Article

Predicting Ethanol Steam Reforming Products of Au-Cu Supported over Nano-Shaped CeO₂ Using the Johnsen Measure in PLS

Chen Zhi ^{1,†}, Muhammad Tahir ^{1,*} and Tahir Mehmood ^{2,†}
¹ School of Statistics, Shanxi University of Finance and Economics, Taiyuan 030000, China; chenzhi@sxufe.edu.cn

² School of Natural Sciences (SNS), National University of Sciences and Technology (NUST), Islamabad 44300, Pakistan; tahir.mehmood@sns.nust.edu.pk

* Correspondence: tahir.muhammad6946@gmail.com

† These authors contributed equally to this work.

Abstract: Hydrogen fuel cells have long been regarded as a more environmentally friendly alternative to traditional fossil fuels. Ethanol steam reforming (ESR) is a promising long-term, safe method of producing carbon-neutral hydrogen. ESR products are (CeCO₂) support generate hydrogen (H₂) with byproducts such as carbon dioxide (CO₂) and carbon monoxide (CO). The researchers are interested in the quantification and estimation of syngas components. The current article introduces the Johnsen index-based measure in partial least squares (PLS) for predicting ESR products with cube, polyhydra, and rod morphologies, based on FTIR. The proposed method makes use of existing filter measures such as loading weights, variable importance on projection, and significant correlation. The proposed PLS measures based on the Johnsen index outperform the existing methods for predicting ESR products based on FTIR spectroscopic data. For (H₂) conversion percent prediction with cube and polyhedra morphologies, the functional compounds (C-O), (C=O), (CH), and (C-H,=CH₂) are common. Similarly, the functional compound (s-RCH=CHR) is frequently used for (H₂) conversion percent prediction with polyhedra and rod morphologies. Moreover, on simulated data, the proposed Johnsen measure in PLS demonstrates higher sensitivity and accuracy. Furthermore, the proposed Johnsen measure in PLS identifies influential wavenumbers that map over the functional compounds.

Keywords: prediction; PLS; hydrogen fuel; environment friendly; FTIR spectroscopy

Citation: Zhi, C.; Tahir, M.; Mehmood, T. Predicting Ethanol Steam Reforming Products of Au-Cu Supported over Nano-Shaped CeO₂ Using the Johnsen Measure in PLS. *Appl. Sci.* **2021**, *11*, 10402. <https://doi.org/10.3390/app112110402>

Academic Editor: Francisco Jesus Fernandez-Morales

Received: 9 September 2021

Accepted: 27 October 2021

Published: 5 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A fuel cell is an electrochemical cell that converts the chemical energy of the fuel and an oxidising agent into electricity. This conversion typically makes use of hydrogen as a fuel and oxygen as an oxidising agent. Hydrogen fuel cells have long been regarded as a more environmentally friendly alternative to traditional fossil fuels. Because water is the only direct byproduct of their energy production, they will significantly reduce waste and man-made greenhouse gases by oxidising molecular hydrogen [1,2]. However, a group of American researchers believes that fuel cells may have a negative impact on the environment. Tracey Tromp of the California Institute of Technology and colleagues used atmospheric simulations to show that the unavoidable pollution emitted by fuel cell technology could cause significant harm to the ozone layer [3]. Among non-fossil feedstocks such as ethanol, methanol, and others, ethanol has many advantages as a chemical (H₂) source for fuel cells due to its storage capabilities, ease of handling, and stable transportation due to its low toxicity and volatility. The production of (H₂) from renewable resources is a critical stage in the evolution of technology. The production of (H₂) from bioethanol derived from agroindustrial wastes, in particular, is an appealing option because bioethanol derived from residual biomass is less expensive than bioethanol derived from food crops and does not jeopardise food safety [4,5]. Ethanol steam reforming (ESR) is a

promising reaction for producing carbon-neutral hydrogen in a long-term, safe manner for this purpose, according to [6,7]. ESR primarily generates hydrogen (H_2), with byproducts such as carbon dioxide (CO_2), carbon monoxide (CO), and methane (CH_4).

The purification of this syngas is critical and is defined by the final use of Hydrogen (H_2), so the quantification of syngas components is of interest to the researcher [8,9]. It appears that using a (CeO_2) catalyst in a system is promising for removing the majority of the (CO) from syngas [10]. Modifying the support nanostructure of (CeO_2), such as polyhedra, rods, and cubes, has been identified as a novel technique for developing more active, selective, and stable catalysts [11].

