

DOI: 10.21767/2470-6973.100023

The Application of Advanced Design of Experiments for the Efficient Development of Chemical Processes

Paul M Murray* and Laura C Forfar

Paul Murray Catalysis Consulting Limited,
67 Hudson Close, Yate, BS37 4NP, United
Kingdom

Abstract

The combination of Principal Component Analysis (PCA) with Design of Experiments (DoE) is a powerful and very efficient tool for optimising chemical processes. This article explains how to apply PCA and DoE, and demonstrates the benefits with three case studies.

Keywords: Design of experiments; Principal component analysis; Quality by design; Process development; Optimisation; Screening; Solvent selection; Ligand selection

Abbreviations: OVAT: One Variable at a Time; sPC: solvent Principal Component; Cp*: 1,2,3,4,5 Pentamethylcyclopentadiene; PC: Principal Component; R²: Measure of fit; Q²: Measure of prediction; DPEPhos: Bis[(2-diphenylphosphino)phenyl] ether; triple cage/^tBu: 2,8,9-Tri-*i*-butyl-2,5,8,9-tetraaza-1-phosphabicyclo[3.3.3] undecane; Dave-Phos: 2-Dicyclohexylphosphino-2'-(*N,N*-dimethylamino)biphenyl; X-Phos: 2-Dicyclohexylphosphino-2',4',6'-triisopropylbiphenyl.

Received: September 12, 2017; **Accepted:** November 15, 2017; **Published:** November 21, 2017

Introduction

Experimental design is a cost- and time-efficient method of investigating and optimising reactions. Using this approach, several factors can be studied simultaneously, as opposed to the traditional experimental method of varying one factor at a time. The method uses statistics to determine the most important factors in a reaction and the interactions between these factors which may affect the outcome. In our experience, DoE is an underutilised technique throughout chemistry, although its benefits are being recognized and it is being adopted more readily in industrial settings [1-3]. One possibility for the slow update of experimental design is that it is a statistical method, however, the mathematics involved is simple and the focus should be on an improved understanding of chemical reactivity. The aim of this paper is to deliver an overview of experimental design, providing case studies to exemplify its use, and empower chemists to add DoE to their toolkit. DoE is a fundamental tool for Quality by Design (QbD), and this is increasing the requirement to use DoE within industry [4,5].

When considering a chemical reaction there is factors which are easily investigated such as concentration, temperature, pressure or stoichiometry of reagents [3]. These are continuous factors

and are only limited by the ability of the equipment to achieve the appropriate setting of the factor (e.g. temperature, pressure, addition rate etc.). In an experimental design, continuous factors are studied at a high (coded as +1) and low (coded as -1) level¹. Centre point experiments are run at the mid-level in duplicate, to allow for estimation of experimental error and variation. There are other parameters which are also important in a chemical reaction, but are more difficult to investigate, such as solvent, catalyst or ligand. These are termed discrete factors and are limited by their existence, which means they are more difficult to relate to each other (e.g. comparing solvent effects of toluene to methanol to acetonitrile). However, it is possible to correlate these factors by looking at their chemical and physical properties, allowing investigation by DoE.

Principal component analysis

Principal Component Analysis (PCA) is a multivariate data analysis tool which can be used to relate discrete factors through their chemical and physical properties, for example, polarity, polarizability and hydrogen-bonding [6]. PCA allows the creation

¹ The high level is assigned +1 and low level assigned -1 to allow the centre point to be 0.

of new principal components from the multitude of original properties, leading to 3 to 4 properties which can explain 70-80% of the variation in the data. The principal components can be selected to suit the reaction under investigation. In its simplest form, PCA generates a map of solvents where materials close to each other on the map will have similar properties and behave similarly but may be structurally very different. To use a PCA map successfully, the range of materials and the chosen properties should be appropriate for the chemistry in question.

The use of a solvent PCA map allows each solvent principal component to be studied in an experimental design just like the standard factors of temperature, concentration and equivalents of base (**Figure 1a**), where each solvent principal component is examined at a low and high value. Three solvent principal components will together provide the location of one solvent (e.g. -1-1-1) on the map. The corners of a cube can be selected to cover a large area of the solvent map, ranging from -1-1-1 to +1+1+1, where each corner codes for a solvent in the solvent map. The diversity of solvents in the solvent PCA map will ensure the 8 solvents are very different from each other. PCA solvent maps have been developed and used for chemical reactions and polymorph screening [7-10]. PCA has also been used successfully for ligand mapping by us in collaboration with the University of Bristol [11-12] as well as for amines, aldehydes, ketones and Lewis acids [13-15].

Combining PCA and DoE

The combination of PCA with DoE allows for a very efficient investigation of the whole chemical space for a given reaction [16]. When PCA and DoE are combined during reaction screening it directs the future experimental effort into the regions of chemical space where the specific chosen reaction occurs most successfully. An example of a five factor DoE combined with PCA is shown in **Figure 2**, combining solvent principal components with the continuous factors of temperature and concentration. The design can then be investigated with different design options to choose from to balance the amount of information gained with the number of experiments performed. This depends on i) the number of potentially significant factors and ii) the number of experiments that can be carried out accurately in a time efficient manner. In practice, a resolution IV² fractional

²In a resolution III design main terms are confounded with two-way interactions. Some 2-factor interactions are confounded with other 2-factor interactions in a resolution IV design. The actual interaction can often be deciphered through chemical interpretation of each possible interaction. The results are then confirmed during the validation experiments. In a resolution V design, main terms and two-way interactions are free from confounding.

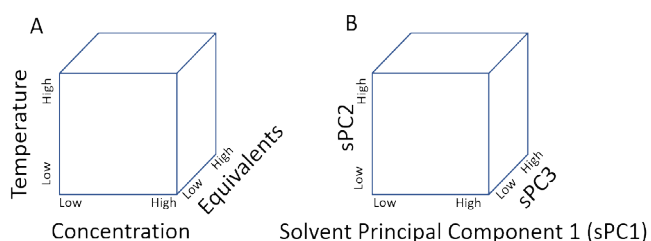


Figure 1 **1a** DoE with three continuous factors. **1b** DoE with three solvent principal components.

factorial design will typically deliver the maximum information for the minimum effort with the understanding that this design will result in confounding of the interactions. Confounding or aliasing are statistical terms describing when the effects of two or more coefficients or interactions are indistinguishable from each other. However, these confounded terms can be separated and explained by carrying out additional experiments.

When a factor is studied in a DoE a suitable high (+1) and low (-1) value for the factor is investigated. Each factor investigated leads to a potential increase in the number of experiments. Two factors requires four experiments while four factors requires sixteen experiments (number of experiments = 2^n where n is the number of factors). When combining PCA with DoE for a straightforward reaction with only 2 reagents, such as an S_N1/S_N2 reaction (**Scheme 1**), a number of factors are likely to be important.

The number of potentially significant parameters for even a seemingly simple reaction can extend into double figures and will include the stoichiometry of the reagents, the concentration and temperature of the reaction, whether an additive, such as a base, is required and the amount of additive, and the solvent. The method for calculating the huge number of reactions required to investigate these factors (6400 experiments, **Figure 3**) has been detailed at length previously [17]. Employing PCA with DoE can reduce this number to just 19 for an initial design. This will identify the significant factors and the optimum ranges and if there are any interactions between the factors. The PCA maps can then be used to identify additional materials in the identified region

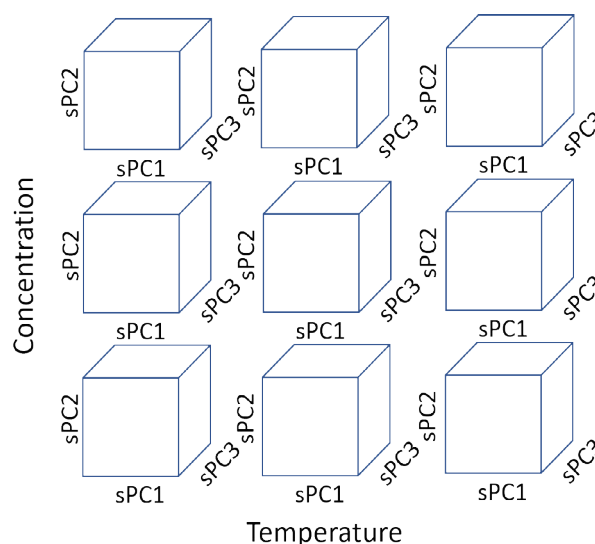
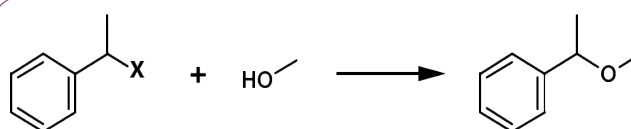


Figure 2 DoE combining three solvent principal components (sPC1, sPC2 and sPC3) with two continuous factors (temperature and concentration).



