

Understanding Catalytic Biomass Conversion Through Data Mining

Erik-Jan Ras · Ben McKay · Gadi Rothenberg

Published online: 3 June 2010

© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract Catalytic conversion of biomass is a key challenge that we chemists face in the twenty-first century. Worldwide, research is conducted into obtaining bulk chemicals, polymers and fuels. Our project centres on glucose valorisation via furfural derivatives using catalytic hydrogenation. We present here new results for a set of 48 bimetallic catalysts supported on silica, and demonstrate the application of data mining tools to identify major trends in the data. These results are combined with a full factorial data set for the hydrogenation of 5-ethoxymethylfurfural over alumina-supported transition metal catalysts. All the catalysts in the combined datasets were synthesized and tested for performance under identical conditions. This, combined with the fact that no combinations of metals were left out, enables the use of advanced data mining tools. The paper describes the data and highlights the relevant trends from a chemist's viewpoint.

Keywords Biomass conversion · Iridium · Palladium · Platinum · Furfural · Bimetallic catalysts · Parallel flow reactors · Sugar conversion · Heterogeneous catalysis · Trickle flow

Electronic supplementary material The online version of this article (doi:10.1007/s11244-010-9563-z) contains supplementary material, which is available to authorized users.

E.-J. Ras · B. McKay
Avantium Technologies B.V., Zekeringstraat 29,
1014BV Amsterdam, The Netherlands

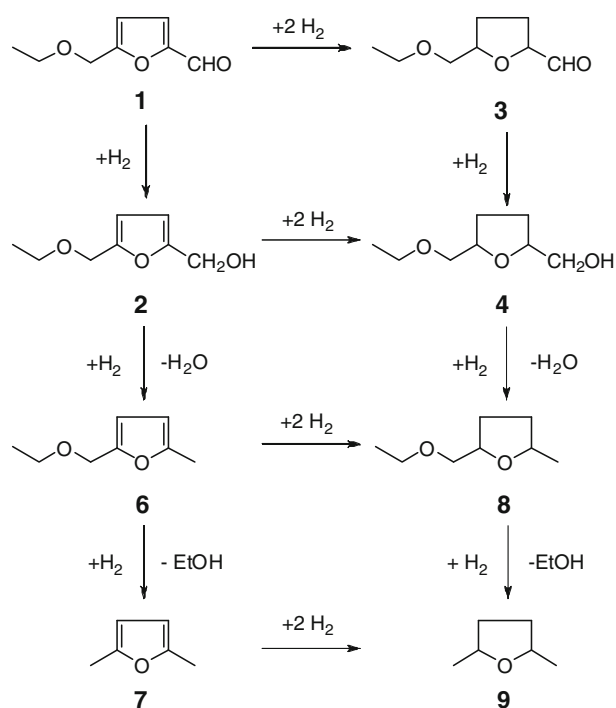
E.-J. Ras · G. Rothenberg (✉)
Van 't Hoff Institute of Molecular Sciences, University
of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam,
The Netherlands
e-mail: g.rothenberg@uva.nl

1 Introduction

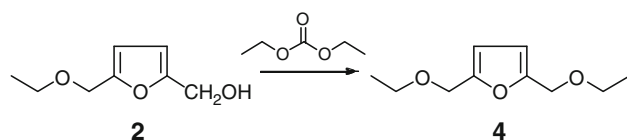
Valorising biomass to bulk chemicals and/or fuel components is a 'hot topic' of research worldwide. Typically, "brute force" reactions are first used to break down the raw biomass into smaller building blocks [1–4]. Once this is done, it's a job for catalysis chemists to turn these building blocks into useful chemicals. The challenge is that biomass differs significantly from crude oil. Crude oil is composed largely of unfunctionalised hydrocarbons. Biomass, on the other hand, is typically over-functionalized. It is a mixture of alcohols, ethers, esters, and carboxylic acids, so biomass refining presents new challenges. Previously, we have studied various catalytic options for valorising fatty acids [5], triglycerides [6], and glycerol derivatives [7]. Here, we extend this work to furfural derivatives, that can be obtained from cellulose, starch, hemicellulose and sugars [8].

