Refuse Derived Fuels: Using Near-infrared Spectroscopy and Machine Learning to Predict Moisture Content and Dry Particle Spectra

J. Fischer*,[1], T. Kunz, K. Treiber** and V. Scherer***

1 Corresponding author, E-Mail: fischer@leat.rub.de

* Ruhr-Universität Bochum, Bochum, Germany

** VDZ Technology gGmbH, Düsseldorf, Germany

Abstract

Refuse derived fuel (RDF) sourced from industrial and municipal waste contains primarily plastics, paper/cardboard and textiles. Since RDF has high heating values and low prices, it is used as alternative fuel to coal in cement production, offering CO_2 reductions. However, the cement industry currently lacks real-time quality control of RDF, resulting in unnoticed reductions in heating value caused by elevated moisture content. Moisture content determination by drying takes up to 4 hours, which does not allow for quick reactions in operation of the cement kiln.

This study addresses the challenge by applying near-infrared spectroscopy (NIRS) to predict the moisture content of paper and cardboard particles, which retains most moisture in RDF. The particles were moisturized by different methods with NIRS measurements at each stage. For the prediction, different approaches ranging from linear regression of selected wavelength absorptions to multiple machine learning regression techniques like ridge, lasso and multilayer perceptron are compared in accuracy and feature representation. The multilayer perceptron showed accurate results and robust feature selection, although also simple linear regression at selected wavelength gave adequate accuracy.

In a second step, moisture content and spectra of wet particles were used to predict the spectra of dry particles. These spectra are essential to correctly identify particle types or predict heating values, since training data consists of dry particle data. While linear models are not reaching acceptable results, both an autoencoder and a random forest regressor yielded spectra with correct shapes, which is sufficient for classification with derivative spectra.

1. Introduction

Cement clinker production is an extremely energy intensive process with a thermal fuel energy consumption of 2787 MJ/t of clinker [1]. Most of the energy is required in the calciner and cement kiln with process temperatures up to 1450 °C. This energy is provided by the direct combustion of fuel. The traditional fuel, powdered coal, is increasingly replaced by alternative fuels like refuse derived fuel (RDF) from waste. In Germany, 71.6 % of the fuel by energy input in 2022 came from alternative fuels, with RDF as biggest contributor [1]. This replacement is not only due to significantly lower fuel prices, but also due to lower CO_2 emissions. The CO_2 emissions for RDF is around 50 g CO_2 per provided MJ energy [2], while hard coal produces 94 g CO_2 / MJ [3] and lignite 111 g CO_2 / MJ [3]. The reason for the lower CO_2 emissions is mainly a lower C/H ratio of RDF. Furthermore, a portion of the RDF is biomass-based (paper, cotton), and the net CO_2 emissions of these biomasses are considered zero, as they are renewable.

Since waste streams vary in composition and quality depending on location, time and supplier, RDF streams also experience similar fluctuations. This variability poses challenges in a) controlling the cement kiln operation to ensure that varying RDF compositions do not influence cement clinker quality and b) further increasing the ratio of alternative fuels, especially RDF, in the cement production. Important RDF quality parameters, such as composition, moisture content or heating value are often either unknown or only available with long time delays.

Moisture content in RDF also varies, typically ranging between 10 and 30 wt% [4–6], but in some cases reaching up to 35 % [7]. In [8], it was found that moisture does not distribute evenly across all RDF fractions. For example, plastic foil or rubber absorb significantly less moisture compared to paper and cardboard. Hence, the paper and cardboard particles in RDF may have significantly higher moisture content than the mean 35 %.

RDF moisture content is usually either measured thermogravimetrically in a drying oven [9], mostly as part of a proximate analysis [7,9–11]. Another method used for moisture determination is Thermogravimetric analysis (TGA) [10,12]. However, both of these methods are both labor-intensive and time-consuming, making it unsuitable for near real-time control. The residence time of RDF from the feeding point to the cement kiln is only few minutes, whereas the drying oven method can take 2-4 hours to deliver results. TGA may be faster, but still takes 30 to 90 minutes, depending on the sample preparation.

Another method to measure the moisture content in particles is near-infrared spectroscopy (NIRS). NIRS is a non-destructive measurements technique in which a sample is exposed to radiation in the near-infrared region (800 nm – 2500 nm). The reflected radiation is then measured by a sensor, e.g. an indium gallium arsenide (InGaAs) sensor [13]. Samples absorb a part of the radiation by anharmonic molecular vibrations. The corresponding absorbance spectra provide information about the molecular structure of the sample. Water has a NIRS absorbance peak at 1450 nm [13,14], where the first overtone of the stretching band of OH (2 $\nu_{1,3}$) is located [14]. Additional peaks for water can be found at 970 nm, 1190 nm and 1940 nm [14]. An example for a pure water NIR spectrum can be found for example at Muncan [14]. NIRS moisture content measurements were successfully applied to, inter alia, coal [15,16], wood [17,18] and nylon [19]. For RDF with an uncommonly low moisture content of only up to 4 %, NIRS was used in combination with a convolutional neural network [20]. In this study, we aim to predict the water content for the paper and cardboard fraction of RDF for higher

moisture contents using NIRS measurements. Additionally, we seek to predict the spectra of dry particles based on the spectra of the wet particles.

2. Experimental

2.1 Sample preparation

For the regression of moisture content in refuse derived fuel, it is assumed that moisture is mainly found in the fraction of paper and cardboard [8]. Therefore, the experiment was performed on 55 model RDF cardboard particles of dimensions 24 x 11 x 1 mm (mean dimensions from sorting analysis [21]). The process diagram of the sample preparation is shown in Figure 1. In the first step, the particles were dried in a drying chamber at 105 °C for three days. Afterwards, their dry weight was determined. The spectra of all 55 particles were measured with the Viavi MicroNIR 1700 ES with an integration time of 9.2 ms and a resolution of 6 nm (resulting in 125 data points per spectrum). For each particle, three orientations were recorded, with an average of 100 measurements per orientation taken against a black background.

To introduce water uniformly into the particles, different methods were selected: In the first method (particles no. 1-25), a pipette was used to drop water at the cardboard surface and distribute it. After measuring the NIR spectrum, the particles were air dried for 3 hours, before another NIR spectrum was measured. This results in two different moisture contents for each particle. In the second method (particles no. 26-55), the particles were held in a chamber with air saturated with water for a defined time of 20 min, 40 min, 1 h, 2 h, 3 h, 4 h and 5 h, yielding higher moisture content than with the pipette method. To gain a second moisture content, the particles of the humidity chamber were submerged in water for 1h.

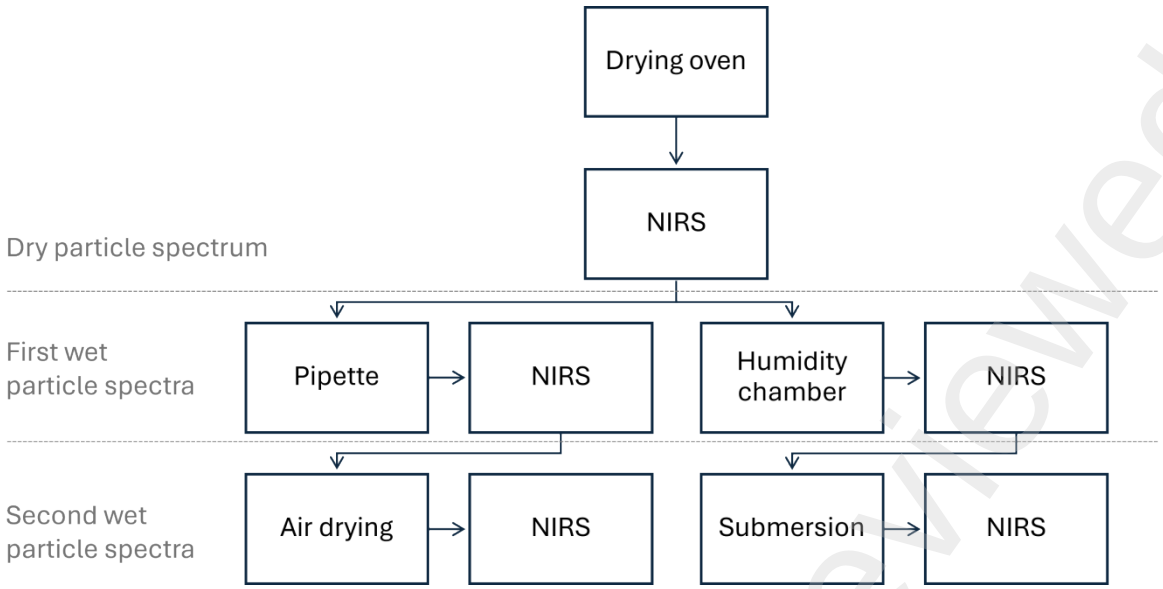


Figure 1: Process diagram of the sample preparation: First, dry particles NIR spectra are measured. Then, different moisture content particles are produced by either pipette and air drying (left) or humidity chamber and submersion (right).

The chamber consists of an ultrasonic nebulizer immersed in a water reservoir, and three propellers to ensure thorough mixing. The particles are held in a wire cage. A sketch of the chamber is shown in Figure 2. In both cases, the added water m_w was measured by back weighing as the difference between the wet and dry particle, (Eq. 1).

$$m_w = m_{wet} - m_{dry}. \tag{1}$$

The moisture content f is then calculated with Eq. (2).

$$f = \frac{m_w}{m_{dry}} \tag{2}$$

The sample preparation results in 55 samples with each 3 orientations x 3 moisture contents (including 1 dry moisture content = 0 %) x 125 wavelength = 1125 datapoints per sample.

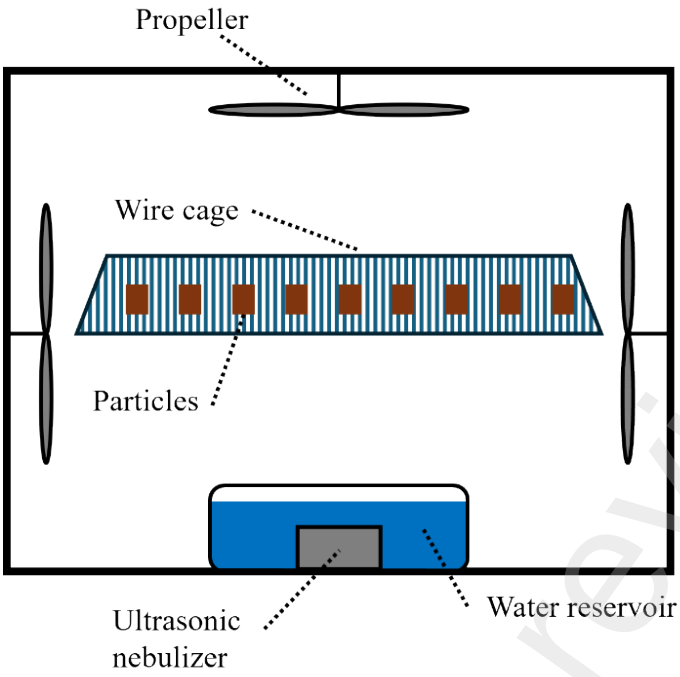


Figure 2: Sketch of the chamber to humidify particles: An ultrasonic nebulizer immersed in water reservoir provides a steady stream of mist, which is mixed in the chamber by three propellers. The particles are held by a wire cage.

