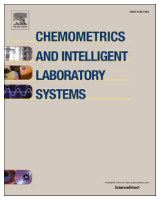
Chemometrics and Intelligent Laboratory Systems 254 (2024) 105242

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemometrics



Applicability domain of a calibration model based on neural networks and infrared spectroscopy

M. Suliany Rodríguez-Barrios [a], Joan Ferré [a,*], M. Soledad Larrechi [a], Enric Ruiz [b]

[a] Department of Analytical Chemistry and Organic Chemistry, Faculty of Chemistry, Universitat Rovira i Virgili, Tarragona, Spain
[b] Repsol Petróleo, Tarragona, Spain

ARTICLE INFO

Keywords:
Infrared spectroscopy
Artificial neural networks
Autoencoder
Decoder
Applicability domain
Diesel

ABSTRACT

Artificial neural networks are used as calibration models in routine analytical determinations that involve spectroscopic data. To ensure that the model will generate reliable predictions for new samples, the applicability domain must be well defined. This article describes a strategy for establishing the limits of the applicability domain when the calibration model is a feed-forward neural network. The applicability domain was defined by two limits: 1) the 0.99 quantile of the squared Mahalanobis distance calculated from the network activations of the training set and 2) the 0.99 quantile of the reconstruction error of the training spectra using either an autoencoder network or a decoder network. A new sample with a squared Mahalanobis distance and/or spectral residuals beyond these limits is said to be outside the applicability domain, and the prediction is questionable. The approach was illustrated by predicting the density of diesel fuel samples from mid-infrared spectra and the fat content in meat from near-infrared spectra. The methodology could correctly detect anomalous spectra in prediction using either the autoencoder or the decoder.

Novelty statement

A new procedure to establish the applicability domain when a feed-forward neural network is the calibration model in analytical determinations based on spectroscopic data.

The squared Mahalanobis distance and the spectral residuals of an autoencoder (or a decoder) were used to define the limits of the applicability domain of a regression network. Beyond these limits, the prediction of a new sample is not considered reliable enough.

The methodology was illustrated with two datasets involving MIR and NIR spectra, respectively.

1. Introduction

The demand for fast, inexpensive, and multiparametric analytical methods has led to the widespread use of infrared spectroscopy with multivariate calibration. Among the calibration models, artificial neural networks (ANNs) are a versatile class that is especially suited to handle nonlinear relationships between the spectrum and the property to be predicted [1]. Some examples that show the relevance of ANNs in analytical determinations can be found in Refs. [1–7].

The use of ANNs in routine analyses requires proper model validation, the specification of the applicability domain, the periodic testing of the model's performance against the reference method, and some means of estimating the prediction uncertainty. While aspects such as prediction uncertainty [8,9] and analytical figures of merit [9–11] have been addressed, the characterization of the applicability domain of a feed-forward neural network has received less attention, and this is the focus of this work. The applicability domain (AD) can be defined as the sample space where the model is expected to predict with a given reliability [12]. In spectroscopic methods, predictions are more reliable when the spectra of the new samples are like the spectra of the training samples, i.e., within the AD. Samples outside the AD are extrapolations and therefore their predictions are less reliable. The reasons why a new sample may fall outside the AD are diverse. Some are related to the sample itself, such as having a new ratio of constituents or different matrix effects, while others are related to gross errors, new measurement conditions, or instrument performance. These sources of discrepancy are discussed regularly in reports that deal with novelty detection, anomaly detection, fault detection, and model updating. In this sense, the specification of the AD is connected to the approaches used to detect prediction outliers, which are those instances that are outside the

* Corresponding author.
E-mail address: joan.ferre@urv.cat (J. Ferré).

https://doi.org/10.1016/j.chemolab.2024.105242
Received 23 April 2024; Received in revised form 26 June 2024; Accepted 2 October 2024
Available online 5 October 2024
0169-7439/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

boundaries of the AD. In summary, the purpose of defining the AD is to answer the question: "Can we trust the prediction from this spectrum?".

Quantitative-structure activity relationship (QSAR) studies offer a convenient initial perspective about some approaches that can be used to describe the AD. The definition of the AD in QSAR is an active area of research [13–18] motivated by the fact that the validation of QSAR models used in regulatory applications must include the specification of the AD [19]. When the modeling algorithms used in analytical determinations coincide with those in QSAR, the principles used to define the AD are common [20,21]. These principles are also found in surveys about outlier detection in classification and regression [22–28], where the AD is not explicitly mentioned, but it is indirectly understood as the space enclosed by the limits used to issue outlier warnings.

There are different approaches to defining the AD of a model. They can be based on variable ranges, distances, densities, ensemble learning, or prediction intervals, to mention one way of grouping them [20,29]. This diversity of methods reflects the variety of scenarios, and the choice of the approach depends on the characteristics of the model and the type and distribution of the available data. Nevertheless, the prevailing situation is the lack of training instances outside the AD, so most methods are unsupervised and the boundaries of the AD are set by learning from good data. Among these methods, distance-based approaches are simple and widely used [13,17,18,30]. The similarity between a new sample and the training samples can be measured by the Euclidean distance [14], Mahalanobis distance [31], Manhattan distance [32], Hotelling's T^2 and leverage [29]. For highly correlated variables, such as spectra, the Mahalanobis distance is especially useful [33,34]. It is assumed that the prediction accuracy decreases as the distance of the new sample to the training samples increases. Hence, a distance above a given threshold indicates that the sample is outside the AD and its prediction is unreliable. The threshold can be set as a quantile of the theoretical distribution that the metric is known to follow or a quantile of the distances calculated for the training samples up to the maximum distance obtained for the training set [13,29,35]. Reconstruction-based methods are also common. They assume that novelties (i.e., spectra of good samples not yet represented by the model) or anomalies (i.e., spectra of bad samples or erroneous spectra of good samples) cannot be reconstructed from projections in a space of fewer dimensions calculated from good samples [25,36]. Therefore, they can be detected by a spectrum reconstruction error above a given threshold [21]. This idea is used in factor-based multivariate models, where the sum of squared spectral residuals (a.k.a. Q-value) is commonly used to detect outliers. In nonlinear scenarios, autoencoders are a type of neural network that can be used for this purpose. These networks are designed to reconstruct the input through a narrow hidden layer, known as the bottleneck. By compressing the input through the narrow layer, the common information relevant to describing the samples is preserved, so the autoencoder learns to reconstruct the training data approximately. Hence, an autoencoder will properly reconstruct the spectrum of a new sample if it is similar to the spectra used to train it. Otherwise, reconstruction will be poor. Therefore, novel and anomalous samples can be identified by having larger spectral residuals than the residuals of the training samples. Autoencoders have been used in multivariate process control, sample classification, and fault diagnosis, among other applications [7,37].

This paper shows a procedure for establishing the AD when a feed-forward neural network is the calibration model of an analytical determination. The AD is enclosed by two limits. One limit is based on the Mahalanobis distance calculated from the projection of the training spectra in the reduced latent space of the regression network. The other limit uses the sum of squared spectral residuals of an autoencoder. Two variations of the latter are presented and compared. One is the classical autoencoder, trained from the raw spectra. The other uses only the decoder part of an autoencoder, trained using the activations from the regression network as input.

The approach was applied to two study cases. The first one is the determination of the density of diesel fuel samples from IR spectra using a feed-forward neural network. The combination of IR spectroscopy and ANNs have been shown to be effective in determining relevant physicochemical properties of diesel fuel samples [38–41]. The second case is the determination of fat content in meat samples using a dataset [42,43] that is commonly used as a benchmark for ANNs.

2. Theory

2.1. Feed-forward neural network

The feed-forward multilayer perceptron (Fig. 1) is the simplest and possibly the most common ANN used in quantitative analysis with spectroscopic data. In this nonlinear regression model, a layer of J nodes processes data from the previous layer according to equation (1):

$$a_j = f\left(\sum_{k=1}^{K} w_{jk}x_k + w_{j0}\right) \tag{1}$$

where a_j is the output of node j of the layer, x_k's are the inputs of node j, w_{jk} and w_{j0} are the weights and bias associated with this node, and f is the so-called activation function. For the first hidden layer, x_k's are the elements of the input spectrum $\mathbf{x} = [x_1, \ldots x_k, \ldots x_K]^{\mathrm{T}}$ and f is a nonlinear function. For the output layer, which in this case has one node only, the x_k's are the outputs of the previous hidden layer, f is a linear function, and the output is the predicted value of the property, $\hat{y}$. The number of hidden layers, the number of nodes in each layer, and the activation functions are selected for the problem at hand. The weights and biases are estimated by backpropagation from the spectrum and property value pairs $\{\mathbf{x}, y\}$ of the training set, using the mean square error (MSE) (optionally with L_2-regularization) as a loss function [44].

