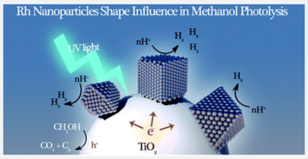
The TEM images-based predictive modeling for differently shaped Rh nanoparticles classification in a hybrid photocatalyst



M.A. Nazarkovsky[1], B. Albuquerque[2], G. Chacón[2], J. Dupont[2]

[1]Chemistry Department, Pontifical Catholic University of Rio de Janeiro, 225 Marques de Sao Vicente Str., Rio de Janeiro 22451-900, Brazil

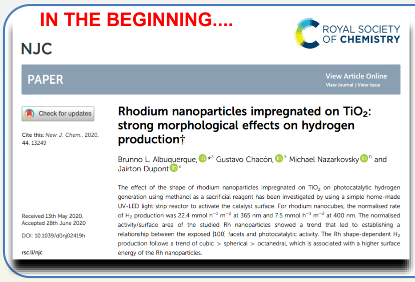
[2]Laboratory of Molecular Catalysis, Institute of Chemistry, UFRGS, 9500 Bento Gonçalves Av., Porto Alegre 91501-970, RS, Brazil

#DATA SCIENCE #INDUSTRY 4.0 #PREDICTIVE MODELING #MACHINE LEARNING #AI for #CHEMISTRY

Development of the four materials science and engineering paradigms

- 1st – Empirical Science (experiments)
- 2nd – Theoretical Science (laws of natural sciences)
- 3rd – Computational Science, Simulations (DFT, molecular dynamics)
- **4th – Big Data-Driven Science (Artificial Intelligence)**

IN THE BEGINNING....



ROYAL SOCIETY OF CHEMISTRY

NJC

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *New J. Chem.*, 2020, **44**, 13249

Rhodium nanoparticles impregnated on TiO_2: strong morphological effects on hydrogen production†

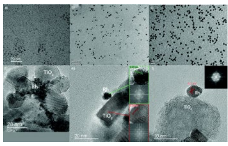
Brunno L. Albuquerque,*[a] Gustavo Chacón,[a] Michael Nazarkovsky[b] and Jairton Dupont[a]

The effect of the shape of rhodium nanoparticles impregnated on TiO_2 on photocatalytic hydrogen generation using methanol as a sacrificial reagent has been investigated by using a simple home-made UV-LED light strip reactor to activate the catalyst surface. For rhodium nanocubes, the normalised rate of H_2 production was 22.4 mmol h^{-1} m^{-2} at 365 nm and 7.5 mmol h^{-1} m^{-2} at 400 nm. The normalised activity/surface area of the studied Rh nanoparticles showed a trend that led to establishing a relationship between the exposed [100] facets and photocatalytic activity. The Rh shape-dependent H_2 production follows a trend of cubic > spherical > octahedral, which is associated with a higher surface energy of the Rh nanoparticles.

Received 13th May 2020,
Accepted 28th June 2020

DOI: 10.1039/d0nj02419h

rsc.li/njc



TEM image of colloidal (a) RhNP, (b) RhNC, and (c) RhOh and HRTEM images of (d) RhNP, (e) RhNC showing the (200) atomic planes and corresponding FFT for Rh (green box) and TiO_2 (red box) and (f) a single RhOh nanoparticle on TiO_2 showing the (111) atomic planes of Rh and corresponding FFT (inset).

...AND THE "WOW" IDEA EMERGED...

The results allow us to distinguish each shape by size distribution profiles from respective TEM microphotographs and make predictive modeling by means of machine learning algorithms!

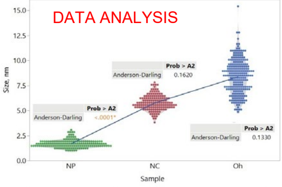
METHODOLOGY

1) data analysis (the size distribution profiles distributions classification, analysis of variances, discriminant analysis);
2) machine learning with training (70%) and validation (30%) of the models stratified by the samples' types;
3) coded in JSL (JMP scripting language) associable with R, Python, Matlab and SAS;
4) scripts, calculators and some other details are available on the repository: https://github.com/Nazarkovsky/Rh-TiO2-classificator