2.2 Post-processing

All spectra measured were then post-processed in the same way:

1) Absorbance below 1000 nm and above 1650 nm are excluded due to too much measurement noise.
2) Outliers equal to the background were removed. This was done manually due to the small dataset.
3) Calculation of derivative with Savitzky-Golay filter (window length =15, polyorder=3, scipy python module [22]).

Together with the values for moisture content *f*, the spectra were used to evaluate different strategies for regression of *f* using either the spectra themselves or the derivative of the spectra. All regression algorithms, except for linear regression (numpy [23]), are part of the sklearn python module [24]. The data was randomly split into training data (80 %) and test data (20 %).

2.3 Moisture content regression

2.3.1 Simple Linear regression

For the simple linear regression, the data were fitted to the form of Eq. (3) with the ordinary least square minimization in Eq. (4).

$$\hat{y} = \beta_0 + \beta_1 x \quad (3)$$

$$min_\beta RSS = min_\beta \sum_i^n (y_i - \hat{y}_i)^2 \quad (4)$$

x symbolizes the absorbance respectively the derivative of the absorbance at one wavelength, β_i are the fitting parameter. The minimization problem is then solved by its analytical form.

2.3.2 Multiple linear regression

For the multiple linear regression, the equations of simple linear regression are transferred to a vector notation [24] in equations (5) and (6):

$$\hat{\boldsymbol{y}} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \quad (5)$$

$$min_{\boldsymbol{\beta}} RSS = min_{\boldsymbol{\beta}} ||\hat{\boldsymbol{y}} - \boldsymbol{y}||_2^2 \quad (6)$$

The minimization problem in Eq. (6) is then solved by singular value decomposition (SVD) [24].

2.3.3 Ridge

In the ridge regression, also known as Tikhonov regularization, equation (6) is expanded by a penalty term of the square of the ℓ_2 norm, Eq. (7), of coefficients $\boldsymbol{\beta}$ to Eq. (8). The penalty term limits the absolute value of the coefficients and prevents overfitting [24]:

$$||\boldsymbol{\beta}||_2 = \sqrt{\beta_1^2 + \beta_2^2 + ... + \beta_n^2} \quad (7)$$

$$min_{\boldsymbol{\beta}}||\hat{\boldsymbol{y}} - \boldsymbol{y}||_2^2 + \alpha||\boldsymbol{\beta}||_2^2 \quad (8)$$

$\alpha \geq 0$ is the regularization parameter and is selected as 1e-3 for the absorbance values and 1e-6 for the derivatives. The minimization problem is also solved with singular value decomposition [24].

2.3.4 Lasso

The idea of the Least Absolute Shrinkage and Selection Operator Regression (lasso) is the same as for ridge regression, but here the penalty term is the ℓ_1 norm (Eq. (9)) of the coefficients β [24], resulting in the minimization problem in Eq. (10).

$$||\beta||_1 = \sum_i^n |\beta_i| \quad (9)$$

$$min_{\boldsymbol{\beta}}\frac{1}{2n}||\hat{\boldsymbol{y}} - \boldsymbol{y}||_2^2 + \alpha||\beta||_1 \quad (10)$$

This results in values of zero for unimportant features. α is selected as 1e-2 for absorbance and 1e-3 for derivatives. The minimization problem in Eq. (10) cannot be solved by SVD since it is not differentiable. Instead, the coordinate descent method is used [24].

2.3.5 Multi-layer perceptron (MLP)

A different approach is the multi-layer perceptron, which is a densely connected neural network with feed forward. In this case, a network structure of input layers, one hidden layer with 125 neurons and an output layer, was used. The Limited Broyden–Fletcher–Goldfarb–Shanno (LBFGS) algorithm is selected with a learning rate of 0.001. Each neuron of the hidden layer uses rectified linear unit (ReLU), shown in Eq. (11) as activation function [24] to reflect non-linear correlations.

$$f(x) = \max(0,x) \tag{11}$$

Regularization in an MLP is, like ridge, an ℓ_2 norm. α was chosen as 1e-4 for absorbance values and 1e-6 for derivative values.

2.3.6 Permutation Importance

Since only absorbance at few wavelengths are expected to be influenced by the presence of water (see section 1), the importance of the features **X** is checked by permutation. Permutation is a two-step procedure: In the first step, the model is fitted to all features and the score of the prediction s (here: R^2 score) is calculated. In a second step, for each feature the training data are shuffled randomly, the model is retained with the shuffled data, and the resulting score is compared to the score of the full model. The feature importance i_j of feature j can then be calculated with Eq. (12) where K is the number repetition of this procedure.

$$i_j = s - \frac{1}{K}\sum_k^K s_{k,j}\,, \tag{12}$$

Here, 125 repeats with all samples are executed and afterwards averaged [24].

2.4 Dry spectra regression

The spectra of the wet particles, along with their moisture content, are used as input variables to predict the spectra of the dry particles. These spectra are necessary for other machine learning applications for RDF, since RDF samples are usually dried to stabilize them biologically, such that all training data are for dry RDF. The applications include, but is not limited to, the classification into different fractions like paper, foil or hard plastics (analogous to [25]), the estimation of heating values or a general fuel quality assessment for the usage in the cement production.

Both inputs (spectra wet particles, moisture content) are combined into an array, resulting in an input vector with a length of 106 (105 wavelengths + 1 moisture content). The output vector has therefore a length of 105. In this task, the algorithm must predict 105 values instead of a single one. For this task, the linear models (multiple linear regression, ridge, lasso) failed. As the problem does not show linear behavior, more complex algorithms were employed, specifically a random forest algorithm and an autoencoder neural network.

2.4.1 Random Forest

A random forest regressor is, as the name implies, an ensemble of many decision trees *T* [26], with this contribution using 50 decision trees. A decision tree consists of connected decision nodes and grows from the root node to the leaf nodes. The last grown nodes, leaf nodes, are responsible for the prediction. For the creation of a node from the available features *N,* the

feature n is selected, which divides the data set M with the lowest mean square error (MSE) into M_0 and M_1. For this purpose, the optimization problem of the Classification and Regression Tree (CART) in eq. (13) is solved at each node [27], such that the weighted mean square errors MSE_0 and MSE_1 are minimal.

$$\min_{n,t_n} \ \frac{m_0}{m} MSE_0 + \frac{m_1}{m} MSE_1 \tag{13}$$

Here, the indices denote the sub data sets of M, where m represents the number of data points in these sub data sets. If the sub data sets M_0 and M_1 contain enough data points, they are again split by solving the optimization problem into M_{00} and M_{01}, respectively M_{10} and M_{11}. Otherwise, the node is considered a leaf, which will produce the prediction value. In a random forest, the available set of features n is selected randomly, hence all decision trees are random. The final prediction value of the random forest is calculated as the average prediction of all decision trees.

2.4.2 Autoencoder

Autoencoder is a type of neural network architecture primarily used to learn a dense representation of the training data. The data enter the network in the input layer, are modified in the network and exit it through the output layer. The network can be separated into three parts: Encoder, latent space and decoder. A scheme of an autoencoder is shown in Figure 3.

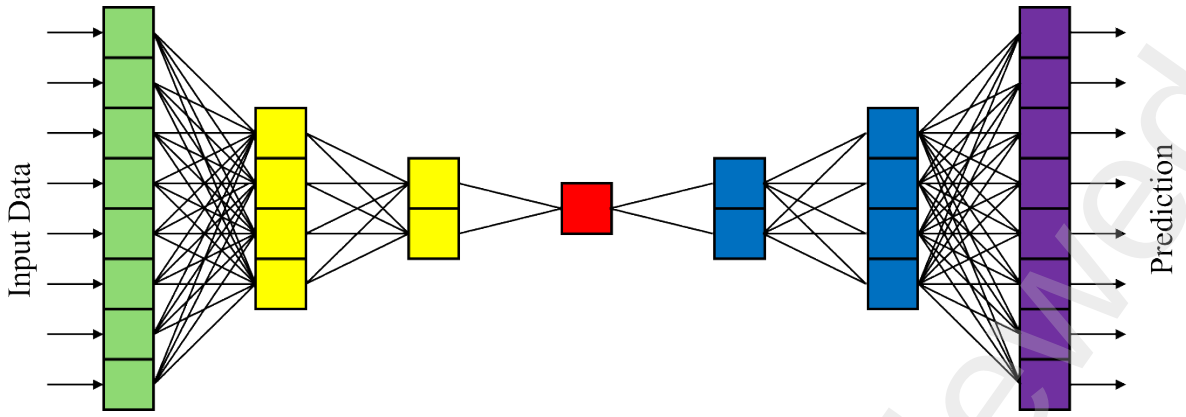


Figure 3: Scheme of an autoencoder. The data flow can be read from left to right: The input data enters the network in the input layer (green), which is part of the encoder. The rest of the encoder (yellow) consists of dense connected layers. In the middle, the latent space (red), the layer size is minimal. After that, the decoder part with more densely connected layers (blue) increases the dimension. The last layer of the decoder part is the output layer (purple), which can produce a prediction.