2.2. Auto-associative neural network

An auto-associative neural network, also known as a replicator neural network or autoencoder, is a feed-forward multilayer perceptron network designed to reproduce the input [44,45]. A key feature is that one hidden layer has fewer neurons than the input layer (Fig. 2) which gives this network the aspect of having a bottleneck. The left-hand side

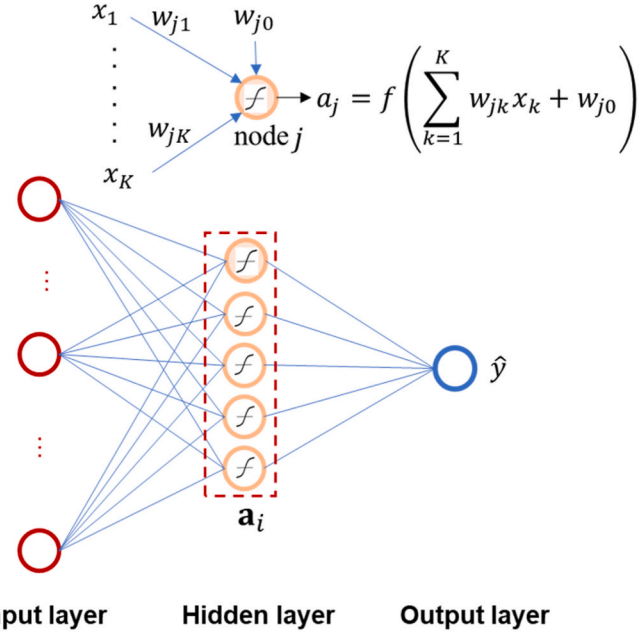


Fig. 1. Feed-forward multilayer perceptron with one hidden layer. The hidden layer activations (dotted square) are used to calculate the applicability domain.

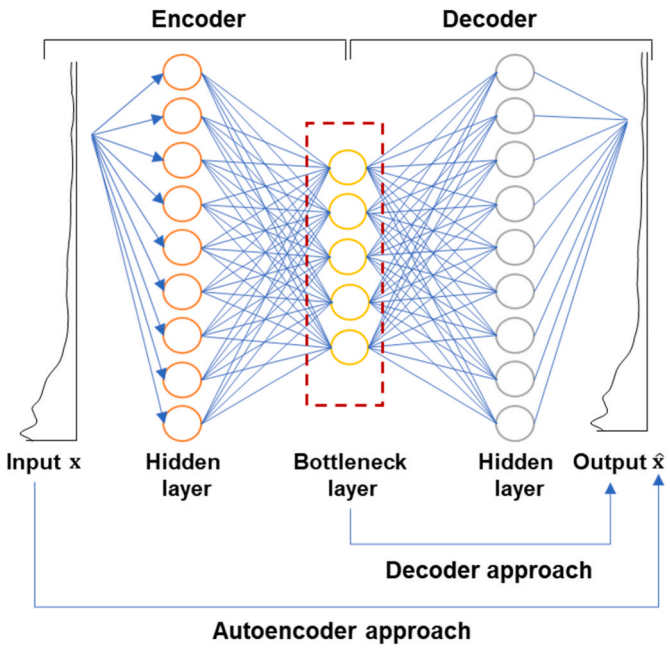


Fig. 2. Autoencoder with three hidden layers. The central layer is the bottleneck layer. The two alternative approaches used to calculate Q_{lim} are indicated.

of the network, from the input layer to the bottleneck layer, performs a nonlinear mapping of the input **x** (e.g., a spectrum) to a reduced latent space so that **x** is represented in fewer dimensions. This block is known as the encoder. The right-hand side of the network, from the bottleneck to the output layer, is the decoder part and uses the compressed representation emerging from the bottleneck to approximately reconstruct the input **x**. The network is trained to reconstruct the input data by minimizing the loss function

$$E = \frac{1}{2}\sum_{m=1}^{M} Q_m \tag{2}$$

where M is the number of training samples and Q_m is the squared reconstruction error of the mth training sample given by [46]

$$Q_m = \sum_{k=1}^{K} (x_{mk} - \hat{x}_{mk})^2 \tag{3}$$

where $\mathbf{x}_m = [x_{m1}, \ldots x_{mk}, \ldots x_{mK}]^{\mathrm{T}}$ is the spectrum of the mth training sample, K is the number of spectral variables, and $\hat{\mathbf{x}}_m = [\hat{x}_{m1}, \ldots \hat{x}_{mk}, \ldots \hat{x}_{mK}]^{\mathrm{T}}$ is the output of the autoencoder. In this work, L_2-norm regularization was added to the loss function in equation (2) to avoid overfitting. With adequate settings, autoencoders can perform, for example, nonlinear principal component analysis [47]. Although autoencoders are usually symmetric [44], symmetry is not strictly necessary, and the encoder and decoder parts may have a different number of layers and nodes.

Forcing the output to be as similar as possible to the input makes autoencoders generalize poorly. This fact can be used to detect anomalies. An autoencoder trained with the spectra that were used for training the regression network will faithfully reconstruct a new sample spectrum only if it resembles the training spectra. An anomalous spectrum containing unmodeled parts will not be reproduced well, and the spectral residuals will be large. A threshold Q_{lim} can be estimated from the spectral residuals of the training spectra so that a new spectrum whose Q value is larger than Q_{lim} is said to be outside the applicability domain of the model. The threshold will depend on the complexity of the autoencoder and the similarity among the training spectra. For a given autoencoder architecture, a training set that consists of very similar spectra will result in a well-fitted autoencoder, small residuals, and a low Q_{lim}. Only very similar new spectra will be reproduced well and the applicability domain will be restricted. A slightly different spectrum will easily be marked as outside the applicability domain. This implies an increased risk of rejecting a good extreme sample (type I error). Conversely, an autoencoder trained with very diverse spectra will have a worse fit and a higher Q_{lim}. The applicability domain will be less tight and more varied spectra will be accepted, increasing the likelihood that a true outlier will go unnoticed (type II error). The number of layers and nodes of the autoencoder will also affect Q_{lim}. For a given training set, a wider bottleneck results in a better fit and a lower Q_{lim} than a narrow bottleneck. Thus, fine-tuning the autoencoder architecture can restrict or widen the applicability domain. The trade-off will depend on the problem at hand.

Note that an autoencoder will reproduce a spectrum independently on the regression network for which we are specifying the applicability domain. In this sense, defining the applicability domain of a regression network using the autoencoder's ability to reproduce spectra is like using the nearest neighbor distance or principal component analysis to define the applicability domain of a partial least squares regression model. None of these methods use the specific form of the regression model to indicate that a spectrum is inside or outside the applicability domain. Since the regression network emphasizes different parts of the spectrum depending on the property to be predicted, the effect of a spectral anomaly on the prediction is different depending on the zone of the spectrum affected. To emphasize the detection of anomalies in the wavelength ranges that are most influential for prediction, an alternative spectral reconstruction approach can be tested. It consists of training only the right-hand block of the autoencoder, the decoder, to reproduce the training spectra. In this case, the input to the decoder for the training sample i can be the activations vector of the hidden layer of the regression network $\mathbf{a}_i$ and the output will be the reconstructed spectrum $\hat{\mathbf{x}}_i$. The decoder is then trained to minimize the reconstruction error given by equation (2). This is a less optimal implementation of an autoencoder, as only the decoder part is trained starting from a loose representation of the spectrum. It results in larger spectral residuals but is an attempt to increase the sensitivity to anomalies at the wavelengths that are relevant for regression.

2.3. Mahalanobis distance

Let $\boldsymbol{X} = [X_1, \ldots, X_p]^{\mathrm{T}}$ be $p \times 1$ random vector with population mean $\boldsymbol{\mu} = E(\boldsymbol{X})$ and covariance matrix $\boldsymbol{\Sigma} = E\left[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^{\mathrm{T}}\right]$. The Mahalanobis distance between $\boldsymbol{X}$ and $\boldsymbol{\mu}$ is given by

$$D(\boldsymbol{X}, \boldsymbol{\mu}) = \left\{(\boldsymbol{X} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu})\right\}^{1/2} \tag{4}$$

This scalar is a generalized distance that measures where a vector $\boldsymbol{X}$ lies with respect to the center of the multivariate space taking into account the correlations among variables. If $\boldsymbol{X}$ is normally distributed, i.e., $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, D^2 follows a chi-square distribution with p degrees of freedom [48]. In practice, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are estimated from the M training samples as $\bar{\mathbf{x}} = \frac{1}{M}\sum_{i=1}^{M} \mathbf{x}_i$ and $\mathbf{S} = \frac{1}{M-1}\sum_{i=1}^{M} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^{\mathrm{T}}$ so the sample version of the squared Mahalanobis distance for observation $\mathbf{x}_i$.

$$D_i^2 = (\mathbf{x}_i - \bar{\mathbf{x}})^{\mathrm{T}}\mathbf{S}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}), i = 1, \ldots, M \tag{5}$$

is only approximately distributed as a χ_p^2 if $\mathbf{x}_i \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ but will asymptotically approach χ_p^2 as M gets larger.