Here we report new results on the selective hydrogenation of 5-ethoxymethylfurfural **1** over silica-supported bimetallic catalysts (Scheme 1). The results are discussed in terms of mechanistic implications and process conditions, using data mining techniques to highlight general trends and explain their chemical meaning. We recently reported the hydrogenation of **1** over alumina supported catalysts [9]. The key results of the alumina-based catalysts are compared here with the silica-based results in terms of major trends.

Our main objective is the selective carbonyl reduction product 5-hydroxyethylfurfuryl alcohol **2**, although high selectivity routes to other products are also interesting. The alcohol **2** can be used as a component in novel diesel fuels. Alternatively, it can be oxidized to 2,5-furandicarboxylic acid and used as a monomer in biobased polyester synthesis [8].



Only using diethyl carbonate solvent



Scheme 1 Chemical reactions involved in the hydrogenation of 5-ethoxymethylfurfural **1**

Combining the silica and the alumina results gives a large and highly consistent data set. Figure 1 outlines the parameter space explored in this data set. The consistency

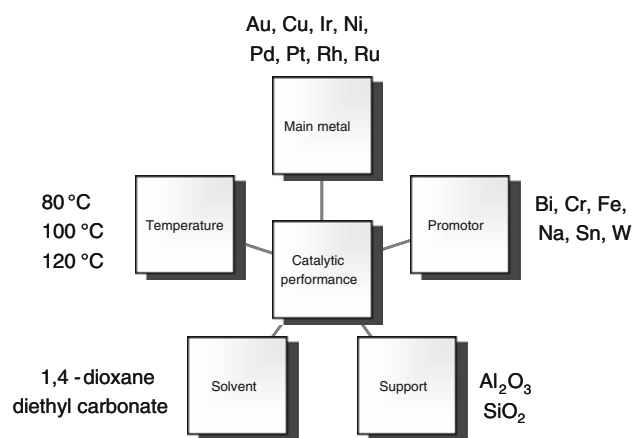


Fig. 1 Parameter space explored in the combined data sets. The full factorial combination of these parameters results in 576 combinations which have all been tested. Additionally, monometallic catalysts and blanks were also tested

comes from using high-throughput methodology, applying a 16-in-parallel flow reactor (<http://www.flowrence.com>) and ensuring that all catalysts are synthesized and tested under identical conditions. To efficiently explore this data set we use Principal Component Analysis (PCA), a technique typically used in data mining for dimension reduction and exploratory data analysis [10–12].

2 Results and Discussion

All catalysts were synthesized using a wet impregnation procedure. The liquid volume was chosen such that all catalysts could be prepared using identical precursor mass concentrations. This gives a rather dilute impregnation slurry, where the volume is determined by the metal precursor with the lowest solubility. In all cases impregnation time was 24 h. Drying of catalysts was performed at 105 °C, followed by calcination in air at 320 °C. All experiments were performed in trickle flow in a commercial 16-reactor setup at 80–120 °C. In a typical experiment, the reactor system is loaded with 15 calcined catalysts at 50 mg each. The remaining reactor is left empty to obtain a reliable measurement of conversion and probe for uncatalysed reactions. The catalysts are first reduced at 300 °C using a mixture of hydrogen and nitrogen. After cooling down to 80 °C, each reactor is supplied with a feed of substrate solution corresponding to a LHSV of 0.032 mol g⁻¹ h⁻¹. The samples are collected in the liquid phase and analyzed offline using GC (Varian #CP9205 column, FID detector). A detailed description of the equipment used and the catalyst synthesis methodology is published elsewhere [9].

2.1 Temperature Effects in Activity and Selectivity

At a first glance, the number of catalysts with desirable performance is fairly low in the SiO₂-supported group when compared with our earlier results on Al₂O₃-supported catalysts. Where in the Al₂O₃ data set a substantial set of catalyst showed high conversions, in the SiO₂ data set only about 10% of the catalysts show conversions higher than 20%. Figure 2 shows the selectivity towards the unsaturated alcohol **2** as a function of the conversion of **1**. The markers are shaded by test temperature. The observed temperature effect differs from the alumina-based catalysts. In this case, a 100% conversion was achieved in many cases already at 80 °C and with increasing temperature the selectivity towards the unsaturated alcohol **2** dropped dramatically. Here we see a more modest temperature effect.

