CatScI

Innovation in Catalysis

The Power of PCA in the Search for Reaction Understanding

Dr. Jonathan Moseley

CatScI Ltd

Introduction

- (Context)
- Introduction to CatScI
- Experimental Design (DoE)
- Principal Component Analysis (PCA)
- A Worked Example
- Further Examples
- Conclusions

About CatScI Ltd

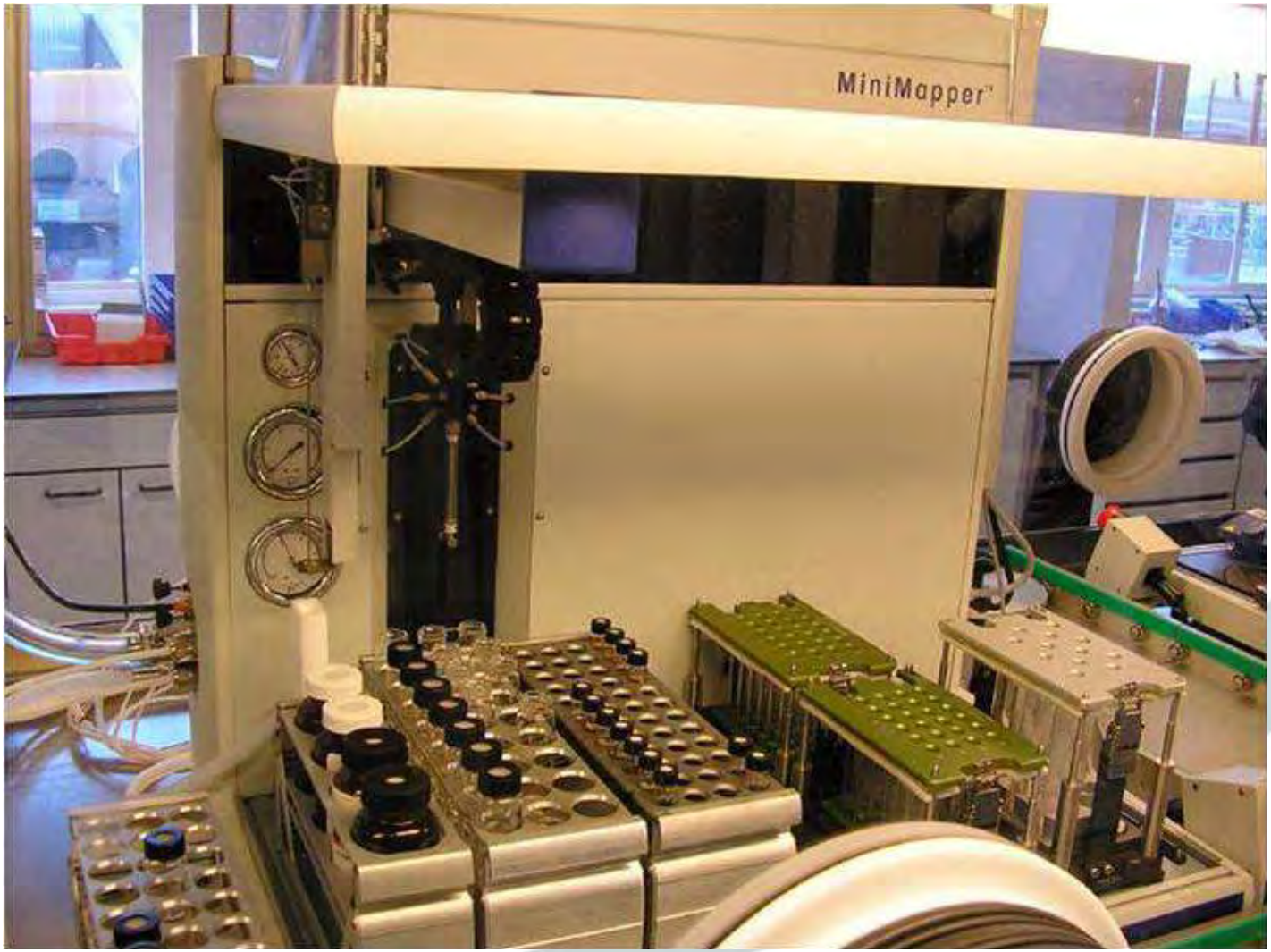
- UK-based SME – spun out of AstraZeneca in Q1 2011
- Highly experienced ex-pharma management team
- State-of-the-art automated screening laboratory
- We specialise in the rational understanding, development and optimisation of catalysed reactions across the range of chemical industries
  - bio- and chemo-catalysis, homo- and hetero-geneous
  - R&D experience at all stages of the project timeline, from discovery chemistry through to full scale manufacturing

CatScI
Innovation in Catalysis

CatScI Facilities – Weighing Station

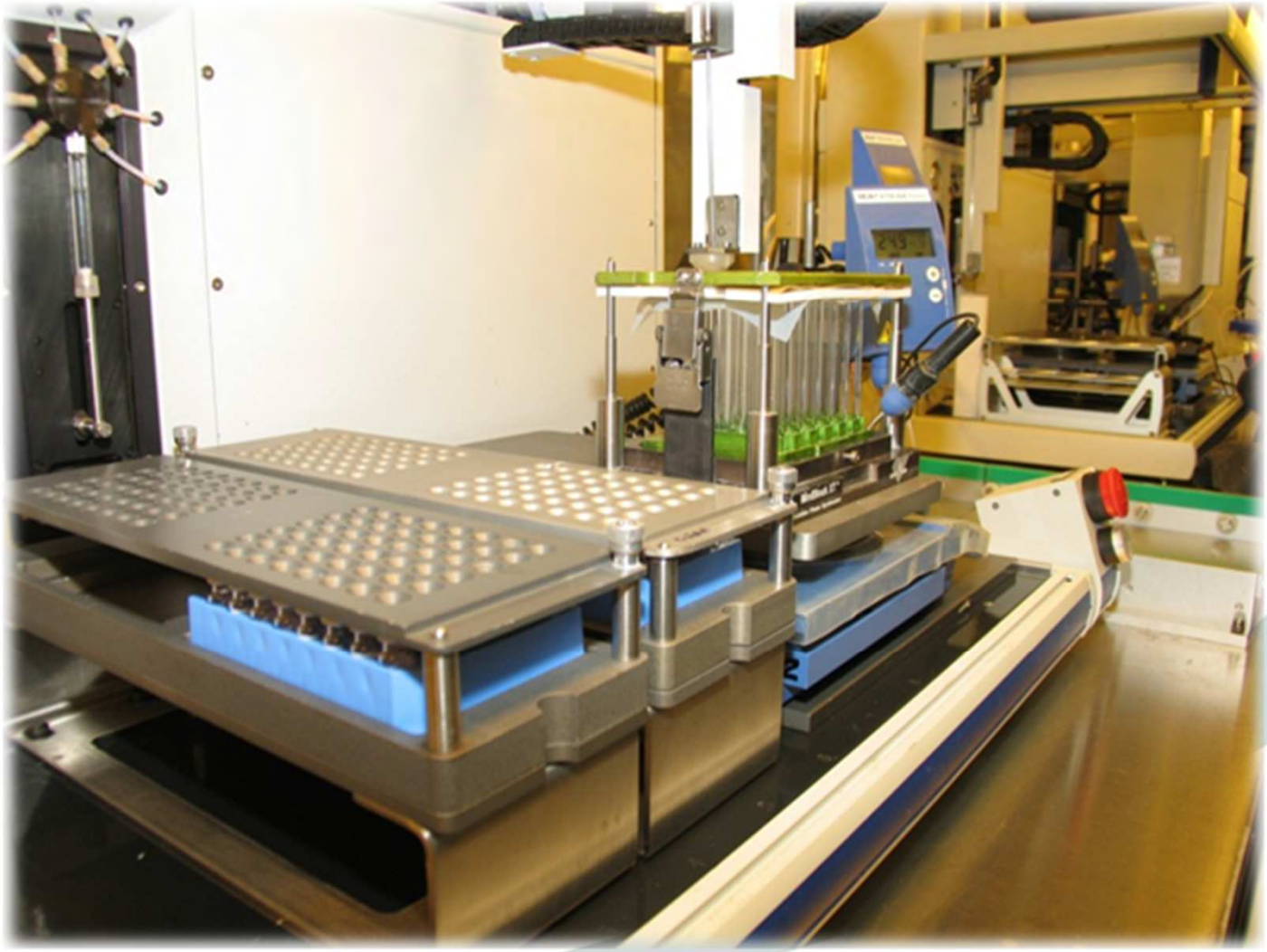
CatScI
Innovation in Catalysis

CatScI Facilities – Liquid Dispensing



CatScI
Innovation in Catalysis

CatScI Facilities – Reaction Station



CatScI
Innovation in Catalysis

CatScI Facilities – Glove Box

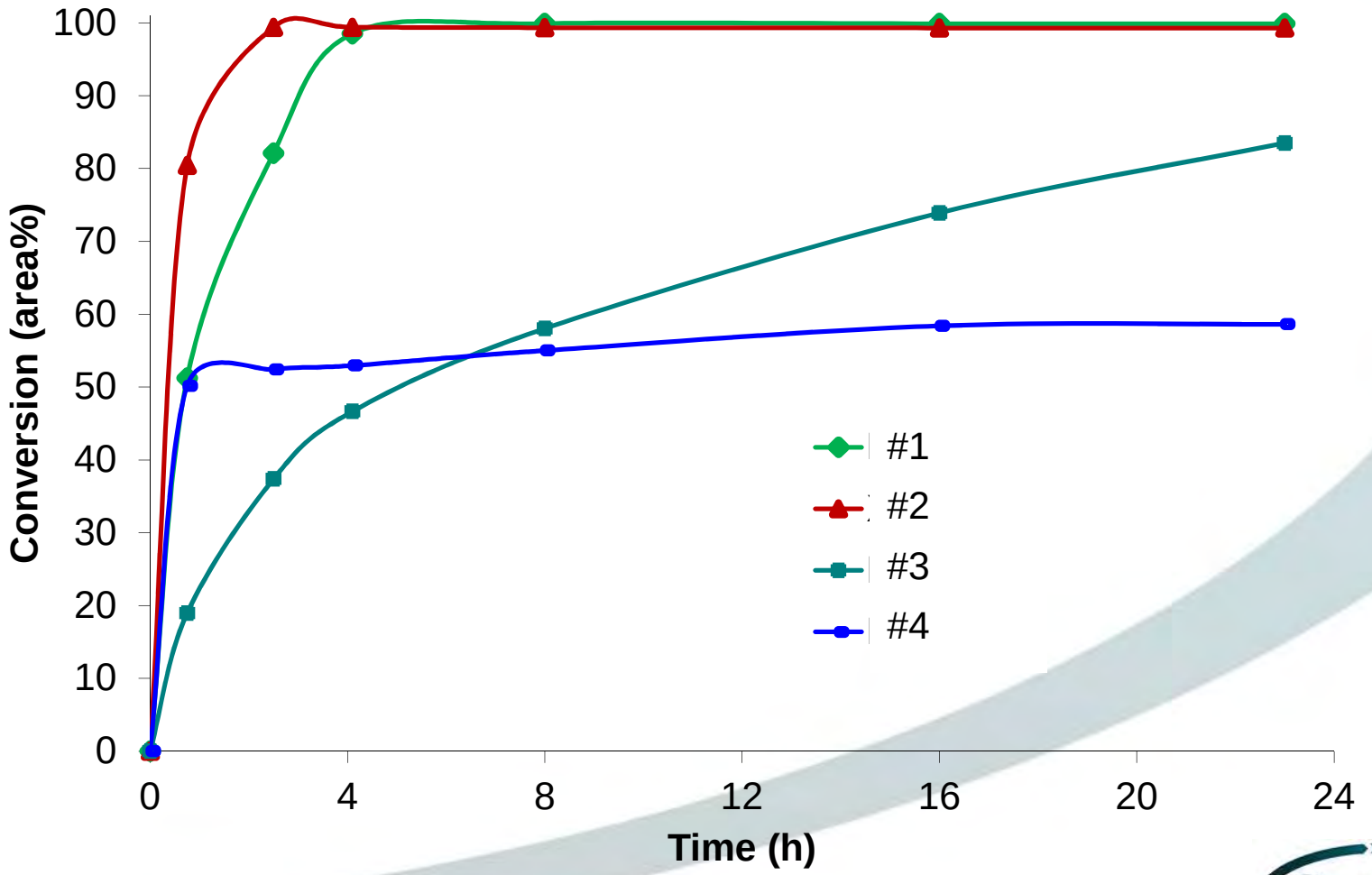
CatScI
Innovation in Catalysis

Analytical Instrumentation

CatSci
Innovation in Catalysis

Reaction Profiling

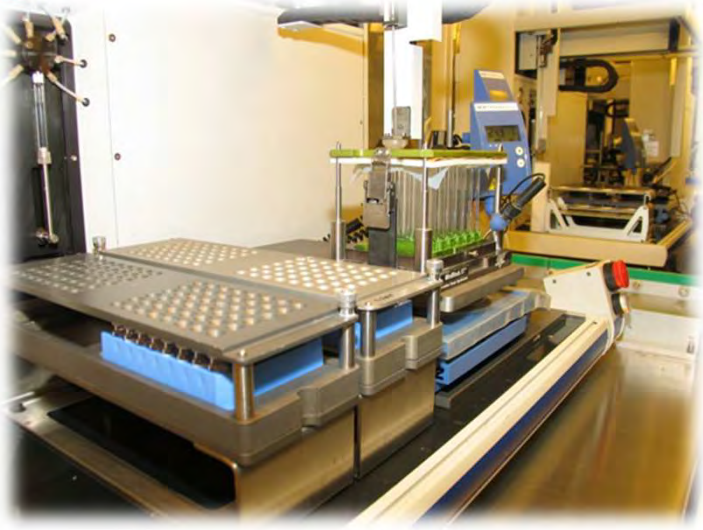
Ligand screening in a Buchwald-Hartwig Reaction



CatSci
Innovation in Catalysis

CatScI Capabilities

- Bio-catalysis
- Homogeneous catalysis
- Heterogeneous catalysis
- Metal recovery/recycle
- Process development
- Solvent/reagent selection
- Training and Consultancy



Powered by our expertise and experience in

- Reaction understanding
- Statistical methods (DoE and PCA)
- Analytical excellence

CatScI
Innovation in Catalysis

Design of Experiments (DoE)

CatSci
Innovation in Catalysis

Experimental Design

- Substrate(s) may be limited
- Experiments take time (limited)
- Reagents and solvents are consumed (cost)
- Analysis takes time (limited)
- Can only run a finite number of experiments
- Your time is more precious than the metal

- Aim to maximise data for the time and cost
  - Leads to greater Reaction Understanding

Design of Experiments (DoE)

- DoE is an efficient, structured way to investigate potentially significant experimental factors…
- …and their cause-and-effect relationship on the experimental outcome (responses)

- DoE uses statistical methods to extract and interpret the relationships between the factors

- Applicable to all chemical reactions and processes

For a Typical Chemical Reaction

Factors

Substrate A
Substrate B
Reagent
Solvent
Order of addition
Type of equipment

} Discrete

Substrate A/B ratio
Reagent quantity
Temperature
Concentration
Rate of addition
pH

} Continuous

Time

Responses

Conversion
Selectivity (regio)
Selectivity (stereo)
Selectivity (enantio)
Rate or rate constant
Equilibrium constant
Purity/Impurities

Cost
Efficiency
Throughput
Sustainability

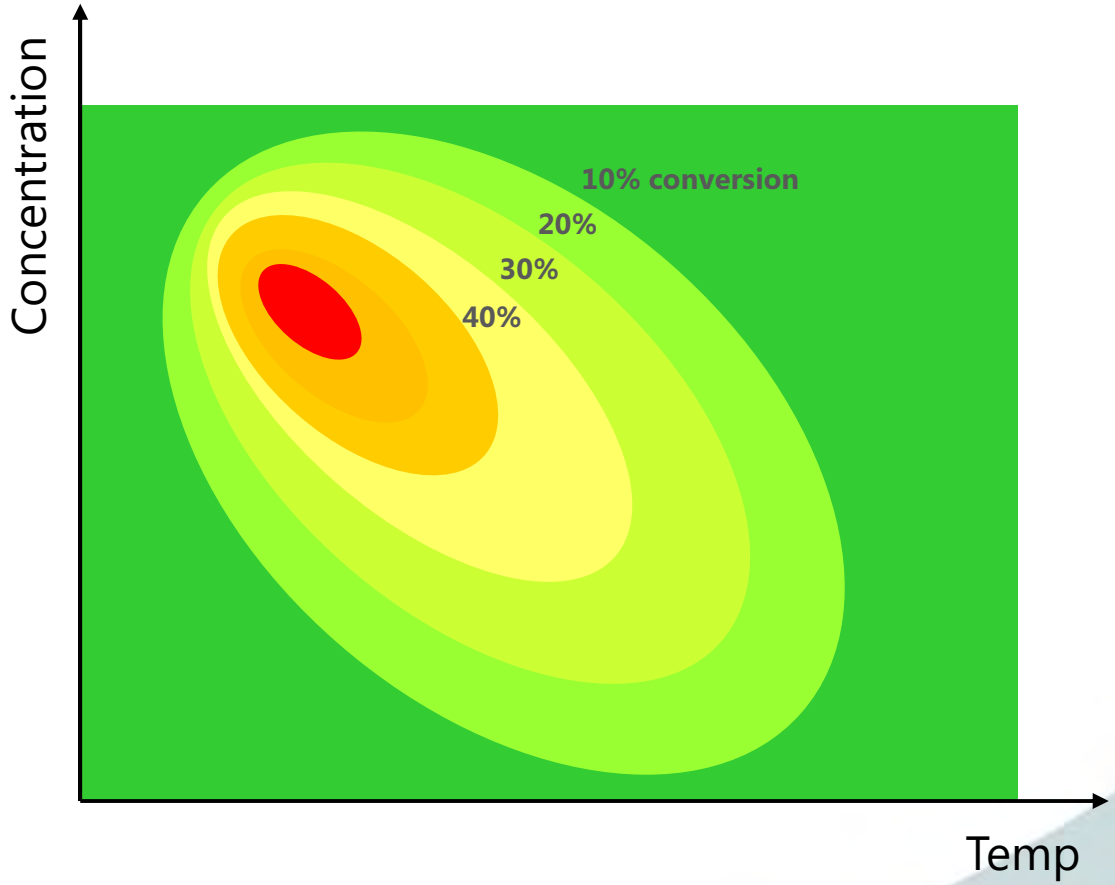
Isolated Yield

CatScI
Innovation in Catalysis

Comparison of DoE with OVAT

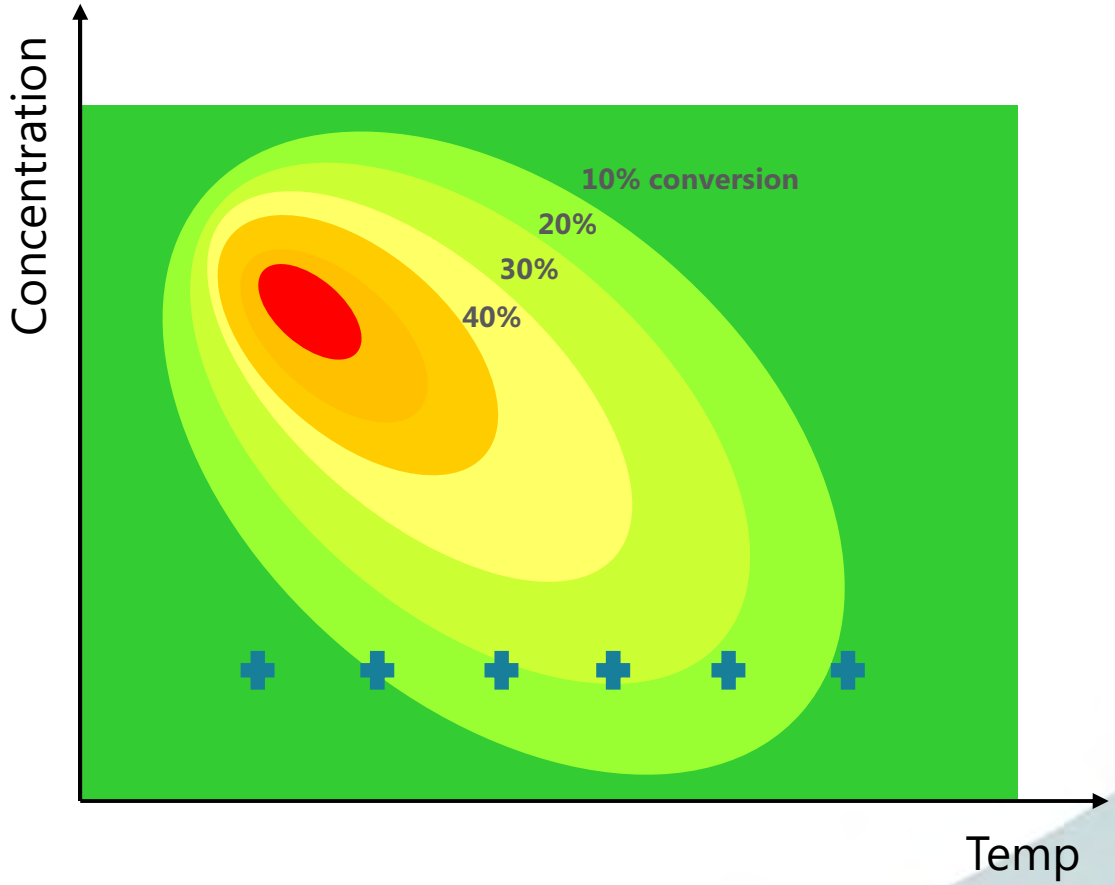
CatSci
Innovation in Catalysis

One Variable at a Time (OVAT)