Scheme 1 Example of an S_N1/S_N2 reaction.

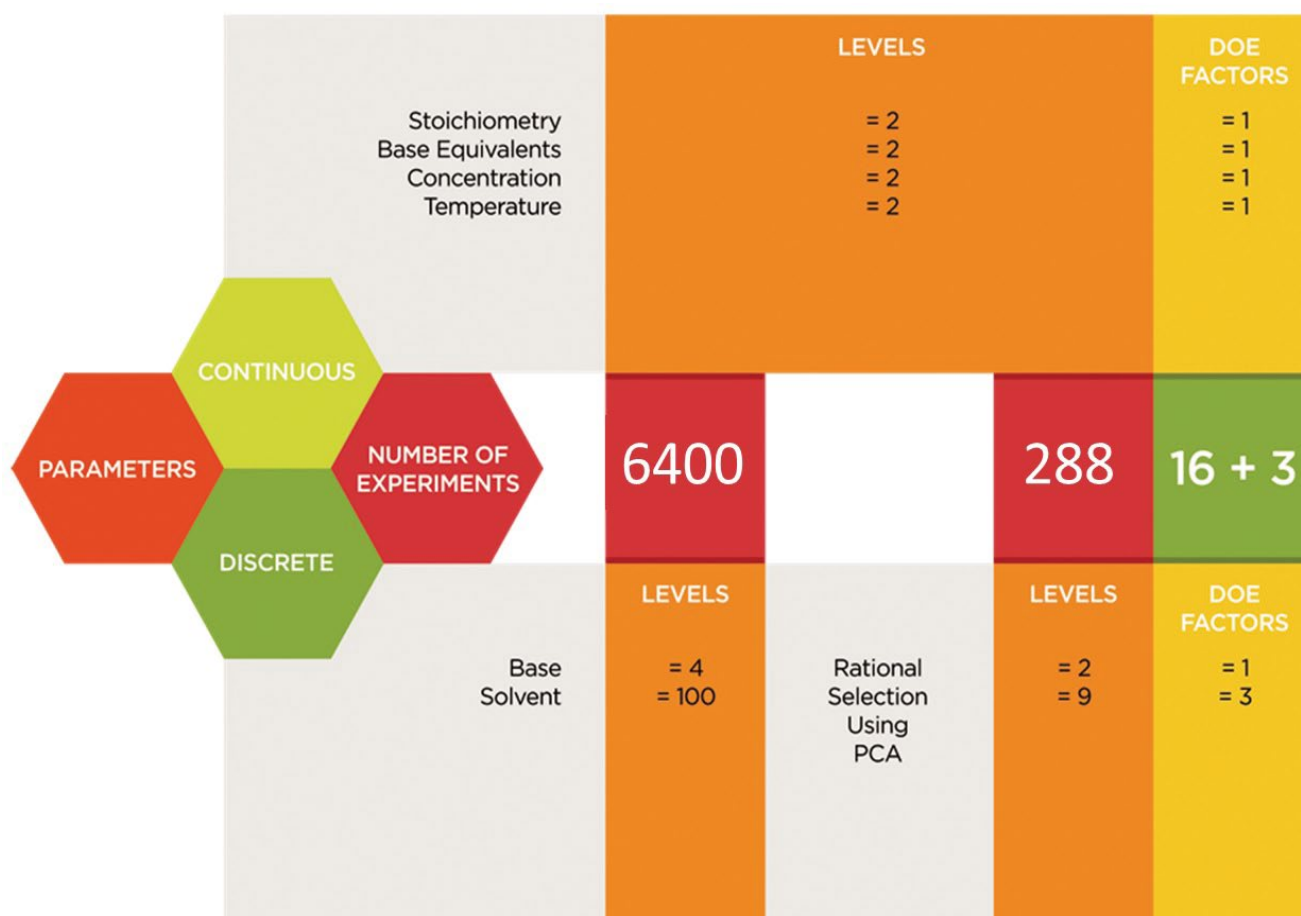


Figure 3 Potentially significant parameters for a S_N2 reaction and the number of potential experiments by OVAT (assuming only two levels of each factor) and by DoE.

of activity, giving further opportunity to improve the reaction. A final design of 11-19 experiments can then be implemented to finalise the optimum conditions.

In a metal catalysed reaction, the number of potentially important factors grows considerably [17]. The choice of catalyst, ligand, amount of catalyst and the ratio of metal to ligand are likely to have a significant effect on the outcome of the reaction in addition to the parameters considered for the S_N2 reaction.

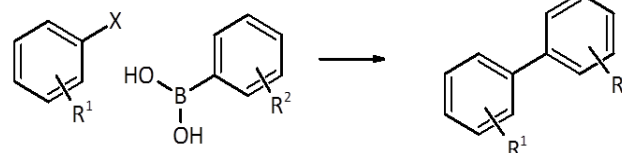
If you consider a palladium catalysed reaction, such as a Suzuki Miyaura reaction (Scheme 2), and accept 500 potentially significant ligands for that reaction, a design of 51.2 million potential experiments would be required to investigate (Figure 4). Using PCA to select 9 diverse ligands and 9 solvents, a DoE of 32 experiments with three repeat centre points can effectively be used to investigate the chemical space. Where this approach has been used during the development of Suzuki Miyaura reactions to generate pharmaceutical intermediates, the selection of 10 additional ligands from the identified region of activity provided increased catalyst activity resulting in a lower catalyst loading in one process.

DoE has been used in combination with PCA to reveal that the choice of solvent was the key to the stability of the reaction, with the chosen solvent preventing decomposition of the starting materials and still allowing a rapid cross coupling reaction. On

selection of the best solvent and ligand for the process from a screening design, a subsequent design provided the optimum reaction conditions, typically based around 19 experiments.

A recent publication from scientists at Merck detailed a high throughput screening approach in which 1536 nanomole-scale reactions were performed in one day with the analysis requiring a further 2 days [18]. This approach did allow for the detailed investigation of a number of discrete variables such as base and ligand, but even these large numbers of reactions are only a small portion of the possible experiments and no information was gained on the reproducibility, any factor interactions and if there were any non-linear responses.

It is clear from Figures 3 and 4 that the combination of DoE and PCA can significantly reduce the number of experiments required to investigate the chemical space around the reaction.



Scheme 2 A Suzuki Miyaura reaction.

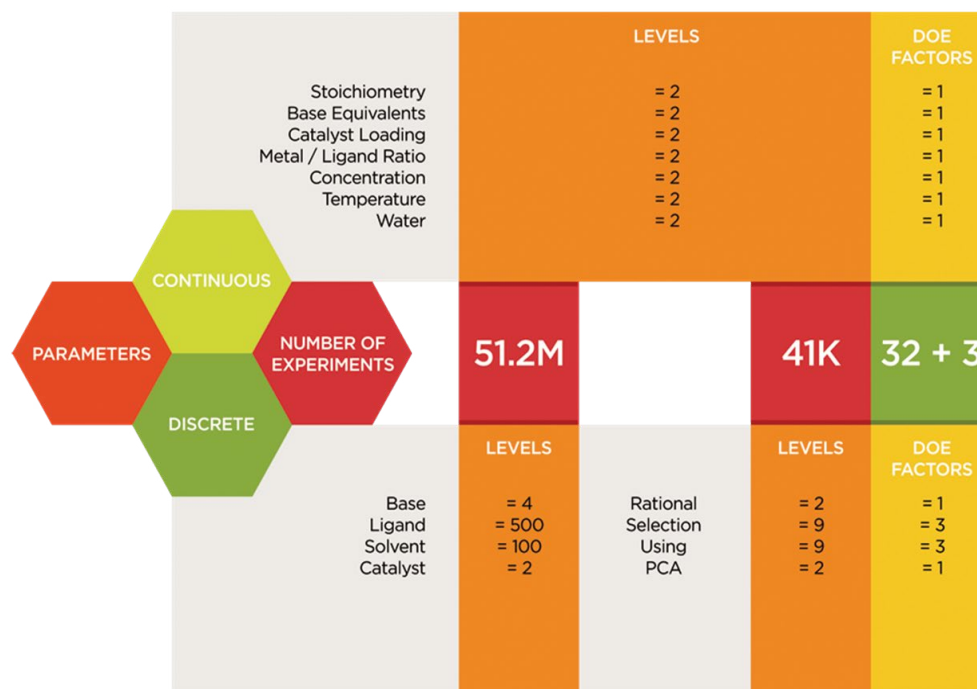


Figure 4 Potentially significant parameters for a catalytic reaction and the number of potential experiments by OVAT (assuming only two levels of each factor) and by DoE.