Fourier-transform infrared spectroscopy (FTIR) is used for characterization because it is a fast, accurate, and inexpensive technique. For example, FTIR has been used to characterise MnO_x - CeO_2 catalysts for low-temperature selective oxidation [12]. FTIR looks into the conditions that occur during CO oxidation over $Ru(x)$ - CeO_2 catalysts [13]. FTIR describe CeO_2 catalysts for diesel pollutant removal [14]. FTIR describes the preparation of copper-promoted CeO_2 catalysts [15]. FTIR investigates the total oxidation of formaldehyde over MnO_x - CeO_2 catalysts [16] and so forth.

Because FTIR produces high-dimensional data with a small sample size, partial least squares (PLS) regression is considered a viable option for modelling the spectrum data [17]. Mehmood et al. recently used PLS to predict the antibacterial activity based on FTIR data [18]. Based on spectroscopic data, PLS assisted in predicting catalyst activity [19] and it aids in determining the composition of bio-oil using infrared spectroscopy [20].

Based on FTIR data, we predicted ESR products such as hydrogen (H_2) with side products such as carbon dioxide (CO_2), carbon monoxide (CO), and methane (CH_4) in this article. Variable selection measures are required for PLS-based characterization. For this purpose, PLS is equipped with loading weight, regression coefficients, variable importance on projection [21] and selectivity ratio [22,23] type filter measures exist [24]. These measures' performance still needs to be improved [25]. For this purpose, we have introduced the Johnsen Index [26,27] based variable selection in PLS. The proposed measure is compared to the reference measure on real data, and these measures are compared for predicting ESR products based on FTIR spectrum data. The functional characterization of ESR products is linked to spectrum-based variable selection.

2. Material and Methods

2.1. Catalyst Data Set

The data set used in this study is taken from [28] and is accessible from [29] where Au-Cu supported over nano-shaped CeO_2 is used as stable catalysts for the carbon monoxide removal from syngas. We only looked at Au-Cu supported over nano-shaped CeO_2 data, and we considered several morphologies such as polyhedra, rods, and cubes. The specific nano shapes were created using a hydrothermal method with continuous stirring for 1 h at 600 rpm. The resulting slurry was heated in an airtight container for 24 h at 100, 120, and 160 °C for polyhedra, rods and cubes, respectively. The precipitate was neutralized with water and calcined at 500 °C for 2 h.

The actual syngas was produced by the ESR. The catalyst activity test was carried out, and CO conversion (%), (CO_2) yield, and (H_2) conversion (%) were monitored from 300 to 100 °C, with temperature decreasing every 30 min in a sequence of around 20 °C.

2.2. Catalyst Characterization

TGA used a thermogravimetric analyzer to measure deposits on catalyst samples (Mettler Toledo, Columbus, OH, USA). The test was carried out at temperatures ranging from 30 to 1000 °C (5 °C/min) in dry air (150 mL/min). The weight loss from the AC samples was subtracted from the weight loss from the spent catalysts. A sampler holder was loaded with 0.02 g of samples. The sample holder was then sealed and aised to prevent interference from the environment using an external flow of Ar (20 mL/min) that remained constant throughout the test. Following that, the sample was flushed with Ar (15 mL/min)

at 50 °C for 30 min. Finally, ten pulses of 30 µL of CO were injected into the sampler holder from a certified 5% CO/Ar mixture; between each pulse, Ar (15 mL/min) was passed for 10 min. The signal was collected between 4000 and 400 cm⁻¹, with a resolution of 2 cm⁻¹, and at a rate of 64 scans per minute.

2.3. Interpolation of Ethanol Steam Reforming Products

The catalyst activity test measures the CO conversion (%), CO₂ yield, and H₂ conversion (%) as a function of temperature from 300 to 100 °C, with temperature decreasing every 30 min by around 20 °C. The spectrum of each catalyst was recorded at various time intervals ranging from 1 to 140 min, and the weight loss was used to map the spectrum to temperature. As a result, catalyst activity and the spectrum of catalyst characterization are taken as a function of temperature. We used an interpolation technique because both catalyst activity and catalyst characterization are performed at different temperatures. First, the second degree polynomial was used to fit catalyst activity as a function of temperature one by one. The temperature measured against the spectrum is then used in conjunction with the fitted polynomial to estimate the catalyst activity. This produces the response matrix $Y_{n \times 3}$ where columns are CO conversion (%), CO₂ yield and H₂ conversion (%), respectively. Moreover, the FTIR spectrum data taken from [28] produces the data matrix $X_{n \times p}$, where n is the sample size and p is the number of FTIR wavenumbers.