- Consider two factors for a typical reaction

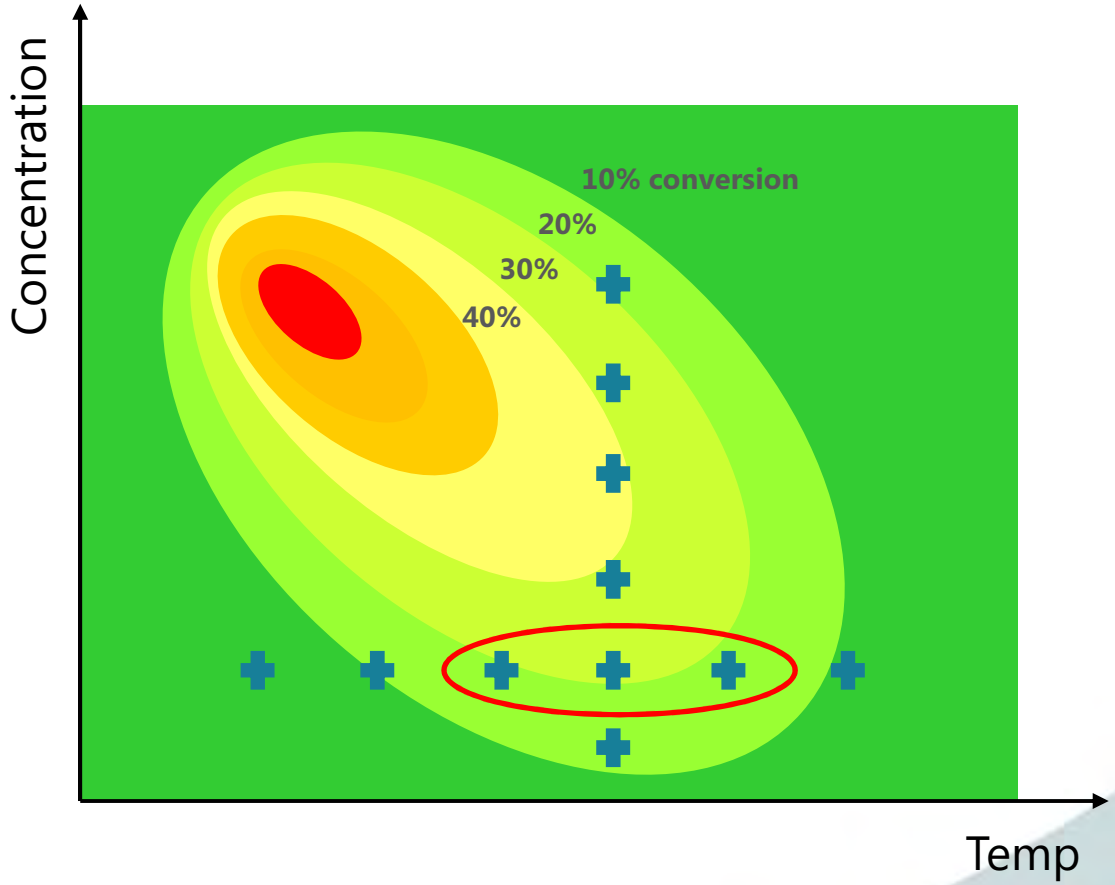
OVAT – First Factor



- Vary one factor from an arbitrary starting point

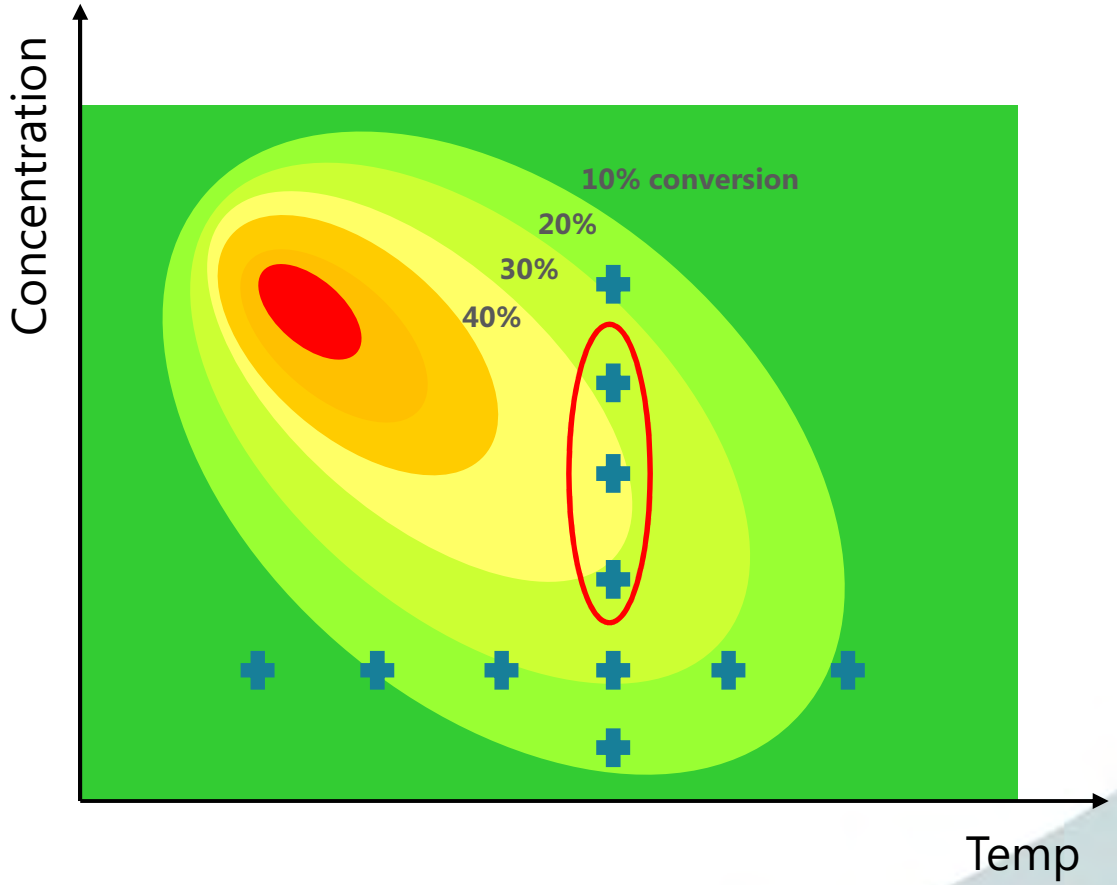
CatScI
Innovation in Catalysis

OVAT – Second Factor



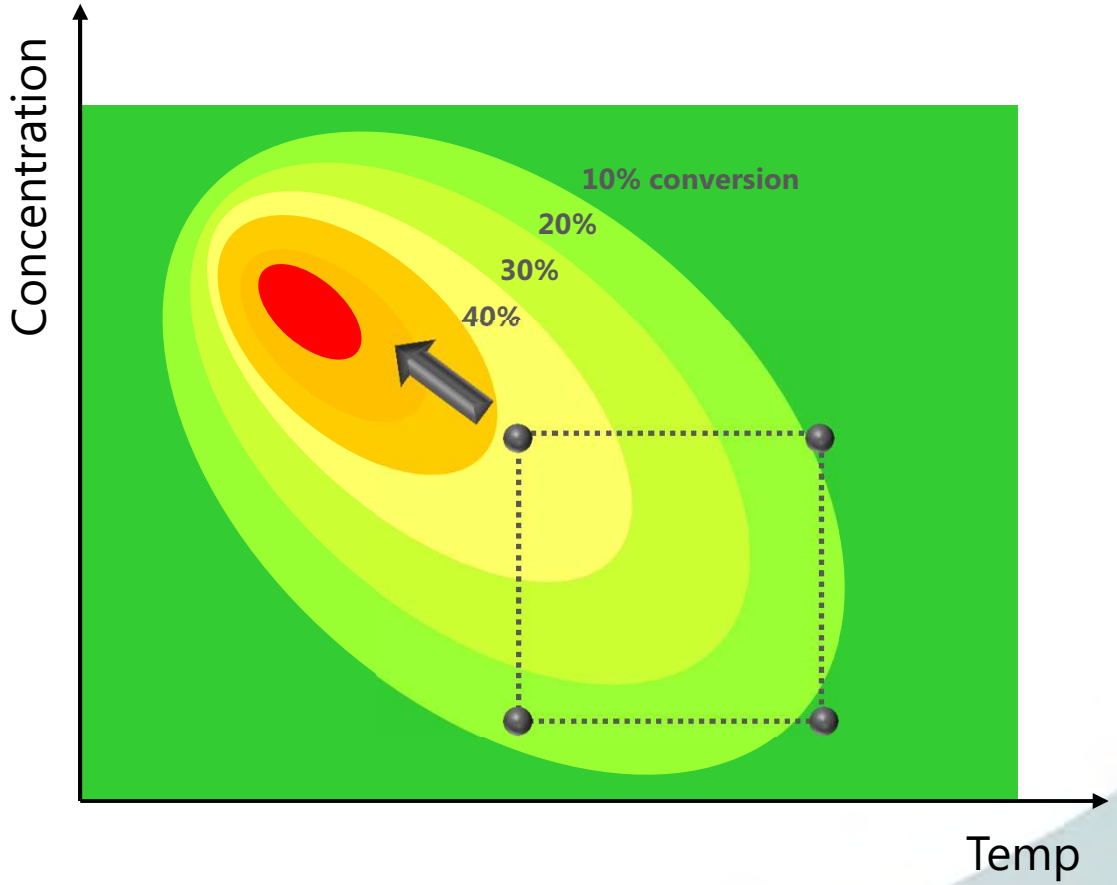
- At the optimum level of the first factor, vary the second factor

OVAT – Final Result



- A local "optimum" is identified…
- …but not the actual optimum

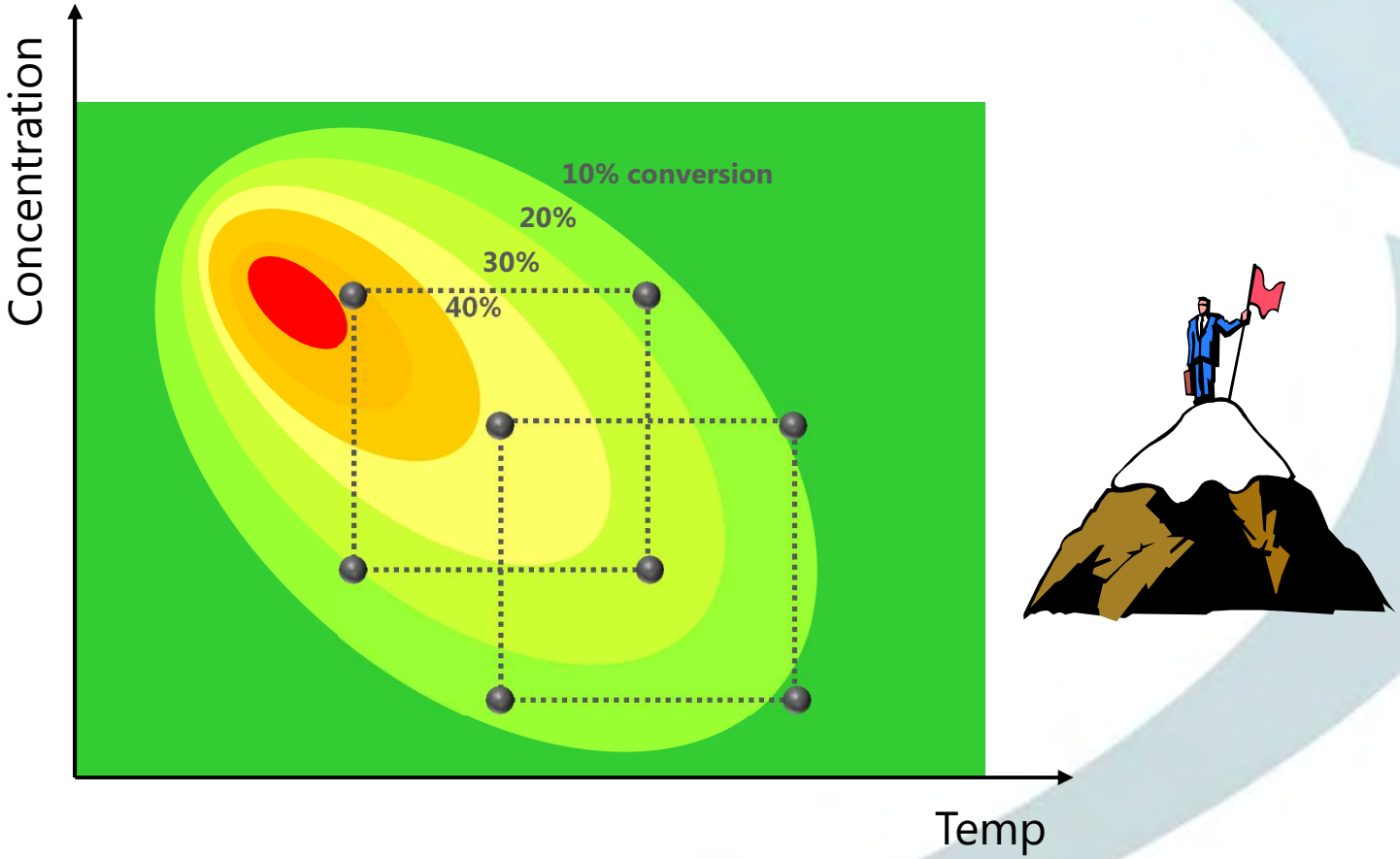
DoE – Screening Design



- A simple screening design in the same region

CatSci
Innovation in Catalysis

DoE – Optimisation Design



- A follow-up DoE design leading to optimisation

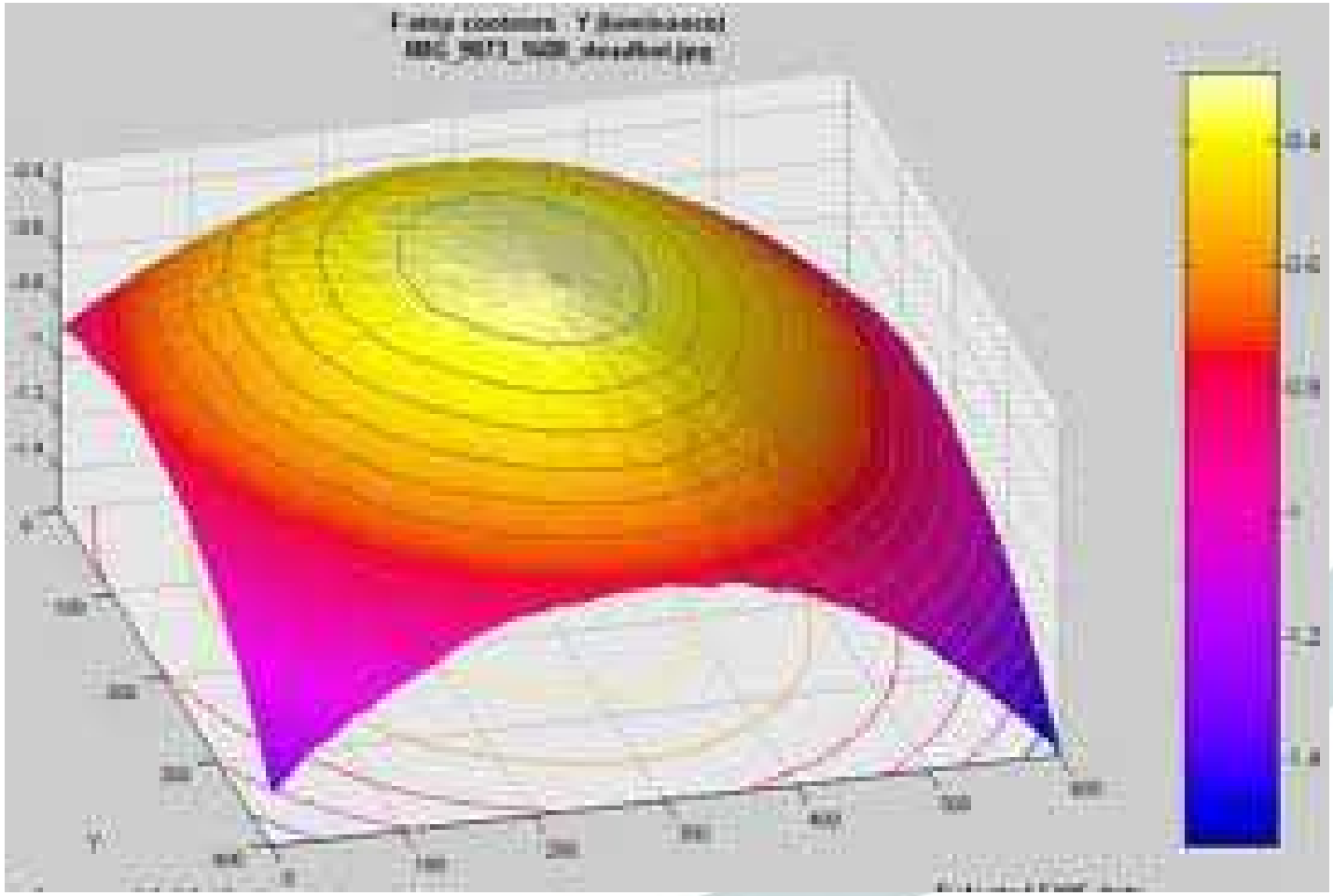
CatScI
Innovation in Catalysis

Comparison of DoE with OVAT

- With OVAT, genuine optimum may be missed
  - the experimental approach may make it impossible to find!
- Inefficient use of resources
  - more experiments gave sub-optimal result
  - better conditions are available
- Limited coverage of chemical reaction space
- No information on interactions
- No measure of variability (experimental error)