Observing Fig. 2 in more detail, one sees similarities between silica and alumina. Generally speaking, using diethyl carbonate as a solvent gives higher conversion

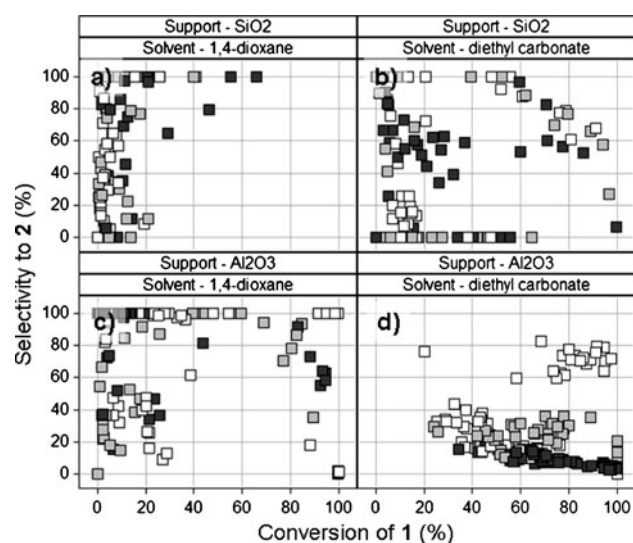


Fig. 2 Selectivity towards the unsaturated alcohol **2** versus conversion of **1**. Subplot **a** contains data for SiO₂ catalysts tested in 1,4-dioxane. Subplot **b** contains data for SiO₂ catalysts tested in diethyl carbonate. Subplot **c** contains data for Al₂O₃ catalysts tested in 1,4-dioxane. Subplot **d** contains data for Al₂O₃ catalysts tested in diethyl carbonate. For all subplots, markers are shaded from light to dark with increasing reaction temperature

compared with using 1,4-dioxane. This is caused by a lower barrier for adsorbing 1,4-dioxane compared with diethyl carbonate on the catalyst surface. When looking

at the best result obtained in this set, two catalysts with full selectivity to the unsaturated alcohol **2** stand out. These are Pt and Rh catalysts with a conversion of 50–65%. Using 1,4-dioxane, this performance is achieved at 120 °C; using diethyl carbonate, this performance is reached at 80 °C or 100 °C. Maintaining high selectivity at high temperature is usually difficult, because at high temperature often excess hydrogenation reactions take over, ultimately resulting in products like dimethyl furan and dimethyl tetrahydrofuran. The results in terms of conversion of **1** and selectivity to **2** obtained for the silica catalysts containing Pd, Pt and Rh catalysts are given in Table 1. The data for all catalysts is given in the supporting information.

2.2 Effect of Main Metals and Promoters

To facilitate interpretation of the effects of main metals and promoters, all catalysts were prepared with a fixed main metal loading of 1 wt% and a fixed promotor loading of 10 mol% relative to the main metal. This 1:10 ratio should be sufficient to identify a promotor effect if present. We selected a diverse set of promotor metals, aiming to create large differences between promoters. The promoters include classical hydrogenation promoters (Cr, Fe, W), an adsorption promotor (Na), a poison (Bi) and a common promotor

Table 1 Key hydrogenation results obtained for SiO₂-supported catalysts containing Pd, Pt or Rh