2.4. PLS Modeling of Ethanol Steam Reforming Products

Ethanol Steam Reforming (ESR) products include CO conversion (%), CO₂ yield, and H₂ conversion (%), which are assembled as $Y_{n \times 3}$ and modelled with FTIR data assembled as $X_{n \times p}$. Because each ESR product has its own unique identification, each response variable must be modelled separately as $y_{n \times 1}$. This data set has a very small sample size in comparison to the number of wavenumbers counted. We were unable to use the traditional regression method to model the ethanol steam performing products in this scenario. Partial least squares regression (PLS) [17] provides the way forward for modeling ethanol steam reforming products. In iterative procedure of PLS loading weights $w_i = X'_{i-1}y_{i-1}$, scores $t_i = X_{i-1}w_i$, X-loadings $p_i = X'_{i-1} \frac{t_i}{t_i't_i}$, y-loading $q_i = y'_{i-1} \frac{t_i}{t_i't_i}$, deflated $X_i = X_{i-1} - t_i p_i'$ and deflated $y_i = y_{i-1} - t_i q_i$ are computed for i th components. The loading weights w_i , scores t_i , X-loadings p_i and y-loadings q_i are assembled in W , T , P and q respectively. These metrics computes the PLS regression coefficients $\hat{\beta} = W(P'W)^{-1}q$.

Characterization of ESR products is required, as is the identification of influential wavenumbers that best explain the variation in ESR product. PLS loading weights, regression coefficients, variable importance on projection [21] and selectivity ratio [22,23] are all used in this case. The variable importance on projection is defined as

$$v_j = \sqrt{p \sum_{i=1}^I [(q_i^2 t_i' t_i)(w_{ij} / \|w_i\|)^2] / \sum_{i=1}^I (q_i^2 t_i' t_i)}.$$

A significance multivariate correlation (C) is defined as:

$$c_j = \frac{MS_{j,PLS_{regression}}}{MS_{j,PLS_{residuals}}} = \frac{\|\hat{\beta}'_j\|^2}{\|x_j - \frac{\hat{\beta}'_j}{\|\hat{\beta}'_j\|^2}\|^2 / (n-2)}$$

where $MS_{j,PLS_{regression}}$ is mean squares explained by PLS regression and $MS_{j,PLS_{residuals}}$ is the residual squares of PLS regression. Hence, the reference methods are PLS_W , PLS_V , PLS_C , which are respectively based on PLS loading weights, variable importance on projection and significance multivariate correlation.

2.5. Proposed Measure for Ethanol Steam Reforming Characterization

It is still necessary to improve the performance of existing filter measures such as loading weight, regression coefficients, variable importance on projection, and selectivity ratio [25]. For this purpose, we have introduced the *Johnsen* Index [26,27] based variable selection in *PLS*. The *Johnsen* index is defined as

$$H_j = \lambda_j^2 \beta_j^2,$$

where λ_j is the eigen value comes out from the spectral value decomposition of X that is, $X = P\Lambda Q^t$. The eigenvalues do not indicate the importance of the respective variable, but they do indicate the importance of the transformed linear combination. Motivated by the *Johnsen* index, we developed the following wave number selection measures: *PLS* loading weights and variable importance, loading weights and significance multivariate correlation, and significance multivariate correlation and variable importance on projection as:

$$Johnsen_{(W\&V)} = w_j^2 v_j^2$$

$$Johnsen_{(W\&C)} = w_j^2 c_j^2$$

$$Johnsen_{(C\&V)} = c_j^2 v_j^2$$

As a result, the proposed methods are PLS_{WV} , PLS_{WC} , and PLS_{CV} , all of which are based on the *Johnsen* index introduced over the combination of *PLS* loading weights, variable importance on projection, and significance multivariate correlation.

3. Simulation Study

The simulation study is carried out to provide numerical support for the proposed *PLS*-based variable selection methods, where we know truly important variables. The data is generated by simulating a known model. $y = X\beta + \epsilon$, where y is response and x is data matrix assumed to follow multivariate normally distribution with mean-vector $\mu = 0$ and covariance matrix Σ [25].

$$X \sim MVN(0, \Sigma).$$

Because the variances of all variables are assumed to be one, Σ is the correlation matrix. Furthermore, H groups of correlated x variables are extracted by applying the group diagonal structure to Σ , that is:

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 & \dots & 0 \\ 0 & \Sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Sigma_L \end{bmatrix}$$