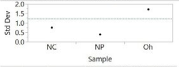
THE MODELS:

K-Nearest Neighbors (KNN) – K = 100, Euclidean distances between the points
Bootstrap Forest - 1 split per sample, learning rate 0.1, 35 trees
Logistic Regression
Classification Tree - 9 splits
Naive Bayes Classificator

DATA ANALYSIS



Tests that the Variances are Equal



Level	Count	Std Dev	MeanAbsDif to Mean	MeanAbsDif to Median
NC	150	0.750018	0.593170	0.588520
NP	150	0.393962	0.298627	0.291800
Oh	225	1.719237	1.387676	1.386444

Test	F Ratio	DFNum	DFDen	Prob > F
O'Brien[.5]	48.5655	2	522	<.0001*
Brown-Forsythe	115.7745	2	522	<.0001*
Levene	116.6925	2	522	<.0001*
Bartlett	160.9747	2	.	<.0001*

Welch's Test

Welch Anova testing Means Equal, allowing Std Devs Not Equal

F Ratio	DFNum	DFDen	Prob > F
2927.9852	2	307.35	<.0001*

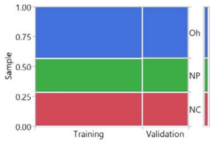
Kruskal-Wallis Tests (Rank Sums)

Level	Count	Score Sum	Expected Score	Score Mean	(Mean-Mean0)/Std0
NC	150	36553.5	39450.0	243.690	-1.844
NP	150	11325.0	39450.0	75.500	-17.911
Oh	225	90196.5	59175.0	400.873	18.035

1-Way Test, ChiSquare Approximation

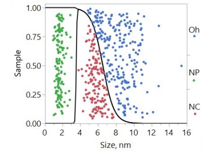
ChiSquare	DF	Prob>ChiSq
417.4483	2	<.0001*

TRAIN-VALIDATION BY SAMPLES



Count Total % Col % Row %	NC	NP	Oh	Total
Training	105 20.00 70.00 28.53	105 20.00 70.00 28.53	158 30.10 70.22 42.93	368 70.10
Validation	45 8.57 30.00 28.66	45 8.57 30.00 28.66	67 12.76 29.78 42.68	157 29.90
Total	150 28.57	150 28.57	225 42.86	525

LOGISTIC REGRESSION



Confusion Matrix

Training

Actual Sample	Predicted Count NC	NP	Oh
NC	84	0	21
NP	0	105	0
Oh	28	0	130

Validation

Actual Sample	Predicted Count NC	NP	Oh
NC	44	0	1
NP	0	45	0
Oh	11	0	56

Whole Model Test

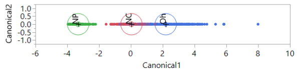
Model	-LogLikelihood	DF	ChiSquare	P
Difference	303.05418	2	606.1084	<.0001*
Full	93.89866			
Reduced	396.95283			

Fit Details

Measure	Training	Validation	Definition
Entropy RSquare	0.7635	0.7975	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.9129	0.9285	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.2552	0.2186	∑ -Log(ρ[j])/n
Misclassification Rate	0.1332	0.0764	∑ (ρ[j]≠ρMax)/n
N	368	157	n

DISCRIMINANT ANALYSIS

Canonical Plot



Score Summaries

Source	Count	Number Misclassified	Percent Misclassified	Entropy RSquare	-2LogLikelihood
Training	525	65	12.3810	0.73058	305.236

Actual Sample	Predicted Count NC	NP	Oh
NC	143	0	7
NP	0	150	0
Oh	58	0	167

BOOTSTRAP FOREST

Overall Statistics

Measure	Training	Validation	Definition
Entropy RSquare	0.8199	0.7636	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.9379	0.9131	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.1943	0.2552	∑ -Log(ρ[j])/n
Misclassification Rate	0.0788	0.0764	∑ (ρ[j]≠ρMax)/n
N	368	157	n

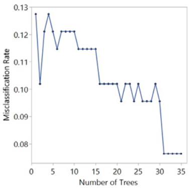
Confusion Matrix

Training

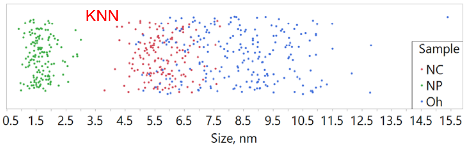
Actual Sample	Predicted Count NC	NP	Oh
NC	94	0	11
NP	0	105	0
Oh	18	0	140