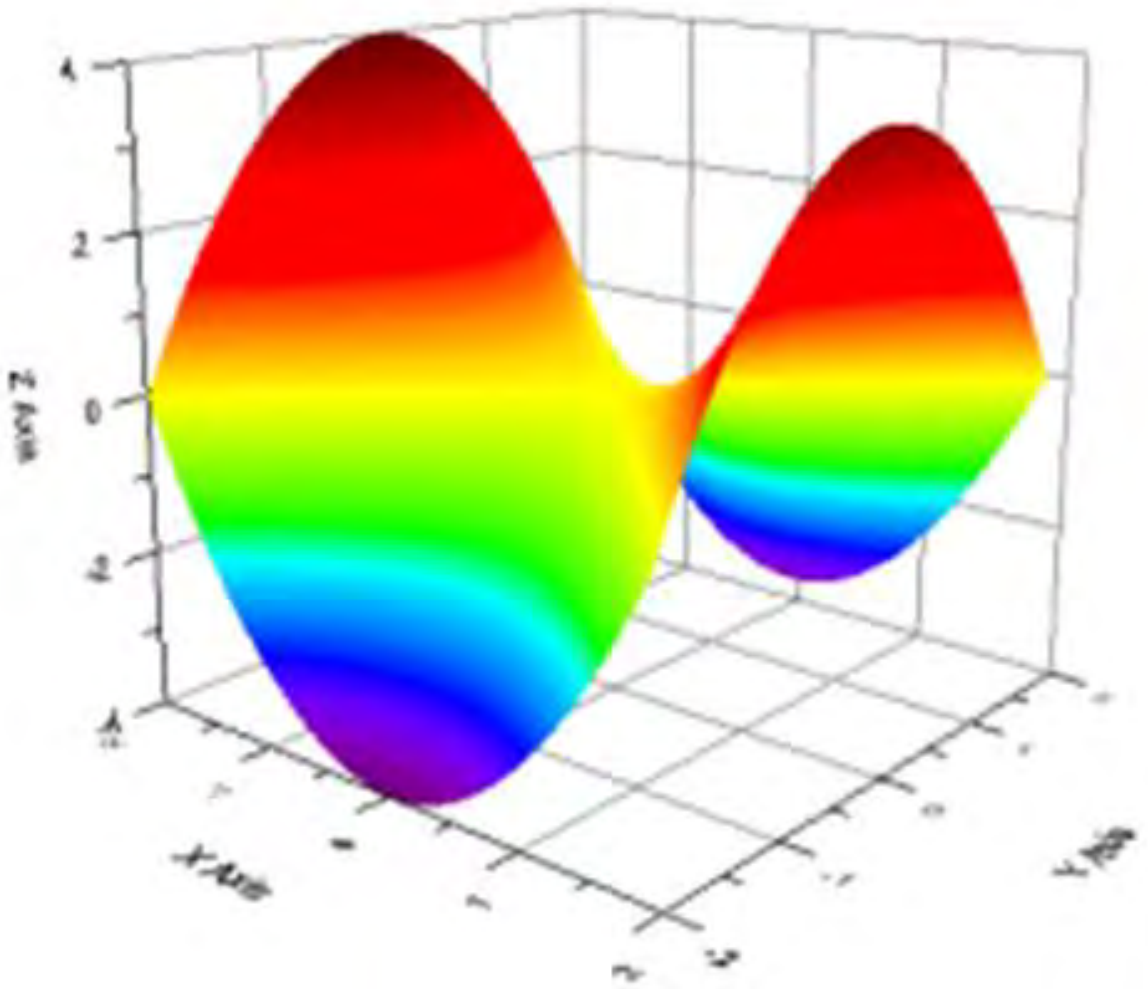
- DoE gives more data from fewer experiments

Response Surface



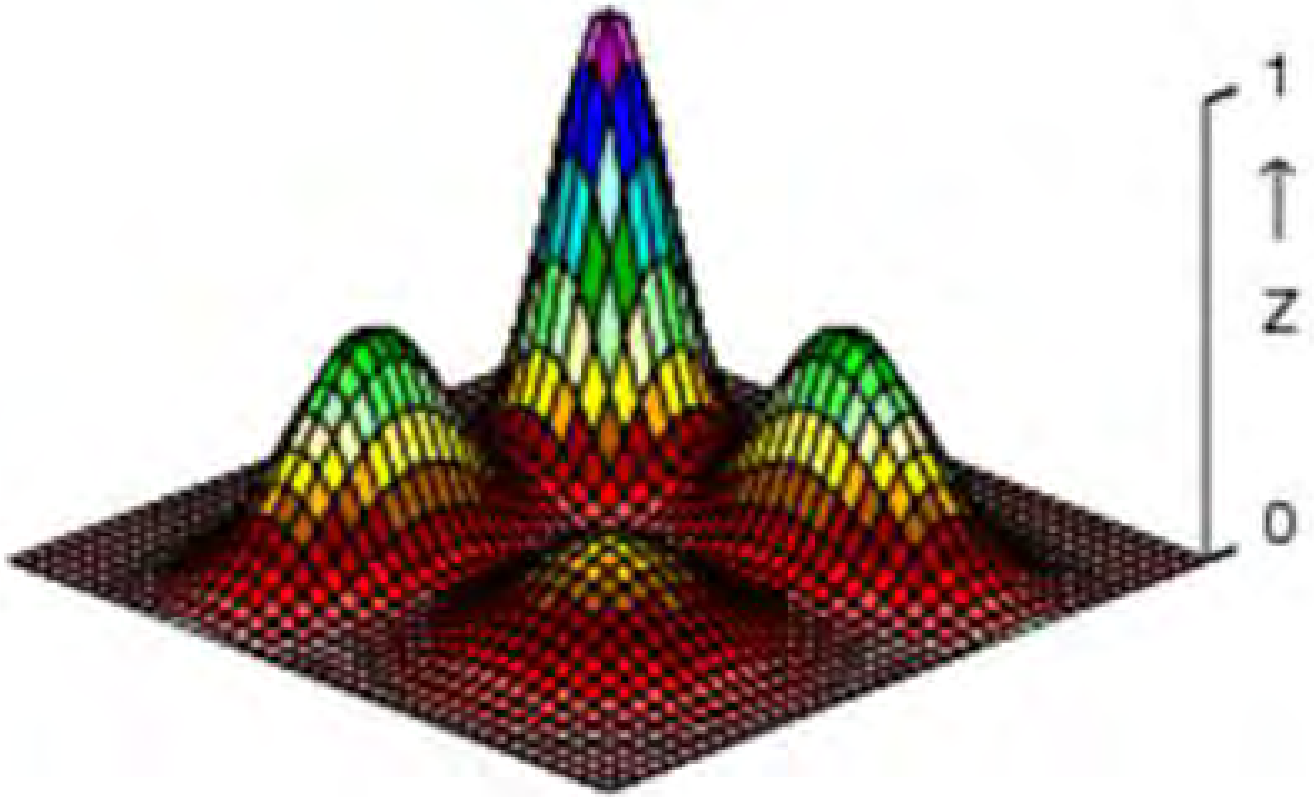
CatSci
Innovation in Catalysis

Response Surfaces



CatSci
Innovation in Catalysis

Response Surfaces

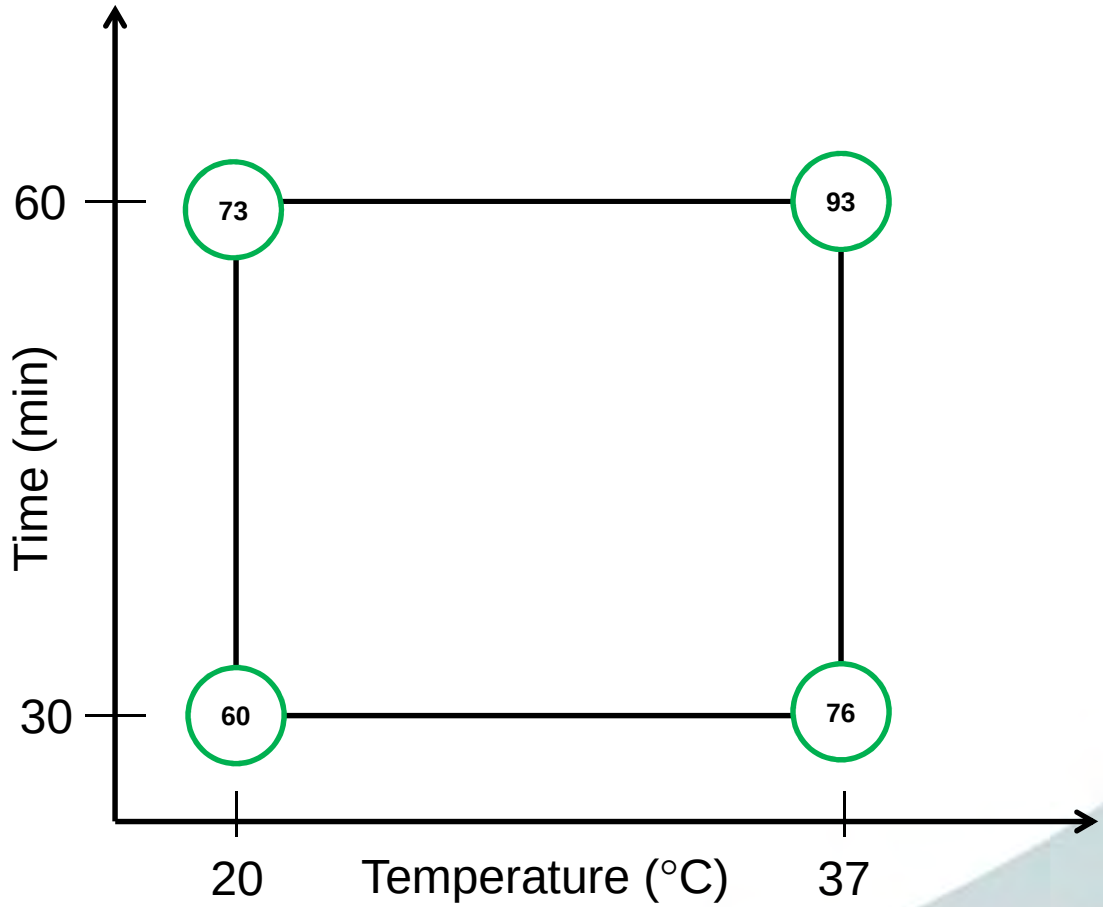


CatScI
Innovation in Catalysis

Types of Design

CatSci
Innovation in Catalysis

A Two Factor, Two Level Design



- Quantitative, not just qualitative
- DoE data is used multiple times

CatSci
Innovation in Catalysis

Quantify the Main Effects

- Average of values at high setting minus average of values at low setting
- Main effect of temperature is:

  $\{(93 + 76)/2\} - \{(73 + 60)/2\} = 84.5 - 66.5 = \mathbf{18}$

- Main effect of time is:

  $\{(93 + 73)/2\} - \{(76 + 60)/2\} = 83 - 68 = \mathbf{15}$

- DoE experimental results are used multiple times
  - more data is extracted from **each** experiment
  - a key characteristic of DoE

Modelling a Two Factor Design

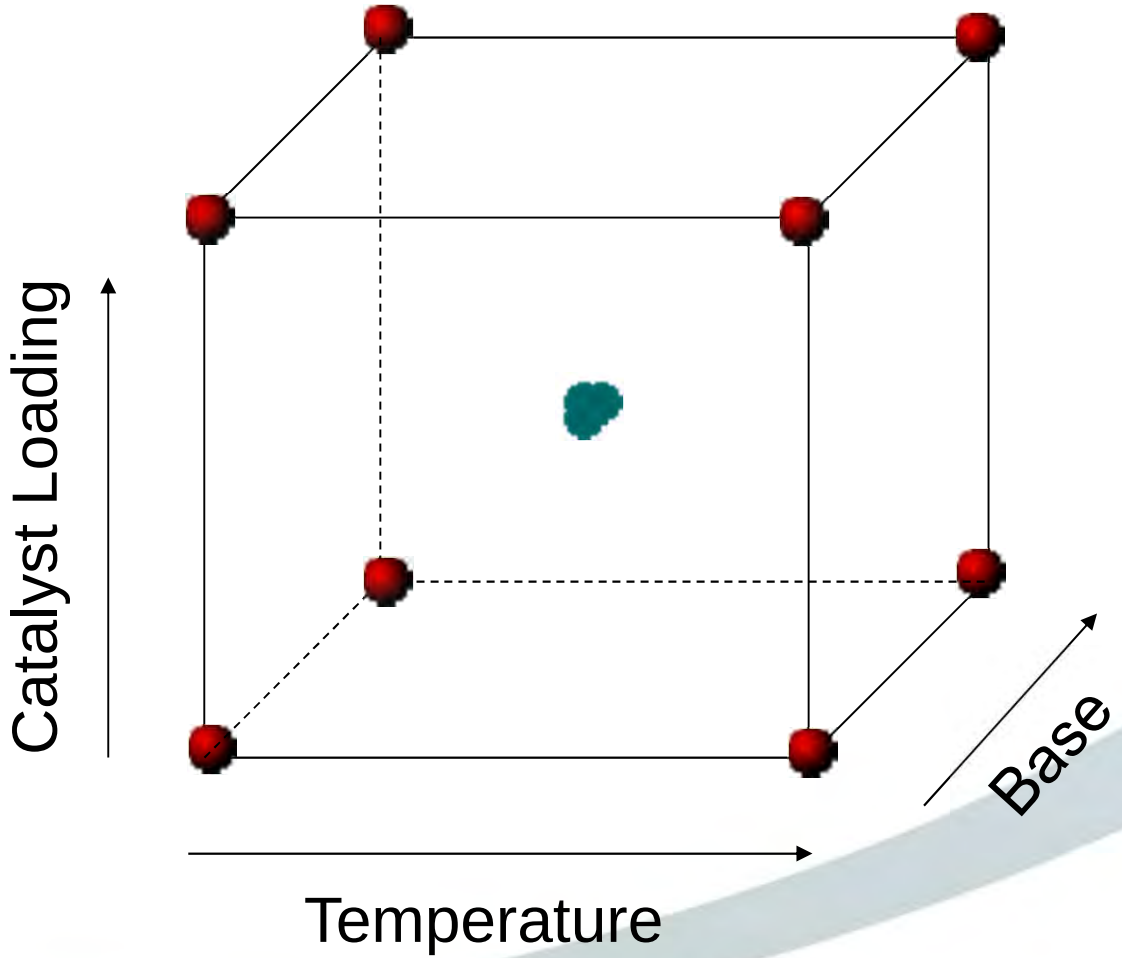
Experiment	Temperature	Time	interaction	Response
	x_1	x_2	x_1x_2	y
1	-1	-1	+1	60
2	-1	+1	-1	73
3	+1	-1	-1	76
4	+1	+1	+1	93
5	0	0	0	80

- $60 = b_0 - b_1 - b_2 + b_{12} + e_1$
- $73 = b_0 - b_1 + b_2 - b_{12} + e_2$
- $76 = b_0 + b_1 - b_2 - b_{12} + e_3$
- $93 = b_0 + b_1 + b_2 + b_{12} + e_4$
- $80 = b_0 + e_5$

5 equations
5 unknowns

this can be solved

A Three Factor, Two Level Design



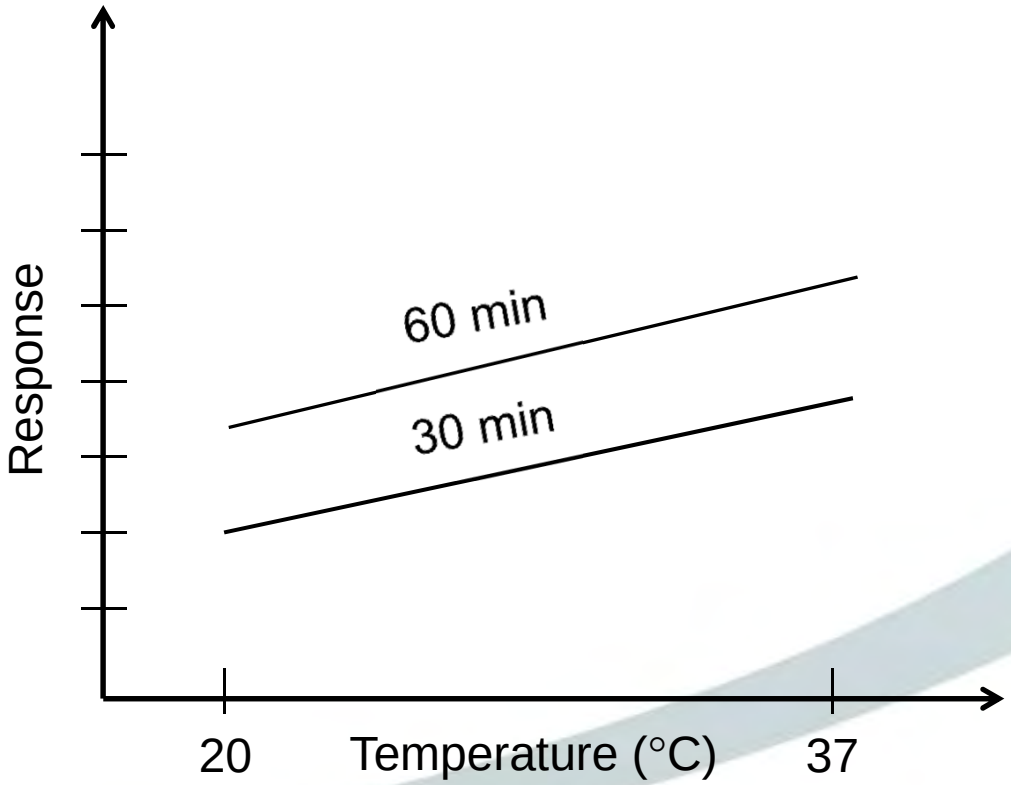
CatScI
Innovation in Catalysis

Three Factor Design Interactions

	A	B	C	AB	AC	BC	ABC
1	-1	-1	-1	+1	+1	+1	-1
2	-1	+1	-1	-1	+1	-1	+1
3	+1	-1	-1	-1	-1	+1	+1
4	+1	+1	-1	+1	-1	-1	-1
5	-1	-1	+1	+1	-1	-1	+1
6	-1	+1	+1	-1	-1	+1	-1
7	+1	-1	+1	-1	+1	-1	-1
8	+1	+1	+1	+1	+1	+1	+1
Control	0	0	0	0	0	0	0