Design types

Resolution IV fractional factorial designs are chosen to maximize the amount of information gained from the initial screening experiments, while minimizing the number of experiments in the first design. The use of resolution IV designs means there will be confounding (aliasing) between two-way interactions, but the main terms are free from this effect. The confounding between interactions can often be interpreted using chemical understanding of the reaction, and can be confirmed in subsequent experimentation, when the number of potential factors is reduced to a smaller set, or by individual confirmation experiments. The use of this type of screening design will provide information about the presence of interactions between the factors investigated in the reaction. It is important to note that the designs are exploring the potentially significant parameters for the chemical reaction and the initial experiments are unlikely to provide the optimum process conditions. However, they will efficiently direct future experimental efforts into the optimum regions of the chemical space.

Potentially significant scale dependent process factors such as mass transfer, heat transfer, heating, cooling, and addition rates should be looked at in more detail during subsequent experimentation as they are easier to measure and control on a larger scale.

D-Optimal designs may also provide a similar reduction in the number of experiments to the resolution IV fractional factorial designs. D-Optimal designs are generated by a computer algorithm and allow parameters to be estimated without bias whilst minimizing the variance of the parameter estimates for a pre-specified model (i.e. linear, interaction or response

surface model). These types of designs are particularly useful when classical designs do not apply, such as looking at more than two levels of a discrete factor or looking at three or more levels of a continuous factor. D-Optimal designs are useful for selecting solvents or ligands from a database based on principal components from a solvent or ligand PCA map. The investigation of additional reaction parameters, or possible effects of ligand and solvent together, requires a candidate set of all the possible combinations of various factor levels under consideration to be initially created, and the experiments in the D-Optimal design are selected from this candidate set.

Discussion

DoE is a very powerful tool when used properly. The generation of reaction profiles, from collecting multiple samples at different time points, for each experiment will deliver increased information about the reaction process. In addition to this, identifying all significant impurities (e.g. greater than 5%) will complement the statistical analysis with chemical understanding.

A single sample from an experiment after a fixed time gives limited information about the chemical process (**Figure 5a**), whereas, the creation of reaction profiles can provide a significant amount of additional information. Understanding what impurities are formed and how they are formed can aid control of the impurity. Understanding when impurities are formed provides powerful insight into the reaction and indicates the ease or difficulty of scaling up the reaction. Identifying reaction conditions which eliminate impurities, generate product more rapidly or give increased selectivity can be very beneficial and may not be identified when investigated in an OVAT format. The additional samples required to generate reaction profiles increases the time

in analysing the reactions but the benefits significantly outweigh the cost of the additional work in most cases. Having collected all the data from the experiments, understood the outcomes of the reactions and when or why impurities are formed, the samples can be analysed by DoE. It is not necessary to generate models for every sample taken. The chemical understanding and the statistical models should be considered together alongside the preliminary kinetic information obtained from the reaction profiles.

Figure 5a shows a reaction with a single sample taken after an overnight reaction. **Figure 5b** shows the additional information gained from the same reactions with reaction profiles showing conversion for the 4 reactions over time. **Figure 5c** shows the progress of a reaction for starting material consumption, impurity formation and product formation while **Figure 5d** is a hypothetical profile for a reaction where an intermediate is initially formed and the intermediate goes on to form both product and the impurity.

The application of DoE and PCA

It is known that the use of non-obvious solvents, such as non-

polar solvents for electron rich Heck reactions, can completely change reaction selectivity [19]. The power of combining PCA and DoE will increase the possibilities for organic and process chemists and is demonstrated in the following Case Studies.

Case study 1: The reaction of benzyl alcohol with 1.2 mol eq. morpholine using 5 mol% IrCp*Cl dimer, 5 mol% potassium carbonate in toluene at 100°C for 4 h showed only trace levels of product (generic reaction in **Scheme 3**) [20]. The same reaction using Ru(p-cymene)Cl dimer with DPEphos showed 13% conversion after 1 hr and full conversion after 24 hrs. When the reaction mechanism is considered, it is not clear why the base is required (either catalytically or in stoichiometric quantities) [21-23]. There is evidence that added base is not required, and in fact, some substrates perform more efficiently with added acid³ or the inclusion of molecular sieves [24]. A DoE was carried out to investigate the ligand and solvent effects on the reaction as well

³ Personal communication with Alan Pettman, Pfizer: the use of a catalytic acid such as TFA provided significant enhancements in the rate of some redox neutral couplings of alcohols and amines, presumably protonating the imine and increasing its rate of reduction.

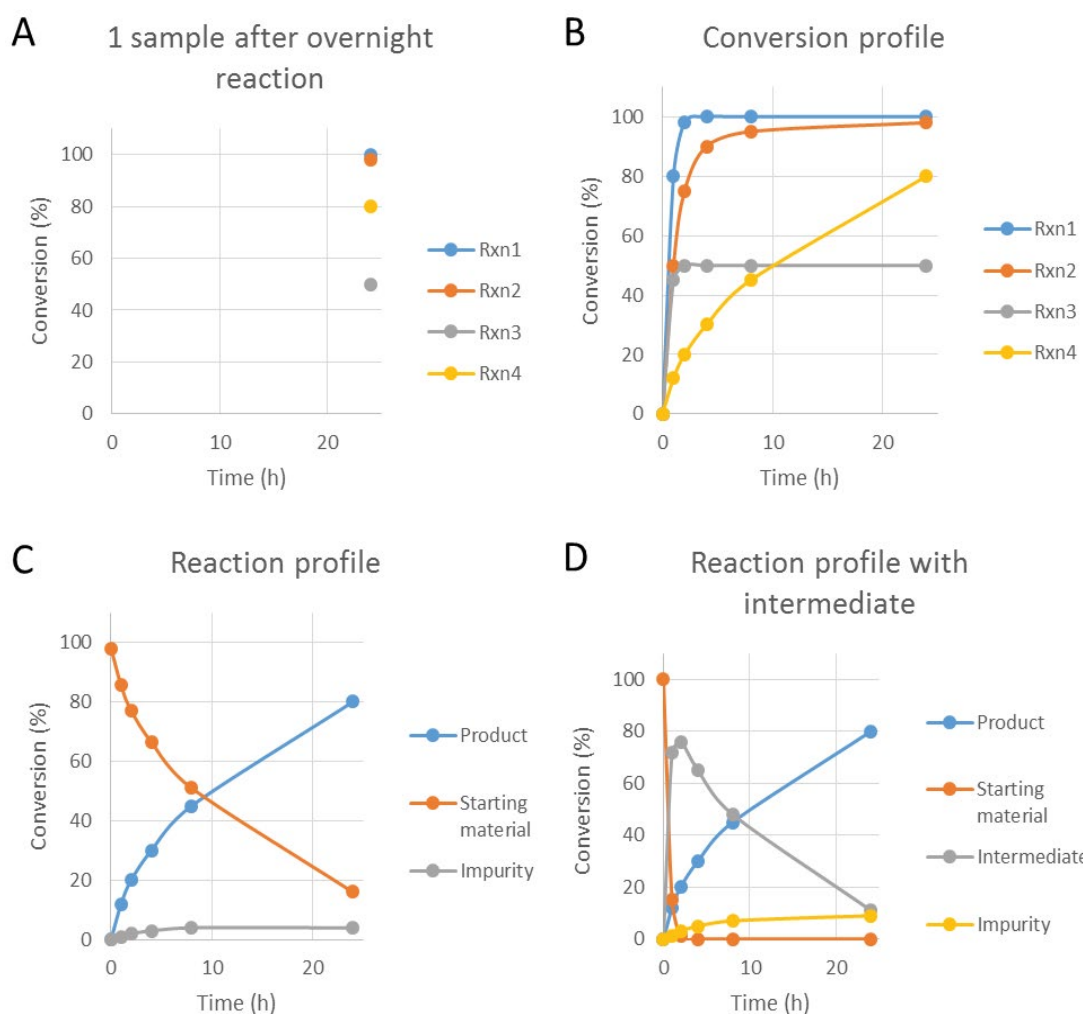
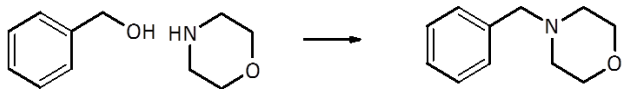


Figure 5 Reaction profiles providing greater insight into reaction pathways.



Scheme 3 Redox neutral N-alkylation of morpholine.

as investigating the effect of the additives⁴. The design focused on monodentate ligands for the Ru catalysed redox neutral coupling. The design also investigated the effect of the pre-catalyst ($[\text{IrCp}^*\text{Cl}]_2$ or $[\text{Ru}(p\text{-cymene})\text{Cl}]_2$) and the additive (acid, base or nothing). A diverse set of solvents (**Table 1**) and ligands (**Table 2**; **Figure 6**) were chosen separately from the appropriate solvent and ligand maps.

A candidate set of experiments was created based on all combinations of ligand, solvent, catalyst and additive for the monodentate ligands. A D-optimal design was used to select 46 experiments from the monodentate ligand set. This design was chosen to allow investigation of the discrete additive at three levels (acid, base or nothing). Two control reactions using toluene and DPEphos were carried out with each design as this was the typical (bidentate) ligand used.