We have used $H = 10$ groups of the equal number of variables. Data are simulated with $n = 100$, with four levels of the number of variables $p = (100, 500, 1000)$, with two levels of within-group correlation $R_1 = (0.35, 0.15, 0.2, 0, 0, 0, 0, 0, 0, 0)$, $R_2 = (0.75, 0.95, 0.50, 0, 0, 0, 0, 0, 0, 0)$ and with the correlation between x and y $C_{xy} = (0.6, -0.5, 0.2, 0, 0, 0, 0, 0, 0, 0)$. For each simulated data set, the first 30 (percentage) variables are considered important. For example, with $(p = 100)$, the variables 1 to 30 are considered important, while the rest are considered unimportant. The first 30 variables will be influential. This results in six different parameter combinations for the simulation. For tuning the *PLS* model's parameters, 100 simulation runs are used, with each run data divided into 70% training and 30% test. The simulated data set is used to fit the proposed *PLS*-based methods for variable selection and reference method. C_{xy} indicates that the first two groups of simulated variables are significant. Different levels of (R) and (p) produce

6 simulation designs. The behaviour of measures used in reference methods and proposed methods can be seen in Figure 1. This indicates the proposed methods have higher specificity and Johnsen index with W , V has higher sensitivity. Because sensitivity indicates the proportion of true variable selection. Table 1 shows the results of a sensitivity, specificity, and accuracy comparison of proposed and reference PLS models against the simulation design. This means that in the majority of simulation designs, our proposed algorithms are more sensitive and accurate, whereas the reference methods are more specific.

Table 1. The sensitivity, specificity, and accuracy of selected variables by reference and proposed method are presented in comparison to all simulation designs considered. The cells with the highest sensitivity, specificity, and accuracy within each simulation design are highlighted in grey.

p	Rohs	Method	Sensitivity (%)	Specificity(%)	Accuracy (%)
100	R1	W	0.86	0.50	0.79
		V	0.81	0.90	0.83
		C	0.99	0.65	0.92
		WV	0.97	0.70	0.92
		WC	0.99	0.65	0.92
		VC	0.99	0.65	0.92
	R2	W	0.96	0.50	0.87
		V	0.95	1.00	0.96
		C	1.00	1.00	1.00
		WV	1.00	1.00	1.00
		WC	1.00	1.00	1.00
		VC	1.00	1.00	1.00
500	R1	W	0.92	0.56	0.85
		V	0.76	0.71	0.75
		C	0.91	0.63	0.86
		WV	0.96	0.52	0.87
		WC	0.98	0.30	0.85
		VC	0.98	0.30	0.85
	R2	W	0.98	0.50	0.88
		V	0.97	1.00	0.97
		C	1.00	1.00	1.00
		WV	1.00	1.00	1.00
		WC	1.00	0.99	1.00
		VC	1.00	1.00	1.00
1000	R1	W	0.92	0.48	0.83
		V	0.85	0.87	0.85
		C	0.96	0.75	0.90
		WV	0.97	0.77	0.92
		WC	1.00	0.52	0.90
		VC	1.00	0.52	0.90
	R2	W	0.90	0.81	0.88
		V	0.89	1.00	0.91
		C	1.00	1.00	1.00
		WV	1.00	1.00	1.00
		WC	1.00	0.96	0.99
		VC	1.00	1.00	1.00

Figure 1. The left panel shows the reference measures loading weights (W), variable importance on projection (V), and significance multivariate correlation (C) extracted from the simulation study, while the right panel shows the proposed measures, which are the Johnsen index as a combination of W , V , and C . The data was generated using a simulation. $R_1 = (0.75, 0.95, 0.50, 0, 0, 0, 0, 0, 0, 0)$ and with the correlation between x and y $C_{xy} = (0.6, -0.5, 0.2, 0, 0, 0, 0, 0, 0, 0)$, $p = 1000$ and $n = 100$.

4. Results

For predicting Ethanol Steam Reforming (ESR) products including CO conversion (%), CO₂ yield and H₂ conversion (%) the Au-Cu supported over nano-shaped CeO₂ is used where three morphologies including polyhedral, rods and cubes are considered. The description of these ESR products is summarized in Table 2. This indicates that the CO conversion is highest with cube morphology and lowest with rods morphology. The CO₂ yield is highest with cubes and polyhedral morphologies, and lowest with rods morphology. Similarly, with cube morphology, the H₂ conversion is at its highest level, while with polyhedral morphology, it is at its lowest.

Table 2. The summary statistics include the average, minimum, maximum, and standard deviation (SD) of ESR products with various morphologies.