In linear calibration models such as multiple linear regression, principal components regression or partial least squares regression, the Mahalanobis distance is related to the leverage, which is a recommended measure to flag that a new sample spectrum is outside the calibration range [49]. The Mahalanobis distance can be used for the same purpose as the leverage, by setting a limit of the AD as a quantile

(e.g., 0.99) of χ^2_p in case of multivariate normality or a quantile of the D^2 values of the training set if $\boldsymbol{X}$ is not normally distributed. This latter approach resembles the ASTM norm recommendation that a leverage larger than the maximum leverage of the calibration set can be used as an indication of extrapolation of the model.

Although the Mahalanobis distance can be calculated for the spectra, it requires the covariance matrix be nonsingular, which limits its use to cases when the number of samples is larger than the number of variables. Moreover, the fitted model is not considered to define its AD because eq (5) only involves the spectra. A less restrictive option is to calculate the Mahalanobis distance using the activations of the hidden layer of the regression network as

$$D_i^2 = (\mathbf{a}_i - \bar{\mathbf{a}})^{\mathrm{T}} \mathbf{S}^{-1} (\mathbf{a}_i - \bar{\mathbf{a}}) \tag{6}$$

where $\mathbf{a}_i$ is the vector of activations of the hidden layer for sample i and $\bar{\mathbf{a}}$ and $\mathbf{S}$ are the average and covariance matrix of the hidden layer activations of the training set respectively. As noted above, if the underlying distribution of the activations is multinormal, then a quantile of χ^2_d where d is equal to the dimension of $\mathbf{a}$ (that is, the number of nodes in the hidden layer) can be used as a limit of the applicability domain. If D^2 does not follow a chi-square distribution, then a limit D^2_{lim} can be set as a quantile (e.g., 0.99) of the D^2 values of the training set. In all cases, a higher D_i^2 than the limit will indicate that a sample is outside the applicability domain and that its prediction cannot be trusted [50]. This work describes the case when the regression network has only one hidden layer. Such a model was enough for the property being modelled.

2.4. Q spectral residual

The root mean square of spectral residuals has been recommended [49] to detect that a new sample contains interferences not present in the calibration samples. Similarly, for an autoencoder that has been trained to reproduce an input spectrum, define the Q value for a sample [46] as the squared reconstruction error of the spectrum:

$$Q = \sum_{k=1}^{K} (x_k - \hat{x}_k)^2 \tag{7}$$

A new sample will be considered to be outside the AD if the sum of squared spectral residuals from the autoencoder is larger than a limit set from the spectral residuals of the training data. If the spectra follow a multinormal distribution, then Q follows a χ^2_K distribution with the number of degrees of freedom equal to the number of spectral variables K. In that case, a limit Q_{lim} can be set as a certain quantile (e.g., 0.99) of the chi-square distribution. Otherwise, Q_{lim} can be set, for example, as the 0.99 quantile of the Q values of the training set. Thus, a new sample is said to be outside the applicability domain if $Q > Q_{\mathrm{lim}}$.

2.5. Applicability domain of a regression network

The applicability domain of the regression neural network can be established as follows. Once a regression network has been trained to predict a property of interest from the infrared spectrum, the vector of activations of each training sample in the hidden layer of the regression network $\mathbf{a}_i$ (Fig. 1) is kept aside. The activations of all the training samples are then used to calculate $\mathbf{S}$ in equation (6), the squared Mahalanobis distance of each training sample, and the limit of the applicability domain D^2_{lim}. Next, two approaches are proposed to calculate the limit of the applicability domain that is based on the spectral residuals, Q_{lim}. The first approach consists of training an autoencoder (Fig. 2) to reconstruct the training spectra using the training spectra as input. The spectral residuals of the training set are used to define Q_{lim}. Alternatively, only the decoder part of Fig. 2 is trained to reproduce the training spectra using as input the activations $\mathbf{a}_i$ that were used to calculate the Mahalanobis distance. The spectral residuals are used to define Q_{lim}. The spectrum of the new sample is submitted to the regression network to obtain the prediction and the Mahalanobis distance is calculated from the activation of the hidden layer. Next, either the spectrum is submitted to the autoencoder to obtain the reconstructed spectrum $\hat{\mathbf{x}}$ and Q (equation (7)) or the hidden layer activations are submitted to the decoder to obtain the reconstructed spectrum $\hat{\mathbf{x}}$ and Q. The spectrum is inside the applicability domain if $Q \leq Q_{\mathrm{lim}}$ and $D^2 \leq D^2_{\mathrm{lim}}$. Otherwise, the prediction is not reliable and must be verified using the reference analytical method.

3. Material and methods

3.1. Datasets

3.1.1. Diesel dataset

A set of 1792 diesel fuel samples were collected at the Repsol oil refinery in La Pobla de Mafumet, Tarragona (Spain) over 35 months. They were mostly samples from the refinery collector and tanks of the finished product. The density of the samples was determined following the ASTM D4052 method [51] with an ANTON PAAR digital densimeter model DMA 4500 M. The values ranged between 820.0 kg/m^3 and 875.0 kg/m^3. Absorbance spectra between 2591.86 and 1785.76 cm^{-1} with a 4 cm^{-1} resolution were acquired with an Analect Diamond 20 FTIR/FT-NIR process lab analyzer using a flow cell of 0.5 mm pathlength. Each spectrum was the average of 64 scans. A new background spectrum was acquired daily to keep up with baseline shifts and environmental fluctuations. The dataset was randomly split into a training set (899 samples), a validation set (451 samples) and a test set (434 samples) which contained 50 %, 25 % and 25 % of the samples analyzed each month over 35 months. Five spectra measured in a second FTIR/FT-NIR instrument of the same manufacturer were added to the test set. Two discordant spectra resulting from an erroneous manipulation of the sample in the instrument, the mislabelled spectrum of a gasoline that had been labelled as diesel, and a flawed background measurement were also added to the test set.

3.1.2. Tecator dataset

This dataset has been previously used for testing the performance of ANN models [52,53] because of its known nonlinearity. It consists of 240 chopped meat samples with known moisture, fat, and protein content values and near-infrared absorbance spectra between 850 and 1050 nm. The dataset was downloaded from http://lib.stat.cmu.edu/datasets/tecator [54]. The content of fat (0.9–49.1 %) was used in this work. We used a training set of 172 samples and a validation set of 43 samples, as suggested in Ref. [54]. Eight extrapolation samples (also indicated in Ref. [54]) were selected as the test set.

3.2. Data analysis

Calculations were carried out in MATLAB (MathWorks Inc., Natick, MA, R2022a) with MATLAB's Deep Learning ToolboxTM and PLS_Toolbox v7 (Eigenvector Research, Manson, WA).

For the diesel dataset, a regression network was trained to predict the density of the diesel samples from the mid-infrared spectra. The architecture of the network was selected after training different networks with one and two hidden layers, with different combinations of number of nodes in each layer up to 25 nodes. The transfer function of the hidden layers was the hyperbolic tangent. The models were trained by backpropagation with 5000 or more epochs to minimize the MSE with L_2-regularization. Training used stochastic gradient descent with momentum, initial learning rate 0.001, and L_2-regularization factor 10^{-4}. A mini-batch size of 16 was used to evaluate the gradient of the loss function and update the weights. The simplest network with a low MSE in the validation set was selected. Different setups of autoencoder with three hidden layers were tested with an increasing number (up to 25) of

neurons each. To reduce the number of possible structures that could be evaluated, the number of neurons in the bottleneck layer was the number of nodes in the hidden layer of the selected regression network. Except L_2-regularization factor 10^{-3}, the rest of the settings for training the autoencoder were the same as those used for the regression network. For the decoder, different setups of decoder were tested. The training parameters for the decoder were the same as those used for the autoencoder. Both in the autoencoder and the decoder, the number of neurons of the output layer was the dimensions of the spectrum.

For the Tecator dataset, a regression network was trained to predict the fat content using the same procedure as for the diesel dataset, using the NIR spectra after the standard normal variate (SNV) transform. The training approaches for the autoencoder and decoder networks were the same as those used for the diesel dataset.

