raman spectra using the KK relation. Then, the imaginary part of the simulated CARS spectrum was extracted using four methods: MEM, KK, CNN, and LSTM. Finally, the model performances at different NRB strengths are evaluated based on correlation and PCA-LDA analysis.

First, the Pearson correlation analysis revealed that the prediction capability is approximately the same for the MEM and KK, followed by CNN and LSTM. Further, the Pearson correlation of the MEM and KK predictions was increased after removing the background from the extracted imaginary part. In contrast, normalizing the input data favored the CNN and LSTM model

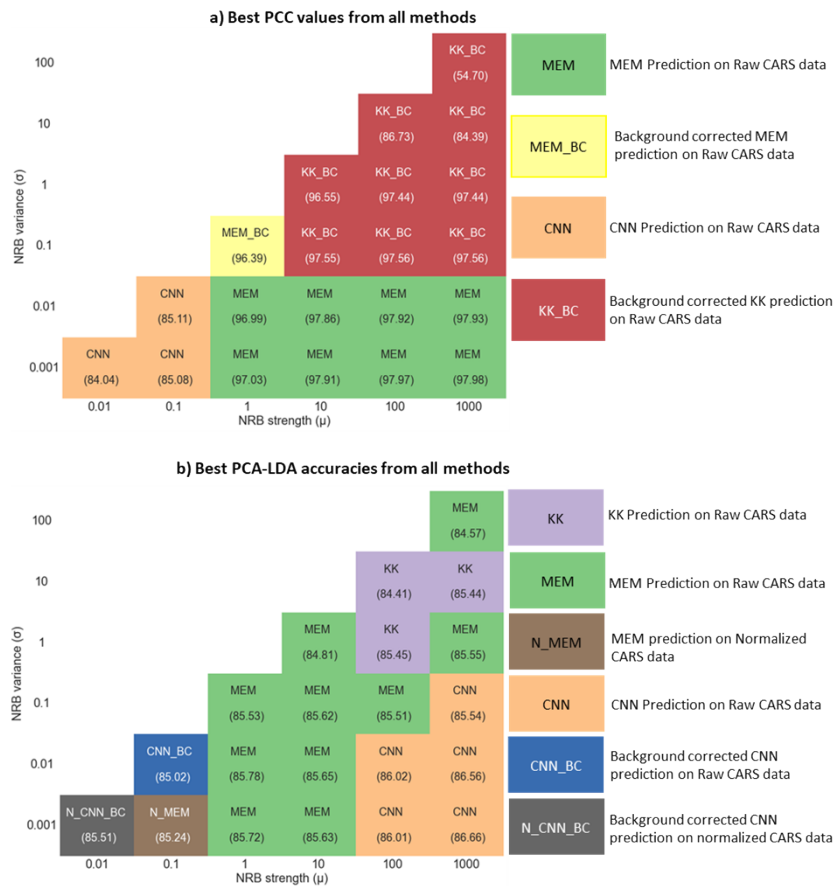


Fig. 7. (a) Represent the best PCC values obtained from all the methods for each NRB strength combination. The KK model predictions with background correction are optimum for higher NRB strengths and variance while the CNN model is suited for very low NRB values. (b) Represents the best classification accuracies obtained from all the methods for each NRB strength combination. Higher accuracies were achieved for MEM and KK predictions on raw CARS data for higher NRB strengths and variance.

performance, whereas it negatively impacted the remaining two methods. Figure 7(a) visualizes the final overview of the best methods suited for each NRB strength combination based on measured PCC values. In which, MEM prediction on raw CARS data is best for NRB strength ≥ 1 and variance < 0.1 , whereas KK predictions with background correction are optimum for higher NRB strengths. The CNN method is suited for very low NRB strengths

PCA-LDA analysis demonstrated that higher accuracies are achieved for MEM and KK, followed by CNN and LSTM. It also revealed that the classification accuracies are not affected by the presence of the background in the predicted imaginary part. Figure 7(b) illustrates the overview of the best methods for each NRB combination based on PCA-LDA accuracies. The MEM and KK models' prediction on raw CARS data is best for NRB strength ≥ 0.1 . The CNN predictions with background correction are optimum for NRB strengths ≤ 1 . Additionally, the LSTM model performs effectively only at lower NRB strengths, likely due to its simple architecture, which limits its applicability to specific cases.

This study systematically investigates how different NRB correction techniques perform if different NRBs (strength and variance) are found in the data. This should give a roadmap of which methods should be applied to the given experimental data at hand. As the simulation was only using linear NRBs, other functions need to be investigated in the future.

Funding. Horizon 2020 Framework Programme (101016923 (CRIMSON)); Bundesministerium für Bildung und Forschung (13N15466, 13N15706); Deutsche Forschungsgemeinschaft (DFG).