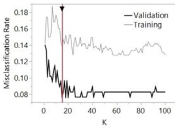
Validation

Actual Sample	Predicted Count NC	NP	Oh
NC	42	1	2
NP	0	45	0
Oh	9	0	58



KNN





Confusion Matrix for Best K=15

Training

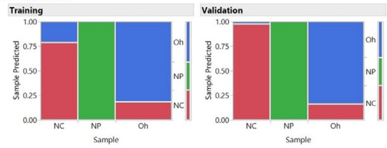
Actual Sample	Predicted Count NC	NP	Oh
NC	83	0	22
NP	0	105	0
Oh	29	0	129

Validation

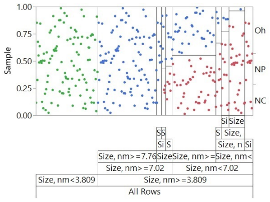
Actual Sample	Predicted Count NC	NP	Oh
NC	44	0	1
NP	0	45	0
Oh	11	0	56

Mosaic Plot

Mosaic Plot for K=15



CLASSIFICATION TREE



Fit Details

Measure	Training	Validation	Definition
Entropy RSquare	0.7894	0.8127	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.9248	0.9350	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.2272	0.2023	∑ -Log(ρ[j])/n
Misclassification Rate	0.1304	0.0828	∑ (ρ[j]≠ρMax)/n
N	368	157	n

Confusion Matrix

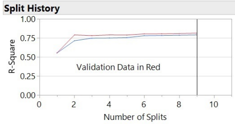
Training

Actual Sample	Predicted Count NC	NP	Oh
NC	97	0	8
NP	0	105	0
Oh	40	0	118

Validation

Actual Sample	Predicted Count NC	NP	Oh
NC	45	0	0
NP	0	45	0
Oh	13	0	54

Split History



NAIVE BAYES

Fit Details

Measure	Training	Validation	Definition
Entropy RSquare	0.7576	0.7948	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.9102	0.9273	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.2615	0.2216	∑ -Log(ρ[j])/n
Misclassification Rate	0.1250	0.0764	∑ (ρ[j]≠ρMax)/n
N	368	157	n

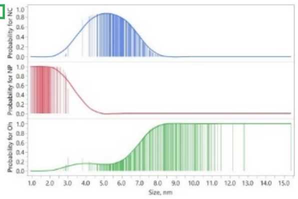
Confusion Matrix

Training

Actual Sample	Predicted Count NC	NP	Oh
NC	93	0	12
NP	0	105	0
Oh	34	0	124

Validation

Actual Sample	Predicted Count NC	NP	Oh
NC	45	0	0
NP	0	45	0
Oh	12	0	55



Conclusions. **NC and Oh are normally distributed by the Anderson-Darling criterion, except NP whose nanoparticles size range is the narrowest among three samples. The variances are revealed to be heteroscedastic by all four tests, the non-parametric Kruskall-Wallis test has shown the non-equality for all three samples. The discriminant analysis at the overall MR of 12.38% has become promising to develop the machine learning algorithms for practical digital recognition of the samples by size. The most precise model is Logistic Regression at the misclassification rate MR is 7.64% with other better metrics (Generalized R^2 and Entropy R^2) than another two models have at the same MR (Naive Bayes and Bootstrap Forest). As a result, an offline HTML-calculator for Logistic Regression was developed.**

Acknowledgements. CAPES (158804/2017-01 and 001; 2013037-31005012005P5–PNPD-PUC Ri), FAPERGS (16/2552-0000 and 18/2551-0000561-4), and CNPq (406260/2018-4, 406750/2016-5 and 465454/2014-3), the National Laboratory of Nanotechnology (LNNano-Campinas) under proposal TEM-24375 for the HRTEM analysis, the Central Laboratory of Electron Microscopy (LCME) at UFSC for the TEM analysis. Special thanks to Dr. David Kirmayer (The Hebrew University of Jerusalem) for consulting and valuable advice.