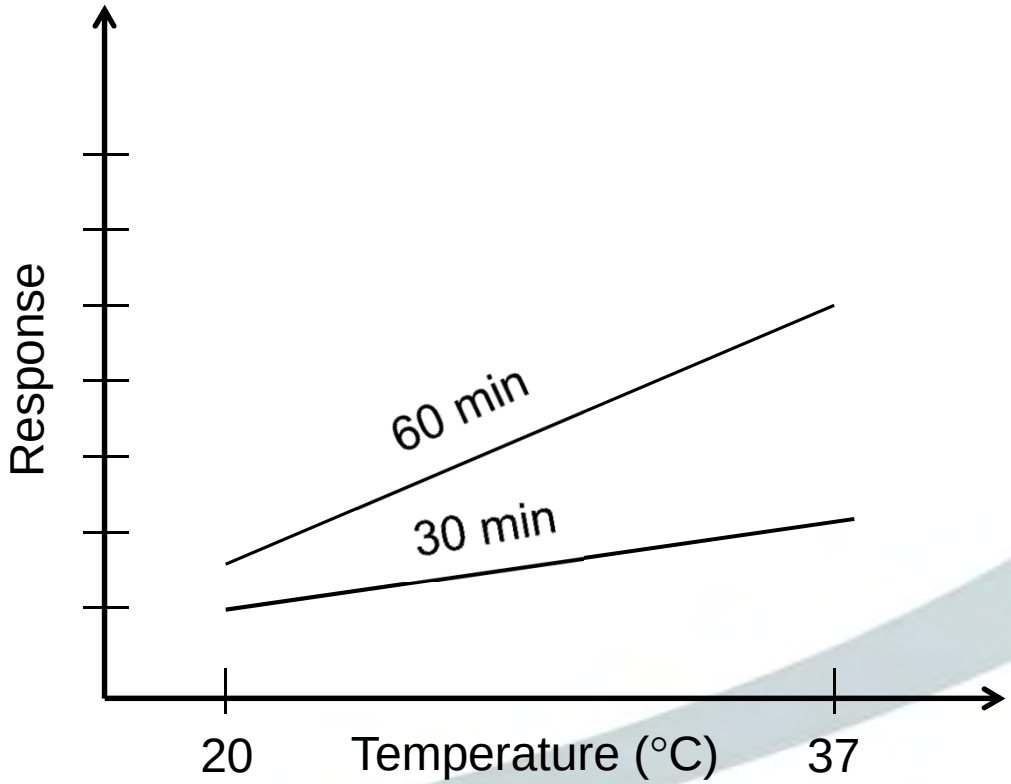
No Interaction

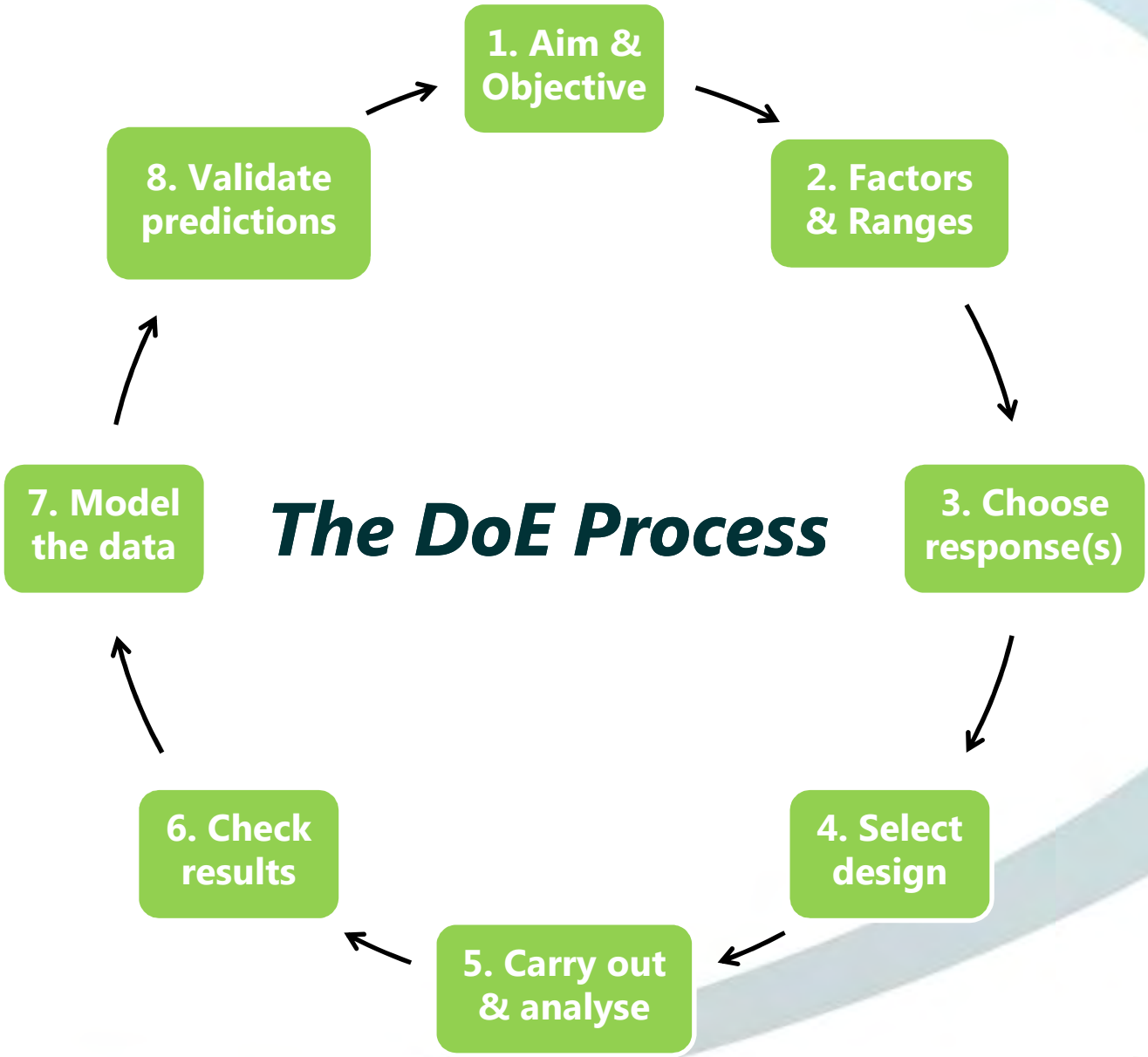
- An interaction exists when differences on one factor depend on the level of another factor



Some Interaction

- An interaction exists when differences on one factor depend on the level of another factor





CatSci
Innovation in Catalysis

Why Factorial Design?

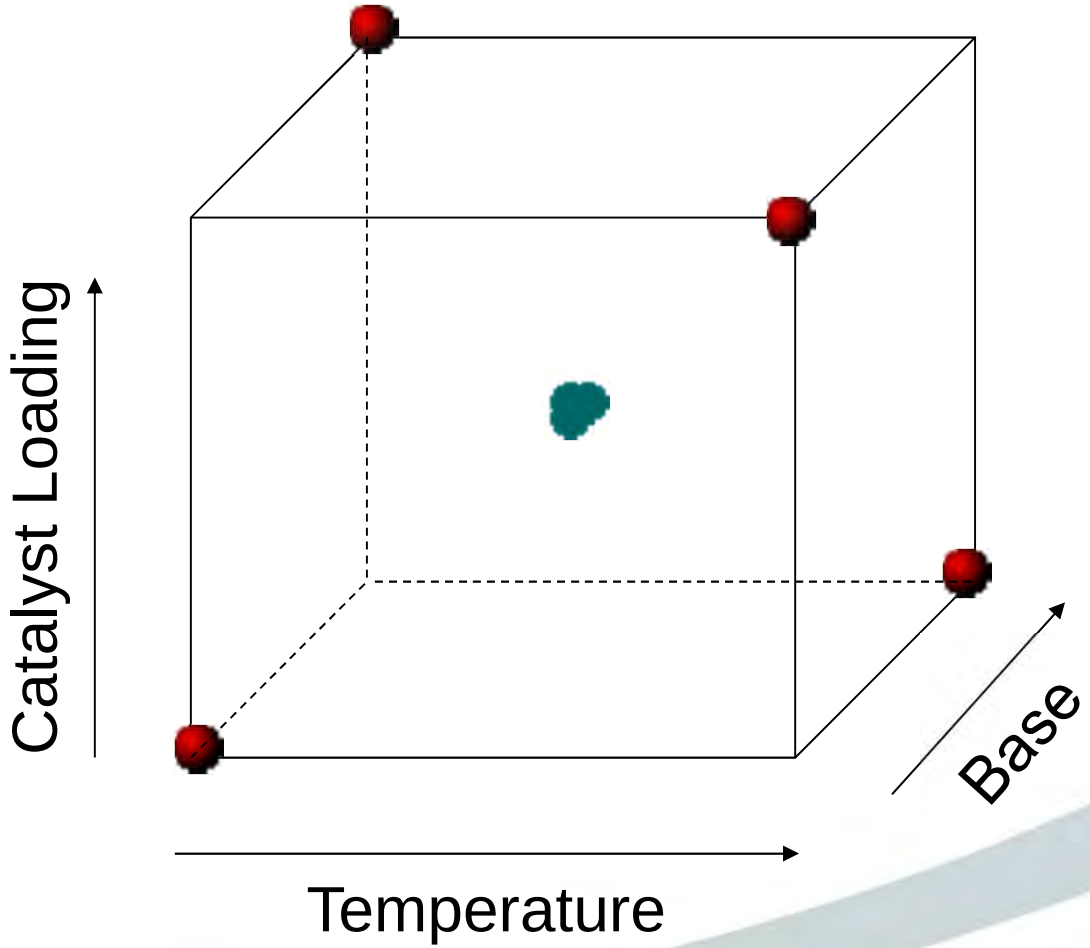
CatSci
Innovation in Catalysis

Factorial Experimental Design

Factors	Full set	Recommended	Factorial design
3	8 + 3		4 + 3
4	16 + 3		8 + 3
5	32 + 3	16 + 3	8 + 3
6	64 + 3	32 + 3	16/8 + 3
9	512 + 3	128 + 3	32/16 + 3
12	4096 + 3	256 + 3	32/16 + 3

- Experiment numbers grow exponentially as the number of factors increase
  - as do the number and complexity of interactions

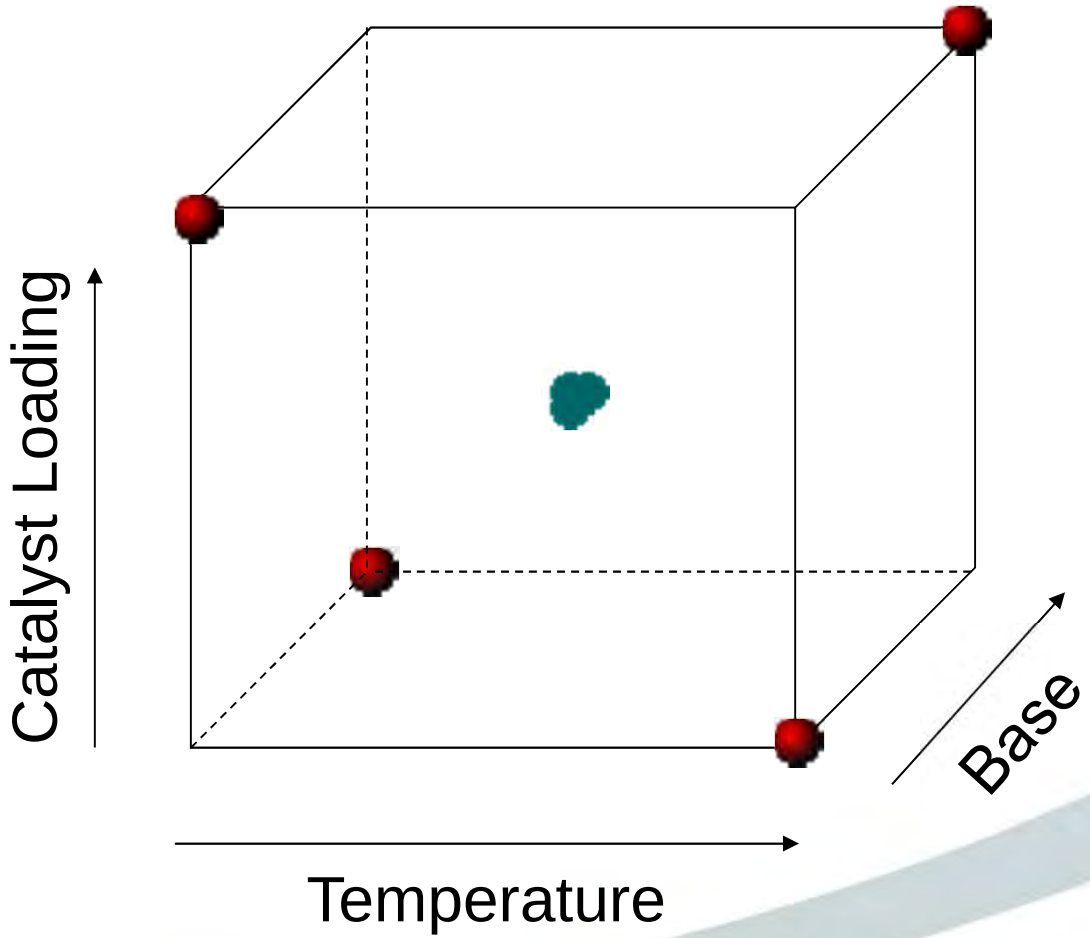
Half Factorial – Design 1



	A	B	C
1	–	–	–
4	+	+	–
6	–	+	+
7	+	–	+

CatScI
Innovation in Catalysis

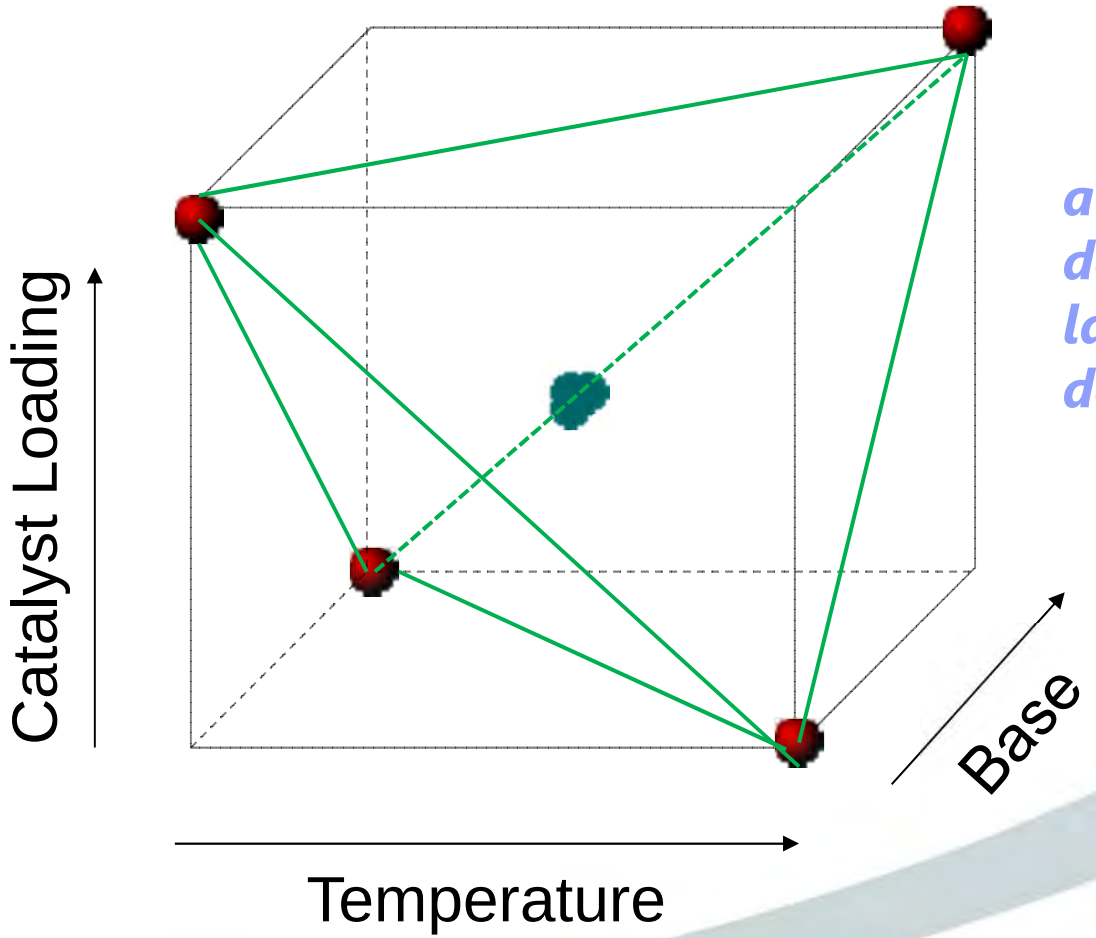
Half Factorial – Design 2



	A	B	C
2	–	+	–
3	+	–	–
5	–	–	+
8	+	+	+

CatScI
Innovation in Catalysis

Half Factorial – Tetrahedron



a tetrahedral fractional design will cover as large a part of the design space as possible

CatScI
Innovation in Catalysis

Resolution of Factorial Designs

Experiments \ Factors	2	3	4	5	6	7	8	9	10	11	12
4	Full	1/2									
8		Full	1/2	1/4	1/8	1/16					
16			Full	1/2	1/4	1/8	1/16	1/32	1/64	1/128	1/256
32				Full	1/2	1/4	1/8	1/16	1/32	1/64	1/128
64					Full	1/2	1/4	1/8	1/16	1/32	1/64
128						Full	1/2	1/4	1/8	1/16	1/32

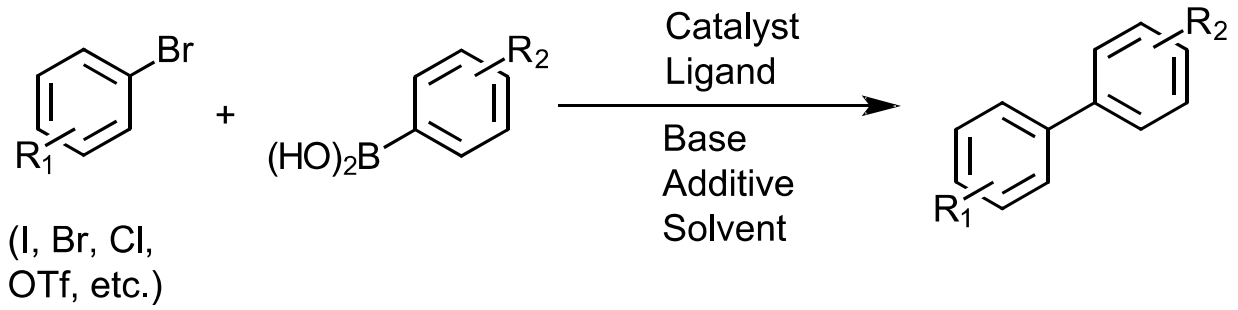
- Factorial experiment designs can provide significant information from fewer experiments
  - need to balance information required with realistic number of experiments

For a Typical Catalytic Reaction

Factors	Levels	Responses
Substrate A	Discrete	Conversion
Substrate B		Selectivity (regio)
Catalyst		Selectivity (stereo)
Ligand		Selectivity (enantio)
Solvent		Rate or rate constant
Base		Equilibrium constant
Additive		Purity/Impurities
Order of addition		
Substrate A/B ratio	Mol. equiv.	Cost
Base quantity	Mol. equiv.	Efficiency
Temperature	K	Throughput
Concentration	Molarity	Sustainability
Rate of addition	Mins	**Yield**
Catalyst quantity	Mol. equiv.	
Catalyst/ligand ratio	Mol. equiv.	
Base/Additive quantity	Mol. equiv.	