The encoder's first layer is the input layer, followed by dense layers with less neurons than the input layer. In the latent space layer, the number of neurons is significantly smaller than that in the input layer. Hence, after successful training, the training data can be represented by only the values in the latent space. Following the latent space, the decoder is symmetrical mirrored to the encoder with increasing neuron numbers in each layer, but with the output layer as the last layer. In the standard formulation, the output layer of the decoder is the same size as the input layer in the encoder. This can be used to remove unnecessary information from the data. The output values of the decoder can afterwards be used as input for other algorithms, e.g., for classification. In this contribution, the input layer has 106 neurons (105 wavelengths and 1 moisture content per sample), while the output layer has only 105 neurons. The decoder part is hence adapted, such that it serves as regression with the latent space values as input values. The topography of the used neural network is shown in Table 1. All layers except the output layer

use a rectified linear unit (ReLU), Eq. (11) as the activation function, i.e., a function to introduce non-linearity. In the output layer, the sigmoid function, Eq. (14), is used.

$$f(x) = \frac{1}{1 + e^{-x}} \tag{14}$$

The implementation was done with Keras [28] and Tensorflow [29]. The data was split into 80 % training data, 10 % validation data and 10 % test data. The loss function, which is minimized during training, was the mean square error, Eq. (15).

$$MSE = \frac{1}{n}\sum_{i}^{n} (\hat{y} - y)^2 \tag{15}$$

Table 1: Topography of the autoencoder

	Encoder			Latent Space		Decoder	
Layer	1 (Input)	2	3	4	5	6	7 (Output)
Neurons	106	64	32	32	32	64	105
Activation function	ReLU	ReLU	ReLU	ReLU	ReLU	ReLU	Sigmoid

3. Results and Discussion

3.1 Sample preparation

In Figure 4, an overview of the moisture content values achieved in the samples are shown. The pipette technique provided lower moisture contents, ranging from 6 % to 25 %, while the moisture chamber provided higher moisture content, ranging from 80 % to 140 %. After being exposed to air for 3 h, the particles 1-25 were measured again. Their moisture content decreased

on average by 13 percent point. A part of the particles (36-45) was again measured after being submerged in water for 1h. These particles reached a moisture content up to 156 %. This provides a good range (4 % - 156 %) of training data, although it is missing values in the range of 50 % to 100 %.

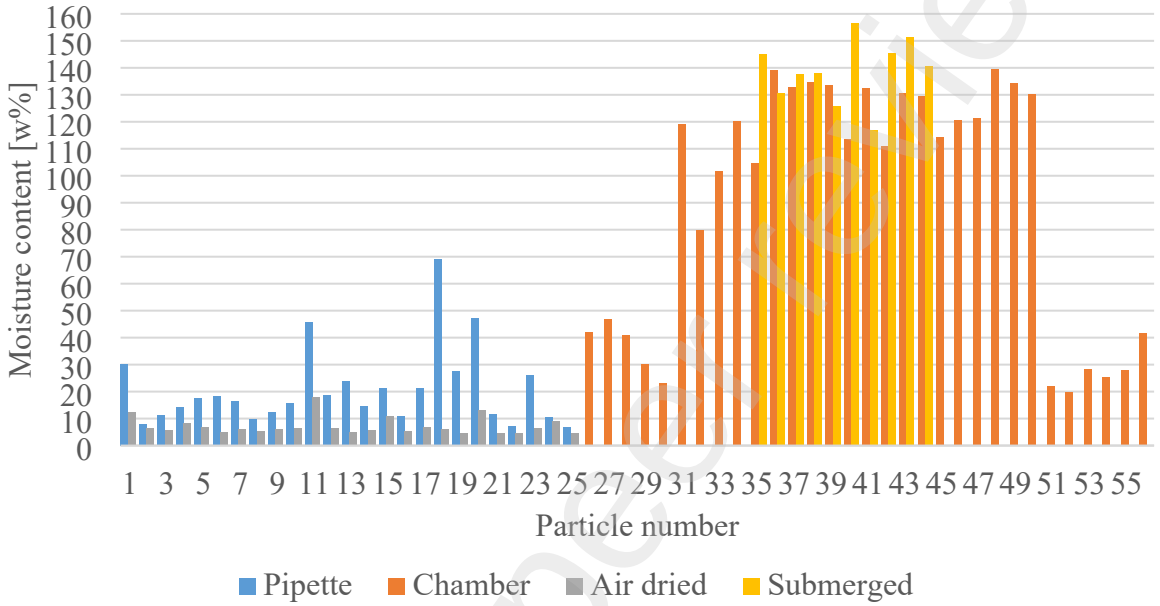


Figure 4: Moisture content of model paper particles by different methods to change moisture content: Applying water drops by pipette (blue), air drying of these particles (grey), humidifier chamber (orange) and submerging to water (yellow).

The absorbance spectra and their derivatives of the dry and wet samples were averaged, with the result shown in Figure 5. At wavelength up to 1300 nm, the spectra do not deviate much in shape, only absolute value. The dry samples spectra show a broad peak in absorbance at wavelength 1400 nm to 1650 nm, while the wet samples spectra show a narrower peak at 1453 nm. The derivatives for wet samples have a pronounced maximum at 1403 nm

(approximately 5 times higher than that observed for the dry samples) and a minimum at 1502 nm (approximately 20 times higher than for the dry samples).

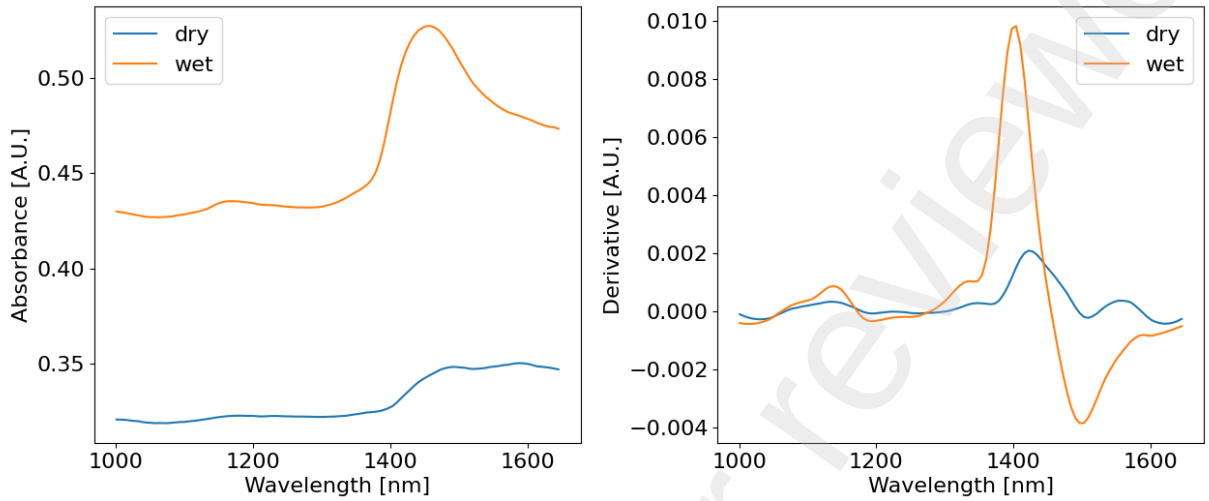


Figure 5: Mean absorbance (left) and derivatives of absorbance (right) over wavelength.

For each sample different orientations were measured, which should be expected to be only marginally different. Therefore, they can be used to estimate an upper bound for the standard deviation. For this, the standard deviation for each sample at each wavelength was calculated with all orientations. The mean of this standard deviation for the dry particles was 0.045 and 0.035 for the wet particles.

3.2 Moisture content regression

3.2.1 Simple Linear Regression

For the simple linear regression, three specific wavelengths were selected based on the peaks of the spectra of the wet sample: 1453 nm for absorbance, and 1385 nm (maximum) and 1471 nm (minimum) for the derivatives. The results are shown in Table 2 with R^2 values from 0.89 up to 0.94. Therefore, it is not absolutely necessary to measure the full NIR spectrum, but only distinct wavelengths, which correspond to the specific wavelength of water at 1450 nm [14].

In Figure 6, the derivatives of the measurement values at wavelength 1471 nm and the corresponding linear regression are shown. Since the Pearson correlation coefficient for all three tested wavelengths is high (0.95, 0.95, -0.96), the correlation is linear, and no polynomial fitting is needed.

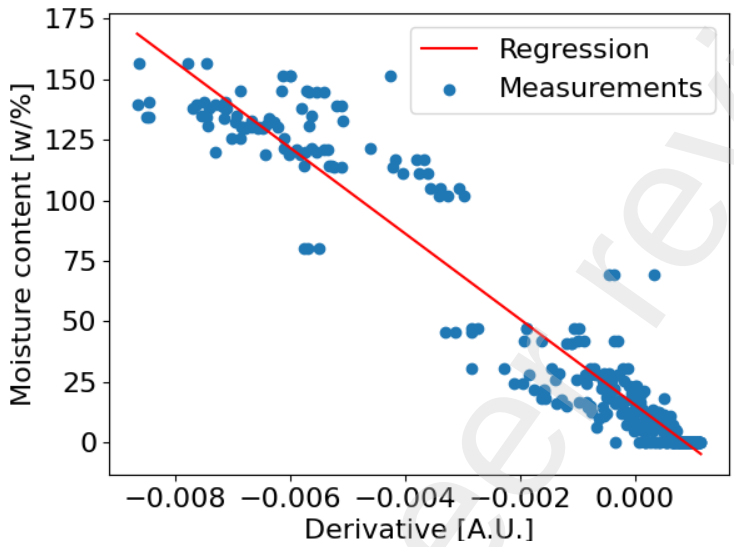


Figure 6: Moisture content over 1. derivative values at 1471 nm, absorbance derived after wavelength by Savitzky-Golay filter (blue) and the corresponding simple linear regression (red).

Table 2: Results for simple linear regression at three manual selected wavelengths.

Wavelength [nm]	Absorbance or Derivative	β_0	β_1	R^2
1453	Absorbance	-83.33	254.30	0.89
1385	Derivative	-1.69	7014.35	0.91
1471	Derivative	15.31	-17741.90	0.94

3.2.2 Multiple Regression

The R^2 score for all algorithms used are shown in Table 3. For the use of the absorbance data, all four algorithms show a high R^2 score in range 0.94 to 0.96 with MLP reaching the best score.

In Figure 7, the feature importances, determined by permutation, are shown. For the linear regression, the feature importance gets to high values up to 92703. This is possible since the R^2 score can be infinitely bad and the coefficients of the multiple linear regression are not restricted in any form, such that the coefficients β_i also reach large values in the range of -71479 to 53443. Surprisingly, the maximum values for feature importance are at 1403 nm and 1535 nm, rather than at the maximum absorbance values located at the center of the peak. For both ridge and MLP regression, the feature importance resembles the shape of the absorbance peaks. Since both use ℓ_2 regularization, the coefficients are not as poorly distributed as for the multiple linear regression, providing a good balance between emphasizing important features and keeping a robustness against noise. For lasso regression, the distribution of feature importance is even more extreme with all values equal zero except 1298-1316 nm, 1422-1440 nm and 1582-1632 nm. This is due to the ℓ_1 regularization of the lasso algorithm.