Acknowledgments. This work is supported by the EU funding program for the project CRIMSON. This work is also supported by the BMBF, funding program Photonics Research Germany (13N15466 (LPI-BT1), 13N15706 (LPI-BT2)) and is integrated into the Leibniz Center for Photonics in Infection Research (LPI). The LPI initiated by Leibniz-IPHT, Leibniz-HKI, Friedrich Schiller University Jena and Jena University Hospital is part of the BMBF national roadmap for research infrastructures. We also acknowledge support by the German Research Foundation Projekt-Nr. 512648189 and the Open Access Publication Fund of the Thuringer Universitaets- und Landesbibliothek Jena.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Supplemental document. See [Supplement 1](#) for supporting content.

References

1. W. M. Tolles, J. W. Nibler, J. R. McDonald, *et al.*, "A review of the theory and application of coherent anti-Stokes Raman spectroscopy (CARS)," *Appl. Spectrosc.* **31**(4), 253–271 (1977).
2. M. Müller and A. Zumbusch, "Coherent anti-Stokes Raman scattering microscopy," *ChemPhysChem* **8**(15), 2156–2170 (2007).
3. D. S. Choi, H. Rhee, M. Cho, *et al.*, "Discriminating congested vibrational peaks of condensed organic materials with time- and frequency-resolved coherent anti-Stokes Raman scattering spectroscopy," *J. Phys. Chem. B* **127**(17), 3958–3968 (2023).
4. S. Yampolsky, D. A. Fishman, S. Dey, *et al.*, "Seeing a single molecule vibrate through time-resolved coherent anti-Stokes Raman scattering," *Nat. Photonics* **8**(8), 650–656 (2014).
5. C. H. Camp and M. T. Cicerone, "Chemically sensitive bioimaging with coherent Raman scattering," *Nat. Photonics* **9**(5), 295–305 (2015).
6. R. Junjuri, M. Calvarese, M. Vafaeinezhad, *et al.*, "Estimation of biological variance in coherent Raman microscopy data of two cell lines using chemometrics," *Analyst* **149**(17), 4395–4406 (2024).
7. S. Li, Y. Li, R. Yi, *et al.*, "Coherent anti-Stokes Raman scattering microscopy and its applications," *Front. Phys.* **8**, 598420 (2020).
8. T. Guerenne-Del Ben, Z. Rajaofara, V. Couderc, *et al.*, "Multiplex coherent anti-Stokes Raman scattering highlights state of chromatin condensation in CH region," *Sci. Rep.* **9**(1), 13862 (2019).
9. C. Krafft, B. Dietzek, Jürgen Popp, *et al.*, "Raman and CARS microspectroscopy of cells and tissues," *Analyst* **134**(6), 1046–1057 (2009).
10. C. H. Camp, Y. J. Lee, J. M. Heddleston, *et al.*, "High-speed coherent Raman fingerprint imaging of biological tissues," *Nat. Photonics* **8**(8), 627–634 (2014).
11. F. Vernuccio, A. Bresci, B. Talone, *et al.*, "Fingerprint multiplex CARS at high speed based on supercontinuum generation in bulk media and deep learning spectral denoising," *Opt. Express* **30**(17), 30135–30148 (2022).
12. H. A. Rinia, K. N.J. Burger, M. Bonn, *et al.*, "Quantitative label-free imaging of lipid composition and packing of individual cellular lipid droplets using multiplex CARS microscopy," *Biophys. J.* **95**(10), 4908–4914 (2008).