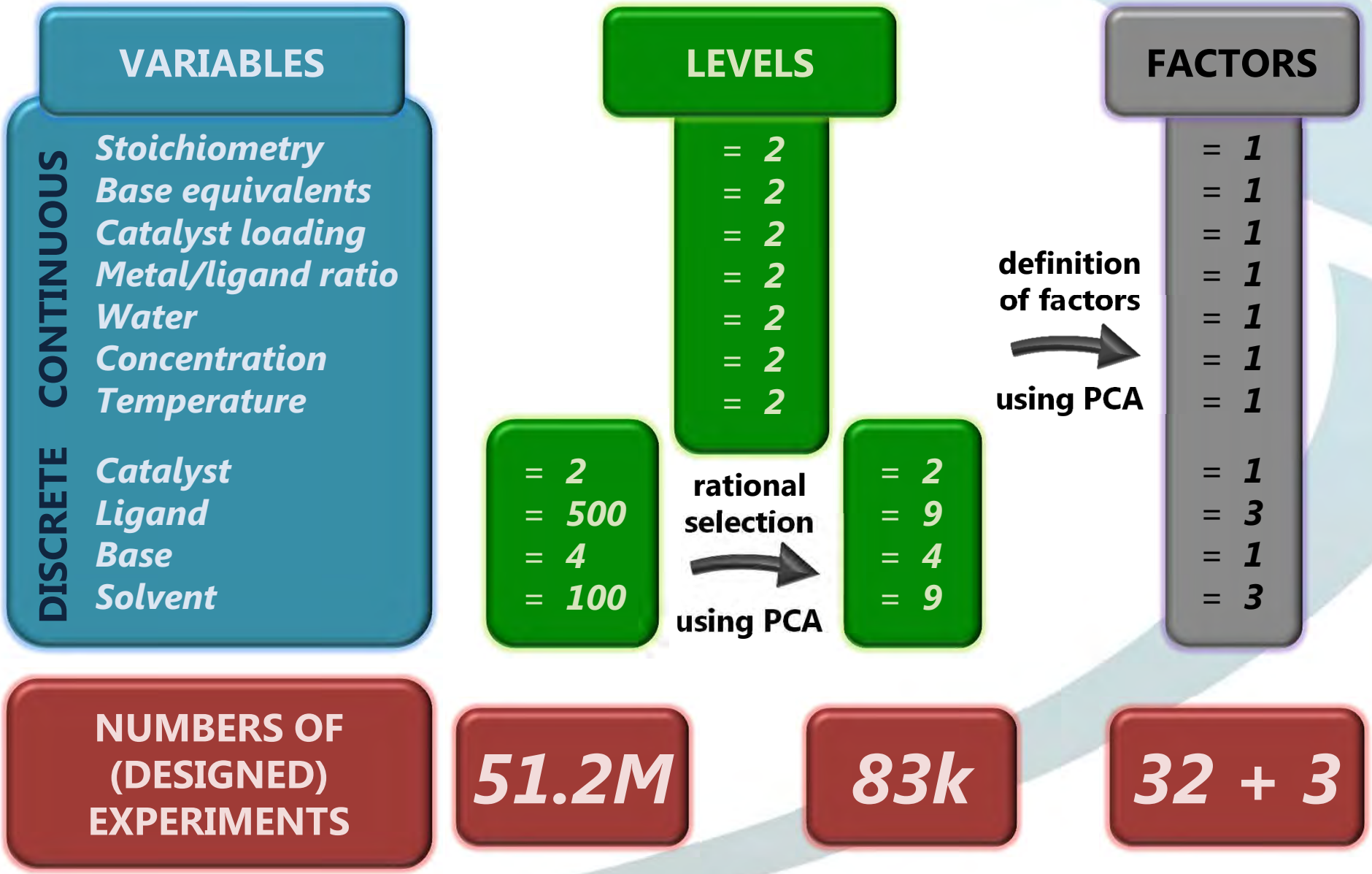
CatSci
Innovation in Catalysis

A Typical Cross-Coupling Reaction



- Suzuki reaction with defined R_1 and R_2
- Consider only one halide or pseudo-halide
- Consider only the boronic acid
- Consider standard factors at 2 levels
- Consider all catalyst/ligand combinations
- Consider all commercially available solvents
- Limit bases to four (1 factor at 4 levels)

CatScI
Innovation in Catalysis



CatSci
Innovation in Catalysis

Concept Article

Beyond the Numbers: Charting Chemical Reaction Space

Paul M. Murray , Simon N. G. Tyler , and Jonathan D. Moseley *
CatSci Ltd., CBTC2, Capital Business Park, Wentloog, Cardiff CF3 2PX, United Kingdom

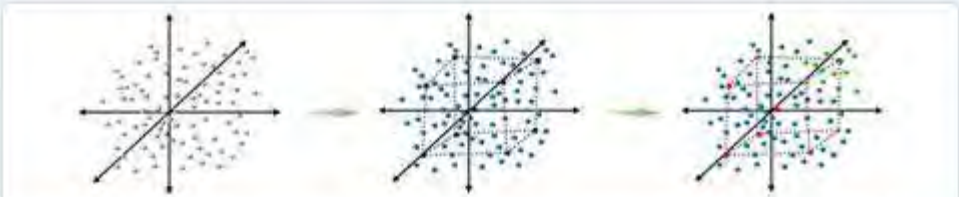
Org. Process Res. Dev., 2013, 17 (1), pp 40-46
DOI: 10.1021/op300275p
Publication Date (Web): December 18, 2012

*Telephone: +44 29 2083 7444. E-mail: jonathan.moseley@catsci.com.

Section: Unit Operations and Processes

ActiveView PDF
PDF [761 KB]
PDF w/ Links [759 KB]
Full Text HTML
Abstract
Figures
Reference QuickView
Add to ACS ChemWorx

Abstract



We present here an informed estimate of the millions of parameter settings that might be required to optimise one typical transition-metal-catalysed reaction. We describe briefly how both Design of Experiments (DoE) and Principal Component Analysis (PCA) techniques may be combined to reduce the number of potential reaction settings to a practical number of experiments without losing critical information. A key feature of this approach is the ability to relate discrete or discontinuous parameters to one another. The methodology is presented so that any reaction may be assessed in a similar way. We believe this represents for the first time an informed estimate of the number of potential permutations that are possible for these types of reactions in particular, and therefore the enormity of the task in optimising them. The powerful combination of DoE and PCA applied systematically and in an experimentally directed approach is beneficial for optimising reactions, particularly challenging transition-metal-catalysed reactions. However, this approach is beneficial to all reactions, especially when dealing with discrete parameters, such as solvents for example.

View: Full Text HTML | Hi-Res PDF | PDF w/ Links

CatSci
Innovation in Catalysis

Dealing with Discrete Factors – Principal Component Analysis

CatSci
Innovation in Catalysis

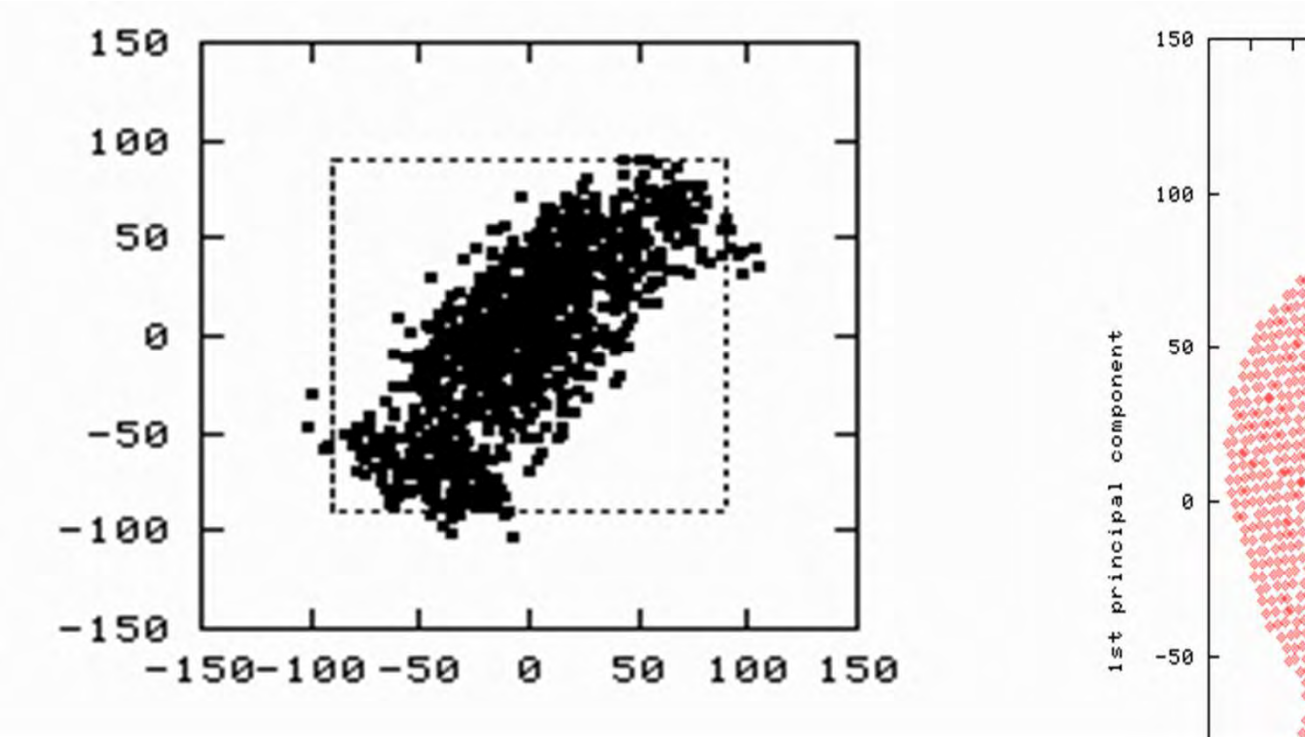
Discrete Factors

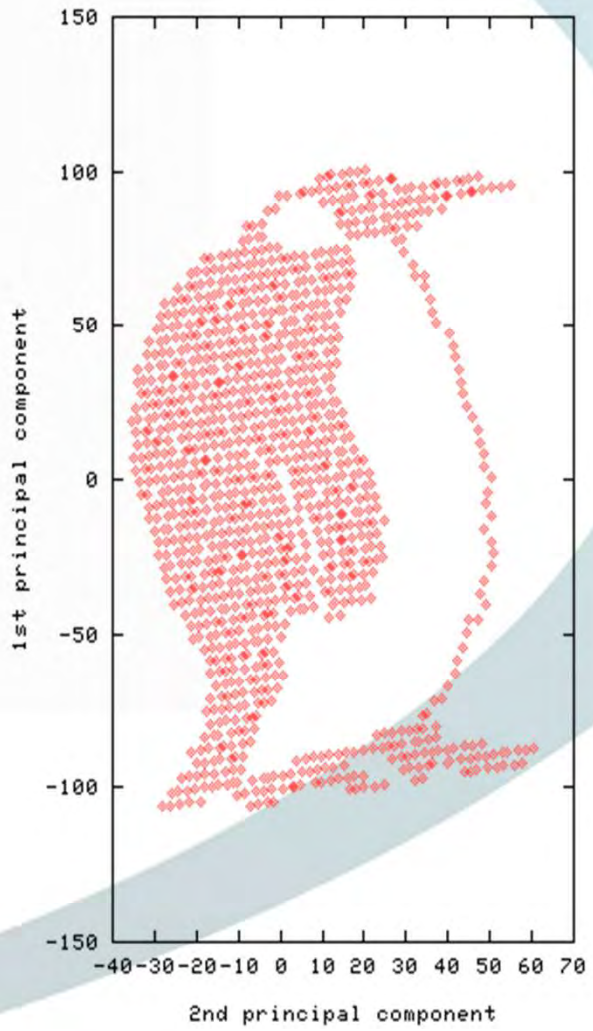
- Solvents
- Amines
- Lewis acids
- Dione ligands
- Phosphine ligands
- Aromatic substituents
- Eluent for chromatography
- Supports for chromatography
- …etc….

Principal Component Analysis (PCA)

- Every chemical can be described by properties
  - Physical (mp, bp, density etc)
  - "Theoretical" (bite angle, bond lengths, orbitals etc)
- A Principal Component (PC) is the sum of one or more of these properties
- PCA is a way of identifying patterns in data
  - expresses data to highlight similarities and differences
- Removes redundancy and reduces dimensions
- See standard textbook
  - Carlson, R.; Carlson, J. E. *Design and Optimization in Organic Synthesis*, 2nd ed.; Elsevier: Amsterdam, 2005 (ISBN 0-444-51527-5)
- Example follows…

PCA – An Example

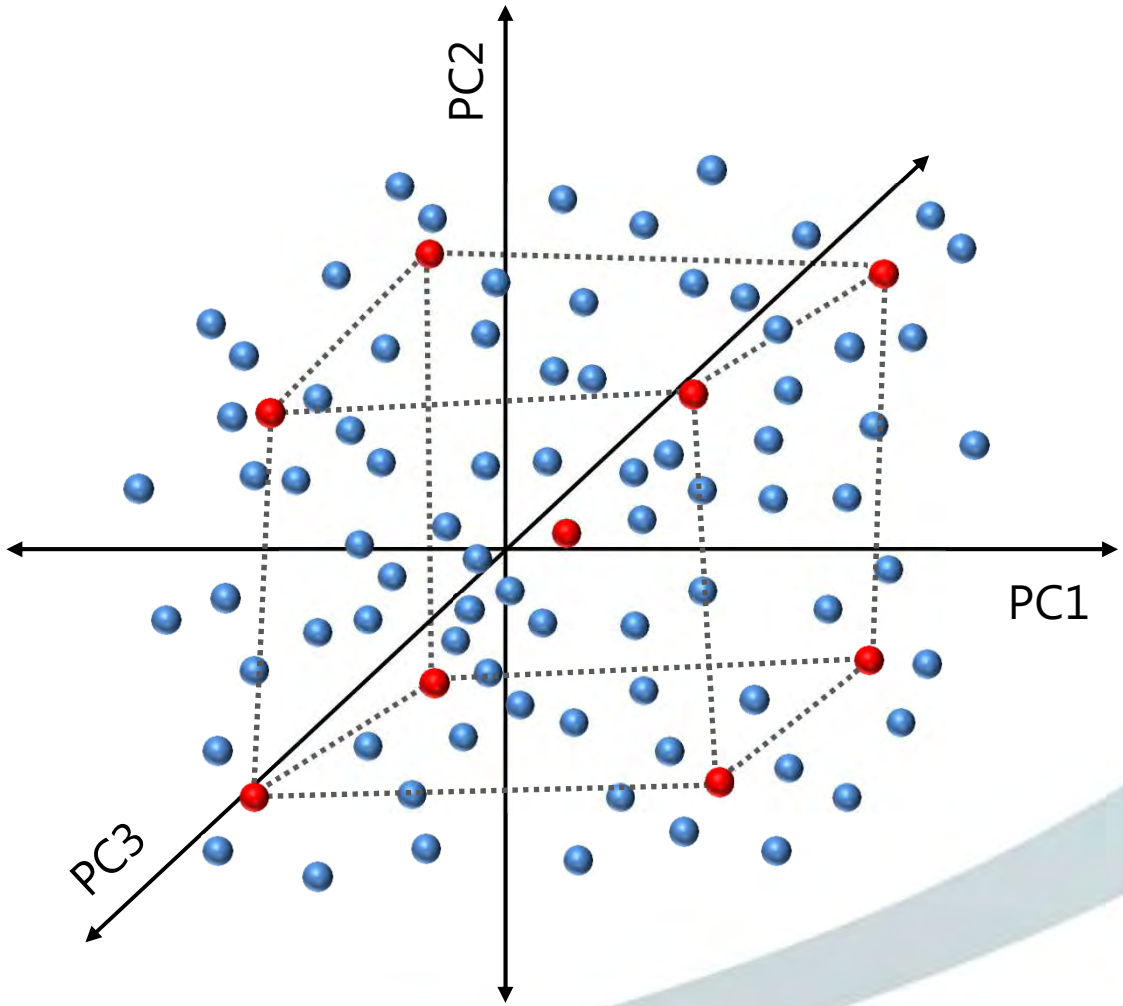




PCA helps show the pattern in the dataset

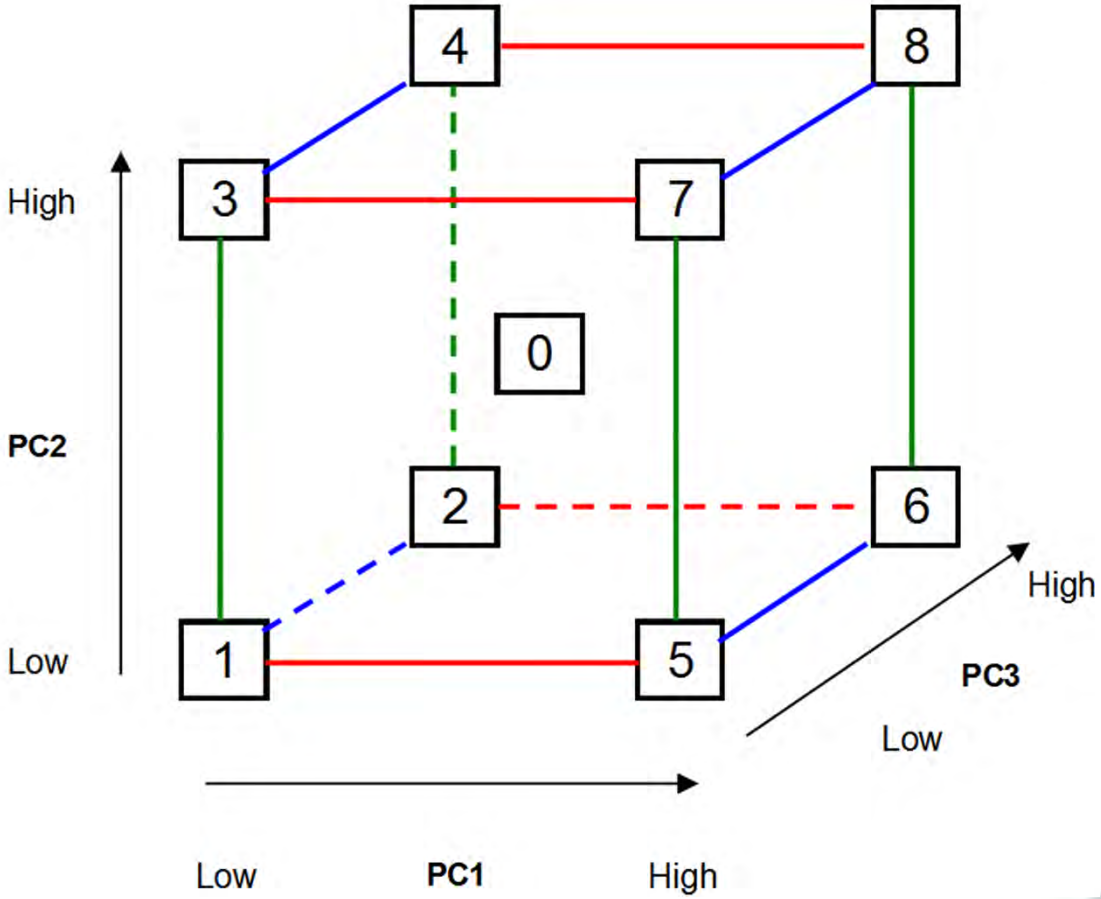
CatSci
Innovation in Catalysis

Reagent Selection in Practice



CatSci
Innovation in Catalysis

Coding of Reagents



Position	PC1	PC2	PC3
0	0	0	0
1	-1	-1	-1
2	-1	-1	1
3	-1	1	-1
4	-1	1	1
5	1	-1	-1
6	1	-1	1
7	1	1	-1
8	1	1	1