In Figure 8, the same is shown for the derivatives. Here, although reaching a high R^2 score, the feature importance of multiple linear regression does not resemble the shape of the derivative spectra at all with even higher values for the coefficients. The lasso regression features are reduced to eight features at the main peaks at 1398 nm and 1470 nm as well as smaller peaks at 1192 nm and 1124 nm. Only the feature 1044 nm does not represent any peak. Lasso can therefore be applied to find promising feature candidates for the simple linear regression. Ridge and MLP again resemble the shape of the derivative peaks. A special detail for ridge can be seen at the peak at 1500 nm, where the feature importance next to the peak is high, but at the peak maximum is near 0. All R^2 scores are in the acceptable range with both absorbance and derivative data. Nevertheless, the distribution of feature importance gives good indication about the robustness. Multiple linear regression overrates certain features, while lasso reduces to very few used features. Both MLP and ridge use ℓ_2 regularization, which results in high, but broad enough distributed feature importances at peaks. This should provide a certain level of

robustness against measurement noise. Tests with the second derivative instead of the first derivative yielded very similar results.

Table 3: Result of different regression algorithms on the absorbance and derivatives data as R^2.

Method	R^2 Test Absorbance	R^2 Test Derivative
Multiple linear regression	0.94	0.94
Ridge	0.95	0.95
Lasso	0.95	0.95
MLP	0.96	0.96

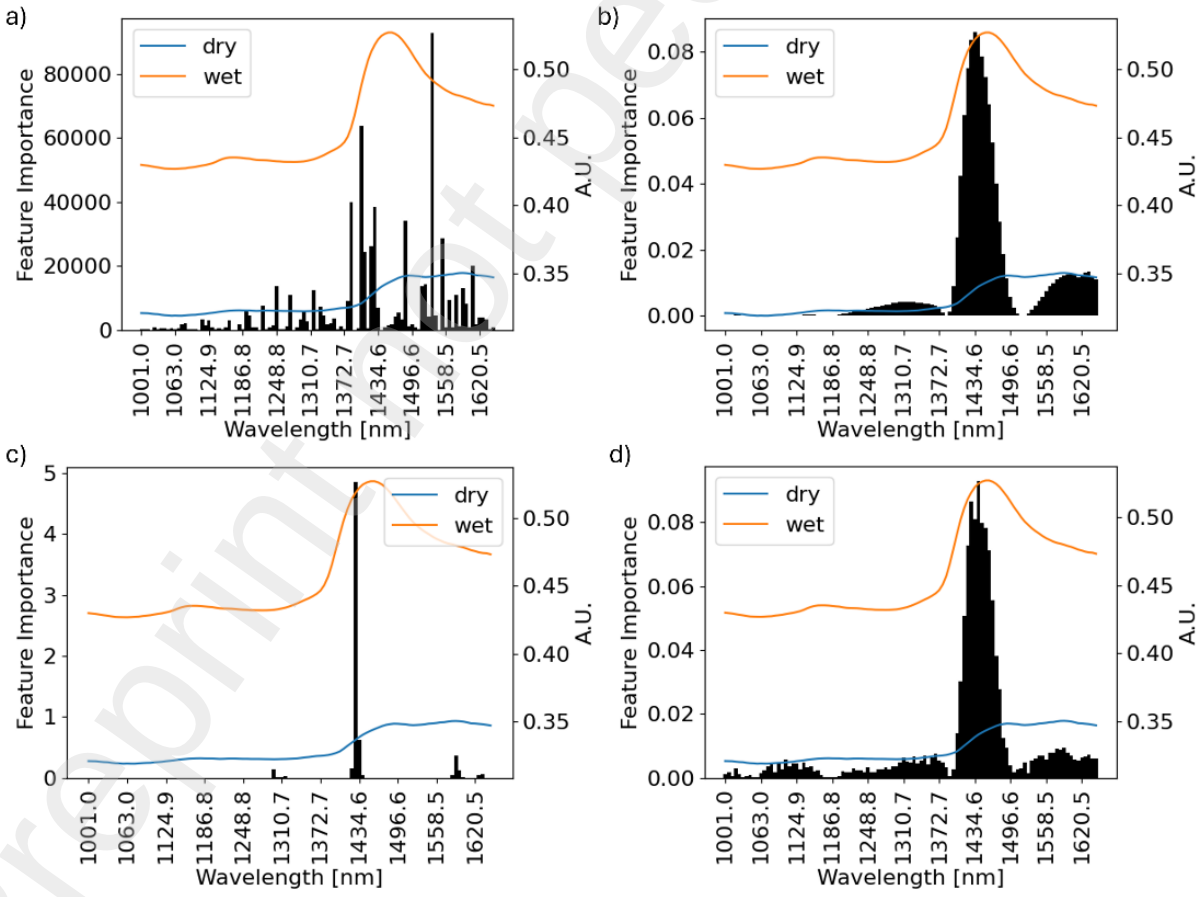


Figure 7: Feature importances for different regression algorithms with the corresponding

absorbances over wavelength: a) Multiple linear regression, b) ridge regression, c) lasso regression and d) MLP regression. Left y axis for feature importance, right y axis for absorbance.

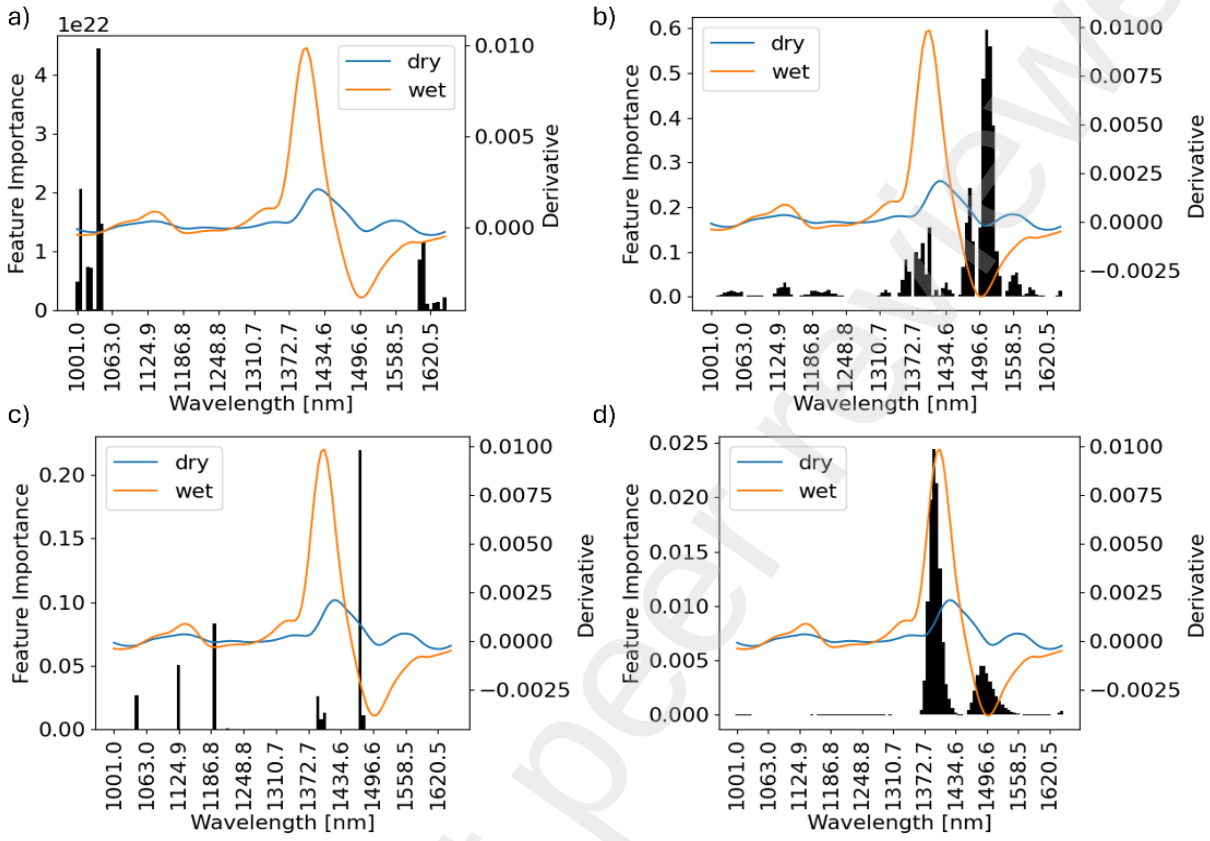


Figure 8: Feature importances for different regression algorithms with the corresponding derivatives of absorbances over wavelength: a) Multiple linear regression, b) ridge regression, c) lasso regression and d) MLP regression. Left y axis for feature importance, right y axis for derivative.

3.3 Dry Spectra regression

In Figure 9, two examples, particles no. 2 and 5, are shown for the regression of the dry spectra from wet spectra and moisture content data with both the random forest and the adapted autoencoder. The first thing to notice is that the overall shape of the dry spectra is well predicted by both examples and algorithms. This can be quantified by the Pearson R between predicted and real wet spectra, which measures the collinearity, for both methods were calculated. Both

yielded high Pearson R, while the random forest with R=0.996 is slightly better than the Autoencoder (R=0.0968). The absolute values of the absorbance are estimated in average 0.015 A.U. too high (autoencoder), respectively only 8e-4 too low (random forest). What is also obvious from the examples is that the random forest produces smoother spectra than the autoencoder. This noisy spectrum can be smoothed by applying a Savitzky-Golay filter, shown also in Figure 9. In Table 4, the mean R^2 scores, the mean squared error and Pearsons R for both algorithms based on the test data are shown. In addition, random forest regressors with less decision trees were tested, namely 10 trees (R^2: 0.69, MSE: 5.9e-5) and 5 trees (R^2: 0.67, MSE: 6.2e-5), which yield slightly less accurate values.

It can be seen that the random forest yields a higher average R^2 score with 0.73 than the autoencoder with 0.35, a smaller MSE and a higher Pearsons R. For this task, therefore, the random forest is suitable, since it yields better results and can easily be parallelized. Autoencoder with different architecture may yield better results.

Table 4: R^2, mean square error and Pearsons R of spectra regression with autoencoder and random forest.