ESR Product	Morphology	Min	Max	Mean	SD
CO Conversion (%)	Cubes	15.22	51.61	37.09	13.81
	Polyhedral	11.11	37.42	30.00	8.15
	Rods	6.56	34.31	25.65	8.87
CO ₂ yield (%)	Cubes	0.11	0.29	0.24	0.07
	Polyhedral	0.02	0.32	0.24	0.10
	Rods	0.05	0.25	0.19	0.06
H ₂ conversion (%)	Cubes	10.90	18.45	13.44	2.54
	Polyhedral	7.90	17.20	10.63	3.15
	Rods	6.26	13.33	11.18	2.27

Since ESR products such as CO conversion (%), CO₂ yield and H₂ conversion (%) are temperature dependent, the catalyst activity and characterization spectrum are also temperature dependent. We used an interpolation technique because both catalyst activity

and catalyst characterization are performed at different temperatures. First, the polynomial equation of degree two was used to fit catalyst activity as a function of temperature one by one. The temperature measured against the spectrum is then used in conjunction with the fitted polynomial to estimate the catalyst activity. The interpolation of CO conversion (%), CO₂ yield and H₂ conversion (%) through polynomial equation of degree two is exemplified for cube morphology is presented in Figure 2.

Figure 2. The interpolation of CO conversion (%), CO₂ yield, and H₂ conversion (%) using a polynomial equation of degree two is demonstrated for cube morphology.

For ESR product prediction, we have proposed the Johnsen index based PLS_{WV} , PLS_{WC} , PLS_{CV} which will be compared with the reference method PLS_W , PLS_V , PLS_C . Hence for each EST product prediction we have to fit 06 PLS models. Since, we have considered 03 ESR products CO conversion (%), CO₂ yield and H₂ conversion (%) the Au-Cu supported over nano-shaped CeO₂ with three morphologies including polyhedral, rods and cubes, hence we have fitted $6 \times 3 \times 3 = 54$ models. Each optimal PLS model is subject to tuning model parameters such as the number of components and the threshold that defines the number of selected variables, that is, the number of wavenumbers. We considered the number of PLS components as well as the threshold levels. Tuning is required for the number of components and the threshold. For this, we divided the data into 70 percent training and 30 percent test, that is, ESR product response y and spectroscopic data X . The training model is used to construct the model, and the performance of the fitted model is evaluated using test data. Furthermore, each model is fitted with a different number of components and thresholds, with the levels that result in the best performance on test data being labeled as optimal. The RMSE value reflects the model's performance. We used a response surface methodology-based curve that results in optimal model parameter levels to achieve a systematic optimal level of model parameters. Figure 3 shows one of the cases from all fitted models. All reference and proposed methods predict ESR products by using the optimal threshold and optimal number of components. The RMSE over training and test data is computed, and the results are shown in Figure 4. The average RMSE on training data for predicting CO conversion (%) with cube morphology is around 1.04 and is 1.03 on test data. On training data, PLS_C has the lowest RMSE, whereas on test data, PLS_{VC} has the lowest RMSE. The average RMSE on training data for predicting CO₂ yield with cube morphology is around 1.035 on training data and 1.019 on test data. On training data, PLS_C has the lowest RMSE, whereas on test data, PLS_{WC} has the lowest RMSE. The average RMSE on training data for predicting H₂ conversion (%) with cube morphology is around 1.014 and is 1.002 on test data. On training data, PLS_C has the lowest RMSE, while on test data, PLS_{VC} has the lowest RMSE. The average RMSE on training data for predicting CO conversion (%) with polyhedra morphology is around 1.035 on training data and is 1.02 on test data. On training data PLS_C has least RMSE, while on test data PLS_{WC} has least RMSE. The average RMSE on training data for predicting CO₂ yield with polyhedra morphology

is around 1.011 on training data and is 1.01 on test data. The training data all have similar RMSE, while the test data PLS_{VC} have the least RMSE. The average RMSE on training data for predicting H_2 conversion (%) with polyhedra morphology is around 1.035 on training data and is 1.022 on test data. On training data PLS_C has the least RMSE, while on the test data PLS_{WC} has the least RMSE. The average RMSE on training data for predicting CO conversion (%) with rod morphology is around 1.035 on training data and is 1.012 on test data. On training data PLS_{VC} has least RMSE, while on test data PLS_{WV} has least RMSE. The average RMSE on training data for predicting CO_2 yield with rod morphology is around 1.05 on training data and is 1.01 on test data. On training data, PLS_C has the least RMSE, while on test data PLS_{WC} has the least RMSE. The average RMSE on training data for predicting H_2 conversion (%) with rod morphology is around 1.02 on training data and is 1.01 on test data. On training data PLS_C has least RMSE, while on the test data PLS_{WV} has the least RMSE.