The experimental results can be seen in the replicate plot (**Figure 7**), which highlights a wide range of activity across the different solvents, ligands, pre-catalysts and additives. Excellent reproducibility was seen for all the control reactions (shown by blue squares). Analysis of the DoE showed that for the monodentate ligands a significant model was obtained with an R^2 of 77% (R^2 shows the model fit, a value of more than 0.5 is desired) and Q^2 of 50% (Q^2 is a test of diverse model problems, a value greater than 0.25 is desired). Refinement of the model revealed three of the eight investigated factors were significant in the reaction, pPC, pP3 and sPC, but interactions between the ligand PCs had the greatest impact on conversion (**Figure 8**). A factor or an interaction can have a positive or negative effect on the model. The three significant factors all have a negative effect on conversion in this model.

A number of monodentate ligands were nearly as effective as DPEphos for the $[\text{Ru}(p\text{-cymene})\text{Cl}]_2$ catalysed process, in particular tri-*i*-butylphosphine, which achieved 91.7% conversion, (**Table 3**) while $[\text{IrCp}^*\text{Cl}]_2$ showed increased reactivity with added ligand (**Table 4**).

The effect of solvents can be seen in **Figure 9**. Solvents in the top left quadrant gave full conversion (marked in green), whereas those in the bottom left quadrant gave good reactivity (marked in light green) but with about 20% lower conversion. More polar solvents on the right are generally less effective with lower levels of conversion. Of particular note was benzonitrile, which seems to inhibit most reactions (<5% conversion for any system investigated)⁵.

⁴ Software: MODDE 10.1; Umetrics, part of Sartorius Stedim Biotech: Umeå, Sweden, 2014, for all experimental designs. SIMCA 13; Sartorius Stedim Data Analytics AB: Umeå, Sweden, 2013, for Principal Component Analysis. <http://umetrics.com>

⁵ Investigations using substrates containing aromatic nitriles have shown no reaction in our hands including for example 4-cyanobenzyl alcohol

From the two-dimensional representation of the monodentate phosphine ligand map (**Figure 10**) it is evident that there is an area of inactivity (pink circle) around which there are varying levels of activity (red as low to green as higher activity) with either Ru or Ir for the ligands initially investigated in the first design. The differing results for the two metal catalysts (Ru, **Figure 11** and Ir, **Figure 12**) show that each have specific requirements for their ligands and the same ligands cannot be used successfully for both metals.

To further investigate the importance of the ligands on the activity of Ir catalyst systems, additional experiments were carried out to determine additional potential monodentate ligands. The selection of these ligands was made to sample specifically around the area of inactivity initially pinpointed in the top right-hand quadrant of **Figure 12** (**Table 4**). Plotting the additional data on the two-dimensional plot of PC1 and PC2, it appears as if some ligands now sit in the area of activity (**Figure 13**). However, if additional principal component plots are utilised, it can be seen that the ligands are further separated (**Figure 14** shows PC1 and PC4). The combined use of PCA with DoE has uncovered monodentate ligands which provide the same high reactivity (>90% conversion) for Ir catalysis as that previously seen with Ru catalysis.

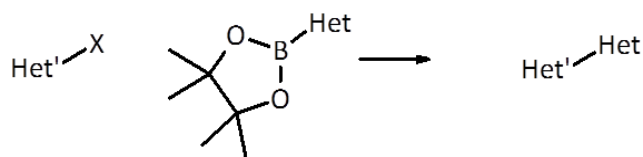
In summary, the combination of PCA and DoE has highlighted: i) ligands effective for Ru are not compatible with Ir; ii) Ir can have the same reactivity as Ru with the correct choice of ligand; iii) solvent can have a significant deleterious effect on the reaction; and iv) the effect of acid, base or molecular sieves was insignificant in the experiments investigated, as evidenced by the absence of these factors in any response models [20].

Case study 2: A Suzuki Miyaura coupling of a heteroaromatic halide with a pinacol boronic ester was used in the late stage of a pharmaceutical intermediate synthesis (**Scheme 4**).

The reaction suffered from poor robustness during manufacture, requiring 5 mol% of a very active catalyst to achieve high yield. A fractional factorial design was chosen to look at nine monodentate ligands, two solvents, and two palladium salts at a resolution V level in 19 experiments. The design provided an excellent model with an R^2 of 96% and a Q^2 of 86%. The replicate plot (**Figure 15**) shows there was a range in conversion across the experiments.

The Coefficient plot (**Figure 16**) identified the ligand PCs as the most significant factors in the design. There was a large interaction between the solvent and the palladium salts, which

with N-methyl piperazine and 3-[(2-hydroxyethyl)(phenyl)amino]propanenitrile with morpholine, (2,2,2-trifluoroethyl)hydrazine, butane-1,4-diamine or 1-cycloheptyl-methanamine (unpublished results).



Scheme 4 Suzuki Miyaura reaction.

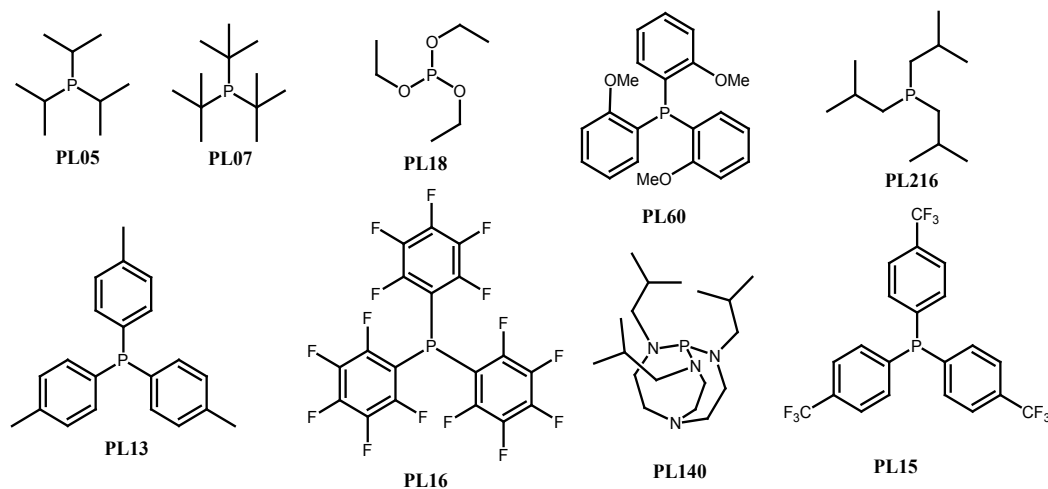

Figure 6 Monodentate ligand structures.

Table 1 A diverse selection of solvents.

Solvents	sPC1	sPC2	sPC3
1,4-Dioxane	0	0	0
Di- <i>n</i> -butyl ether	-1	-1	-1
<i>n</i> -Butyl acetate	-1	-1	1
Tetralin	-1	1	-1
Chlorobenzene	-1	1	1
Pentan-3-one	1	-1	-1
Propionitrile	1	-1	1
<i>N,N</i> -Dimethylacetamide	1	1	-1
Benzonitrile	1	1	1

Table 2 A diverse selection of monodentate ligands.

Monodentate P ligands	pPC1	pPC2	pPC3
Tri- <i>p</i> -tolylphosphine	0	0	0
Tris(pentafluorophenyl)phosphine	-1	-1	-1
Tris[4-(trifluoromethyl)phenyl]phosphine	-1	-1	1
Triethyl phosphite	-1	1	-1
Tri- <i>i</i> -propylphosphine	-1	1	1
Tri- <i>t</i> -butylphosphine (HBF ₄ salt)	1	-1	-1
Tris(2-methoxyphenyl)phosphine	1	-1	1
2,8,9-Tri- <i>i</i> -butyl-2,5,8,9-tetraaza-1-phosphabicyclo[3,3,3]undecane	1	1	-1
Tri- <i>i</i> -butylphosphine	1	1	1

Table 3 Selected results from the reactions catalysed by [Ru(*p*-cymene)Cl]₂.

Phosphine	Solvent	Additive	conv/9h%
(C ₆ F ₅) ₃ P (PL16)	Di- <i>n</i> -butylether	K ₂ CO ₃	87.5
(OEt) ₃ P (PL18)	<i>N</i> -Butyl Acetate	TFA	10.4
(<i>o</i> -OMePh) ₃ P (PL60)	Di- <i>n</i> -butylether	none	0.7
(<i>p</i> -CF ₃ Ph) ₃ P (PL15)	Pentan-3-one	none	72.6
^t Bu ₃ P (PL216)	Tetralin	TFA	91.7
^t Bu ₃ P (PL07)	Chlorobenzene	K ₂ CO ₃	26.7
triple cage/ ^t Bu (PL140)	Chlorobenzene	TFA	8.9
DPE Phos	Toluene	K ₂ CO ₃	93.4
DPE Phos	Toluene	K ₂ CO ₃	95.5

Table 4 Selected results from the reactions catalysed by $[\text{IrCp}^*\text{Cl}]_2$ (^ for additional ligand selection).