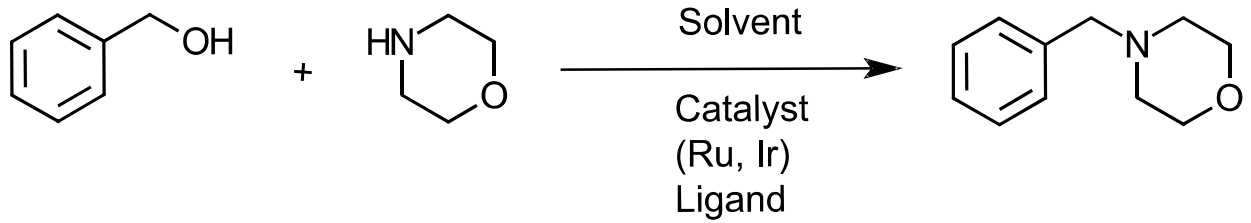
CatScI PCA Maps

- Solvents
  - 500 solvents, 20 properties/descriptors
- Monodentate phosphine ligands
  - 378 ligands, 29 properties/descriptors
- Bidentate phosphine ligands
  - 346 ligands, 28 properties/descriptors
- NHC ligands – available
- Mixed ligands – available
- Dione ligands – available
- Amines – available
- (Lewis acids – available)

Case Study

CatSci
Innovation in Catalysis

"Borrowing Hydrogen" Example



- Challenging reaction to give a good spread of data
  - Reactions works well for Ru (100%)
    - But not very fast so can differentiate ligand effects
  - Reaction is poor for Ir (~30%)
    - Reaction improves with ligand added (not standard)
- A good reaction to investigate catalyst, ligand and solvent effects

CatSci
Innovation in Catalysis

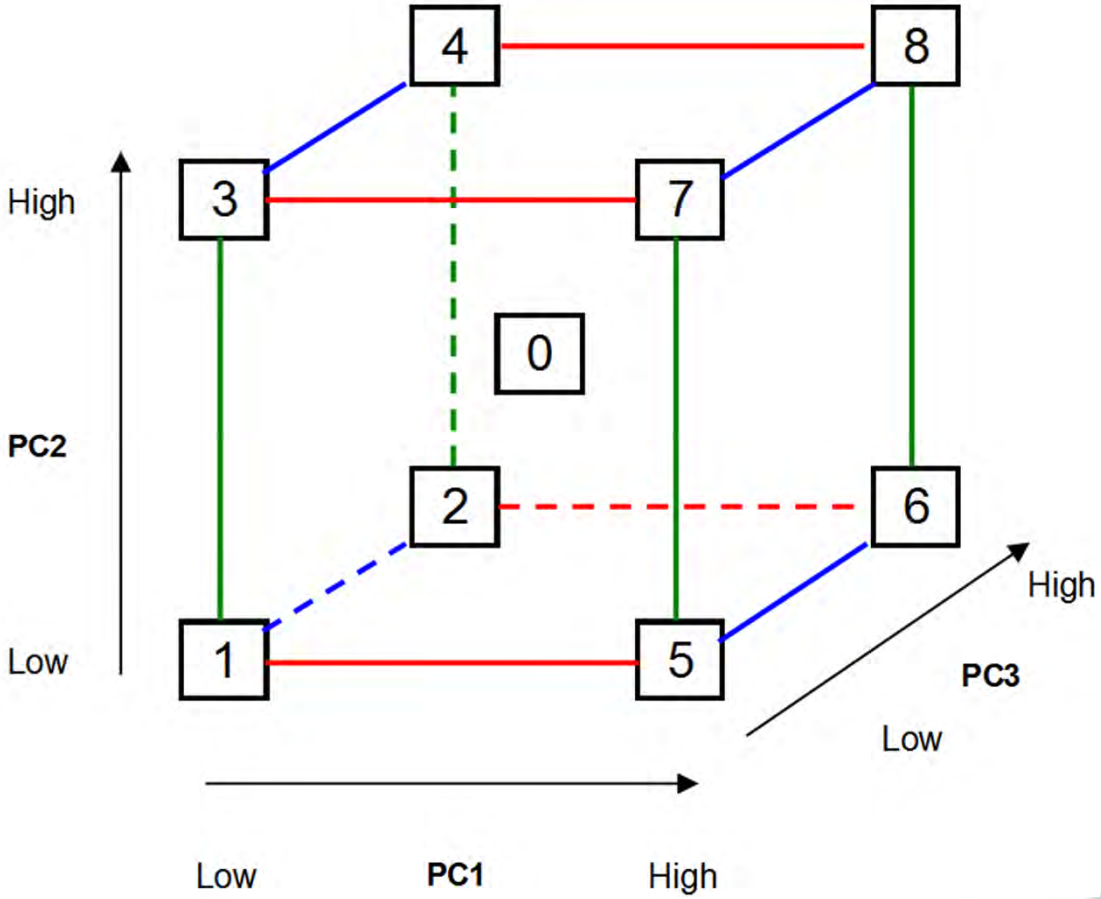
Diverse Range of Ligands

Monodentate P ligands	**PC1**	**PC2**	**PC3**
Tri-p-tolylphosphine	0	0	0
Tris(pentafluorophenyl)phosphine	-1	-1	-1
Tris[4-(trifluoromethyl)phenyl] phosphine	-1	-1	1
Triethyl Phosphite	-1	1	-1
Triisopropylphosphine	-1	1	1
Tri-tert-butylphosphine (HBF4 salt)	1	-1	-1
Tris(2-methoxyphenyl)phosphine	1	-1	1
2,8,9-triisobutyl-2,5,8,9-tetraaza-1-phosphabicyclo[3,3,3]undecane	1	1	-1
Triisobutylphosphine	1	1	1

Bidentate PP and PN ligands	**PC1**	**PC2**	**PC3**
Racemic-2,2'-bis(diphenylphosphino)-1,1'-binaphthyl	0	0	0
(R,R)-Kelliphite	-1	-1	-1
2-diphenylphosphino-1-methyl-1H-imidazole	-1	-1	1
Bis(dimethylphosphino)methane	-1	1	-1
Biphephos	-1	1	1
Nixantphos	1	-1	-1
^{t}Bu-Xantphos	1	-1	1
1,2-Bis(dicyclohexylphosphino)ethane	1	1	-1
1,1'-Bis(di-i-propylphosphino)ferrocene	1	1	1

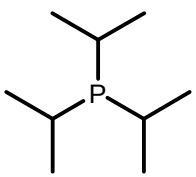
CatScI
Innovation in Catalysis

Coding for Nine Ligands

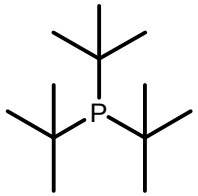


Position	PC1	PC2	PC3
0	0	0	0
1	-1	-1	-1
2	-1	-1	1
3	-1	1	-1
4	-1	1	1
5	1	-1	-1
6	1	-1	1
7	1	1	-1
8	1	1	1