Catalyst		Solvent: 1,4-dioxane						Solvent: diethyl carbonate					
		80 °C		100 °C		120 °C		80 °C		100 °C		120 °C	
Entry	Composition ^a	C ^b	S ^c	C ^b	S ^c	C ^b	S ^c	C ^b	S ^c	C ^b	S ^c	C ^b	S ^c
M01	Pd	1	43	1	47	4	39	79	79	89	66	77	56
M02	Pt	5	50	6	65	2	100	6	76	6	100	17	58
M03	Rh	–	–	3	100	13	75	–	–	–	–	9	50
B01	Pd/Bi	–	–	–	–	2	82	20	72	16	69	26	34
B02	Pd/Cr	3	55	7	35	12	45	76	77	79	77	71	60
B03	Pd/Fe	2	71	4	77	11	69	52	92	74	70	60	53
B04	Pd/Na	2	26	2	28	3	38	61	88	62	89	37	59
B05	Pd/Sn	8	63	17	77	46	79	81	61	96	27	100	6
B06	Pd/W	–	–	3	39	11	36	91	68	94	58	86	52
B07	Pt/Bi	2	100	4	74	4	100	0	100	9	100	27	55
B08	Pt/Cr	24	100	41	100	66	100	55	100	40	100	60	97
B09	Pt/Fe	20	100	19	100	21	100	21	100	14	100	28	63
B10	Pt/Na	3	86	1	100	3	100	8	58	6	100	21	44
B11	Pt/Sn	16	100	11	100	20	100	48	100	52	100	70	83
B12	Pt/W	4	59	13	23	5	79	6	100	5	100	23	62
B13	Rh/Bi	–	–	1	100	14	12	–	–	–	–	5	26
B14	Rh/Cr	7	65	8	100	21	97	3	100	5	100	16	60
B15	Rh/Fe	26	100	40	100	55	100	13	100	6	100	19	51
B16	Rh/Na	4	30	12	23	10	86	–	–	–	–	12	55
B17	Rh/Sn	8	34	4	100	11	98	0	100	3	100	–	–
B18	Rh/W	9	57	14	79	29	65	6	100	3	100	32	39

^a Each catalyst is prepared by co-impregnation of 1 wt% main metal and 10 mol% promotor (if any) relative to the main metal content

^b C denotes conversion in % of 5-ethoxymethylfurfural (**1**) and is calculated as $100\% \cdot ([1]_{\text{feed}} - [1]_{\text{effluent}}) / [1]_{\text{feed}}$. The concentration of **1** in the feed is determined from a blank reactor

^c S denotes selectivity in % to 5-ethoxymethylfurfuryl alcohol (**2**) and is calculated as $100\% \cdot [2]_{\text{effluent}} / ([1]_{\text{feed}} - [1]_{\text{effluent}})$

for the hydrogenation of aldehydes (Sn). The main metals are mostly typical noble-metal catalyst components (Au, Ir, Pd, Pt, Rh, Ru), but Ni is also included as a typical general purpose hydrogenation catalyst. Cu is included since it is commonly used in the gas phase hydrogenation of furfural to furfuryl alcohol [13, 14]. Most of the main metals used in our work were used before in the hydrogenation of (unsaturated) aldehydes with varying degrees of success. Platinum [15–20] and palladium [21] are the most common metals used.

The performance comparison in Fig. 3 clearly shows that Pd outperforms the other metals at all temperatures in terms of conversion. At 120 °C, only Pt comes close. Surprisingly, Ir only shows marginal activity when supported on SiO₂. Since ultimately selectivity is more important than conversion, both Pt and Rh should be considered for a closer investigation due to their high average selectivities.

2.3 Applying Data Mining Methodology

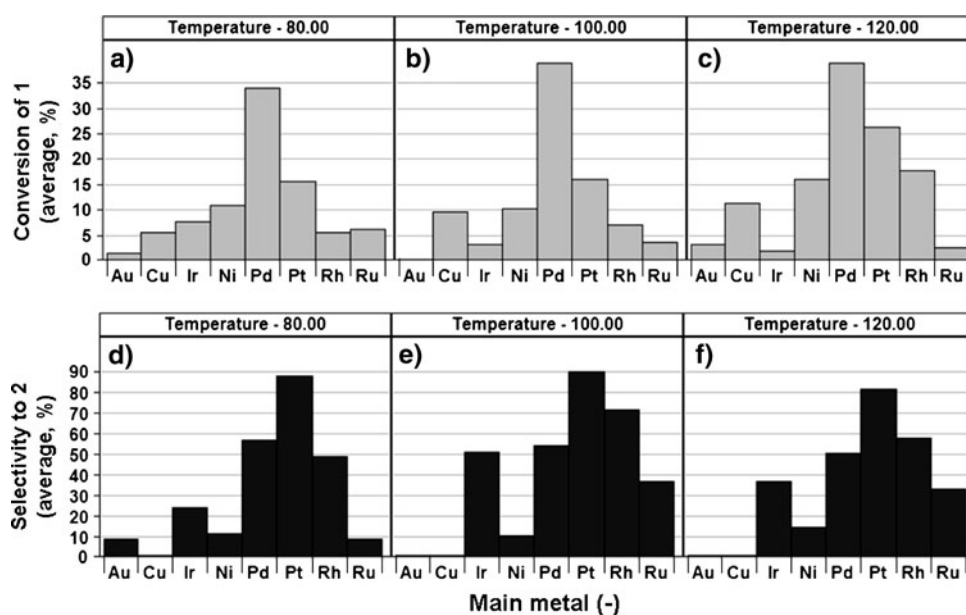
Principal Component Analysis (PCA) is often used to analyze large data sets in many disciplines. To accommodate the non-statistician reader, we include a short introduction below (adapted from [22] with permission of the author). Detailed descriptions of PCA and data mining in general are given elsewhere [23, 24]. The remainder of this section outlines the application of PCA to the combined data set reported here and in our previous work. We demonstrate the identification of major trends by analyzing the yields of identified components in the catalytic test and link these trends to the catalyst compositions and test conditions used. Please note that this type of modeling does