Method	R^2	MSE	Pearson R
Random Forest	0.73	5.10e-04	0.996
Autoencoder	0.35	1.24e-03	0.968

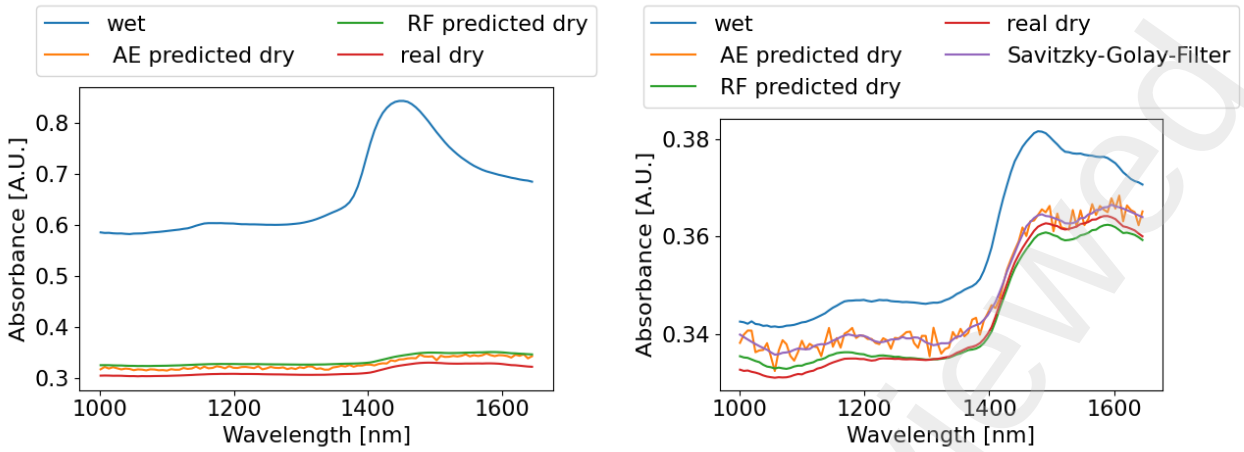


Figure 9: Two examples of predictions of a spectrum of a dry spectrum based on the wet spectrum and the moisture content with an autoencoder (AE) and a random forest (RF). Particle no. 5 (moisture content 17.8 %) on the left and particle no. 2 (moisture content 7.8 %) on the right. Since the AE prediction for particle 2 is oscillating, a Savitzky-Golay filter was applied.

4. Conclusion

Refuse derived fuel (RDF) moisture content, usually measured thermogravimetrically, can also be effectively determined through near-infrared spectroscopy (NIRS). This study demonstrates that NIRS provides accurate moisture content measurements in RDF, even across a wide range of moisture content (4 % - 156 %). Both absorbance data and its derivate can be used to predict moisture contents with good results for the algorithms multiple linear regression, ridge, lasso and multilayer perceptron. Nevertheless, predictions from ridge and multilayer perceptron appear to be more robust than those of the other algorithms, as their feature importances closely resemble the shape of the absorbance and derivative curves. Also, linear regression at selected distinct wavelengths delivers precise predictions of the moisture content of the particles. In a second step, the spectra of dry particles were predicted based on the spectra of wet particles with both an adapted autoencoder and a random forest regressor. The random forest regressor provided more accurate and smoother spectra, which could potentially be used in the future to correctly classify RDF fractions. To further improve the models, a larger sample size with

different shaped and sized particles should be examined. For more robust measurements, the influence of the temperature to the regression should be calibrated, since the OH group peaks slightly shift with the temperature [30]. Also, the influence of the presence of other RDF fractions like plastics requires more investigation.

Nomenclature

m	mass
f	moisture content
y	output value
x	input value
β	regression coefficient
RSS	residual sum of squares
α	regularization parameter
ε	regression bias
MSE	mean squared error
i	feature importance
s	score of prediction
N	set of features
M	subset of N
K	repetitions feature importance

Subscripts

w	water
wet	wet particle
dry	dry particle

Superscripts

ˆ predicted

Acknowledgements

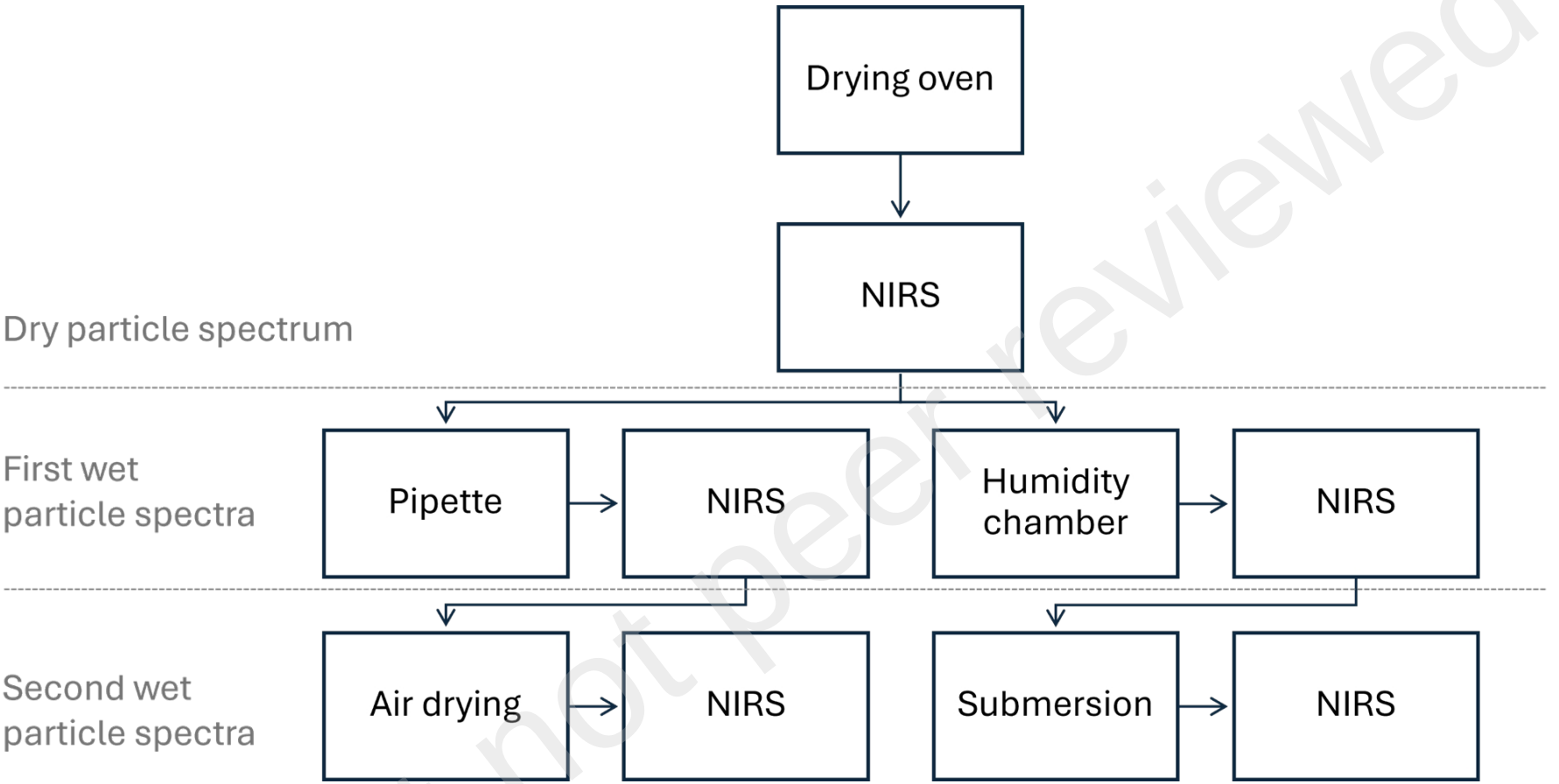
The IGF Project (01IF22676N) is supported within the program for promoting the Industrial Collective Research (IGF) of the German Ministry of Economic Affairs and Energy.