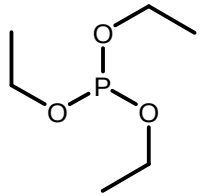
Monodentate Ligands



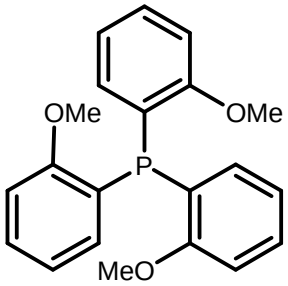
PL05



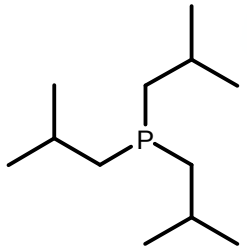
PL07



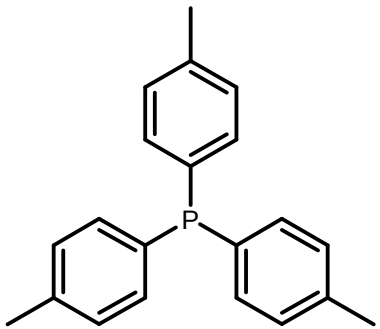
PL18



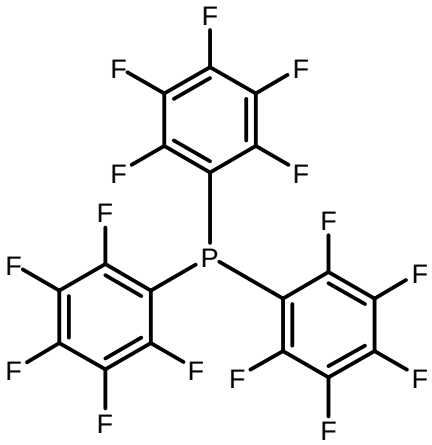
PL60



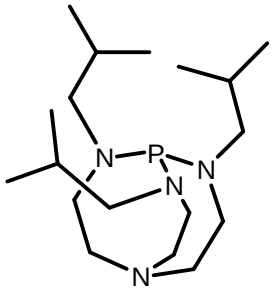
PL216



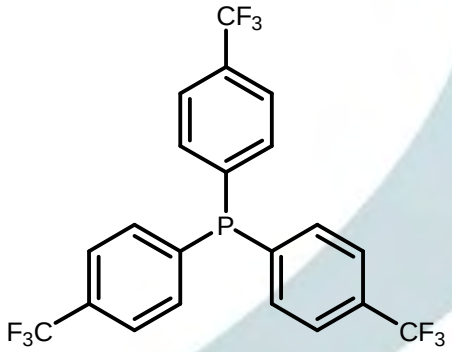
PL13



PL16



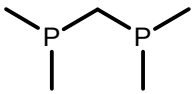
PL140



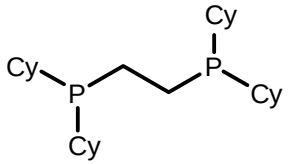
PL15

CatSci
Innovation in Catalysis

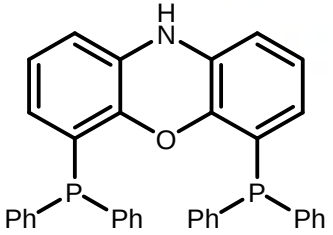
Bidentate Ligands



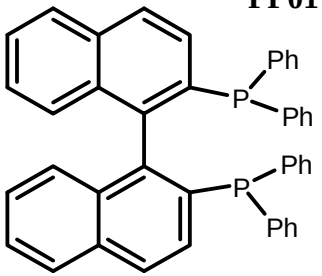
PP01



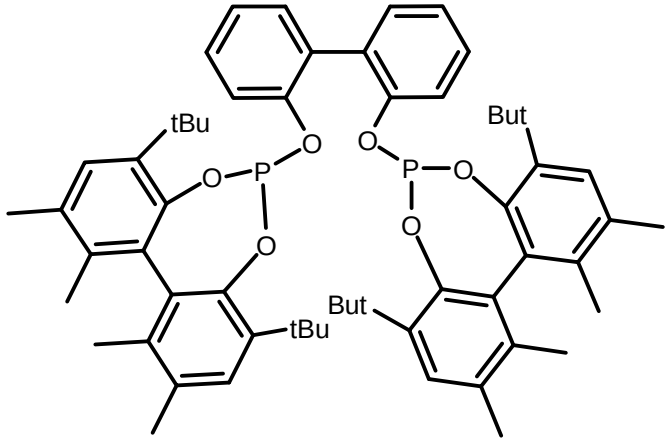
PP40



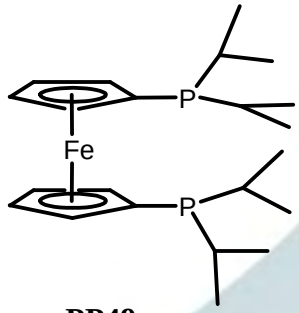
PP57



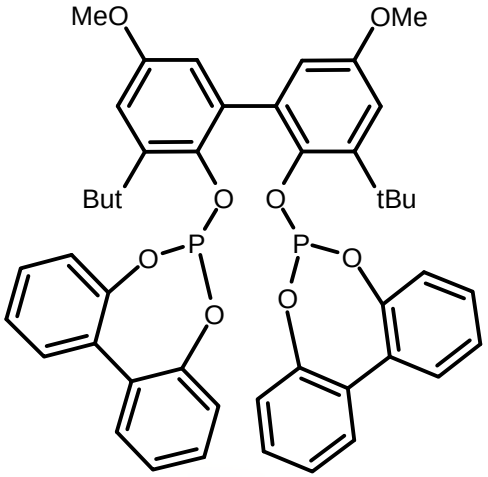
PP45



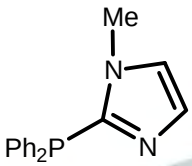
PP81



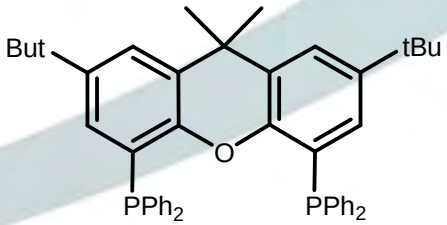
PP49



PP72



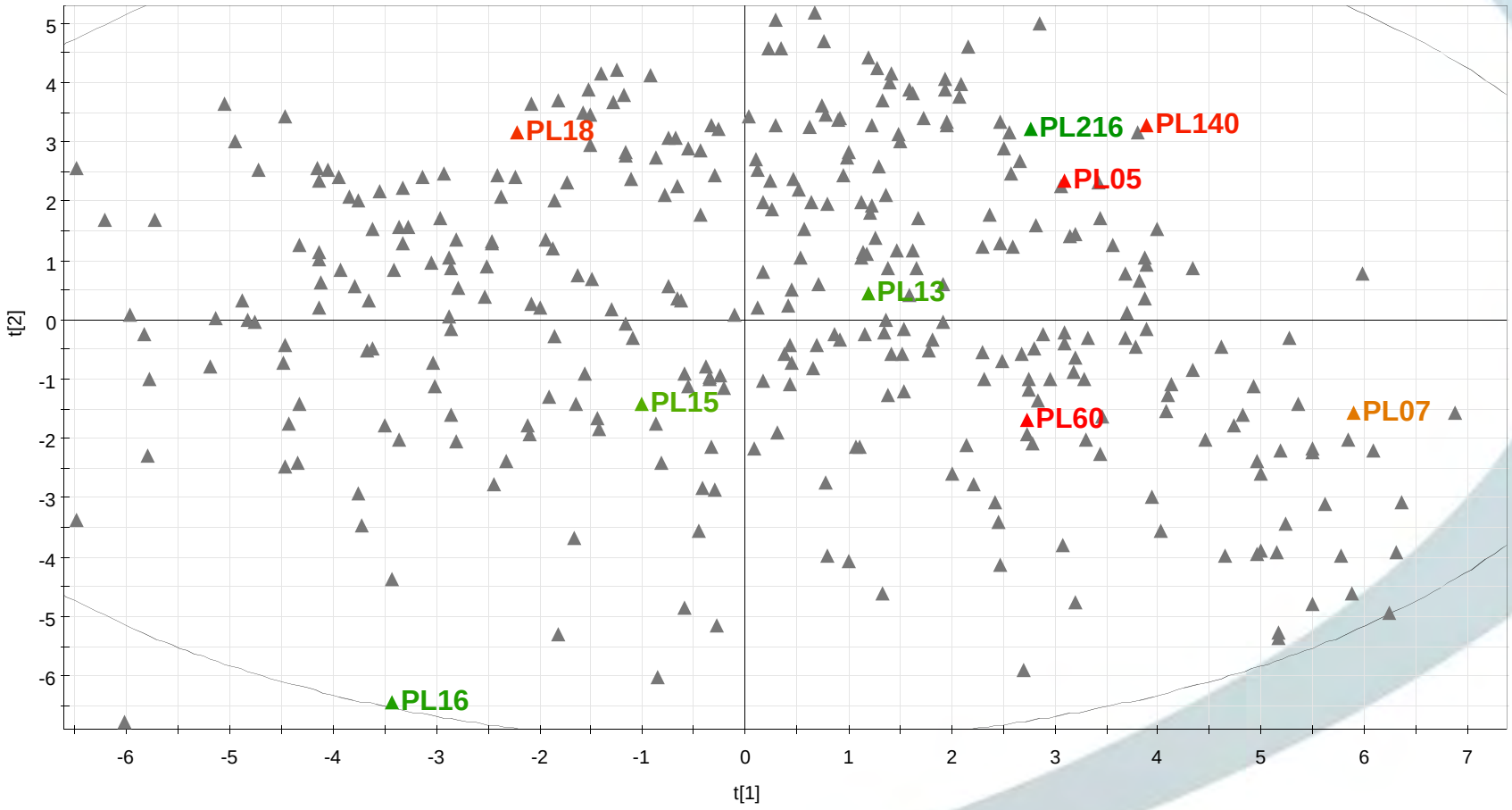
PN16



PP83

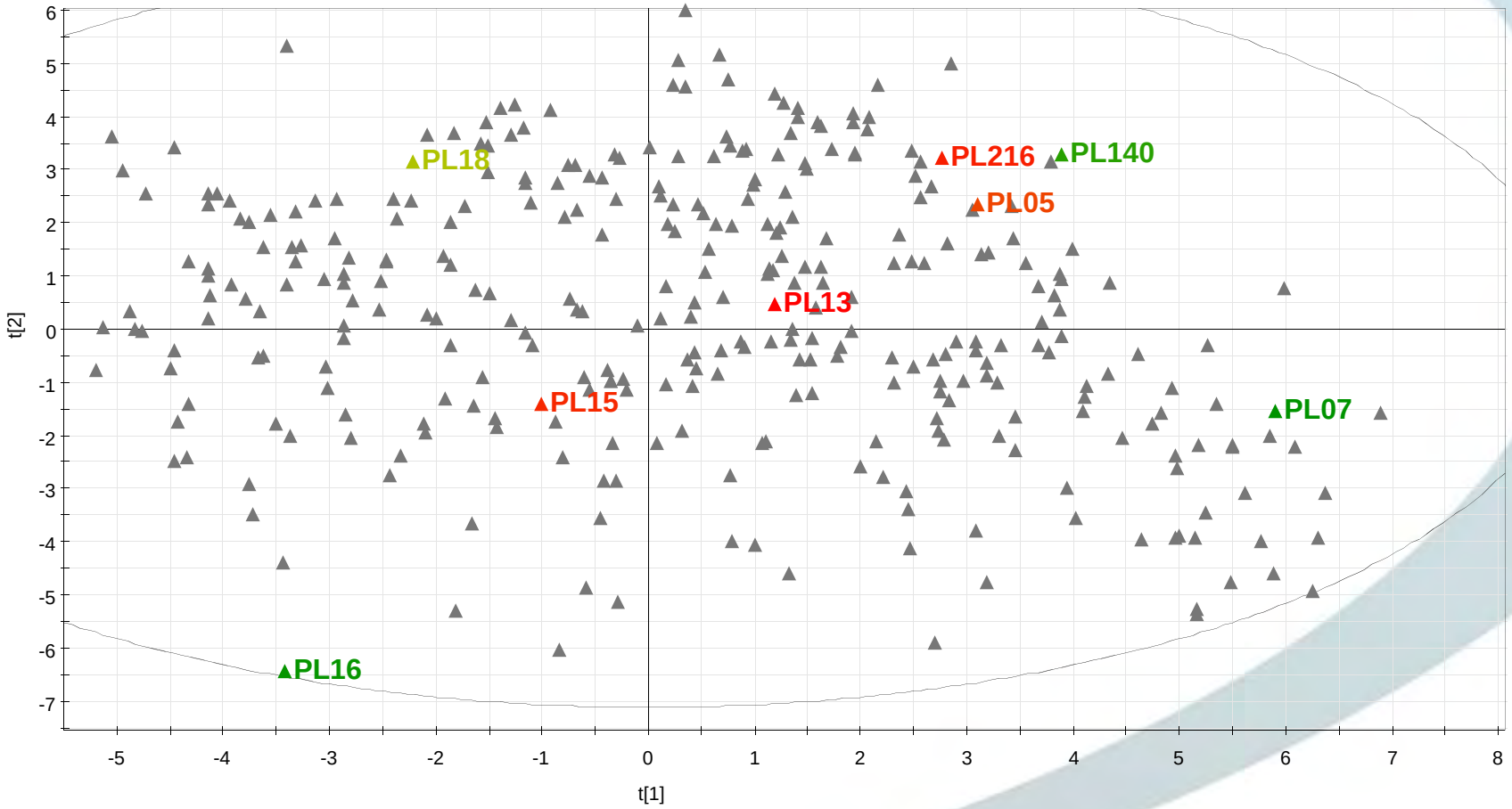
CatSci
Innovation in Catalysis

Ligand Activity for Ru



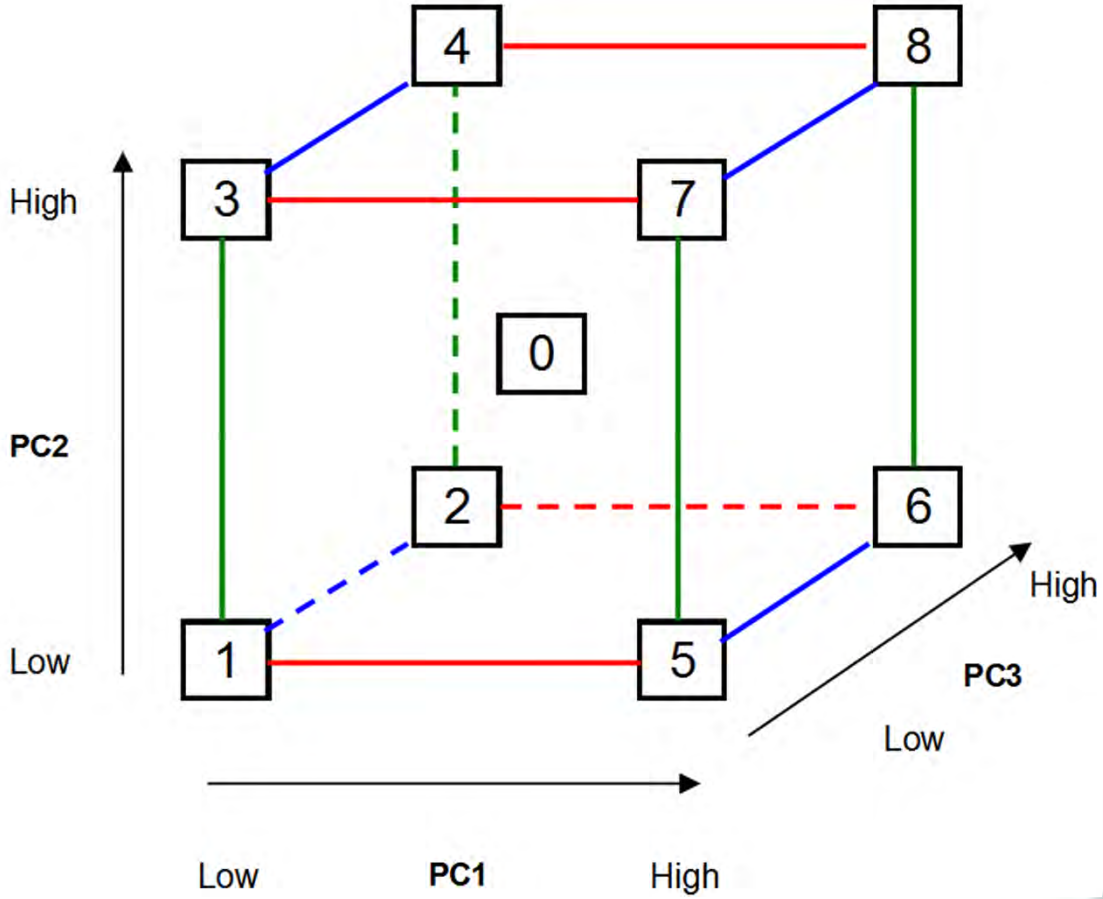
CatSci
Innovation in Catalysis

Ligand Activity for Ir



CatSci
Innovation in Catalysis

Coding for Nine Solvents



Position	PC1	PC2	PC3
0	0	0	0
1	-1	-1	-1
2	-1	-1	1
3	-1	1	-1
4	-1	1	1
5	1	-1	-1
6	1	-1	1
7	1	1	-1
8	1	1	1

Comparison of Tetralin and Toluene

Exp No	Ligand	Code	Catalyst	Additive	Solvent	Conv/9h
10	D^{i}PPF	PP049	Ir	None	Toluene	96.7
4	D^{i}PPF	PP049	Ir	None	Tetralin	77.1
9	D^{i}PPF	PP049	Ru	None	Toluene	98.7
3	D^{i}PPF	PP049	Ru	None	Tetralin	95.8
2	DPE-Phos	PP038	Ir	None	Tetralin	32.1
8	DPE-Phos	PP038	Ir	None	Toluene	31.1
7	DPE-Phos	PP038	Ru	None	Toluene	99.7
1	DPE-Phos	PP038	Ru	None	Tetralin	97.3
5	$^{i}Bu_3P$	PL216	Ru	None	Tetralin	100
11	$^{i}Bu_3P$	PL216	Ru	None	Toluene	100
6	$^{t}Bu_3P$	PL007	Ir	None	Tetralin	65.3
12	$^{t}Bu_3P$	PL007	Ir	None	Toluene	61.2

- Toluene and tetralin are both good solvents

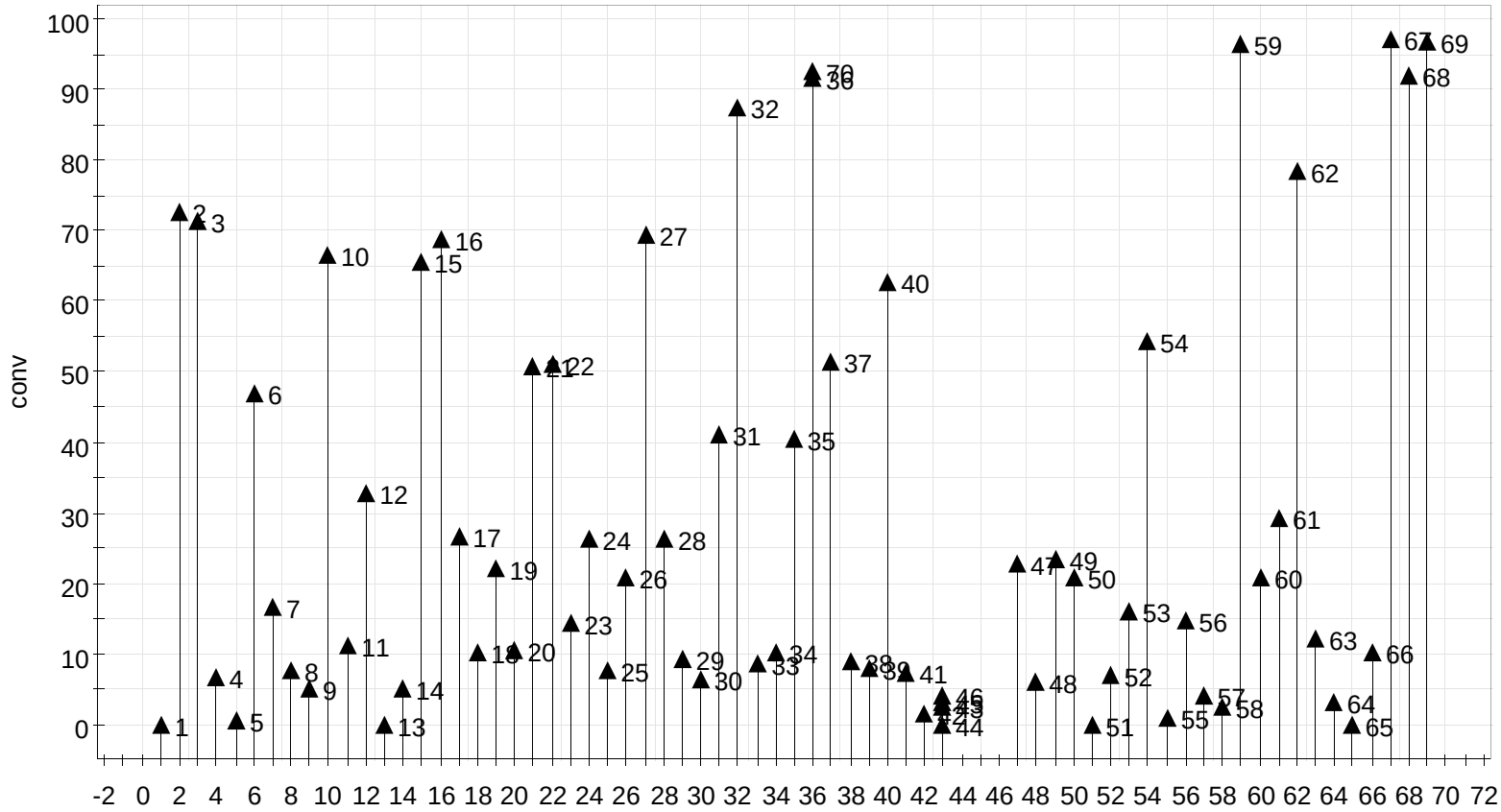
CatSci
Innovation in Catalysis

Monodentate Ligand Analysis

Using statistical tools

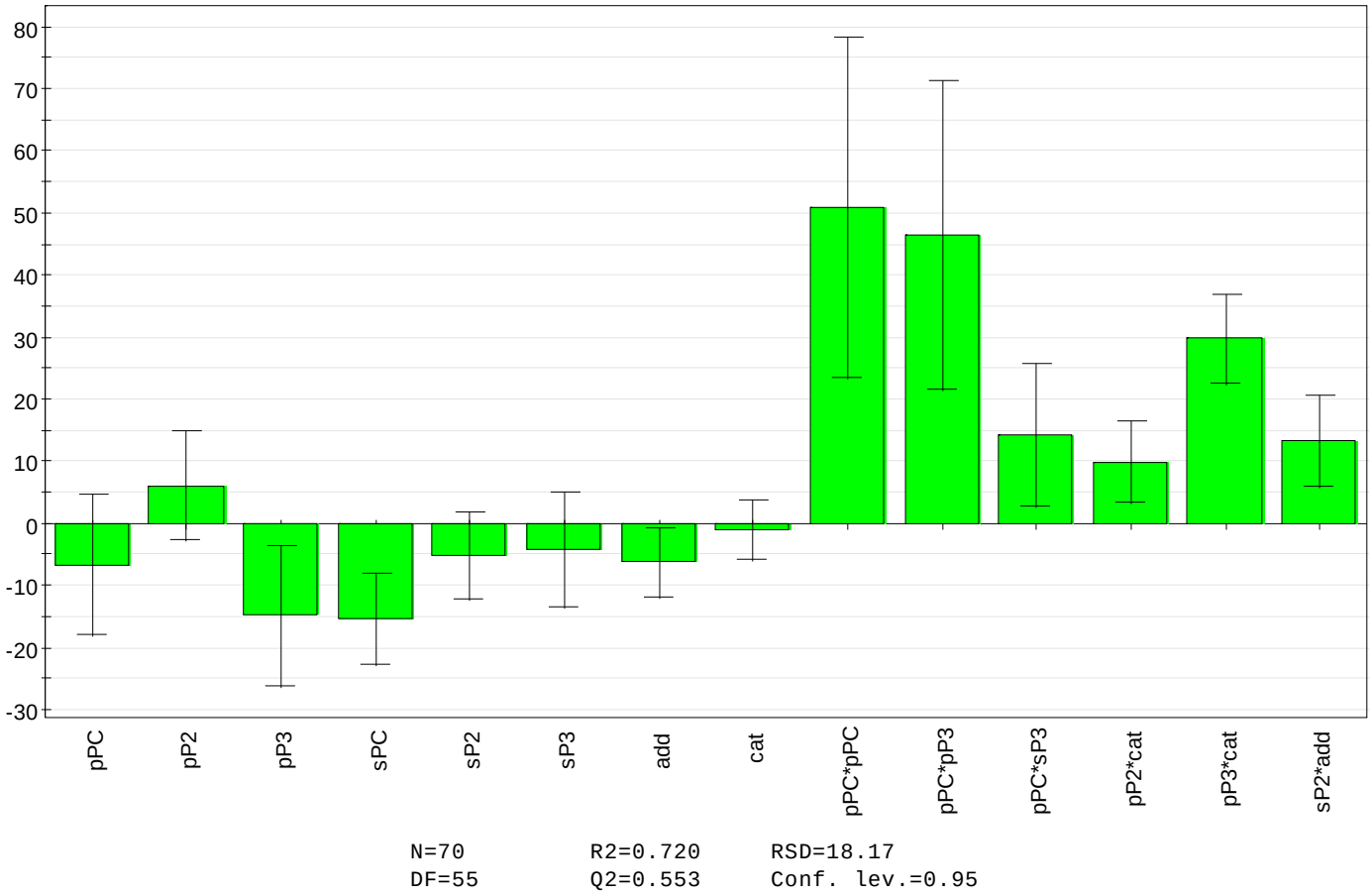
CatSci
Innovation in Catalysis

Replica Plot



CatSci
Innovation in Catalysis

Important Parameters

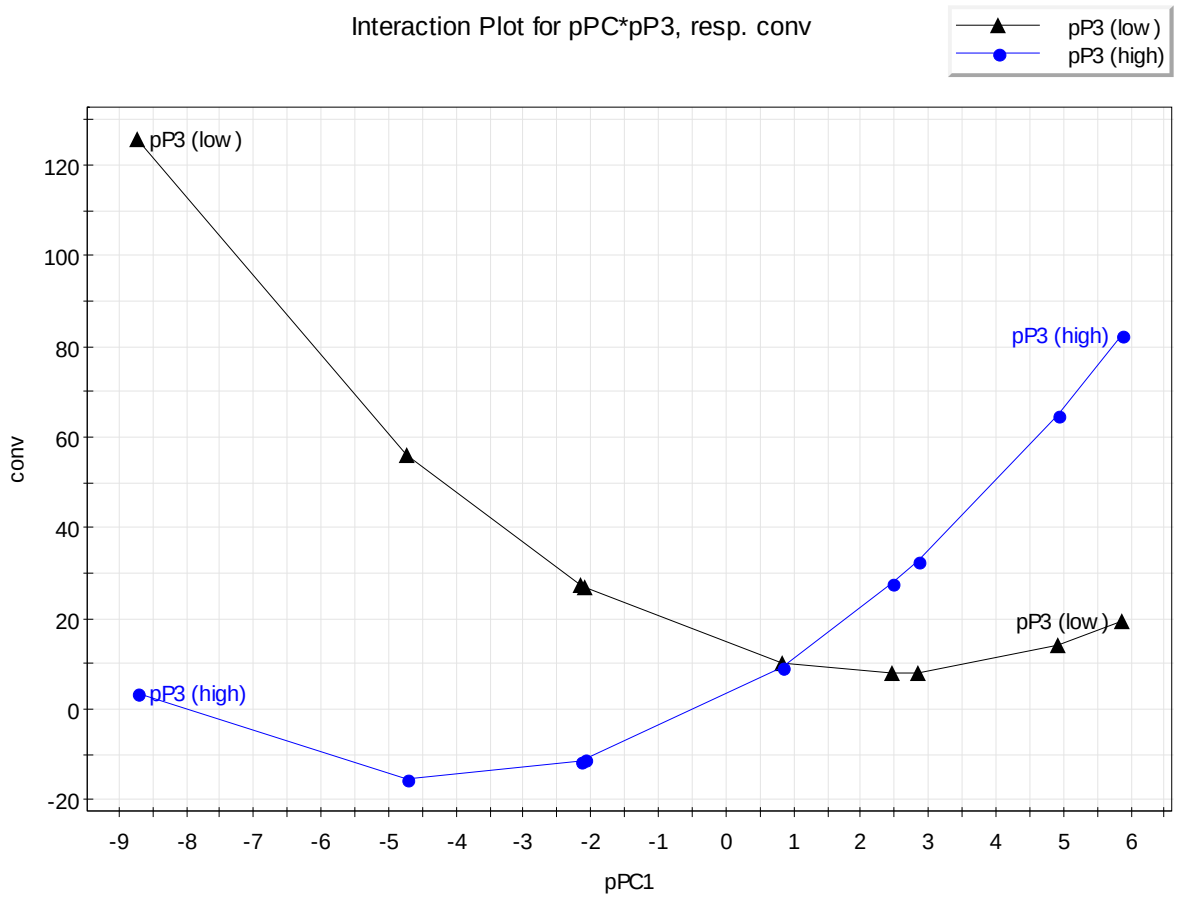


pPC	Phosphine PC1
pP2	Phosphine PC2
pP3	Phosphine PC3
sPC	Solvent PC1
sP2	Solvent PC2
sP3	Solvent PC3
Add	K_2CO_3,0, TFA
Cat	Ru, Ir