4. Results and discussion

4.1. Diesel dataset

4.1.1. Neural network models

Fig. 3 shows the MIR region used to establish the regression model. The range 2591.86–1785.76 cm^{-1} contains the fundamental vibrations of groups C–H and C═O present in most compounds in diesel samples [55]. Regression networks with one or two hidden layers with a variable number of nodes were trained to predict density from MIR spectra. The selected model was the network with one hidden layer (Fig. 1) and ten neurons, as it was the simplest model with an acceptable prediction error of the validation set. Only small differences in the validation performance were found by using more than ten neurons or a second hidden layer. Such improvements fluctuated depending on the randomness of the iterative training process based on minibatches. Fig. 4 shows the predicted density for the calibration and validation sets using the selected network. The root mean squared error (RMSE) for the calibration and validation sets was 0.56 kg/m^3 and 0.62 kg/m^3 respectively. The determination coefficient (R^2) of the linear fit between the reference density and the predictions was 0.98 and 0.98 for the training and validation sets, respectively. The prediction ability of the network was comparable to previously reported results [39] with ANNs or other multivariate calibration methods calculated with smaller sets of samples covering shorter production periods. For comparison, the best PLS model performed slightly worse, with an RMSEP of the validation set of 0.70 kg/m^3.

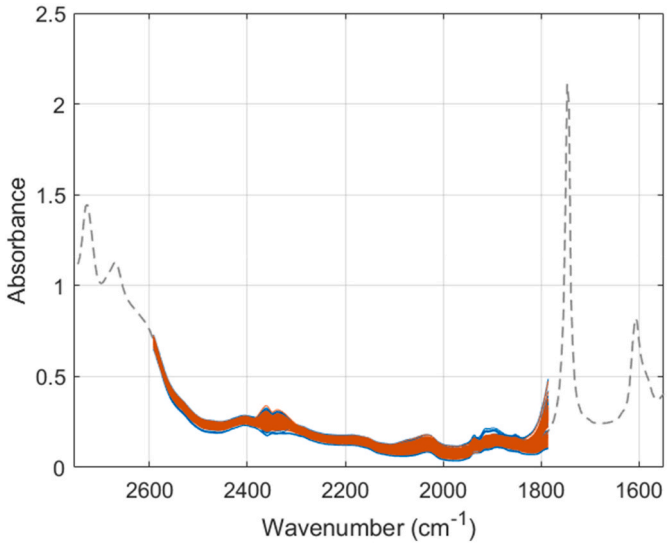


Fig. 3. MIR spectra of the diesel samples: training set (blue) and validation set (orange). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

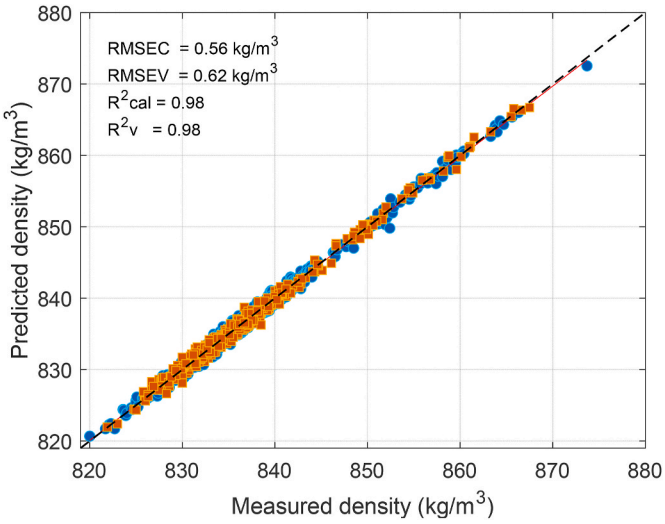


Fig. 4. Predicted versus reference density for the training (blue) and validation (orange) samples. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Different autoencoder architectures with one or three hidden layers were tested to reproduce the training spectra. To keep the number of combinations to be tested low, the number of neurons in the bottleneck was fixed, and it was the same as the number of neurons in the hidden layer in the regression network. The selected autoencoder had 15, 10 and 15 neurons in the first, middle (bottleneck) and third hidden layers respectively. The RMSE for the training and validation sets in the selected model was $8.21 \cdot 10^{-4}$ and $8.47 \cdot 10^{-4}$, respectively. Fig. 5 shows the reconstructed spectra $\hat{\mathbf{x}}$ and the spectral residuals of the calibration and validation samples. For both the calibration and validation sets, the correlation coefficient between each spectrum and the reconstructed spectrum ranged from 0.9992 to 1, confirming that the autoencoder could successfully reproduce the spectra.

The decoder was trained using as inputs the hidden layer activations of the training set from the regression network (a 10×1 data vector per sample). Like in the autoencoder, the output layer returned the reconstructed spectrum and equation (2) with L_2-regularization was the loss function to be minimized. Fifteen different architectures of the decoder were tested with one and two hidden layers, with an increasing number of neurons, up to 25. The simplest network with one hidden layer and 15 neurons was selected as it provided the lowest RMSE for the validation set. The reconstruction error did not improve by using two hidden layers. The RMSE for the training and validation sets in the selected model was $1.43 \cdot 10^{-3}$ and $1.42 \cdot 10^{-3}$, respectively. Fig. 6 shows the spectra estimated by the decoder for the training and validation sets, and the spectral residuals. For both the calibration and validation sets, the correlation coefficient between each spectrum and the estimated spectrum ranged between 0.9986 and 1, confirming that the decoder could also correctly reproduce the original spectra from the ten activation values obtained from the regression network. The spectral residuals are larger than the residuals produced by the autoencoder, which is consistent with the fact that the autoencoder has a more complex structure with more coefficients and is thus expected to fit the training spectra better.

4.1.2. Applicability domain of the regression network

The squared Mahalanobis distance of calibration, validation and test spectra was calculated from the activations of the hidden layer of the regression network (equation (6)). The quantile-quantile plot [56] of the D^2 values of the training samples (Fig. 7) indicated that D^2 does not follow a χ^2_{10} distribution, so the limit of the applicability domain D^2_{lim}

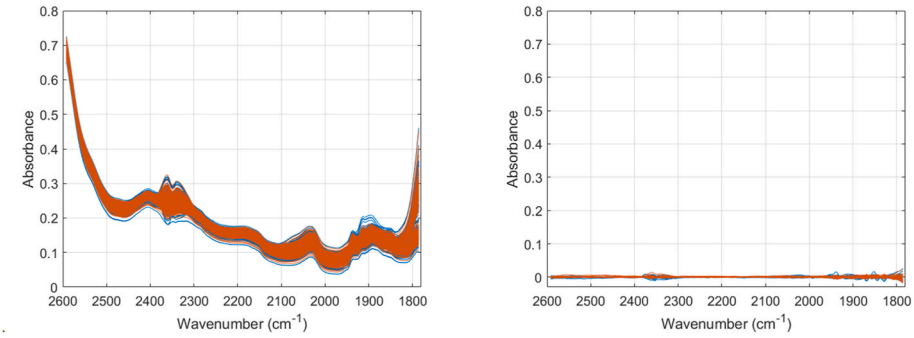


Fig. 5. Reconstructed training spectra and spectral residuals from the autoencoder.

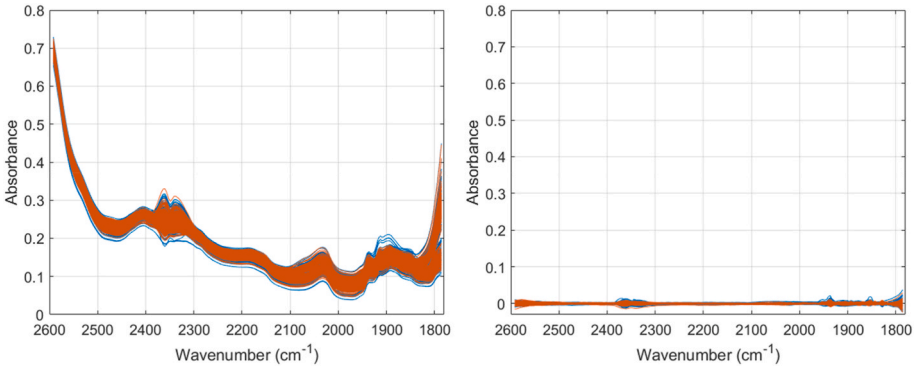


Fig. 6. Reconstructed training spectra and spectral residuals from the decoder.