not result in a predictive model. It will not predict what new combination of variables to use to obtain a better result. It merely aims to explain the existing data in a coherent manner.

2.4 A Simple Explanation of the Concept of PCA

Suppose that you have an experimental data matrix that contains the concentration profiles of 12,500 reactions, performed using 50 different catalysts tested with 50 different substrates under five different conditions, with each profile made of 10 points. This matrix contains $50 \times 50 \times 5 \times 10 = 125,000$ data points. It merits some serious thinking about data mining, because it is unlikely that you will see anything useful just by looking at 125,000 numbers. PCA can reduce this large matrix into two smaller matrices which are easier to examine and interpret. Using PCA, you can extract the key factors. These are the principal components, or PCs (sometimes also called the latent variables). Each PC is a linear combination of the original variables, but unlike the original variables, which may be correlated with each other, the PCs are orthogonal (i.e., uncorrelated, independent of one another). Mathematically speaking, if X is an $(I \times J)$ matrix that contains J variables for I reactions, PCA divides this matrix into a systematic part TP^T (the PCA model), and a residuals part E . T ($I \times R$), and P ($J \times R$) are two smaller matrices, the size of which depends on R , the number of significant PCs. T is the scores matrix. It represents the spread of the reactions within the model space. P is the loadings matrix. It describes the relationships between the variables. The geometrical representation of PCA is outlined in Fig. 4 below for further clarification.

Fig. 3 Average conversions (top row) and selectivity towards **2** (bottom row) obtained for the main metals tested. The averages are calculated over all catalysts containing a given main metal including all variations on promotor and solvent. The data is split from left to right for the three test temperatures 80 °C, 100 °C, and 120 °C



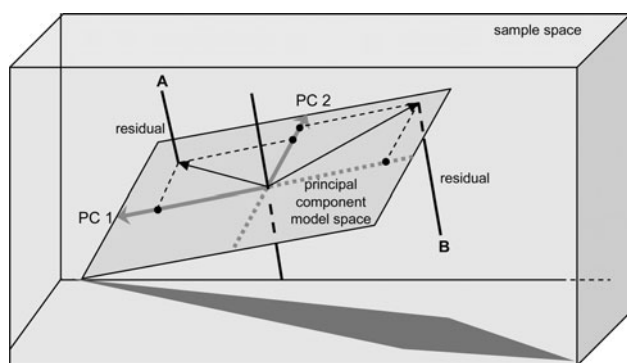


Fig. 4 PCA reduces the dimensionality of the problem by projecting the original dataset onto a lower-dimension PC model, in which the new variables are orthogonal to each other. The distance from point A to the PCA model space equals the residual value for catalyst A (reproduced with permission from Ref. [22])

2.5 Description of the Data Set

The data set contains data for all binary combinations of metals (8 main metals and 6 promotor metals) supported on silica and alumina. This set of 96 catalysts is augmented with the unpromoted main metals, again supported on silica and alumina, resulting in a set of 112 catalysts in total. Each catalyst is tested in two solvents and at three different temperatures. All other conditions were maintained constant. The yields of components 2–9 are used as variables for the PCA analysis. Prior to model generation, the yields were scaled to zero mean and unit variance, to eliminate inherent differences in scale.