Phosphine	Solvent	Additive	conv/9h%
$(\text{C}_6\text{F}_5)_3\text{P}$ (PL16)	Tetralin	TFA	69.4
$(\text{OEt})_3\text{P}$ (PL18)	Tetralin	K_2CO_3	41.3
$(p\text{-CF}_3\text{Ph})_3\text{P}$ (PL15)	Benzonitrile	TFA	10.7
$(p\text{-MePh})_3\text{P}$ (PL13)	dioxane	none	4.1
$i\text{Pr}_3\text{P}$ (PL05)	Tetralin	K_2CO_3	14.6
$t\text{Bu}_3\text{P}$ (PL07)	Chlorobenzene	none	71.4
triple cage/ $i\text{Bu}$ (PL140)	Propionitrile	K_2CO_3	62.7
Triphenylphosphite(PL20)^	Tetralin	None	31.5
Tris[3,5-di(trifluoromethyl)phenyl]phosphine (PL32) ^	Tetralin	None	5.7
Tris(2-methoxyphenyl)phosphine (PL60) ^	Tetralin	None	4.4
X-Phos (PL146) ^	Tetralin	None	42.3
Dave-Phos (PL147) ^	Tetralin	None	93.2
t -Butyldicyclohexylphosphine (PL152) ^	Tetralin	None	91.1
2-(Dicyclohexylphosphino)biphenyl (PL155) ^	Tetralin	None	97.4
$i\text{Bu}_3\text{P}$ (PL216) ^	Tetralin	None	29.7

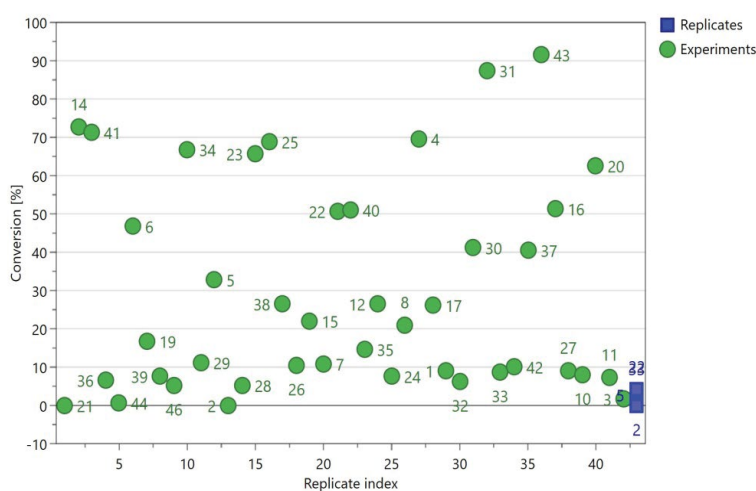


Figure 7 Replicate plot showing range in conversion of benzyl alcohol with monodentate ligands after 9 hrs.

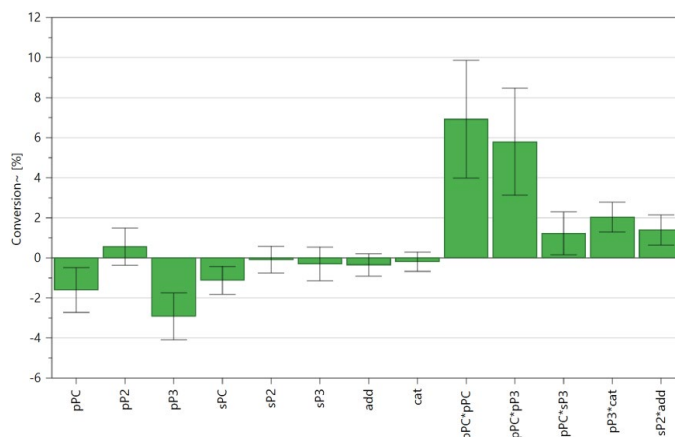


Figure 8 Coefficient plot from the model for conversion in the DoE investigating monodentate phosphines [20].

Abbreviations: pPC is Phosphine PC1; pP2 is Phosphine PC2; pP3 is Phosphine PC3; sPC is Solvent PC1; sP2 is Solvent PC2; sP3 is Solvent PC3; Add is K_2CO_3 , none (0) or Trifluoroacetic acid, Cat is $\text{Ru}(p\text{-cymene})\text{Cl}$ dimer or IrCp^*Cl dimer; Pd(Acetate) is $\text{Pd}(\text{OAc})_2$; Pd(dba) is $\text{Pd}(\text{dba})_3$; Sol(A) is Solvent A and Sol(B) is Solvent B.

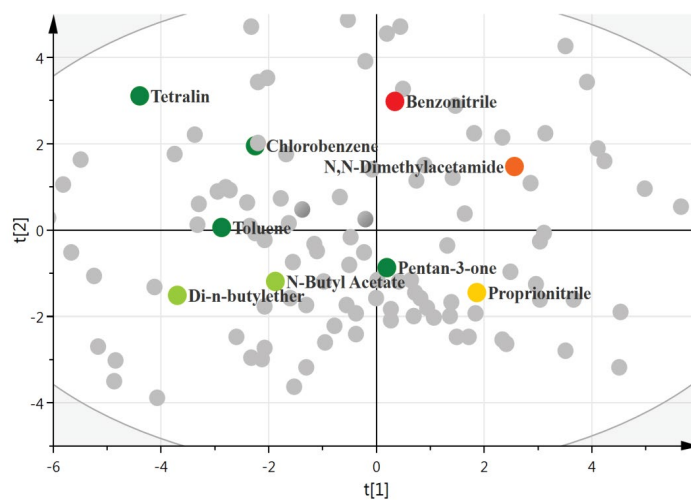


Figure 9 Solvent map using PC1 and PC2 highlighting extent of conversion - coloured red (0%) to green (100%).

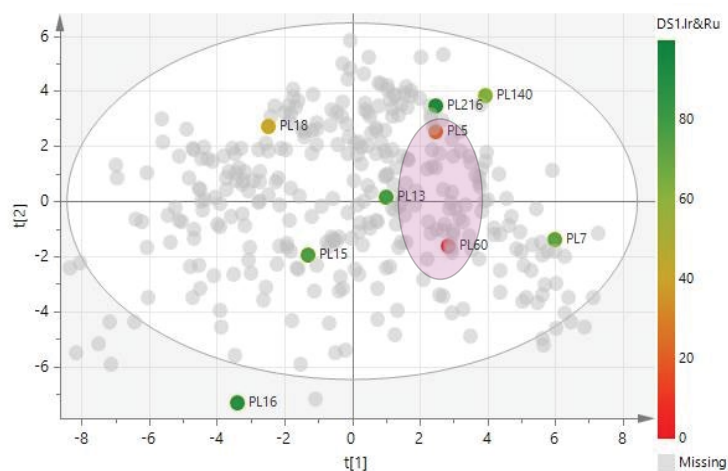


Figure 10 Monodentate ligand map using PC1 and PC2 highlighting the maximum catalyst activity [20].

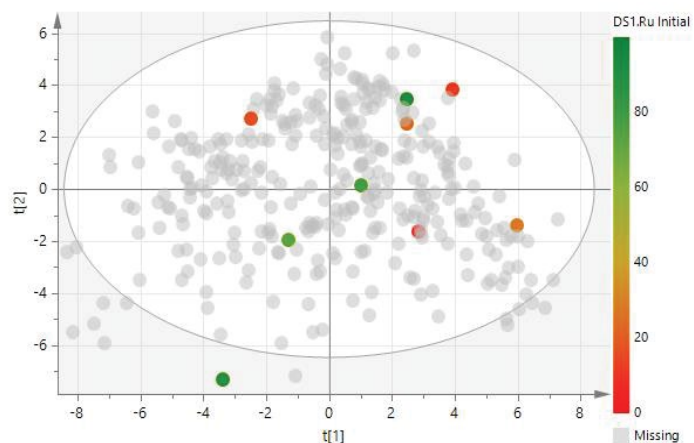


Figure 11 Monodentate ligand map using PC1 and PC2 highlighting activity for experiments with Ru as catalyst.

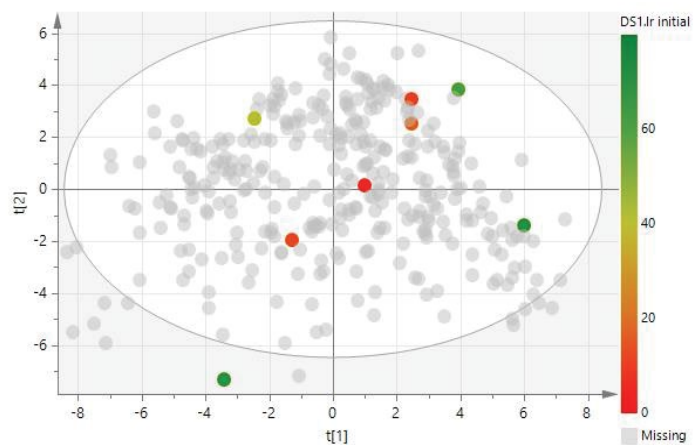


Figure 12 Monodentate ligand map using PC1 and PC2 highlighting activity for experiments with Ir as catalyst.