Figure 3. The RMSE on the test data measuring the prediction capability of CO_2 yield with cube morphology is plotted against the number of PLS_{WV} components and the threshold. The fitted response surface indicates that as the number of components and threshold level increase, the RMSE decreases.

According to the comparison above, the Johnsen index-based PLS variants outperform the test data in predicting ESR products with different morphologies. Furthermore, it has been discovered that ESR product prediction on test data is superior to prediction on training data. Notably, the FTIR spectra used for modeling is baseline corrected. Baselines are frequently set based on a visual examination of their effect on specific spectra. It offers a more objective process for selecting baseline correction algorithms and their parameter values for use in statistical analysis. For good prediction, the spectra must be objective and reproducible if they are to be employed in a statistical analysis. We have used asymmetric least square (ASL) [30], where parameters are tuned for optimal ESR product prediction. For this, we used the original, and the corrected FTIR spectra used in polyhedra is presented here for illustration in Figure 5. The best model for each ESR product prediction is then used to determine the influential wavenumbers. The influential wavenumbers were then mapped to functional compounds for each optimal model. Figure 6 depicts the influential wavenumber indicating the intensity of their selection and respective functional compound. By using the PLS_{VC} , the CO conversion (%) with cube morphology is predicted, which results in six influential wavenumbers. These wavenumbers are mapped to C-O, C=O, O-H, CH_3-CH_2 , OH and C-H= CH_2 . By using the PLS_{WC} CO_2 yield with cube morphology is predicted, which results in 6 influential wavenumbers. These wavenumbers are mapped

to C-O, CH₂-CH₃, C=O, C-H,=CH₂, O-H, and CH. By using the PLS_{VC} H₂ conversion (%) with cube morphology is predicted, which results in 6 influential wavenumbers. These wavenumbers are mapped to C-O, CH₂-CH₃, C=O, C-H,=CH₂, O-H and CH. By using the PLS_{WC} CO conversion (%) with polyhedra morphology is predicted, which results in four influential wavenumbers. These wavenumbers are mapped to C-O, C=O, O-H, and C-H,=CH₂. By using the PLS_{VC} CO₂ yield with polyhedra morphology is predicted, which results in five influential wavenumbers. These wavenumbers are mapped to C=O, C equiv C, CH, C-H,=CH₂ and O-H. By using the PLS_{WC} H₂ conversion (%) with polyhedra morphology is predicted, which results in five influential wavenumbers. These wavenumbers are mapped to C-O, N-H, C=O, C=C and C-H. By using the PLS_{VC} , the CO conversion (%) with rod morphology is predicted, which results in three influential wavenumbers. These wavenumbers are mapped to s-RCH=CHR, C-O and C equiv C. The PLS_{WC} CO₂ yield with rod morphology is predicted, resulting in one influential wavenumber. This wavenumber is unlabeled because it does not belong to any functional compound. By using the PLS_{WV} , the H₂ conversion (%) with rod morphology is predicted, which results in one influential wavenumber. This wavenumber is mapped to s-RCH=CHR. Notably, the C-O is the common functional compound for CO conversion % prediction with all types of morphologies. Moreover, the functional compounds C-O, C=O, O-H and C-H,=CH₂ are common for CO conversion prediction with cube and polyhedra morphologies. Similarly, the functional compounds C=O, CH, C-H,=CH₂ and O-H are common for CO₂ yield prediction with cube and polyhedra morphologies. The unlabeled wavenumbers are further required to investigate for mapping relevant functional compound. The functional compounds C-O, C=O, CH and C-H,=CH₂ are common for H₂ conversion prediction with cube and polyhedra morphologies. Along the same lines, the functional compound s-RCH=CHR is common for H₂ conversion % prediction with polyhedra and rod morphologies. Notably, we have used the ASL method for baseline correction and the PLS based method for ESR predictions based on IR. The contribution is a mixture of statistical learning and chemometrics, moreover the study is new; hence further studies are required to discover the IR peaks with chemical properties. Hence, findings may or may not be proper until confirmed.

Figure 4. Cont.

Figure 4. Cont.

Figure 4. For predicting ESR products, the RMSE computed over the test and training data from all *PLS* models is compared. The data were divided into 70% train and 30% test for this purpose. This division was repeated 100 times, with the RMSE being recorded after each iteration.