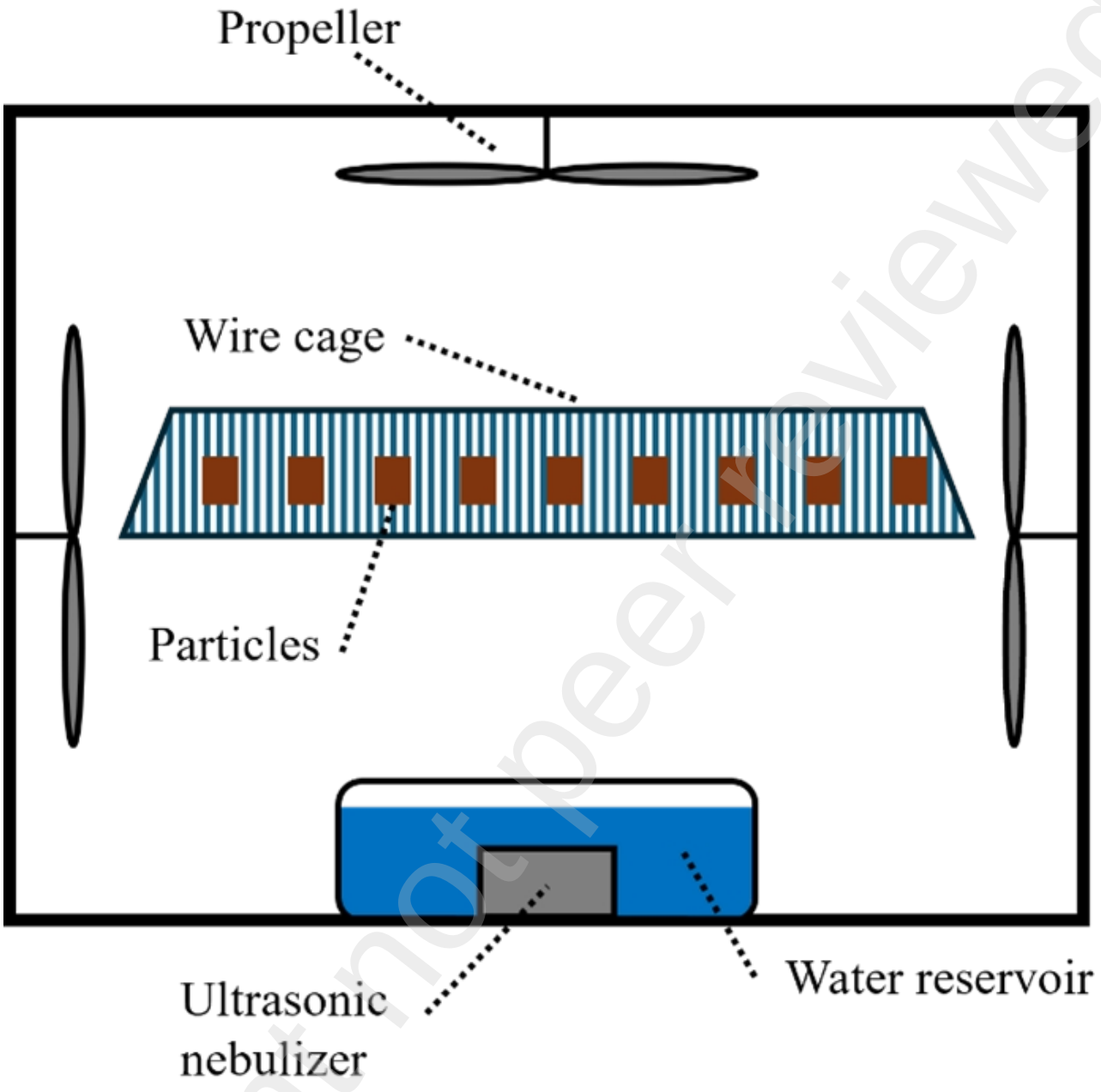
References

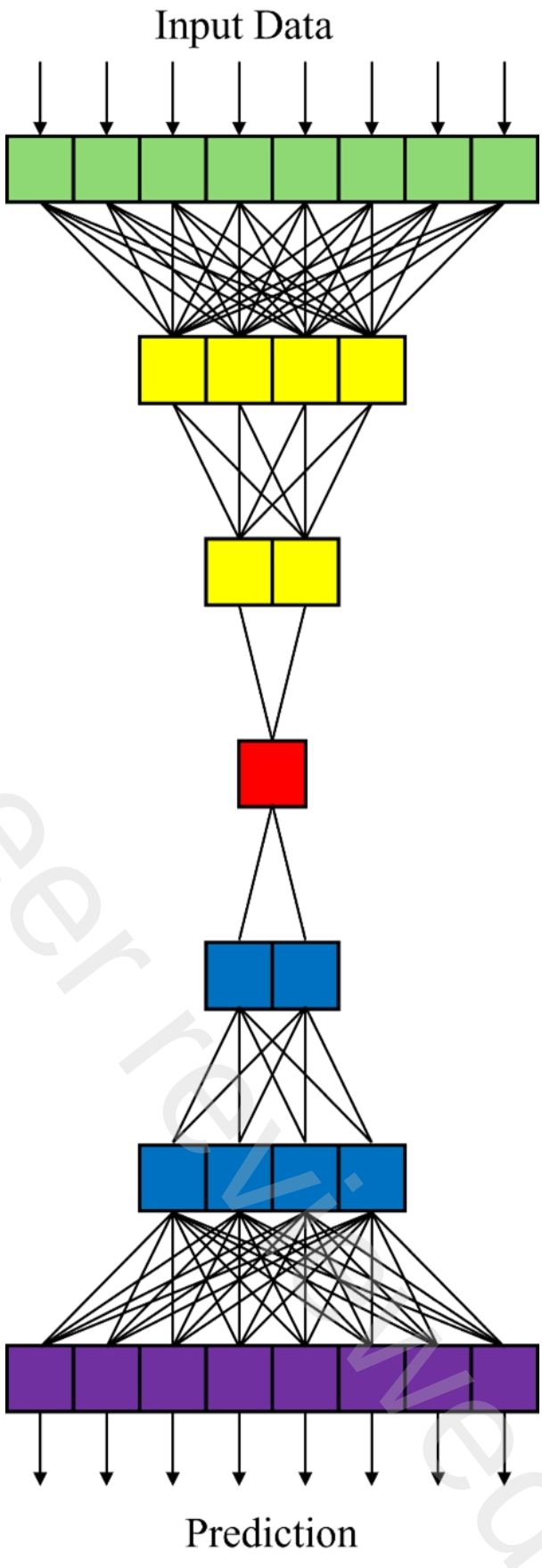
[1] Verein Deutscher Zementwerke e.V. Umweltdaten der deutschen Zementindustrie 2022. Düsseldorf; 2023.
[2] Pohl M, Becker G, Heller N, Birnstengel B, Zotz F. Auswirkungen des nationalen Brennstoffemissionshandels auf die Abfallwirtschaft; 2022.
[3] Juhrich K. CO2-Emissionsfaktoren für fossile Brennstoffe: CLIMATE CHANGE 27/2016. Dessau-Roßlau; 2016.
[4] Zhao L, Giannis A, Lam W-Y, Lin S-X, Yin K, Yuan G-A et al. Characterization of Singapore RDF resources and analysis of their heating value. Sustainable Environment Research 2016;26(1):51–4. https://doi.org/10.1016/j.serj.2015.09.003.
[5] Liedmann B.:. Simulation der thermischen Umsetzung flugfähiger Ersatzbrennstoffe in Drehrohröfen der Zementindustrie [Dissertation]: Ruhr-Universität Bochum; 2018.
[6] El-Salamony A-HR, Mahmoud HM, Shehata N. Enhancing the efficiency of a cement plant kiln using modified alternative fuel. Environmental Nanotechnology, Monitoring & Management 2020;14:100310. https://doi.org/10.1016/j.enmm.2020.100310.
[7] Gallardo A, Carlos M, Bovea MD, Colomer FJ, Albarrán F. Analysis of refuse-derived fuel from the municipal solid waste reject fraction and its compliance with quality standards. Journal of Cleaner Production 2014;83:118–25. https://doi.org/10.1016/j.jclepro.2014.07.085.
[8] VDZ Technology gGmbH. Verfahren zur Schnellprüfung der Qualität von flugfähigen Ersatzbrennstoffen für den Einsatz im Klinkerbrennprozess: Schlussbericht vom 04.08.2023 zu IGF-Vorhaben Nr. 20898 N.
[9] Abd Halim NY, Muhammad N. Correlation of Moisture Content and Heating Value of Refuse-Derived Fuel. J. Phys.: Conf. Ser. 2025;3003(1):12019. https://doi.org/10.1088/1742-6596/3003/1/012019.
[10] Seong Chae J, Wan Kim S, Ohm T in. Combustion Characteristics of Solid Refuse Fuels from DifferentWaste Sources. Journal of Renewable Materials 2020;8(7):789–99. https://doi.org/10.32604/jrm.2020.010023.
[11] Dianda P, Mahidin, Munawar E. Production and characterization refuse derived fuel (RDF) from high organic and moisture contents of municipal solid waste (MSW). IOP Conf. Ser.: Mater. Sci. Eng. 2018;334:12035. https://doi.org/10.1088/1757-899X/334/1/012035.
[12] Gerassimidou S, Velis CA, Williams PT, Komilis D. Characterisation and composition identification of waste-derived fuels obtained from municipal solid waste using

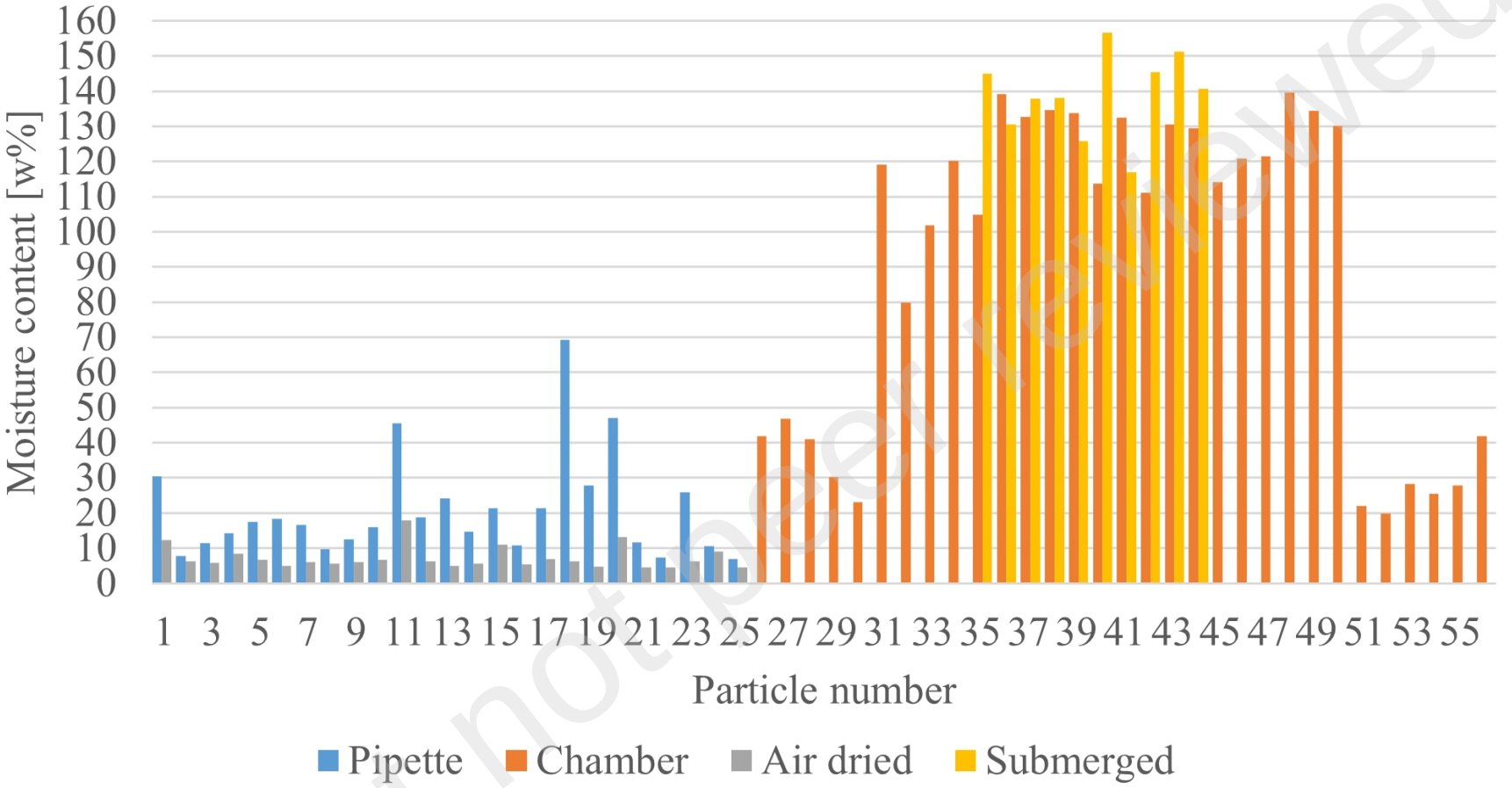
thermogravimetry: A review. Waste Manag Res 2020;38(9):942–65. https://doi.org/10.1177/0734242X20941085.
[13] Ozaki Y, Morisawa Y. Principles and Characteristics of NIR Spectroscopy. In: Ozaki Y, Huck C, Tsuchikawa S, Engelsen SB, editors. Near-Infrared Spectroscopy. Singapore: Springer Singapore; 2021, p. 11–35.
[14] Muncan J, Tsenkova R. Aquaphotomics-From Innovative Knowledge to Integrative Platform in Science and Technology. Molecules 2019;24(15). https://doi.org/10.3390/molecules24152742.
[15] Zhou Y, Zheng Q, Hu X, Liang H, Dong C. Research on coal moisture analysis model based on near infrared spectroscopy. IOP Conf. Ser.: Earth Environ. Sci. 2022;983(1):12092. https://doi.org/10.1088/1755-1315/983/1/012092.
[16] Wang Y, Yang M, Wei G, Hu R, Luo Z, Li G. Improved PLS regression based on SVM classification for rapid analysis of coal properties by near-infrared reflectance spectroscopy. Sensors and Actuators B: Chemical 2014;193:723–9. https://doi.org/10.1016/j.snb.2013.12.028.
[17] Toscano G, Leoni E, Gasperini T, Picchi G. Performance of a portable NIR spectrometer for the determination of moisture content of industrial wood chips fuel. Fuel 2022;320:123948. https://doi.org/10.1016/j.fuel.2022.123948.
[18] Leblon B, Adedipe O, Hans G, Haddadi A, Tsuchikawa S, Burger J et al. A review of near-infrared spectroscopy for monitoring moisture content and density of solid wood. The Forestry Chronicle 2013;89(05):595–606. https://doi.org/10.5558/tfc2013-111.
[19] Camacho W, Vallés-Lluch A, Ribes-Greus A, Karlsson S. Determination of moisture content in nylon 6,6 by near-infrared spectroscopy and chemometrics. J of Applied Polymer Sci 2003;87(13):2165–70. https://doi.org/10.1002/app.11595.
[20] Günkaya Z, Özkan M, Özkan K, Bekgöz BO, Yorulmaz Ö, Özkan A et al. Prediction of the proximate analysis parameters of refuse-derived fuel based on deep learning approach. Environ Sci Pollut Res Int 2023;30(7):17327–41. https://doi.org/10.1007/s11356-022-23272-6.
[21] Liedmann B., Arnold W., Krüger B., Becker A., Krusch S., Wirtz S. et al. Approach to model the thermal conversion and flight behaviour of Refuse Derived Fuel. Fuel 2017(200):252–71.
[22] Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods 2020;17:261–72. https://doi.org/10.1038/s41592-019-0686-2.
[23] Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D et al. Array programming with NumPy. Nature 2020;585(7825):357–62. https://doi.org/10.1038/s41586-020-2649-2.
[24] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 2011;12:2825–30.
[25] Fischer J, Wirtz S, Scherer V. Random forest classifier and neural network for fraction identification of refuse-derived fuel images. Fuel 2023;341:127712. https://doi.org/10.1016/j.fuel.2023.127712.
[26] Breiman L. Random Forests. Machine Learning 2001;45(1):5–32. https://doi.org/10.1023/A:1010933404324.
[27] Géron A. Praxiseinstieg Machine Learning mit Scikit-Learn und TensorFlow: Konzepte, Tools und Techniken für intelligente Systeme. 1st ed. Heidelberg: O'Reilly; 2018.
[28] Chollet F, others. Keras; 2015.
[29] TensorFlow Developers. TensorFlow. Zenodo; 2024.