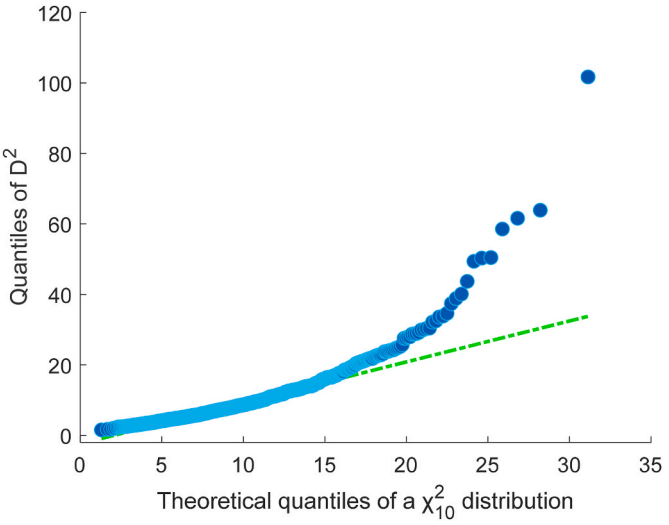


Fig. 7. Quantiles of D^2 vs. theoretical quantiles of a χ^2_{10} distribution.

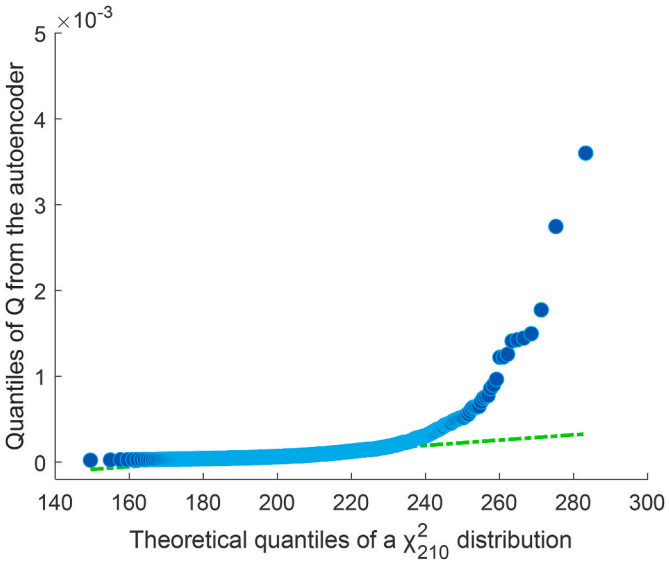


Fig. 8. Quantiles of Q from the autoencoder vs. theoretical quantiles of a χ^2_{210} distribution.

was set at the 0.99 quantile of the D^2 values of the training set, which was 39.49.

The Q values were calculated from the spectral residuals of the autoencoder and the decoder. The quantile-quantile plot indicated that Q for both the autoencoder (Fig. 8) and the decoder (Fig. S1 in the supplementary information) do not follow a chi-square distribution. For the autoencoder, the largest Q value was $3.60{\cdot}10^{-3}$ and the limit $Q_{\text{lim}-\text{AE}}$ was set at $1.20{\cdot}10^{-3}$, the 0.99 quantile of the Q values of the training set. For the decoder, the largest Q value of the training set was $6.60{\cdot}10^{-3}$ and $Q_{\text{lim}-\text{EN}}$ was set at $2.70{\cdot}10^{-3}$, the 0.99 quantile of the Q values of the training set. Note that $Q_{\text{lim}-\text{AE}}$ is lower than $Q_{\text{lim}-\text{EN}}$, which is consistent with the smaller spectral residuals of the autoencoder observed above.

Fig. 9 shows the limits of the AD of the regression network defined by the squared Mahalanobis distance and the spectral residuals of the autoencoder. 1 % of the training samples (that is, 9 samples) had D^2 values higher than D^2_{lim} and 1 % of the training samples had Q values higher than $Q_{\text{lim}-\text{AE}}$. In total, 15 training samples (1.7 % of the set) were outside the AD. It is important to note that these samples were valid samples used in the training step and that their position outside the AD was simply the consequence of using 0.99 quantiles to set the limits of the AD. This indicates that the limits of the applicability domain should

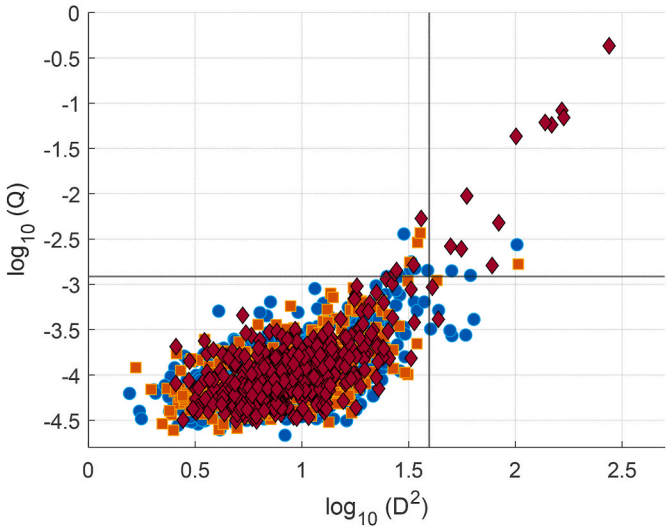


Fig. 9. $\log_{10}(Q)$ of the autoencoder vs. $\log_{10}(D^2)$ for the training (blue), validation (orange) and test diesel samples (red). The limits of the applicability domain of the regression network are shown. The logarithmic scale was used to facilitate the visualization. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

not be interpreted as rigid boundaries that separate good samples from outliers with unreliable predictions. Rather, these limits are the borders of an inner region where the predictions can be confidently accepted.

Only four validation samples had Q values higher than $Q_{\text{lim}-\text{AE}}$ and one of them also had D^2 value higher than D^2_{lim}. Most of the test samples were also within the AD, except 19 samples that had D^2 or Q values exceeding the limits. Four of them stood out only for their high spectral residual ($Q > Q_{\text{lim}-\text{AE}}$) while their D^2 values were lower than D^2_{lim}. Two samples stood out for their high D^2 values ($D^2 > D^2_{\text{lim}}$) while having a low spectral residual ($Q < Q_{\text{lim}-\text{AE}}$). The thirteen remaining samples had simultaneously D^2 and Q values exceeding the limits. Fig. 10 shows the spectra of these thirteen highly discordant samples, compared with the spectra of the training samples. Among these thirteen, five were the ones that had been measured with the second instrument and three were the rare spectra that had been added to the test set. This showed the ability of the AD to flag discordant spectra.

Fig. 11 shows the empirical cumulative distribution function of the

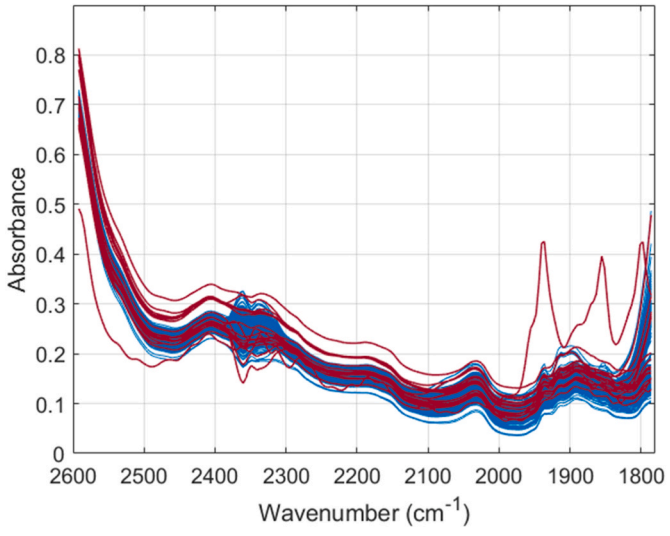


Fig. 10. Spectra of the samples outside the applicability domain compared to the range of the spectra of the training samples.

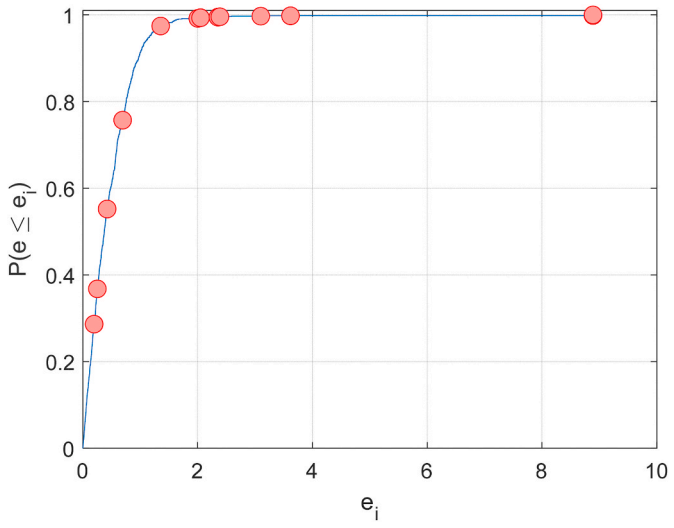


Fig. 11. Empirical cumulative distribution function of the absolute prediction error for density.

absolute prediction error of the density for each sample $e_i = |y_i - \hat{y}_i|$ of the training and validation set together. As expected for regular spectra, there is not a high correlation between the absolute prediction error and the D^2 and Q values as they all are part of the modelled variability of the training set. Hence, a range of prediction errors are possible, as indicated by the cumulative curve. On the other hand, anomalous spectra that have high values of Q and D^2 (that is, they are outside the AD) are more susceptible to produce large prediction errors. This is what is observed in Fig. 11 with the thirteen test samples that were outside both AD limits simultaneously. Nine of them had absolute errors larger than 1.35 kg/m^3, that is, worse than 97 % of the errors of the training and validation sets. The other four test samples that were outside the AD had small absolute errors, in accordance with the idea that the AD cannot be regarded as the space outside which we have the absolute certainty that the prediction errors will be large in all the cases, but as a zone outside which there is a higher risk of large prediction errors and thus, the predictions should not be trusted.