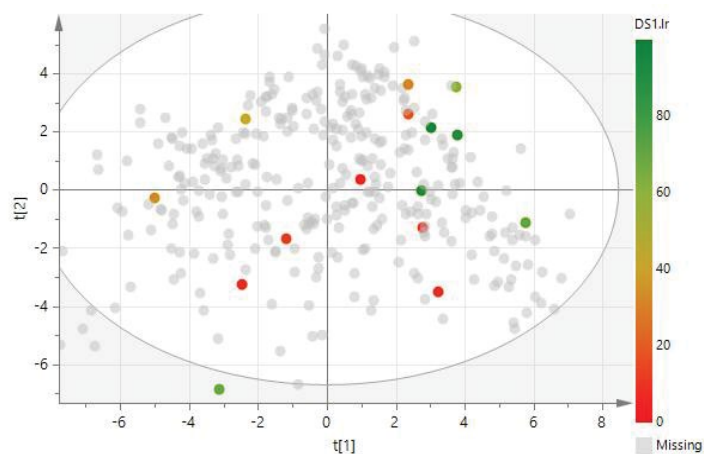


Figure 13 Monodentate ligand map with additional ligands using PC1 and PC2 highlighting ligands with Ir as catalyst.

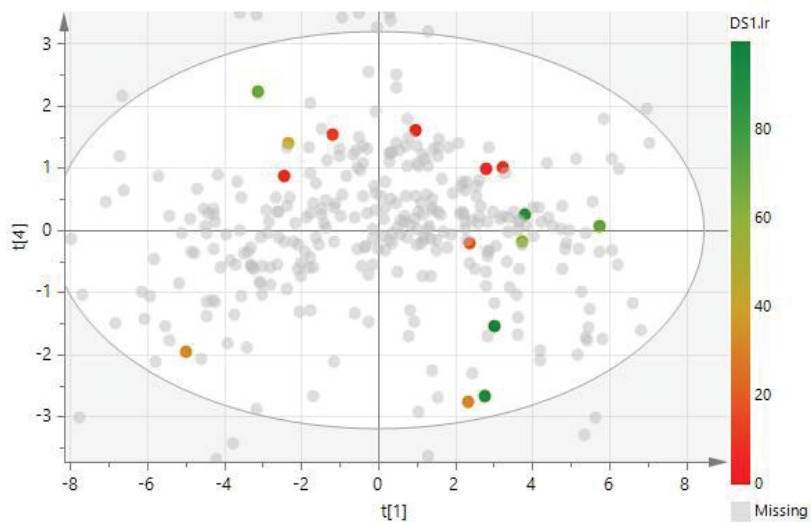


Figure 14 Monodentate ligand map with additional ligands using PC1 and PC4 highlighting ligands with Ir as catalyst.

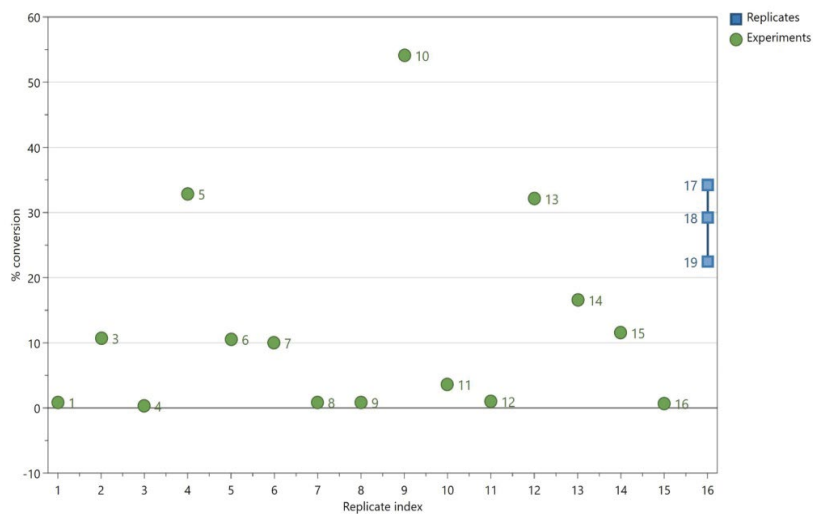


Figure 15 Replicate plot for Suzuki Miyaura design.

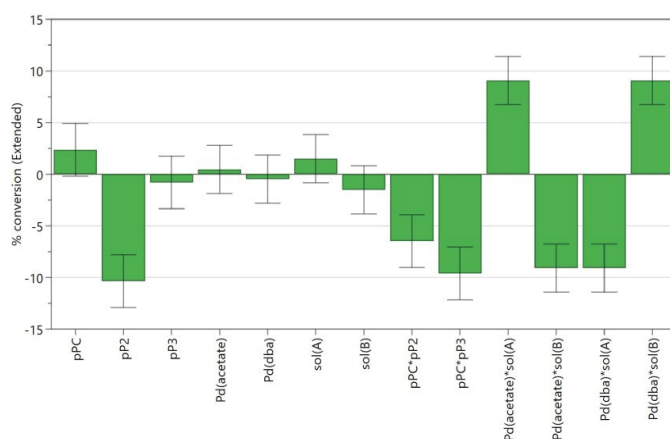


Figure 16 Coefficient plot from the model for conversion in the DoE investigating monodentate ligands in the Suzuki Miyaura reaction. Abbreviations: pPC is Phosphine PC1; pP2 is Phosphine PC2; pP3 is Phosphine PC3; sPC is Solvent PC1; sP2 is Solvent PC2; sP3 is Solvent PC3; Add is K₂CO₃, none (0) or Trifluoroacetic acid, Cat is Ru(p-cymene)Cl dimer or IrCp*Cl dimer; Pd(Acetate) is Pd(OAc)₂; Pd(dba) is Pd2dba₃; Sol(A) is Solvent A and Sol(B) is Solvent B.

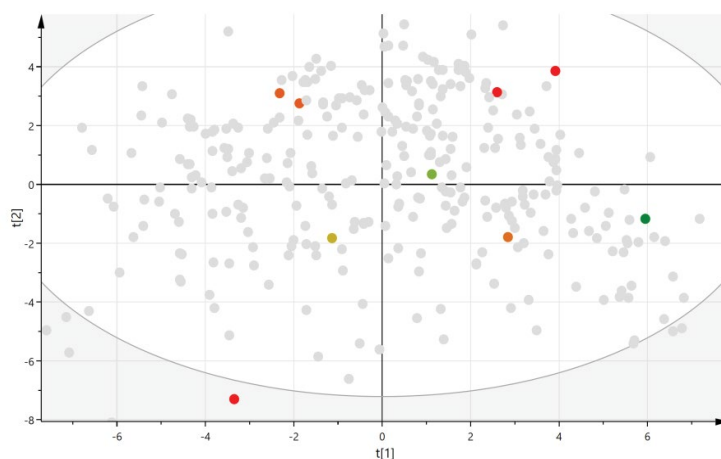


Figure 17 PCA map of monodentate ligands for initial design for the Suzuki Miyaura reaction coloured red (0%) to green (100%).

highlights the importance of choosing the right pre-catalyst and solvent combination.

The maximum yield was achieved after one hour of reaction. A study of the reaction profiles from each experiment highlighted the rapid decomposition of the boronic ester under the reaction conditions in the reaction vessel. However, the boronic ester was completely stable in the reaction solvent even at elevated temperatures, therefore the controlled addition of the boronic ester, over a six-hour period as a solution in reaction solvent, allowed the development of a robust reaction with a lower (1 mol%) catalyst loading.

Subsequently, a study was undertaken to investigate the ligand effects on the reaction using PCA. A fractional factorial DoE design was created by selecting ligands from a monodentate phosphine ligand PCA map. The results immediately identified a new monodentate ligand that performed as well as the original catalyst at a sixth of the cost (**Figure 17**).

Analysis of the results from the initial screening identifies a wider region of ligand activity. The plot in **Figure 17** is of PC1 and PC2. Subsequent principal components separate the unsuccessful ligands in the region of activity. Additional ligands were selected to evaluate the region around the 'sweet spot' of activity. Two iterations of additional ligand selections were made to identify alternative suitable ligands with the final results showing a clear area of activity (**Figure 18**).