Figure 5. The FTIR spectra were adjusted for baseline correction. The baseline correction is carried out of spectra reference to variability in response, since we have modeled 9 responses, i.e., ESR product hence we got 9 baseline correction used in their respective models. The baseline correction presented here is for the model CO conservation with rod morphology. The original and corrected FTIR spectra is presented here for illustration.

Figure 6. The best model for each ESR product prediction is then used to identify the influential wavenumbers. The influential wavenumbers were further mapped to functional compounds for each optimal model. The influential wavenumbers indicate the intensity of their selection, and the corresponding functional compound is presented.

5. Conclusions

The current paper introduces a partial least squares (PLS)-based measure based on the Johnsen index for predicting ESR products with cube, polyhydra, and rod morphologies. The proposed method makes use of existing filter measures such as loading weights, variable importance on projection, and significant correlation. The proposed PLS measures based on the Johnsen index outperform the existing methods for predicting ESR products based on FTIR spectroscopic data. On simulated data, the proposed Johnsen measure in PLS demonstrates a higher sensitivity and accuracy. The functional compounds C-O, C=O, CH and C-H=CH₂ are common for H₂ conversion prediction with cube and polyhedra morphologies. In a similar vein, the functional compound s-RCH=CHR is frequently used for H₂ conversion prediction with polyhedra and rod morphologies.

Author Contributions: Methodology, T.M.; formal analysis, C.Z., M.T. and T.M.; supervision, C.Z.; project administration, C.Z. All authors contributed equally. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in doi:10.17632/38c6gy3t4r.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Appleby, A. Fuel cells and hydrogen fuel. *Int. J. Hydrogen Energy* **1994**, *19*, 175–180. [\[CrossRef\]](#)
- Cook, B. Introduction to fuel cells and hydrogen technology. *Eng. Sci. Educ. J.* **2002**, *11*, 205–216. [\[CrossRef\]](#)
- Tromp, T.K.; Shia, R.L.; Allen, M.; Eiler, J.M.; Yung, Y.L. Potential environmental impact of a hydrogen economy on the stratosphere. *Science* **2003**, *300*, 1740–1742. [\[CrossRef\]](#)
- Sarkar, N.; Ghosh, S.K.; Bannerjee, S.; Aikat, K. Bioethanol production from agricultural wastes: An overview. *Renew. Energy* **2012**, *37*, 19–27. [\[CrossRef\]](#)
- Kuzhaeva, A.; Dzhevaga, N.; Berlinskii, I. Modernization of catalyst systems for the processes of hydrocarbon conversion to synthesis gas. *ARPN J. Eng. Appl. Sci.* **2019**, *14*, 3535–3543.
- Ogo, S.; Sekine, Y. Recent progress in ethanol steam reforming using non-noble transition metal catalysts: A review. *Fuel Process. Technol.* **2020**, *199*, 106238. [\[CrossRef\]](#)
- Zhang, R.; Huang, C.; Zong, L.; Lu, K.; Wang, X.; Cai, J. Hydrogen production from methanol steam reforming over TiO₂ and CeO₂ pillared clay supported Au catalysts. *Appl. Sci.* **2018**, *8*, 176. [\[CrossRef\]](#)
- Azizan, M.T.; Aqsha, A.; Ameen, M.; Syuhada, A.; Klaus, H.; Abidin, S.Z.; Sher, F. Catalytic reforming of oxygenated hydrocarbons for the hydrogen production: An outlook. *Biomass Convers. Biorefin.* **2020**, 1–24. [\[CrossRef\]](#)
- Kim, C.; Choi, S.; Choi, M.J.; Lee, S.A.; Ahn, S.H.; Kim, S.Y.; Jang, H.W. Photoelectrochemical reduction of CO₂ to syngas by reduced Ag catalysts on Si photocathodes. *Appl. Sci.* **2020**, *10*, 3487, doi:10.3390/app10103487. [\[CrossRef\]](#)
- Ayastuy, J.; Fernández-Puertas, E.; González-Marcos, M.; Gutiérrez-Ortiz, M. Transition metal promoters in CuO/CeO₂ catalysts for CO removal from hydrogen streams. *Int. J. Hydrogen Energy* **2012**, *37*, 7385–7397. [\[CrossRef\]](#)
- Si, R.; Raitano, J.; Yi, N.; Zhang, L.; Chan, S.W.; Flytzani-Stephanopoulos, M. Structure sensitivity of the low-temperature water-gas shift reaction on Cu–CeO₂ catalysts. *Catal. Today* **2012**, *180*, 68–80.
- Qi, G.; Yang, R.T. Characterization and FTIR studies of MnOx-CeO₂ catalyst for low-temperature selective catalytic reduction of NO with NH₃. *J. Phys. Chem. B* **2004**, *108*, 15738–15747. [\[CrossRef\]](#)
- Gonzalez-A, E.; Rangel, R.; Solís-García, A.; Venezia, A.; Zepeda, T. FTIR investigation under reaction conditions during CO oxidation over Ru(x)-CeO₂ catalysts. *Mol. Catal.* **2020**, *493*, 111086. [\[CrossRef\]](#)
- Milt, V.; Querini, C.; Miró, E.; Ulla, M. Abatement of diesel exhaust pollutants: NOx adsorption on Co, Ba, K/CeO₂ catalysts. *J. Catal.* **2003**, *220*, 424–432. [\[CrossRef\]](#)
- Jobbagy, M.; Marino, F.; Schönbrod, B.; Baronetti, G.; Laborde, M. Synthesis of copper-promoted CeO₂ catalysts. *Chem. Mater.* **2006**, *18*, 1945–1950. [\[CrossRef\]](#)
- Quiroz, J.; Giraudon, J.M.; Gervasini, A.; Dujardin, C.; Lancelot, C.; Trentesaux, M.; Lamonier, J.F. Total oxidation of formaldehyde over MnO_x-CeO₂ catalysts: The effect of acid treatment. *ACS Catal.* **2015**, *5*, 2260–2269.
- Martens, H.; Naes, T. Multivariate calibration. In *Chemometrics*; Springer: Berlin/Heidelberg, Germany, 1984; pp. 147–156.