2.6 Modelling Results

PCA analysis reduces the yields in the original parameter space to two PCs together explaining 68% of the overall variance in the original data. Practically this means we can

describe almost 70% of what is going on in our experiments by looking at two factors rather than at the original eight. The remainder of this section outlines the key findings from interpretation of the model. A complete set of plots and the full scores and loadings tables are provided in the supporting information.

Looking at Fig. 6a, we see that the loadings for all yields are positive in the first PC. This simply means that this first PC is directly correlated with catalyst activity. Every observation with a high score on the first PC will have a high conversion and every observation with a low score on the first PC will have a low conversion. Translating this observation to the scores plot (Fig. 5) implies that generally conversion increases from left to right. This means we can effectively explain a large portion of the differences in conversion using the first PC. Looking at the loadings on the second PC (Fig. 6b), we see large differences in effect size (magnitude of the bar). Moreover, the effects have different directions. This gives information about the relationship between the variables. For example, the yield of the unsaturated aldehyde **3** is anti-correlated with the yield of the etherification product **5**. This means that if a high yield of **3** is found, a low yield of **5** is expected and vice versa. From a chemistry point of view this makes sense, because if the furan ring has been hydrogenated to produce **3**, the formation of the unsaturated component **5** is no longer possible. More generally, the second PC gives information about the selectivity of each observation.

Combining these two PCs allows us to quickly find the maximum yield of a given component in the scores plot. For example, a high yield of the saturated aldehyde **3** is expected at a moderately positive score on the first PC based on the loading for $y(3)$ in Fig. 6a and a strongly positive score on the second PC based on the loading for $y(3)$ in Fig. 6b. Following the same reasoning, we can

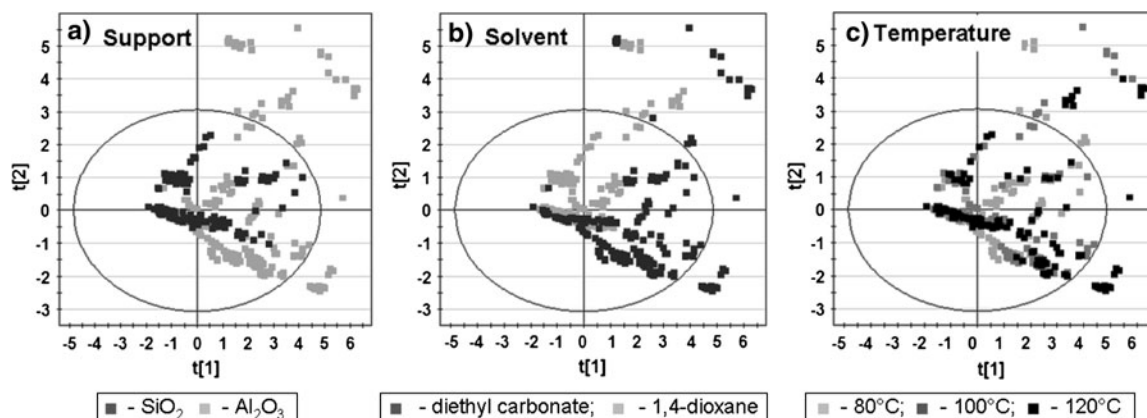
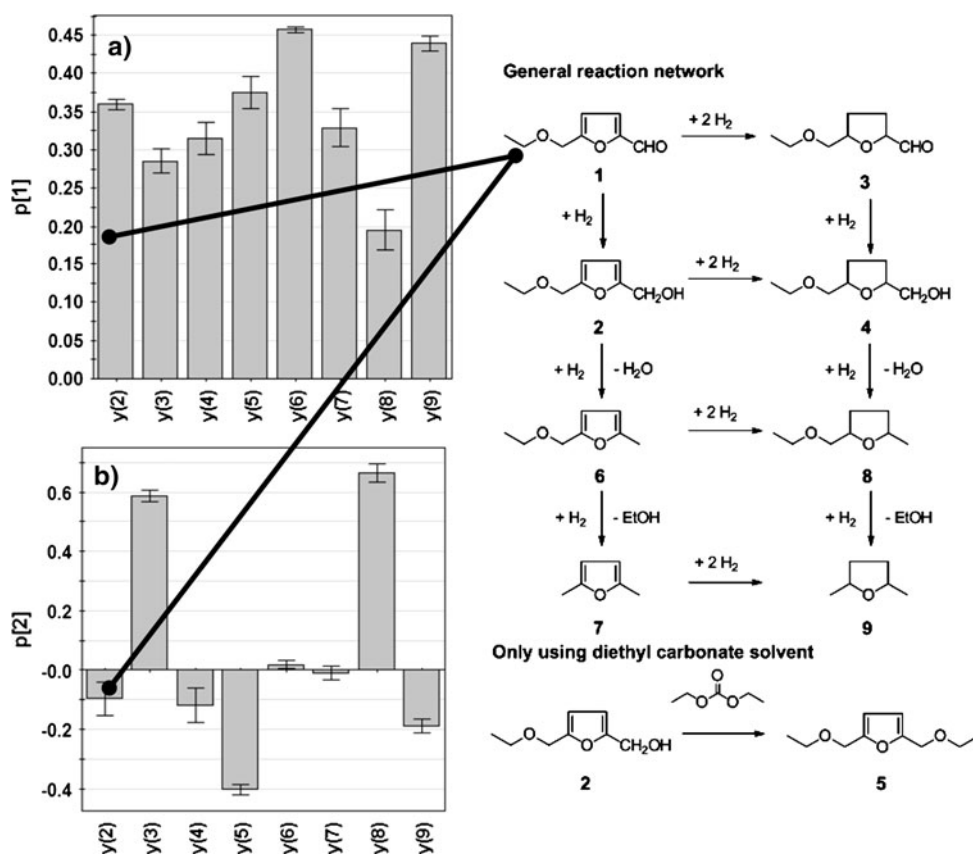