As an alternative to the autoencoder, which is trained independently on the regression network, the spectral residual limit of the AD was also calculated from the decoder part of an autoencoder, calculated from the activations of the hidden layer of the regression network. The decoder reconstructs the spectra worse than the autoencoder, resulting in larger residuals and a larger Q_{lim}. The AD limits set from the Mahalanobis distance and the decoder also detected the thirteen discordant spectra in the test set with a Q value larger than Q_{lim} that were also flagged by the autoencoder (Fig. S2 in the supplementary information) but we did not observe any apparent improvement in the detection of outlying samples by using the decoder as compared to using the autoencoder. The reason was that the decoder did not necessarily had the largest weights at the nodes that received the largest (in absolute value) activations from the hidden layer of the regression network. Therefore, small anomalies at the important wavelengths, that resulted in defective activations (and hence, could be detected by the Mahalanobis distance) did not translate into large spectral residuals in a more sensitive way than for the autoencoder.

4.2. Tecator dataset

Fig. 12 shows the spectra of training and test sets with the spectra of the extrapolation samples highlighted. The optimal regression model consisted of one hidden layer with 16 neurons. Fig. 13 shows the predicted fat content for the calibration, validation, and test sets using this network. The RMSE for the calibration and validation sets were 0.76 %

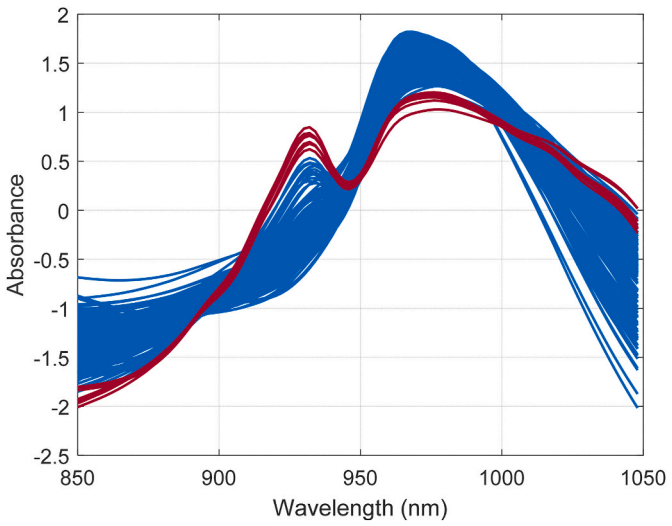


Fig. 12. Near-infrared spectra of Tecator samples: training set (blue), and test set (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

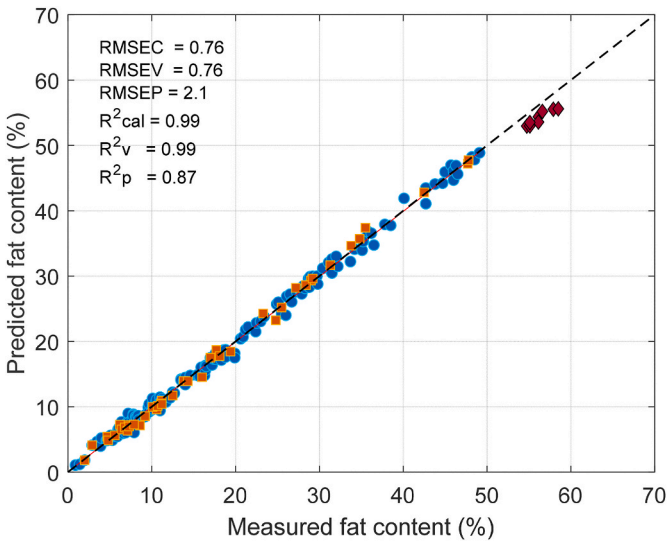


Fig. 13. Predicted versus reference fat content for the training (blue), validation (orange) and test (green) samples. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

($R^2 = 0.99$) and 0.76 % ($R^2 = 0.99$), respectively, which are similar to those reported in previous studies [9,57,58]. The RMSE for the test set was 2.1 % ($R^2 = 0.87$) because the spectra represent an extrapolation of the model.

Fig. 14 shows the empirical cumulative distribution function of the absolute prediction error for the fat content of the training and validation samples. As can be seen, the test set spectra had absolute errors larger than 1.40 %, worse than 90 % of the errors of the training and validation sets. For these samples, a high correlation between the absolute prediction error and the D^2 and Q values is expected. This is observed in Fig. 15, which shows the limits of the AD of the model defined by D^2 and Q calculated from the activations of the regression network and autoencoder (with 32, 16, and 32 neurons in the hidden layers), respectively. As can be seen, all test samples were outside both AD limits, as one could expect since the samples were known to be extrapolations. Very similar results were obtained using the AD limits set from D^2 and the decoder network of one hidden layer with 32 neurons

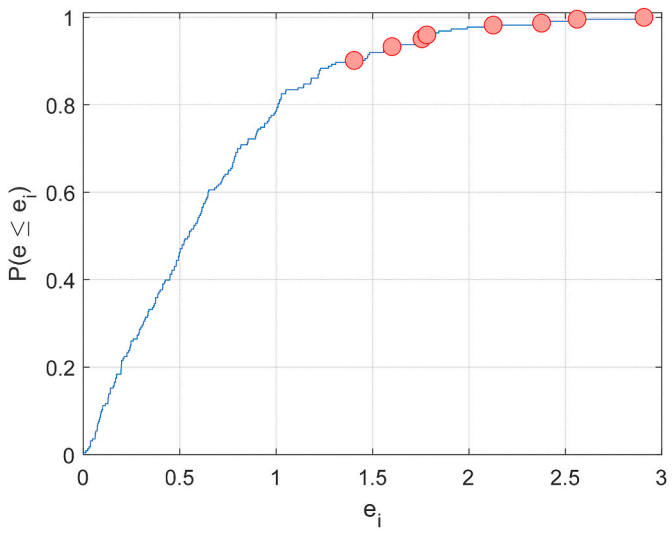


Fig. 14. Empirical cumulative distribution function of the absolute prediction error for fat content.

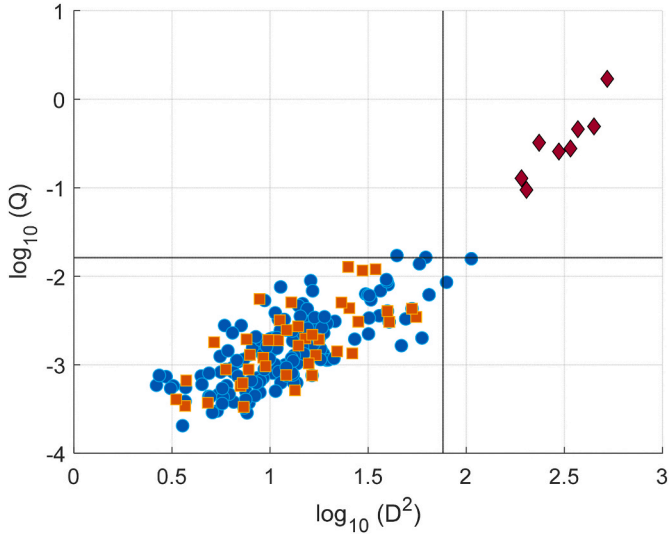


Fig. 15. $\log_{10}(Q)$ of the autoencoder vs. $\log_{10}(D^2)$ for the training (blue), validation (orange) and test meat samples (red). The limits of the applicability domain of the regression network are shown. The logarithmic scale was used to facilitate the visualization. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

(Fig. S3 in the supplementary information).