Ligand screening using PCA identified a number of alternative ligands. The modified experimental procedure enabled a greater than 5-fold reduction in the catalyst loading while providing a more robust process without the degradation of starting material. The identification of alternative catalysts, such as tri-*tert*-butylphosphine/palladium acetate, offered a reduction in total catalyst and ligand cost of approximately 5-fold while maintaining reactivity. The modified procedure allowed the use of less active ligands and triphenylphosphine has been shown to be a viable ligand for this reaction, but with a longer reaction time. The use of PCA allowed informed rational decision making, and provided a more robust process with a 10-fold overall reduction in cost.

Case study 3: DoE and PCA were used in combination to find an alternative ligand for a Buchwald-Hartwig sulfamidation of a heteroaromatic chloride with a sulfonimine (**Scheme 5**) [17]. The initial reaction conditions gave good conversion but used an expensive ligand. Thus, a more affordable ligand, which

maintained or exceeded the performance of the original ligand without generating new impurities, was sought.

Initially, a screening fractional factorial DoE of 35 experiments was conducted to look at nine ligands, the metal to ligand ratios (1:1 and 1:2), and the palladium source ($\text{Pd}(\text{OAc})_2$ and $\text{Pd}_2(\text{dba})_3$)⁶. All other parameters were kept constant.

Analysis of the data shows that over half of the reactions had very low extent of conversion (<10%) whilst 6 ligands gave near complete conversion in the same timeframe (**Figure 19**). A significant model was calculated from the data, with an R^2 of 60% and a Q^2 of 44%, however, the model is not excellent in fit and predictiveness due to the wide-ranging results.

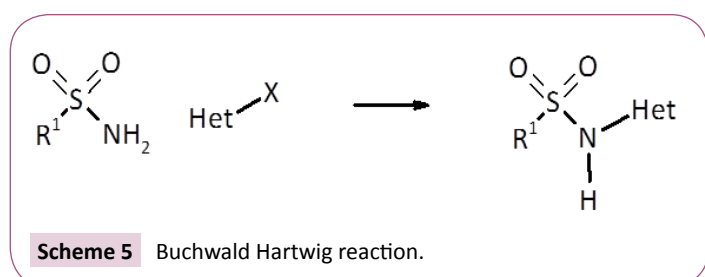
Analysis of the coefficient plot showed that the palladium salt and the metal:ligand ratio had no effect on the response and these factors were therefore removed from the model. The influencing factors, the ligand PCs, are shown in the coefficient plot (**Figure 20**). The Contour plot (**Figure 21**) shows how only a small section of the ligand map results in efficient product formation where the coordinates are pPC +1, pPC2 -1, pPC3 -1.

The two new catalyst systems (i.e., metal:ligand combinations) which were identified to give very good conversion are shown in **Figure 22** highlighted in green. One of these new ligands was half the cost of the initially employed ligand.

A second iteration of 12 reactions focussed on the region of optimum ligand space, as identified in **Figure 22**, for 1 principal component. From this, four more ligands which gave good or complete conversion were found. Another screen of 12 reactions looked at a different principal component in the area of optimum ligand performance which identified a further four effective ligands. **Figure 23** highlights the region of successful ligands identified for this Buchwald-Hartwig sulfamidation.

Of the promising ligands, one was favoured as it was not bound by IP and was therefore more than 10-fold cheaper than the original ligand. This ligand was carried through to the optimisation process, where a design of 19 experiments was performed to determine the best reaction conditions. The factors investigated included metal loading, ligand loading, and stoichiometry of the second starting material.

The use of PCA and DoE to explore alternative catalysts and ligands for this Buchwald Hartwig reaction identified a number of alternative catalytic systems with a potential to reduce the costs for the process by more than 14-fold. The use of PCA allowed informed rational decision making and provided a more cost-effective process.



⁶ Software: MODDE 10.1; Umetrics, part of Sartorius Stedim Biotech: Umeå, Sweden, 2014, for all experimental designs. SIMCA 13; Sartorius Stedim Data Analytics AB: Umeå, Sweden, 2013, for Principal Component Analysis. <http://umetrics.com>

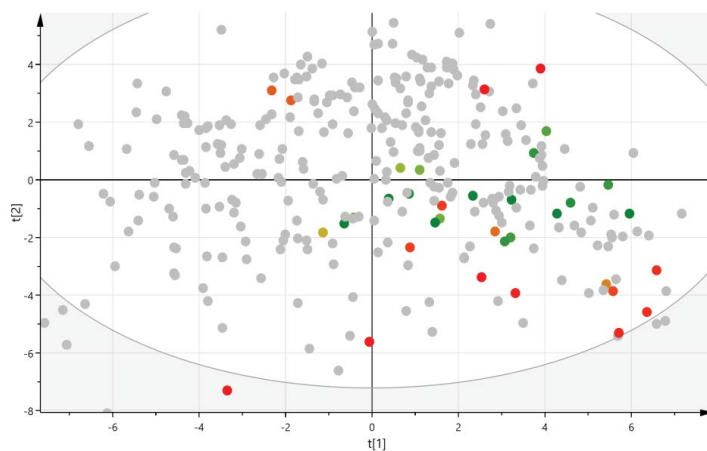


Figure 18 PCA map of monodentate ligands for all experiments from the Suzuki Miyaura reaction coloured red (0%) to green (100%).

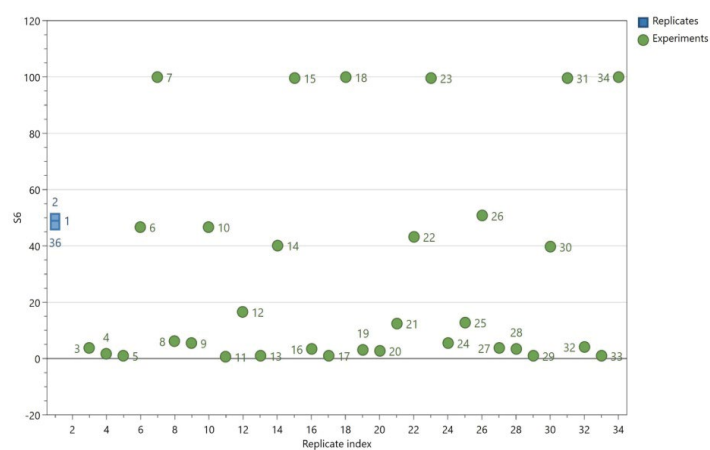


Figure 19 Replicate plot for Buchwald Hartwig reaction.

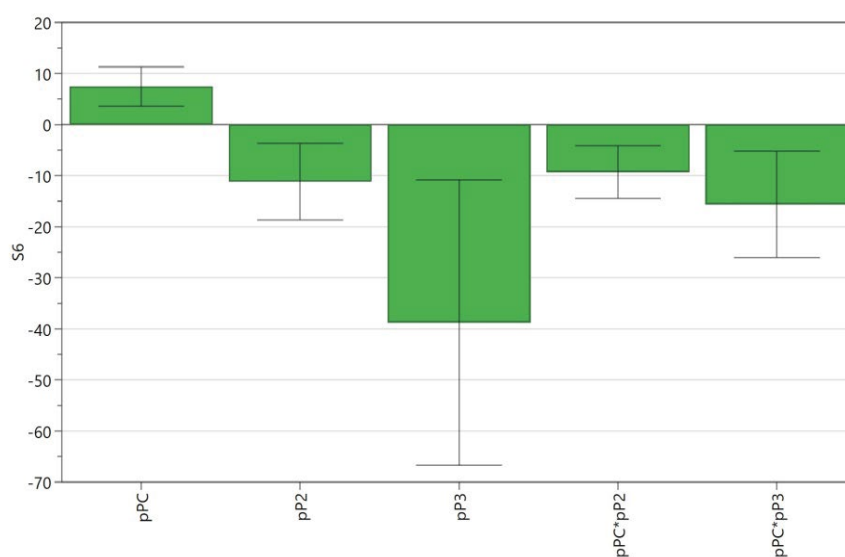


Figure 20 Coefficient plot for the model for conversion at the final sample (S6) in the DoE investigating bidentate phosphines in the Buchwald Hartwig reaction.

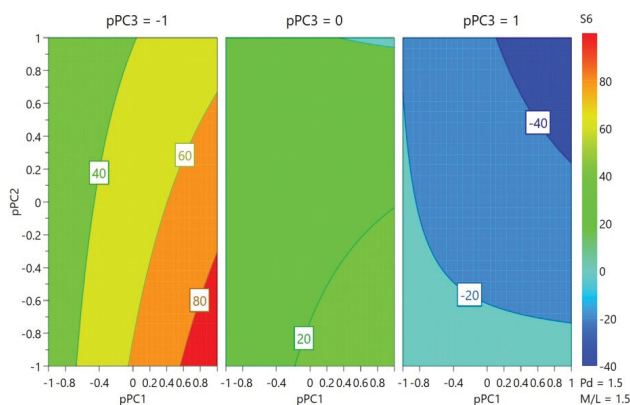


Figure 21 Contour plots for Buchwald Hartwig reaction.