Fig. 5 Scores plot of the two PC model colored by **a** support, **b** solvent and **c** temperature. In all cases, the x -axis corresponds to the first PC $t[1]$, explaining 49% of variance in the data and the y -axis corresponds to the second PC $t[2]$, explaining 19% of variance in the data

Fig. 6 Loadings plots for the first PC **a** and second PC **b**, accompanied by the reaction scheme for indicating the identity of components. The error bars show the 95% confidence interval on the loadings. Only **y(7)** and possibly **y(6)** could be considered insignificant since the confidence interval either includes or is very close to zero



expect a high yield of the etherification product **5** at a strongly positive score on the first PC and a strongly negative score on the second PC. Subtle trends can also be identified. An example of this is the saturated alcohol **4**, which only occurs in small amounts (a maximum of 5% yield and an average of only 0.1% yield across all data points). This component may easily be overlooked by conventional visual data analysis. In our case, the PCA model shows in Fig. 6b that a “high” yield of **4** always coincides with a high yield of etherification product **5**. This is explained by the fact that high amounts of **5** can only be formed if first substantial quantities of its direct precursor **2** are formed. Instead of etherification to **5**, a small portion of the unsaturated alcohol **2** could be further hydrogenated to the saturated alcohol **4**. The reason only a minor amount of the unsaturated alcohol follows this path is because it requires a change in adsorption mode (ring vs. oxygen).

Turning our attention to the effect of the metals used, the first thing that should be noted is that the two most successful silica-supported catalysts in the data set are not explicitly identified by the model at hand. This is not an error, but simply a confirmation of the most important attribute of PCA models: only trends valid across the majority of the data are identified. Less systematic trends, like two odd catalysts outperforming the other catalysts will not be identified. Clearly, this implies that PCA

analysis should only be used together with conventional data analysis and not as its replacement. One of the major trends that can easily be identified is the high selectivity of Pd catalysts towards the unsaturated alcohol **3**. From Fig. 6 we learn that a high yield **y(3)** coincides with a high score on both the first and second PC. This translates to the top-right quadrant in the scores plot (Fig. 5), which is almost exclusively populated with Pd catalysts.