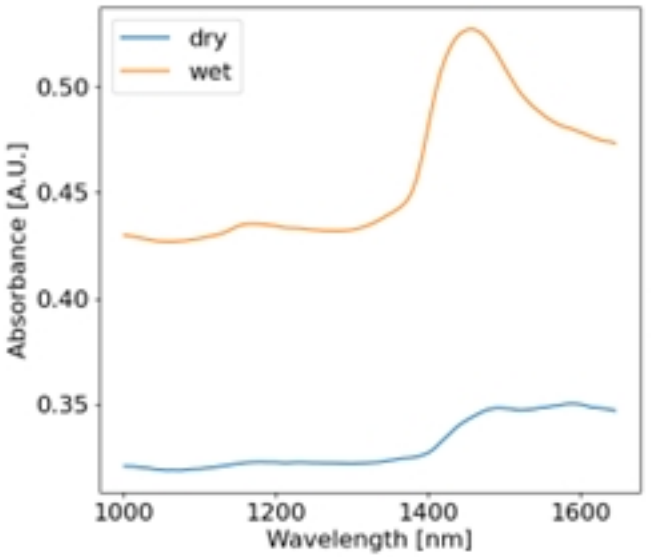
[30] Workman JJ, Weyer L. Practical Guide to Interpretive Near-Infrared Spectroscopy. CRC Press; 2007.