18. Mehmood, T.; Iqbal, M.; Hassan, R. Prediction of antibacterial activity in ionic liquids through FTIR spectroscopy with selection of wavenumber by PLS. *Chemom. Intell. Lab. Syst.* **2020**, *206*, 104124. [[CrossRef](#)]
19. Li, C.; Shi, X.; Bao, L.; Yang, J.; Damirin, A.; Zhang, J. The correlation between multiple variable factors and the autocatalytic properties of cerium oxide nanoparticles based on cell viability. *New J. Chem.* **2018**, *42*, 9975–9986.
20. García Martínez, T.; Veses Roda, A.; López Sebastián, J.M.; Puértolas Lacambra, B.; Pérez-Ramírez, J.; Callén Romero, M. Determining bio-oil composition via chemometric tools based on infrared spectroscopy. *Acs Sustain. Chem. Eng.* **2017**, *5*, 8710–8719. [[CrossRef](#)]
21. Tenenhaus, M. *La Régression PLS: Théorie et Pratique*; Editions Technip: Paris, France, 1998.
22. Kvalheim, O.M.; Karstang, T.V. Interpretation of latent-variable regression models. *Chemom. Intell. Lab. Syst.* **1989**, *7*, 39–51. [[CrossRef](#)]
23. Kvalheim, O.M. Interpretation of partial least squares regression models by means of target projection and selectivity ratio plots. *J. Chemom.* **2010**, *24*, 496–504.
24. Mehmood, T.; Liland, K.H.; Snipen, L.; Sæbø, S. A review of variable selection methods in partial least squares regression. *Chemom. Intell. Lab. Syst.* **2012**, *118*, 62–69.
25. Mehmood, T.; Sæbø, S.; Liland, K.H. Comparison of variable selection methods in partial least squares regression. *J. Chemom.* **2020**, *34*, e3226. [[CrossRef](#)]
26. Johnson, J.W. A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivar. Behav. Res.* **2000**, *35*, 1–19. [[CrossRef](#)] [[PubMed](#)]
27. Johnson, R.M. The minimal transformation to orthonormality. *Psychometrika* **1966**, *31*, 61–66.
28. Cifuentes, B.; Bustamante, F.; Araiza, D.G.; Diaz, G.; Cobo, M. Hydrogen purification of actual syngas streams for energy applications: Au-Cu supported over nano-shaped CeO₂ as stable catalysts for the carbon monoxide removal. *Appl. Catal. A Gen.* **2020**, *598*, 117568. [[CrossRef](#)]
29. Cifuentes, B.; Cobo, M.; Bustamante, F. Activity, selectivity and characterization data of nano-shaped CeO₂ supports for carbon monoxide removal from an actual syngas stream. *Mendeley Data* **2019**, V2. [[CrossRef](#)]
30. Liland, K.H.; Almøy, T.; Mevik, B.H. Optimal choice of baseline correction for multivariate calibration of spectra. *Appl. Spectrosc.* **2010**, *64*, 1007–1016. [[CrossRef](#)] [[PubMed](#)]