3 Conclusions

We found that Pt and Rh supported on silica are efficient catalysts for the selective hydrogenation of **1** to the unsaturated alcohol **2**, obtaining 100% selectivity for conversions up to 66%. We also found substantial promoter effects for Pd, Pt and Rh. Especially in the case of Rh, the unpromoted SiO₂/Rh catalyst shows negligible conversion at all test temperatures, while the SiO₂-Rh/Fe catalyst shows a conversion of 55% at 120 °C. Using PCA, we showed that the major trends in a large data set can be easily identified using standard data mining methodology. Effects of support, temperature and solvent are readily found and effects of metals can also be identified.

Comparing the results of the silica and alumina-supported catalysts, one sees some major differences. The

most obvious is the much lower activity typically found for the silica catalysts. This can be explained by the much lower background activity of the unmodified support. Where bare alumina already gives 50% conversion at 80 °C, silica gives no conversion. From a process point of view, the Ir catalysts on alumina are promising due to their near full conversion and selectivity at low temperature. However, from an economic perspective the Pt catalysts on silica are more attractive on the basis metal price and availability. Full selectivity in combination with moderate conversion provides a basis for a recycle process.

4 Supporting Information Available

The full table of results for both alumina and silica data sets is provided in the supporting information as well as the scores and loadings of the PCA model.

Acknowledgements We thank the Netherlands Organization for Scientific Research (NWO) for funding received under the Casimir programme.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Nowakowski DJ, Jones JM (2008) *J Anal Appl Pyrolysis* 83:12
2. Gallezot P (2007) *Catal Today* 121:76
3. Mascal M, Nikitin B (2009) *ChemSusChem* 2:859
4. Lee S, Eom M, Yoo K, Kim N, Jeon J, Park Y, Song B, Lee S (2008) *J Anal Appl Pyrolysis* 83:110
5. Kiss AA, Dimian AC, Rothenberg G (2006) *Adv Synth Catal* 348:75
6. Kiss AA, Dimian AC, Rothenberg G (2008) *Energy Fuels* 22:598
7. Gordillo A, Durán Pachón L, de Jesus E, Rothenberg G (2009) *Adv Synth Catal* 351:325
8. Moreau C, Belgacem MN, Gandini A (2004) *Top Catal* 27:11
9. Ras EJ, Maisuls S, Haesackers P, Gruter GJ, Rothenberg G (2009) *Adv Synth Catal* 351:3175
10. Berglund A, DeRosa MC, Wold S (1997) *J Comput Aided Mol Des* 11:601
11. Janné K, Pettersen J, Lindberg NO, Lundstedt T (2001) *J Chemom* 15:203
12. Kettaneh N, Berglund S, Wold S (2005) *Comput Stat Data Anal* 48:69
13. Hubaut R, Daage M, Bonnelle JP (1986) *Appl Catal* 22:231
14. Rao R, Dandekar A, Baker RTK, Vannice MA (1997) *J Catal* 253:406
15. Serrano-Ruiz JC, Lopez-Cudero A, Solla-Gullon J, Sepulveda-Escribano A, Aldaz A, Rodriguez-Reinoso F (2008) *J Catal* 253:159
16. Wu J, Shen Y, Liu C, Wang H, Geng C, Zhang Z (2005) *Catal Commun* 6:33
17. Mahata N, Goncalves F, Pereira MFR, Figueiredo JL (2008) *Appl Catal A* 339:159
18. Silvestre-Albero J, Serrano-Ruiz JC, Sepúlveda-Escribano A, Rodríguez-Reinoso F (2008) *Appl Catal A* 351:16
19. Shirai M, Tanaka T, Arai M (2001) *J Mol Catal A: Chem* 168:99
20. Zhao F, Ikushima Y, Shirai M, Ebina T, Arai M (2002) *J Mol Catal A: Chem* 180:259
21. Tanaka S, Mizukami F, Niwa S, Toba M, Tasi G, Kunimori K (2002) *Appl Catal A* 229:175
22. Rothenberg G (2008) *Catalysis: concepts and green applications*. Wiley-VCH, Weinheim. ISBN 978-3-527-31824-7
23. Jackson JE (1981) *J Qual Tech* 13:341
24. Rothenberg G (2008) *Catal Today* 137:2