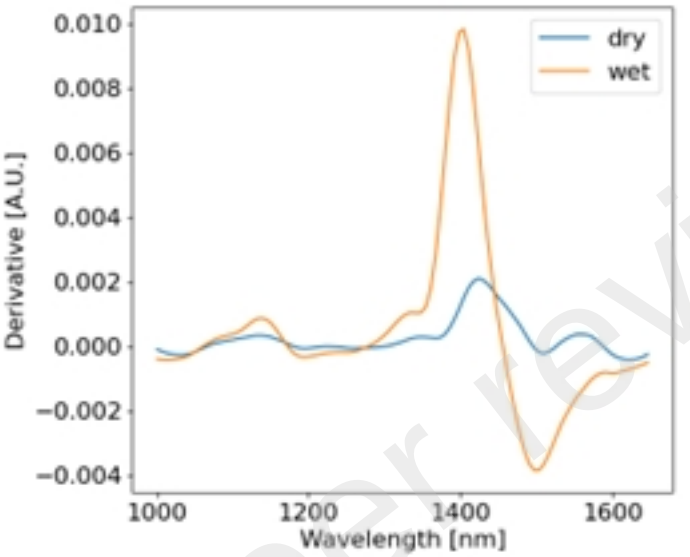


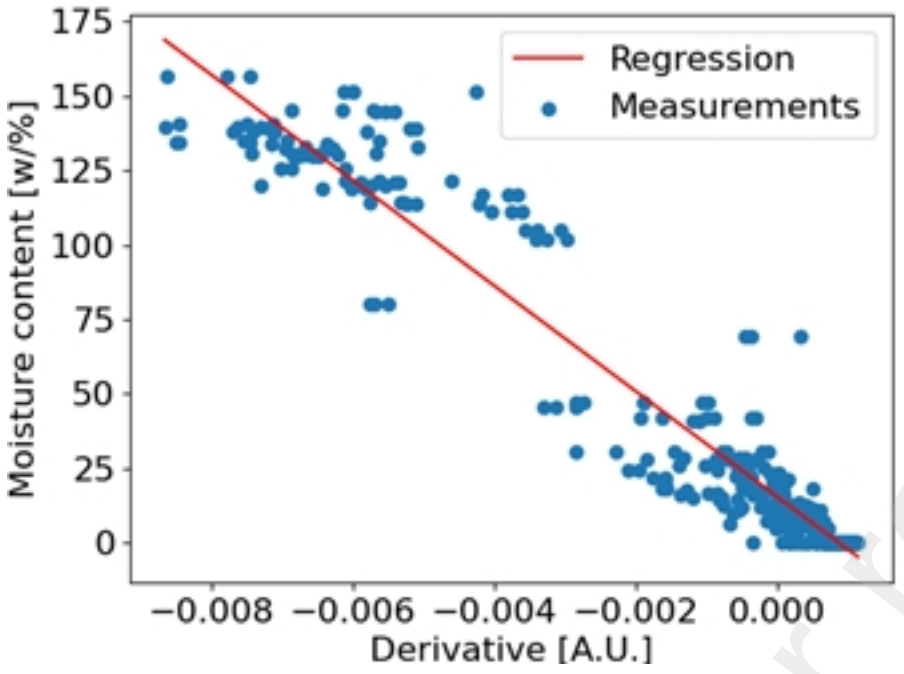


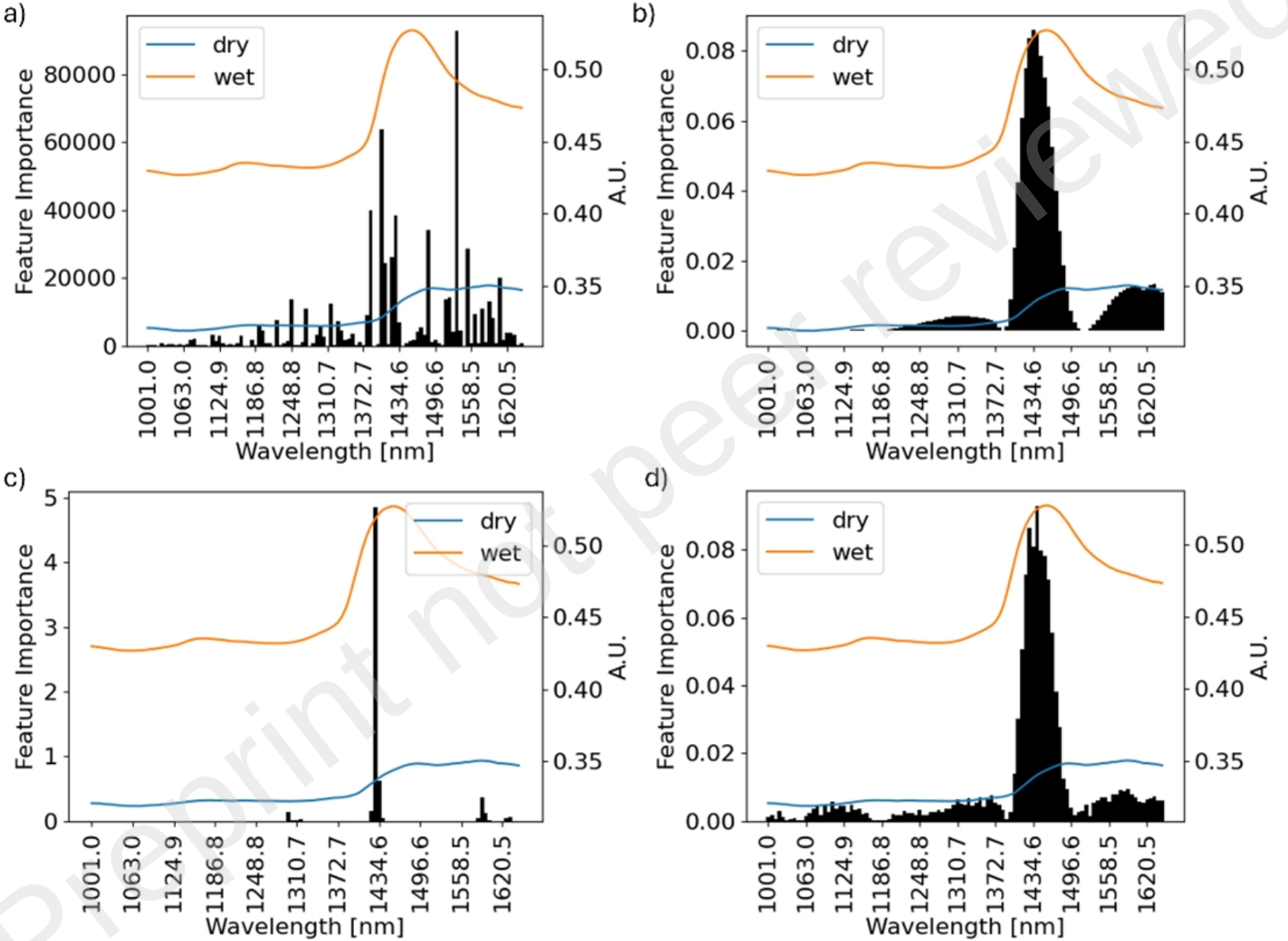


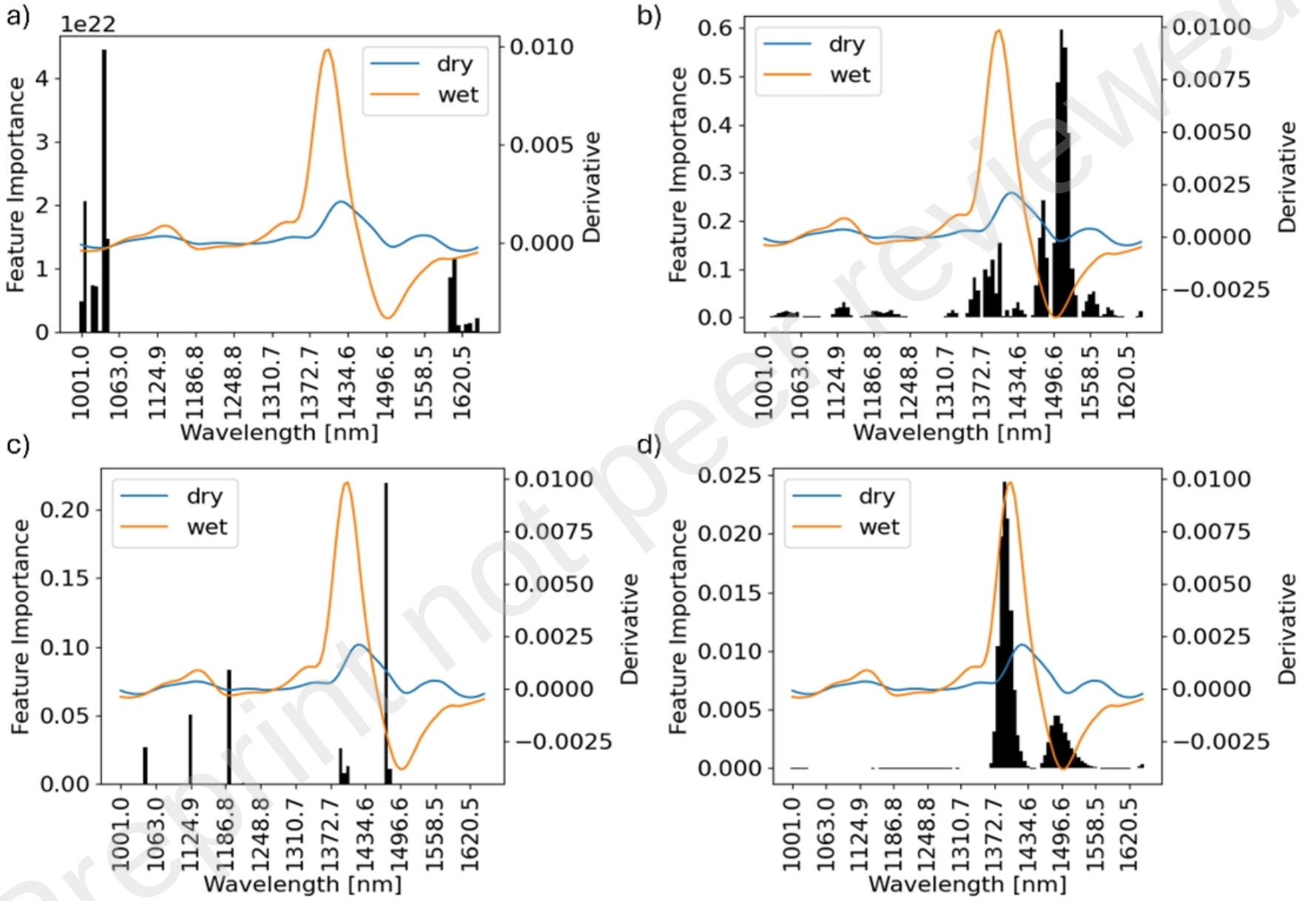


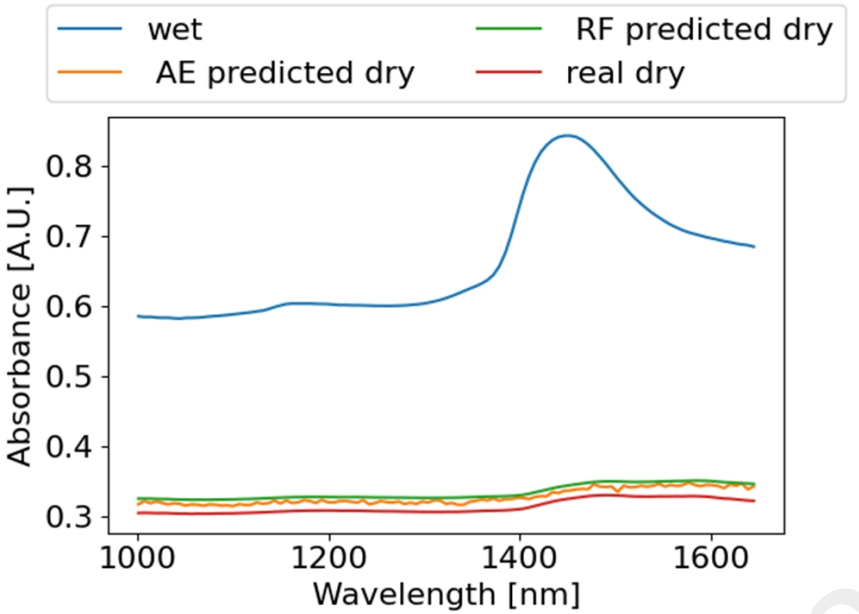












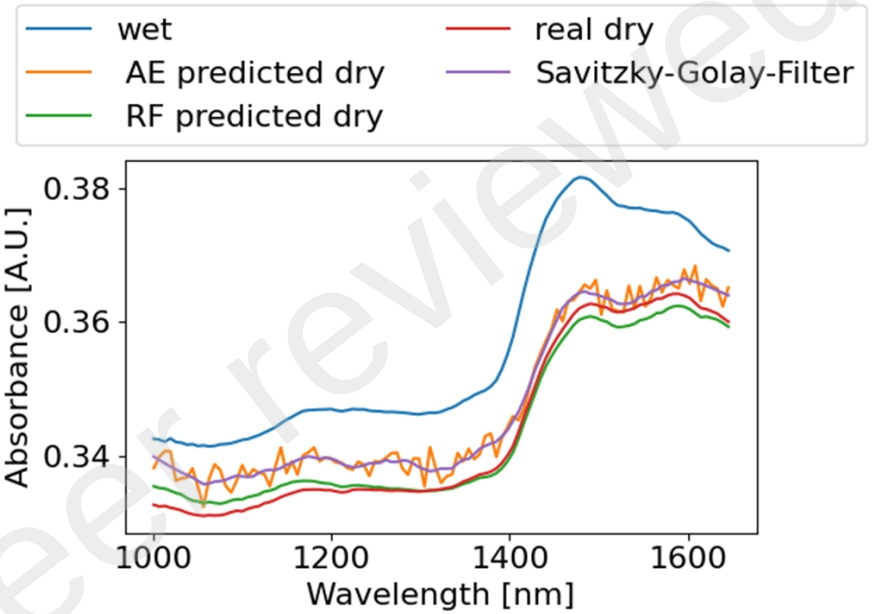


Figure 1: Process diagram of the sample preparation: First, dry particles NIR spectra are measured. Then, different moisture content particles are produced by either pipette and air drying (left) or humidity chamber and submersion (right).

Figure 2: Sketch of the chamber to humidify particles: An ultrasonic nebulizer immersed in water reservoir provides a steady stream of mist, which is mixed in the chamber by three propellers. The particles are held by a wire cage.

Figure 3: Scheme of an autoencoder. The data flow can be read from left to right: The input data enters the network in the input layer (green), which is part of the encoder. The rest of the encoder (yellow) consists of dense connected layers. In the middle, the latent space (red), the layer size is minimal. After that, the decoder part with more densely connected layers (blue) increases the dimension. The last layer of the decoder part is the output layer (purple), which can produce a prediction.

Figure 4: Moisture content of model paper particles by different methods to change moisture content: Applying water drops by pipette (blue), air drying of these particles (grey), humidifier chamber (orange) and submerging to water (yellow).

Figure 5: Mean absorbance (left) and derivatives of absorbance (right) over wavelength.

Figure 6: Moisture content over 1. derivative values at 1471 nm, absorbance derived after wavelength by Savitzky-Golay filter (blue) and the corresponding simple linear regression (red).

Figure 7: Feature importances for different regression algorithms with the corresponding absorbances over wavelength: a) Multiple linear regression, b) ridge regression, c) lasso regression and d) MLP regression. Left y axis for feature importance, right y axis for absorbance.

Figure 8: Feature importances for different regression algorithms with the corresponding

derivatives of absorbances over wavelength: a) Multiple linear regression, b) ridge regression, c) lasso regression and d) MLP regression. Left y axis for feature importance, right y axis for derivative.

Figure 9: Two examples of predictions of a spectrum of a dry spectrum based on the wet spectrum and the moisture content with an autoencoder (AE) and a random forest (RF). Particle no. 5 (moisture content 17.8 %) on the left and particle no. 2 (moisture content 7.8 %) on the right. Since the AE prediction for particle 2 is oscillating, a Savitzky-Golay filter was applied.