5. Conclusions

Routine quantitative determinations based on multivariate spectroscopy and multivariate calibration require safeguards to ensure that the accepted size of the prediction errors is maintained during the use of the model. One of these safeguards is the characterization of the applicability domain of the model. Beyond the limits of this domain, the prediction of a sample is not considered to be reliable enough. This does not necessarily mean that the sample will produce a large prediction error, but it is susceptible to it. This work has presented an approach to finding these limits in a feed-forward neural network, arguably the most common and simple network used in chemical analysis. The limits were set as the 0.99 quantiles of two metrics calculated from the training data. One was the squared Mahalanobis distance calculated from the activations of the hidden layer of the regression network. This measured the

distance of the new sample to the center of the model. The second was the sum of squared spectral residuals obtained by reconstructing the spectrum from a lower dimensional space. For the latter, two approaches were studied, either using an autoencoder or a decoder. The autoencoder was trained independently on the regression network, while the decoder used the activations from the hidden layer of the regression network. Both approaches indicated how well the prediction spectrum could be reproduced from the training data. Large spectral residuals indicated a spectrum outside the applicability domain and a questionable prediction.

When applied to the prediction of the density of diesel fuel samples from their MIR spectra using a neural network, thirteen spectra were identified outside the AD with Q and D^2 values higher than the limits simultaneously. Nine of them were found to have high prediction errors. Although their prediction should not be trusted for the remaining samples, the prediction errors were not abnormally high. The second study case was the prediction of the fat content of meat samples from the NIR spectra. All test spectra, known as extrapolations, were outside the AD and had high prediction errors. While the initial hypothesis that the decoder might outperform the autoencoder could not be confirmed, both methodologies had a consistent behavior.

CRediT authorship contribution statement

M. Suliany Rodríguez-Barrios: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Joan Ferré:** Writing – review & editing, Writing – original draft, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **M. Soledad Larrechi:** Writing – review & editing, Supervision, Methodology, Investigation. **Enric Ruiz:** Writing – review & editing, Resources, Funding acquisition, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

The authors acknowledge the financial support of REPSOL PETRÓLEO, S.A., (contract T19221S) and AGAUR's 2021-SGR00705. M. Suliany acknowledges the financial support from Universitat Rovira i Virgili for providing Marti Franqués Research Fellowship (2021PMF-PIPF-1).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemolab.2024.105242.

References

[1] F. Despagne, D.L. Massart, Neural networks in multivariate calibration, Analyst 123 (1998) 157R–178R, https://doi.org/10.1039/A805562I.
[2] D.A. Cirovic, Feed-forward artificial neural networks: applications to spectroscopy, Trends Anal. Chem. 16 (1997) 148–155, https://doi.org/10.1016/S0165-9936(97)00007-1.
[3] D. Pérez-Marín, A. Garrido-Varo, J.E. Guerrero, Non-linear regression methods in NIRS quantitative analysis, Talanta 72 (2007) 28–42, https://doi.org/10.1016/j.talanta.2006.10.036.
[4] F. Marini, Artificial neural networks in foodstuff analyses: trends and perspectives. A review, Anal. Chim. Acta 635 (2009) 121–131, https://doi.org/10.1016/j.aca.2009.01.009.
[5] J. Yang, J. Xu, X. Zhang, C. Wu, T. Lin, Y. Ying, Deep learning for vibrational spectral analysis: recent progress and a practical guide, Anal. Chim. Acta 1081 (2019) 6–17, https://doi.org/10.1016/j.aca.2019.06.012.
[6] Y. Chen, L. Song, Y. Liu, L. Yang, D. Li, A review of the artificial neural network models for water quality prediction, Appl. Sci. 10 (2020) 5776–5825, https://doi.org/10.3390/app10175776.
[7] P. Mishra, D. Passos, F. Marini, J. Xu, J.M. Amigo, A.A. Gowen, J.J. Jansen, A. Biancolillo, J.M. Roger, D.N. Rutledge, A. Nordon, Deep learning for near-infrared spectral data modelling: hypes and benefits, Trends Anal. Chem. 157 (2022) 116804–116829, https://doi.org/10.1016/j.trac.2022.116804.
[8] L. Yang, T. Kavli, M. Carlin, S. Clausen, P.F. De Groot, An evaluation of confidence bound estimation methods for neural networks, in: Advances in Computational Intelligence and Learning. International Series in Intelligent Technologies, vol. 18, Springer, Dordrecht, The Netherlands, 2002, pp. 71–84, https://doi.org/10.1007/978-94-010-0324-7_5.
[9] F. Allegrini, A.C. Olivieri, Sensitivity, prediction uncertainty, and detection limit for artificial neural network calibrations, Anal. Chem. 88 (2016) 7807–7812, https://doi.org/10.1021/acs.analchem.6b01857.
[10] F.A. Chiappini, F. Allegrini, H.C. Goicoechea, A.C. Olivieri, Sensitivity for multivariate calibration based on multilayer perceptron artificial neural networks, Anal. Chem. 92 (2020) 12265–12272, https://doi.org/10.1021/acs.analchem.0c01863.
[11] K. Shariat, D. Kirsanov, A.C. Olivieri, H. Parastar, Sensitivity and generalized analytical sensitivity expressions for quantitative analysis using convolutional neural networks, Anal. Chim. Acta 1192 (2022) 338697–338706, https://doi.org/10.1016/j.aca.2021.338697.
[12] N. Fjodorova, M. Novič, A. Roncaglioni, E. Benfenati, Evaluating the applicability domain in the case of classification predictive models for carcinogenicity based on the counter propagation artificial neural network, J. Comput. Aided Mol. Des. 25 (2011) 1147–1158, https://doi.org/10.1007/s10822-011-9499-9.
[13] F. Sahigara, K. Mansouri, D. Ballabio, A. Mauri, V. Consonni, R. Todeschini, Comparison of different approaches to define the applicability domain of QSAR models, Molecules 17 (2012) 4791–4810, https://doi.org/10.3390/molecules17054791.
[14] N. Minovski, Š. Župerl, V. Drgan, M. Novič, Assessment of applicability domain for multivariate counter-propagation artificial neural network predictive models by minimum Euclidean distance space analysis: a case study, Anal. Chim. Acta 759 (2013) 28–42, https://doi.org/10.1016/j.aca.2012.11.002.
[15] M. Mathea, W. Klingspohn, K. Baumann, Chemoinformatic classification methods and their applicability domain, Mol. Inform. 35 (2016) 160–180, https://doi.org/10.1002/minf.201501019.
[16] P. Žuvela, J. David, M.W. Wong, Interpretation of ANN-based QSAR models for prediction of antioxidant activity of flavonoids, J. Comput. Chem. 39 (2018) 953–963, https://doi.org/10.1002/jcc.25168.
[17] R. Liu, H. Wang, K.P. Glover, M.G. Feasel, A. Wallqvist, Dissecting machine-learning prediction of molecular activity: is an applicability domain needed for quantitative structure-activity relationship models based on deep neural networks? J. Chem. Inf. Model. 59 (2019) 117–126, https://doi.org/10.1021/acs.jcim.8b00348.
[18] Y. Tian, S. Zhang, H. Yin, A. Yan, Quantitative structure-activity relationship (QSAR) models and their applicability domain analysis on HIV-1 protease inhibitors by machine learning methods, Chemometr. Intell. Lab. Syst. 196 (2020) 103888–103902, https://doi.org/10.1016/j.chemolab.2019.103888.
[19] Organisation for Economic Co-operation and Development, Guidance document on the validation of (quantitative) structure-activity relationship [(Q)SAR] models, OECD Series on Testing and Assessment 69 (2014) 1–154, https://doi.org/10.1787/20777876.
[20] T.I. Netzeva, A.P. Worth, T. Aldenberg, R. Benigni, M.T.D. Cronin, P. Gramatica, J. S. Jaworska, S. Kahn, G. Klopman, C.A. Marchant, G. Myatt, N. Nikolova-Jeliazkova, G.Y. Patlewicz, R. Perkins, D.W. Roberts, T.W. Schultz, D.T. Stanton, J. J.M. van de Sandt, W. Tong, G. Veith, C. Yang, Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. The report and recommendations of ECVAM Workshop 52, Altern. Lab. Anim. 33 (2005) 155–173, https://doi.org/10.1177/026119290503300209.
[21] M.A.F. Pimentel, D.A. Clifton, L. Clifton, L. Tarassenko, A review of novelty detection, Signal Process. 99 (2014) 215–249, https://doi.org/10.1016/j.sigpro.2013.12.026.
[22] M. Markou, S. Singh, Novelty detection: a review—part 1: statistical approaches, Signal Process. 83 (2003) 2481–2497, https://doi.org/10.1016/j.sigpro.2003.07.018.
[23] M. Markou, S. Singh, Novelty detection: a review: part 2: neural network based approaches, Signal Process. 83 (2003) 2499–2521, https://doi.org/10.1016/j.sigpro.2003.07.019.
[24] V.J. Hodge, J. Austin, A survey of outlier detection methodologies, Artif. Intell. Rev. 22 (2004) 85–126, https://doi.org/10.1007/s10462-004-4304-y.
[25] V. Chandola, A. Banerjee, V. Kumar, Anomaly detection: a survey, ACM Comput. Surv. 41 (2009) 1–58, https://doi.org/10.1145/1541880.1541882.
[26] H. Wang, M.J. Bah, M. Hammad, Progress in outlier detection techniques: a survey, IEEE Access 7 (2019) 107964–108000, https://doi.org/10.1109/ACCESS.2019.2932769.
[27] A. Boukerche, L. Zheng, O. Alfandi, Outlier detection: methods, models, and classification, ACM Comput. Surv. 55 (2020) 1–37, https://doi.org/10.1145/3381028.