Abbreviations: pPC is Phosphine PC1; pP2 is Phosphine PC2; pP3 is Phosphine PC3; sPC is Solvent PC1; sP2 is Solvent PC2; sP3 is Solvent PC3; Add is K₂CO₃, none (0) or Trifluoroacetic acid, Cat is Ru(*p*-cymene)Cl dimer or IrCp*Cl dimer; Pd(Acetate) is Pd(OAc)₂; Pd(*dba*) is Pd2*dba*3; Sol(A) is Solvent A and Sol(B) is Solvent B.

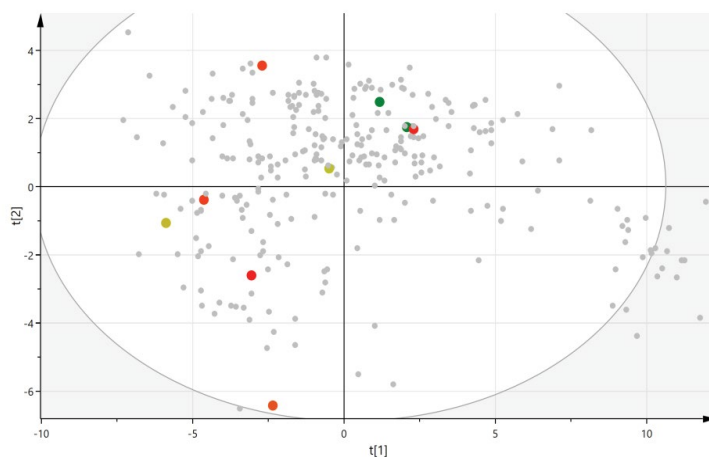


Figure 22 PCA map of bidentate ligands for initial design for the Buchwald-Hartwig sulfamidation coloured red (0%) to green (100%).

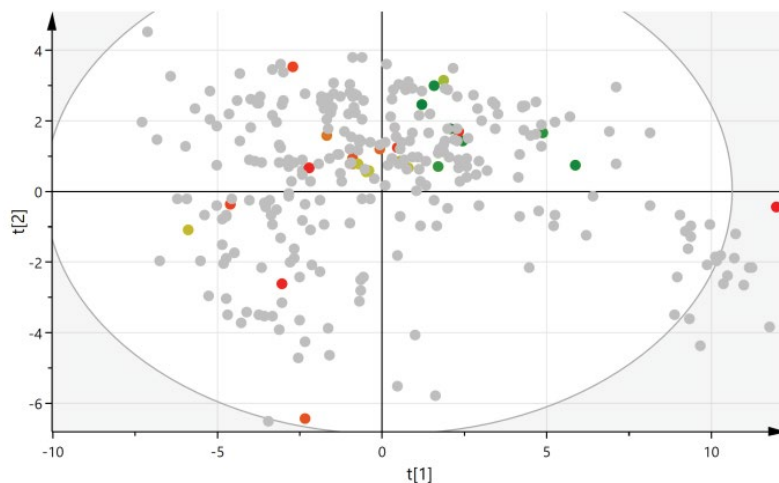


Figure 23 PCA map of ligands for the Buchwald-Hartwig sulfamidation coloured red (0%) to green (100%).

Conclusion

DoE is an efficient tool for the development of chemical processes. In conjunction with PCA, DoE is able to investigate discrete factors such as solvents and ligands in a semi-continuous manner, allowing links to be made between the chosen materials. Applying DoE with PCA can significantly reduce the number of experiments in a rational and focused manner. Depending on complexity, a fully optimised process can be defined with typically 2 to 4 iterations of a design. An initial larger design of 19 to 35 experiments is typical with subsequent refinements using 7 to

19 experiments each to define the “sweet spot”. DoE combined with PCA will enable the rational development of complex chemical processes in fewer experiments than other approaches. DoE should not be used without considering the chemistry, the products, the impurities and the ultimate scale up of a final commercial process.

Supporting Information

A more detailed explanation of PCA and how to combine it with DoE is included in the supporting information along with an explanation of the terminology used in DoE.

References

- 1 Weissman SA, Anderson NG (2015) Design of Experiments (DoE) and Process Optimization. A Review of Recent Publications. *Org Process Res Dev* 19: 1605-1633.
- 2 Leardi R (2009) Experimental design in chemistry: A tutorial. *Anal Chim Acta* 652: 161-172.
- 3 Murray P M, Bellany F, Benhamou L, Bučar DK, Tabor AB, et al. (2016) The application of design of experiments (DoE) reaction optimisation and solvent selection in the development of new synthetic chemistry. *Org Biomol Chem* 14: 2373-2384.
- 4 ICH Harmonised tripartite guideline pharmaceutical development Q8 (R2) (2009), Available from: https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q8_R1/Step4/Q8_R2_Guideline.pdf.
- 5 Yu LX (2008) Pharmaceutical Quality by Design: Product and Process Development, Understanding, and Control. *Pharm Res* 25: 781-791.
- 6 Carlson R, Carlson JE (2005) Design and Optimization in Organic Synthesis. 2nd edition in *Data Handling in Science and Technology*, vol. 24, Elsevier, Amsterdam.
- 7 Carlson R, Lundstedt T, Albano C (1985) Screening of Suitable Solvents in Organic Synthesis. *Strategies for Solvent Selection*. *Acta Chem Scand* 39: 79-91.
- 8 Chastrette M, Rajzmann M, Chanon M, Purcell KF (1985) Approach to a general classification of solvents using a multivariate statistical treatment of quantitative solvent parameters. *J Am Chem Soc* 107: 1-11.
- 9 Allesø M, van den Berg F, Cornett C, Jørgensen FS, Halling-Sørensen B, et al. (2008) Solvent diversity in polymorph screening. *Pharm Sci* 97: 2145-2159.
- 10 Allesø M, Rantanen J, Aaltonen J, Cornett C, van den Berg F (2008) Solvent subset selection for polymorph screening. *J Chemom* 22: 621-631.
- 11 Jover J, Fey N, Harvey JN, Lloyd Jones GC, Orpen AG et al. (2010) Expansion of the Ligand Knowledge Base for Monodentate P-Donor Ligands (LKB-P). *Organometallics* 29: 6245-6258.
- 12 Jover J, Fey N, Harvey JN, Lloyd Jones GC, Orpen AG, et al. (2012) Expansion of the Ligand Knowledge Base for Chelating P,P-Donor Ligands (LKB-PP). *Organometallics* 31: 5302-5306.
- 13 Carlson R, Lundstedt T, Nordahl A, Prochazka M (1986) Lewis Acids in Organic Synthesis. Approach to a Selection Strategy for Screening Experiments. *Acta Chem Scand* 40: 522-533.
- 14 Carlson R, Prochazka MP, Lundstedt TB (1988) Principal properties for synthetic screening: amines. *Acta Chem Scand Ser* 42: 157-165.
- 15 Carlson R, Prochazka MP, Lundstedt, T (1988) Principal Properties for Synthetic Screening: Ketones and Aldehydes. *Acta Chem Scand Ser* 42: 145-156.
- 16 Carlson R, Carlson JE (2005) Principal Properties and Designs for Discrete Variations. *Org Process Res Dev* 9: 680-689.
- 17 Murray PM, Tyler SNG, Moseley JD (2013) Beyond the Numbers: Charting Chemical Reaction Space. *Org Process Res Dev* 17: 40-46.
- 18 Santanilla AB, Regalado EL, Pereira T, Shevlin M, Bateman K, et al. (2015) Nanomole-scale high-throughput chemistry for the synthesis of complex molecules. *SD Science* 347: 49-53.
- 19 Murray PM, Bower JF, Cox DK, Galbraith EK, Parker JS, et al. (2013) A robust first-pass protocol for the Heck-Mizoroki reaction. *Org Process Res Dev* 17:397-405.
- 20 Moseley JD, Murray PM (2014) Ligand and solvent selection in challenging catalytic reactions. *J Chem Technol Biotechnol* 89: 623-632.
- 21 Williams JMJ, Hamid MHSA, Allen CL, Lamb GW, Maxwell AC, et al. (2009) Ruthenium-catalyzed N-alkylation of amines and sulfonamides using borrowing hydrogen methodology. *J Am Chem Soc* 131:1766-1774
- 22 Kawahara R, Fujita K, Yamaguchi R (2010) Multialkylation of Aqueous Ammonia with Alcohols Catalyzed by Water-Soluble Cp*Ir-Ammine Complexes. *J Am Chem Soc* 132: 15108-15111.
- 23 Hollmann D, Tillack A, Michalik D, Jackstell R, Beller (2007) An Improved Ruthenium Catalyst for the Environmentally Benign Amination of Primary and Secondary Alcohols. *M Chem Asian J* 2: 403-410.
- 24 Williams JMJ, Hamid MHSA (2007) Ruthenium catalysed N-alkylation of amines with alcohols. *Chem Commun* pp: 725-727.