13. B. von Vacano, L. Meyer, M. Motzkus, *et al.*, “Rapid polymer blend imaging with quantitative broadband multiplex CARS microscopy,” *J Raman Spectroscopy* **38**(7), 916–926 (2007).
14. J. F. Verdieck, R. J. Hall, J. A. Shirley, *et al.*, “Some applications of gas phase CARS spectroscopy,” *J Chem Educ* **59**(6), 495–503 (1982).
15. S. Roy, J. R. Gord, A. K. Patnaik, *et al.*, “Recent advances in coherent anti-Stokes Raman scattering spectroscopy: Fundamental developments and applications in reacting flows,” *Prog. Energy Combust. Sci.* **36**(2), 280–306 (2010).
16. F. Vestin, M. Afzelius, P.-E. Bengtsson, *et al.*, “Development of rotational CARS for combustion diagnostics using a polarization approach,” *Proc. Combust. Inst.* **31**(1), 833–840 (2007).
17. N. Coluccelli, G. Galzerano, P. Laporta, *et al.*, “Standoff detection of bacterial spores by field deployable coherent Raman spectroscopy,” *Sci. Rep.* **13**(1), 2634 (2023).
18. D. Polli, V. Kumar, C. M. Valensise, *et al.*, “Broadband coherent Raman scattering microscopy,” *Laser Photonics Rev* **12**(9), 1800020 (2018).
19. O. Burkacky, A. Zumbusch, C. Brackmann, *et al.*, “Dual-pump coherent anti-Stokes-Raman scattering microscopy,” *Opt. Lett.* **31**(24), 3656–3658 (2006).
20. J.-X. Cheng, L. D. Book, X. Sunney Xie, *et al.*, “Polarization coherent anti-Stokes Raman scattering microscopy,” *Opt. Lett.* **26**(17), 1341–1343 (2001).
21. M. Jurna, J. P. Korterik, C. Otto, *et al.*, “Background free CARS imaging by phase sensitive heterodyne CARS,” *Opt. Express* **16**(20), 15863–15869 (2008).
22. B. Littleton, P. Shah, T. Kavanagh, *et al.*, “Three-color coherent anti-Stokes Raman scattering with optical nonresonant background removal,” *J. Raman Spectrosc.* **50**(9), 1303–1310 (2019).
23. E. M. Vartiainen, H. A. Rinia, M. Müller, *et al.*, “Direct extraction of Raman line-shapes from congested CARS spectra,” *Opt. Express* **14**(8), 3622–3630 (2006).
24. V. Lucarini, *Kramers-Kronig relations in optical materials research* (Springer Science & Business Media, 2005), Vol. 110.
25. Y. Kan, L. Lensu, G. Hehl, *et al.*, “Wavelet prism decomposition analysis applied to CARS spectroscopy: a tool for accurate and quantitative extraction of resonant vibrational responses,” *Opt. Express* **24**(11), 11905–11916 (2016).
26. M. T. Cicerone, K. A. Aamer, Y. J. Lee, *et al.*, “Maximum entropy and time-domain Kramers–Kronig phase retrieval approaches are functionally equivalent for CARS microspectroscopy,” *J. Raman Spectrosc.* **43**(5), 637–643 (2012).
27. C. H. Camp, J. S. Bender, Y. J. Lee, *et al.*, “Real-time and high-throughput Raman signal extraction and processing in CARS hyperspectral imaging,” *Opt. Express* **28**(14), 20422–20437 (2020).
28. C. H. Camp, Y. J. Lee, M. T. Cicerone, *et al.*, “Quantitative, comparable coherent anti-Stokes Raman scattering (CARS) spectroscopy: correcting errors in phase retrieval,” *J. Raman Spectrosc.* **47**(4), 408–415 (2016).
29. C. H. Camp, “Raman signal extraction from CARS spectra using a learned-matrix representation of the discrete Hilbert transform,” *Opt. Express* **30**(15), 26057–26071 (2022).
30. C. M. Valensise, A. Giuseppi, F. Vernuccio, *et al.*, “Removing non-resonant background from CARS spectra via deep learning,” *APL Photonics* **5**(6), 061305 (2020).
31. Z. W. Wang, K. O’ Dwyer, R. Muddiman, *et al.*, “VECTOR: Very deep convolutional autoencoders for non-resonant background removal in broadband coherent anti-Stokes Raman scattering,” *J. Raman Spectrosc.* **53**(6), 1081–1093 (2022).
32. R. Junjuri, A. Saghi, L. Lensu, *et al.*, “Evaluating different deep learning models for efficient extraction of Raman signals from CARS spectra,” *Phys. Chem. Chem. Phys.* **25**(24), 16340–16353 (2023).
33. Z. Y. Luo, X. Xu, D. Lin, *et al.*, “Removing non-resonant background of CARS signal with generative adversarial network,” *Appl. Phys. Lett.* **124**(26), 1 (2024).
34. R. Houhou, P. Barman, M. Schmitt, *et al.*, “Deep learning as phase retrieval tool for CARS spectra,” *Opt. Express* **28**(14), 21002–21024 (2020).
35. R. Muddiman, K. O’ Dwyer, C. H. Camp, *et al.*, “Removing non-resonant background from broadband CARS using a physics-informed neural network,” *Anal. Methods* **15**(32), 4032–4043 (2023).
36. R. Junjuri, A. Saghi, L. Lensu, *et al.*, “Convolutional neural network-based retrieval of Raman signals from CARS spectra,” *Opt. Continuum* **1**(6), 1324–1339 (2022).
37. A. Saghi, R. Junjuri, L. Lensu, *et al.*, “Semi-synthetic data generation to fine-tune a convolutional neural network for retrieving Raman signals from CARS spectra,” *Opt. Continuum* **1**(11), 2360–2373 (2022).
38. R. Junjuri, A. Saghi, L. Lensu, *et al.*, “Effect of non-resonant background on the extraction of Raman signals from CARS spectra using deep neural networks,” *RSC Adv.* **12**(44), 28755–28766 (2022).
39. T. Härkönen, E. M. Vartiainen, L. Lensu, *et al.*, “Log-Gaussian gamma processes for training Bayesian neural networks in Raman and CARS spectroscopies,” *Phys. Chem. Chem. Phys.* **26**(4), 3389–3399 (2024).
40. N. Ali, S. Girnus, P. Rösch, *et al.*, “Sample-size planning for multivariate data: a Raman-spectroscopy-based example,” *Anal. Chem.* **90**(21), 12485–12492 (2018).
41. C. M. Valensise, “SpecNet” (2020), retrieved 2024, <https://github.com/Valensicv/SpecNet>.
42. C. W. Park, I. Lee, S.-H. Kwon, *et al.*, “Classification of CARS spectral phase retrieval combined with principal component analysis,” *Vib. Spectrosc.* **117**, 103314 (2021).
43. S. Guo, T. Bocklitz, U. Neugebauer, *et al.*, “Common mistakes in cross-validating classification models,” *Anal. Methods* **9**(30), 4410–4417 (2017).