[28] G. Pang, C. Shen, L. Cao, A. Van Den Hengel, Deep learning for anomaly detection: a review, ACM Comput. Surv. 54 (2021) 1–36, https://doi.org/10.1145/3439950.
[29] J. Jaworska, N. Nikolova-Jeliazkova, T. Aldenberg, QSAR applicabilty domain estimation by projection of the training set descriptor space: a review, Altern. Lab. Anim. 33 (2005) 445–459, https://doi.org/10.1177/026119290503300508.
[30] N. Nikolova, J. Jaworska, Approaches to measure chemical similarity - a review, QSAR Comb. Sci. 22 (2004) 1006–1026, https://doi.org/10.1002/qsar.200330831.
[31] M. Toplak, R. Močnik, M. Polajnar, Z. Bosnić, L. Carlsson, C. Hasselgren, J. Demšar, S. Boyer, B. Zupan, J. Stålring, Assessment of machine learning reliability methods for quantifying the applicability domain of QSAR regression models, J. Chem. Inf. Model. 54 (2014) 431–441, https://doi.org/10.1021/ci4006595.
[32] L. Shen, D. Cao, Q. Xu, X. Huang, N. Xiao, Y. Liang, A novel local manifold-ranking based K-NN for modeling the regression between bioactivity and molecular descriptors, Chemom. Intell. Lab. Syst. 151 (2016) 71–77, https://doi.org/10.1016/j.chemolab.2015.12.005.
[33] F. Sahigara, D. Ballabio, R. Todeschini, V. Consonni, Defining a novel k-nearest neighbours approach to assess the applicability domain of a QSAR model for reliable predictions, J. Cheminform. 5 (2013) 1–9, https://doi.org/10.1186/1758-2946-5-27.
[34] R. Todeschini, D. Ballabio, V. Consonni, F. Sahigara, P. Filzmoser, Locally centred Mahalanobis distance: a new distance measure with salient features towards outlier detection, Anal. Chim. Acta 787 (2013) 1–9, https://doi.org/10.1016/j.aca.2013.04.034.
[35] I. Sushko, S. Novotarskyi, R. Körner, A.K. Pandey, A. Cherkasov, J. Li, P. Gramatica, K. Hansen, T. Schroeter, K.R. Müller, L. Xi, H. Liu, X. Yao, T. Öberg, F. Hormozdiari, P. Dao, C. Sahinalp, R. Todeschini, P. Polishchuk, A. Artemenko, V. Kuz'Min, T.M. Martin, D.M. Young, D. Fourches, E. Muratov, A. Tropsha, I. Baskin, D. Horvath, G. Marcou, C. Muller, A. Varnek, V.V. Prokopenko, I. V. Tetko, Applicability domains for classification problems: benchmarking of distance to models for Ames mutagenicity set, J. Chem. Inf. Model. 50 (2010) 2094–2111, https://doi.org/10.1021/ci100253r.
[36] S.Y. Shin, H. Kim, Extended autoencoder for novelty detection with reconstruction along projection pathway, Appl. Sci. 10 (2020) 4497, https://doi.org/10.3390/app10134497.
[37] P.S. Vasafi, O. Paquet-Durand, K. Brettschneider, J. Hinrichs, B. Hitzmann, Anomaly detection during milk processing by autoencoder neural network based on near-infrared spectroscopy, J. Food Eng. 299 (2021) 110510, https://doi.org/10.1016/j.jfoodeng.2021.110510.
[38] Z. Boger, Selection of quasi-optimal inputs in chemometrics modeling by artificial neural network analysis, Anal. Chim. Acta 490 (2003) 31–40, https://doi.org/10.1016/S0003-2670(03)00349-0.
[39] V.O. Santos Jr., F.C.C. Oliveira, D.G. Lima, A.C. Petry, E. Garcia, P.A.Z. Suarez, J. C. Rubim, A comparative study of diesel analysis by FTIR, FTNIR and FT-Raman spectroscopy using PLS and artificial neural network analysis, Anal. Chim. Acta 547 (2005) 188–196, https://doi.org/10.1016/j.aca.2005.05.042.
[40] N. Pasadakis, S. Sourligas, C. Foteinopoulos, Prediction of the distillation profile and cold properties of diesel fuels using mid-IR spectroscopy and neural networks, Fuel 85 (2006) 1131–1137, https://doi.org/10.1016/j.fuel.2005.09.016.
[41] H.A.G. Al-kaf, K.S. Chia, N.A.M. Alduais, A comparison between single layer and multilayer artificial neural networks in predicting diesel fuel properties using near infrared spectrum, Pet. Sci. Technol. 36 (2018) 411–418, https://doi.org/10.1080/10916466.2018.1425717.
[42] C. Borggaard, H.H. Thodberg, Anal. Chem. 64 (1992) 545–551, https://doi.org/10.1021/ac00029a018.
[43] H.H. Thodberg, IEEE Trans. Neural Network. 7 (1996) 56–72, https://doi.org/10.1109/72.478392.
[44] Ch C. Aggarwal, Neural Networks and Deep Learning, Springer, Cham, 2018, https://doi.org/10.1007/978-3-319-94463-0.
[45] J. Schmidhuber, Deep Learning in neural networks: an overview, Neural Network. 61 (2015) 85–117, https://doi.org/10.1016/j.neunet.2014.09.003.
[46] S. Hawkins, H. He, G. Williams, R. Baxter, Outlier detection using replicator neural networks, in: Proceedings of the 4th International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2000), LNCS 2454, Springer Berlin Heidelberg, Berlin, Heidelberg, 2002, pp. 170–180, https://doi.org/10.1007/3-540-46145-0_17.
[47] M.A. Kramer, Nonlinear principal component analysis using autoassociative neural networks, AIChE J. 37 (1991) 233–243, https://doi.org/10.1002/aic.690370209.
[48] K.V. Mardia, J.T. Kent, J.M. Bibby, Multivariate Analysis, first ed., Academic Press London, 1979.
[49] ASTM E1655-17, Standard Practices for Infrared Multivariate Quantitative Analysis, ASTM Int., 2017, pp. 1–29, https://doi.org/10.1520/E1655-17.
[50] H. Moeini, F.M. Torab, Comparing compositional multivariate outliers with autoencoder networks in anomaly detection at Hamich exploration area, east of Iran, J. Geochemical Explor. 180 (2017) 15–23, https://doi.org/10.1016/j.gexplo.2017.05.008.
[51] D40052-15, Standard Test Method for Density, Relative Density , and API Gravity of Liquids by Digital Density Meter, ASTM Int., 2013, pp. 1–8, https://doi.org/10.1520/D4052-18A.2.
[52] T. Chen, J. Morris, E. Martin, Chemometrics and Intelligent Laboratory System 87 (2007) 59–71, https://doi.org/10.1016/j.chemolab.2006.09.004.
[53] N. Hernández, I. Talavera, A. Dago, R.J. Biscay, M.M.C. Ferreira, D. Porro, J. Chemometr. 22 (2008) 686–694, https://doi.org/10.1002/cem.1168.
[54] Carnegie Mellon University, "Tecator dataset," [Online]. Available: http://lib.stat.cmu.edu/datasets/tecator.
[55] R.M. Silverstein, F.X. Webster, D.J. Kiemle, Spectrometric identification of organic compounds, Microchem. J. 21 (4) (2005) 496.
[56] M.B. Wilk, R. Gnanadesikan, Probability plotting methods for the analysis of data, Biometrika 55 (1968) 1–17, https://doi.org/10.2307/2334448.
[57] W. Ni, L. Nørgaard, M. Mørup, Non-linear calibration models for near infrared spectroscopy, Anal. Chim. Acta 813 (2014) 1–14, https://doi.org/10.1016/j.aca.2013.12.002.
[58] S. Malek, F. Melgani, Y. Bazi, One-dimensional convolutional neural networks for spectroscopic signal regression, J. Chemom. 32 (2018) 1–17, https://doi.org/10.1002/cem.2977.