CatSci
Innovation in Catalysis

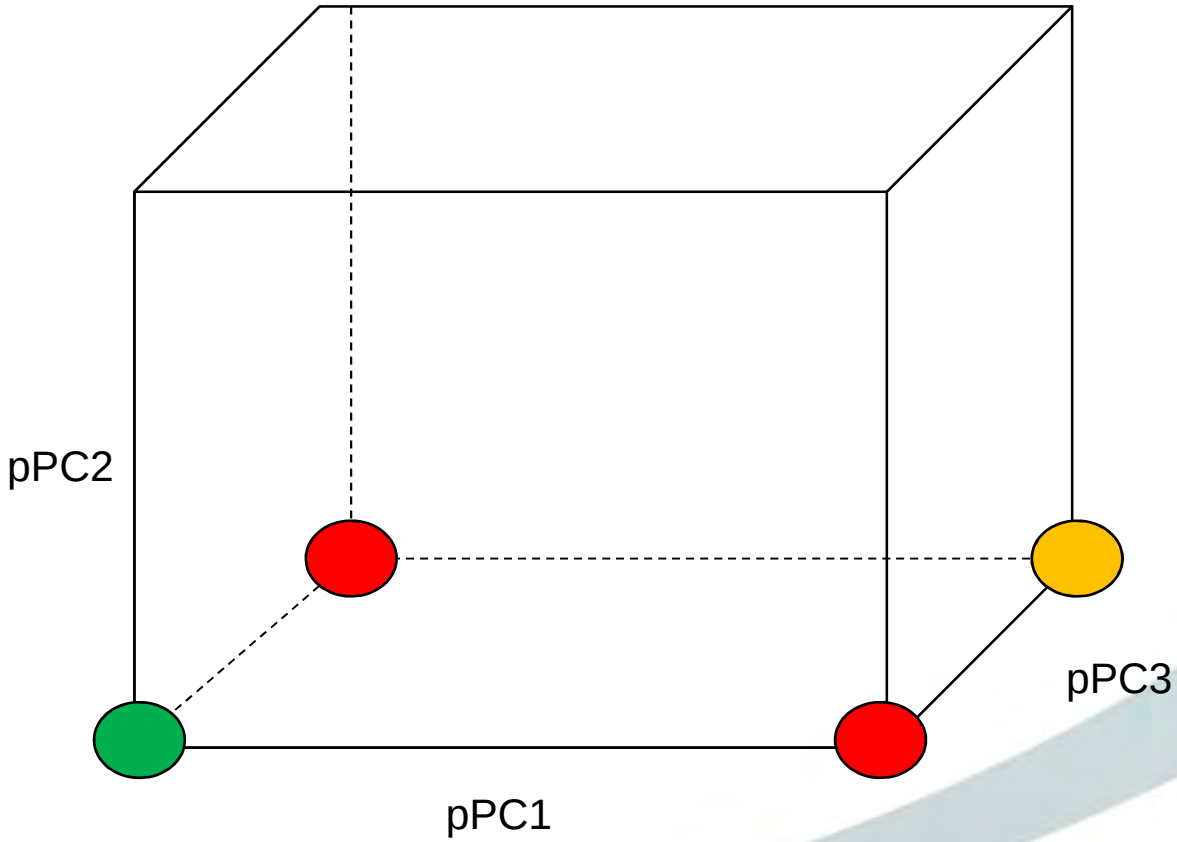
Interaction: Phosphine PC1 and PC3

LHS of monodentate ligand map is better (pPC1 = -1)



CatScI
Innovation in Catalysis

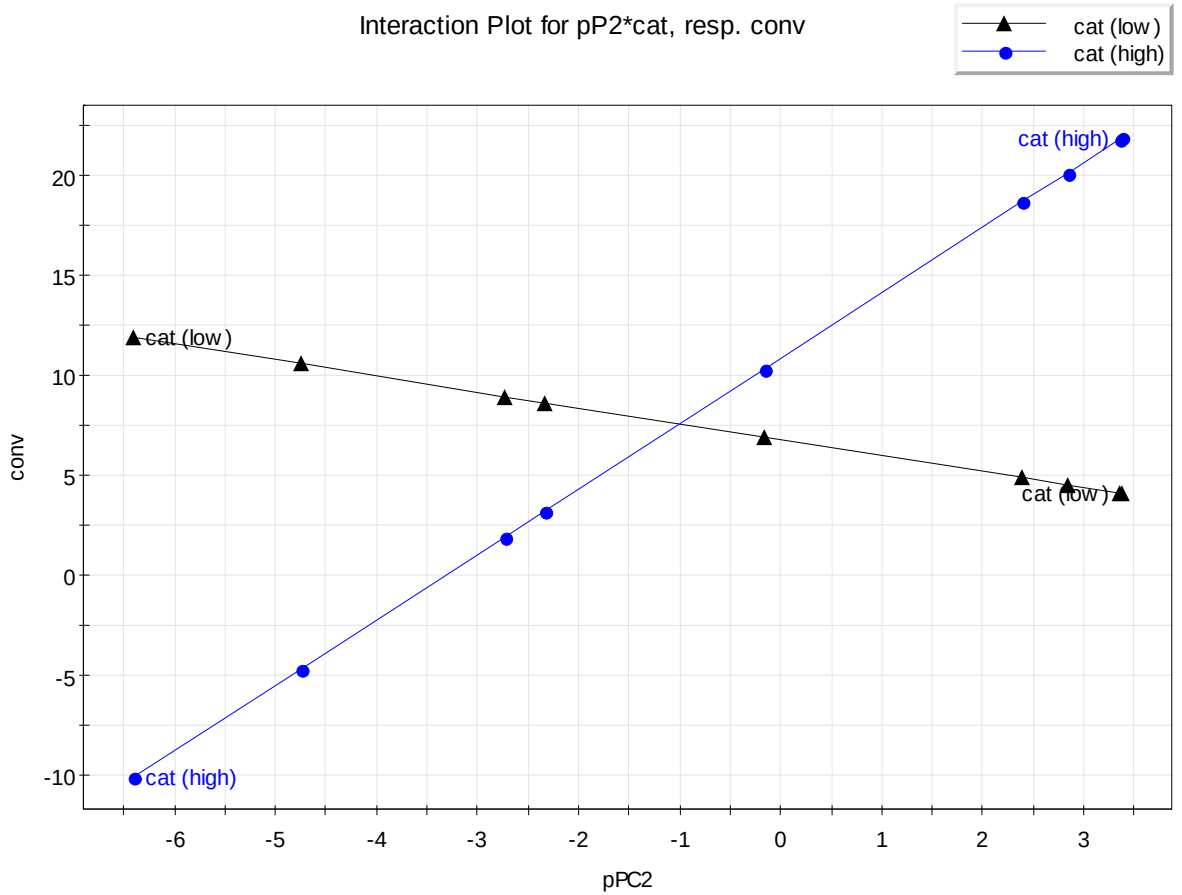
Phosphine Chemical Space: PC1.PC3



CatSci
Innovation in Catalysis

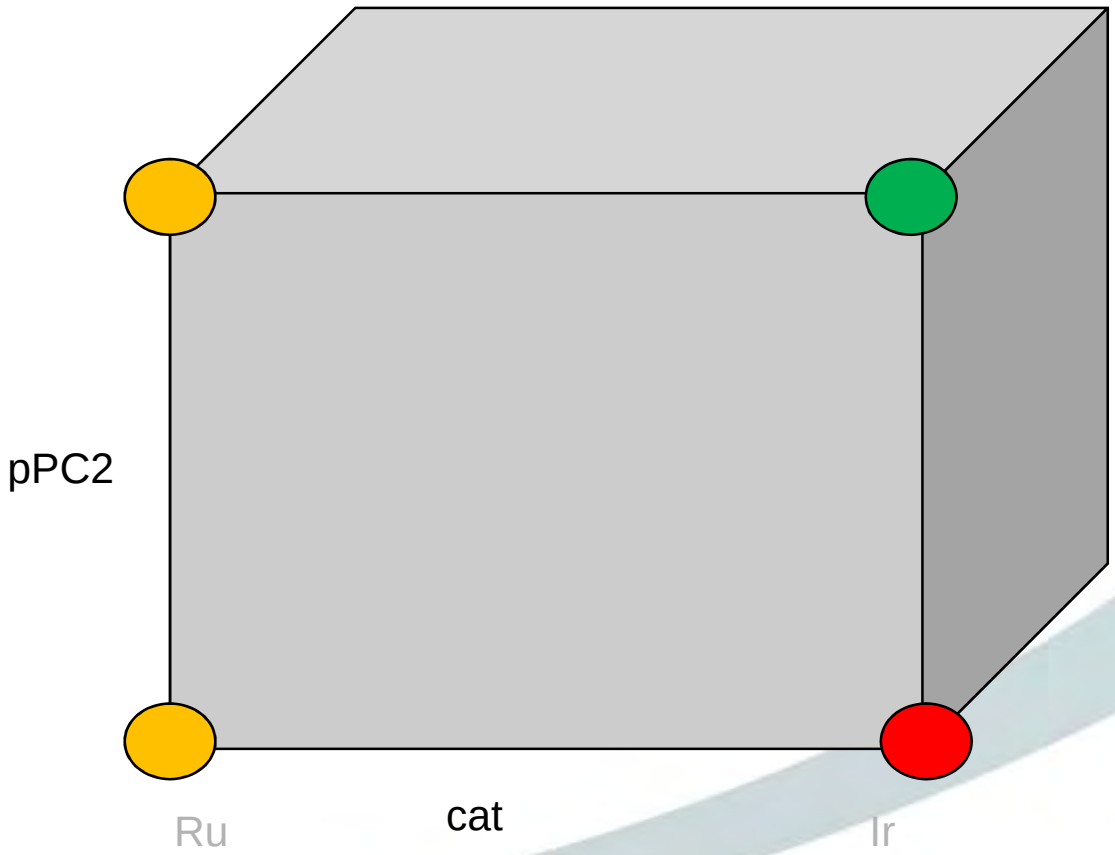
Interaction: Phosphine PC2 and Cat.

Ir-catalysed reactions sensitive to pPC2; Ru-catalysed reactions insensitive



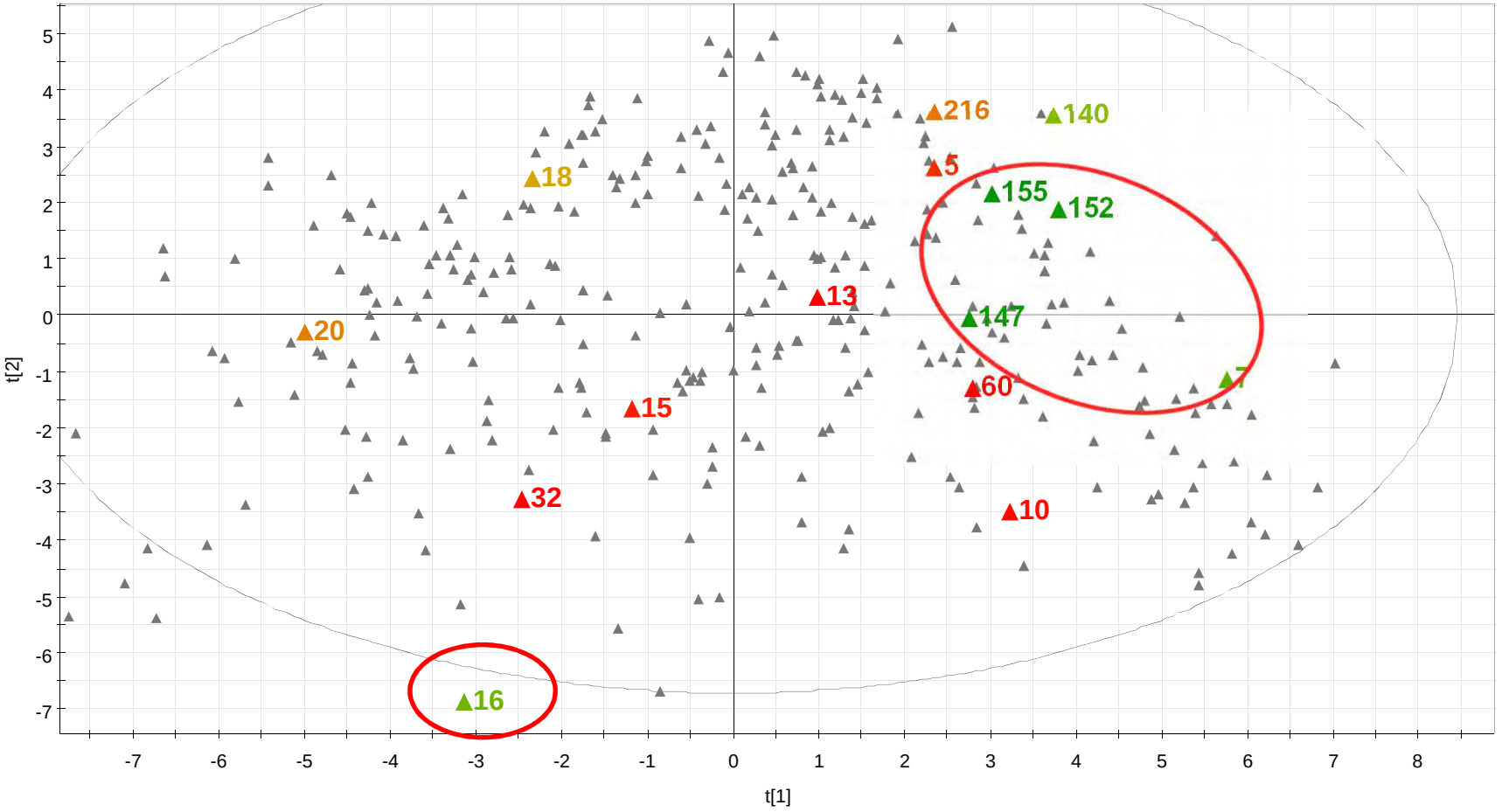
CatScI
Innovation in Catalysis

Phosphine Chemical Space: PC2.Cat



CatSci
Innovation in Catalysis

IrCp* with added ligand (results)



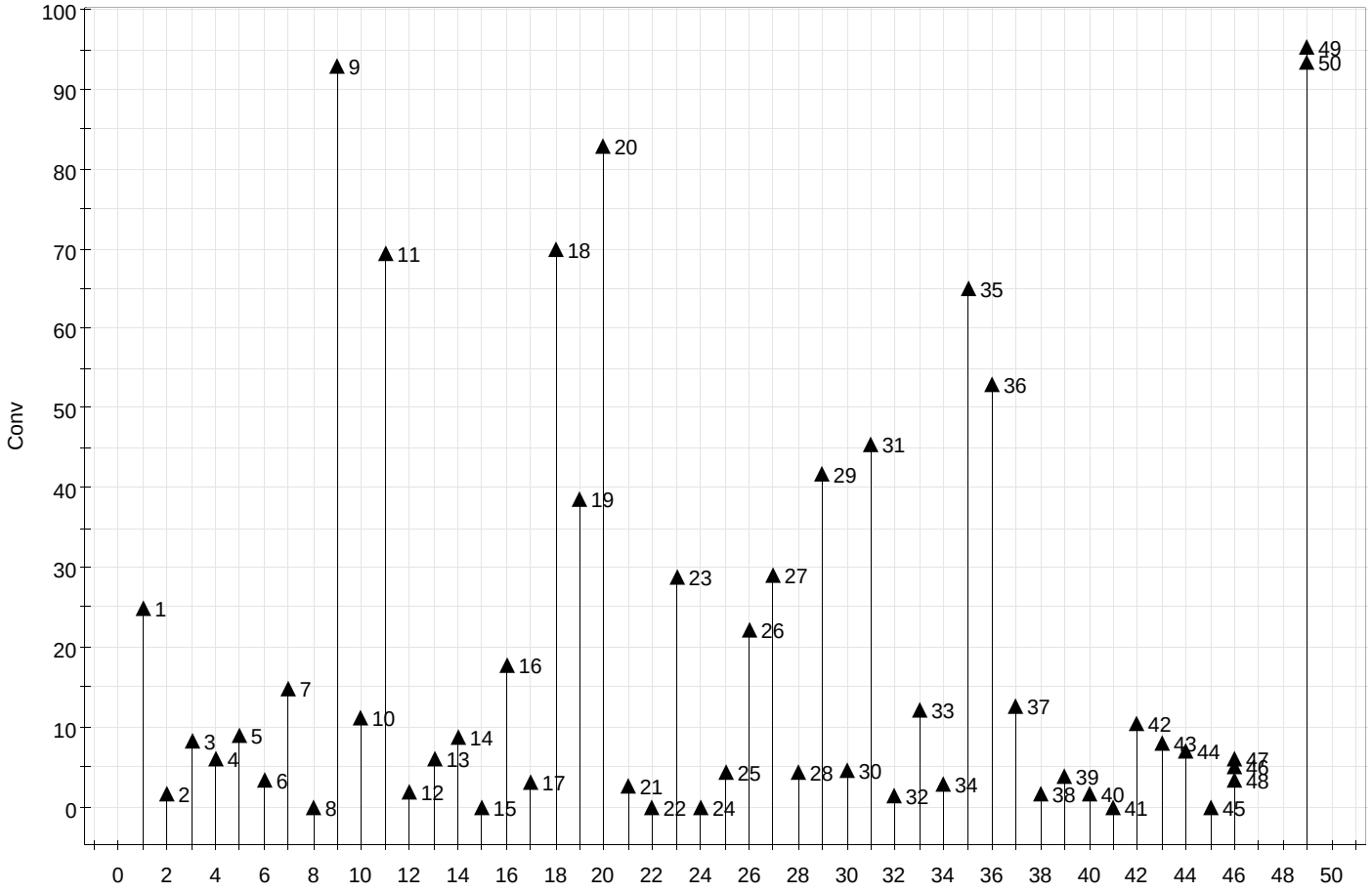
CatScI
Innovation in Catalysis

Bidentate Ligand Analysis

Using statistical tools

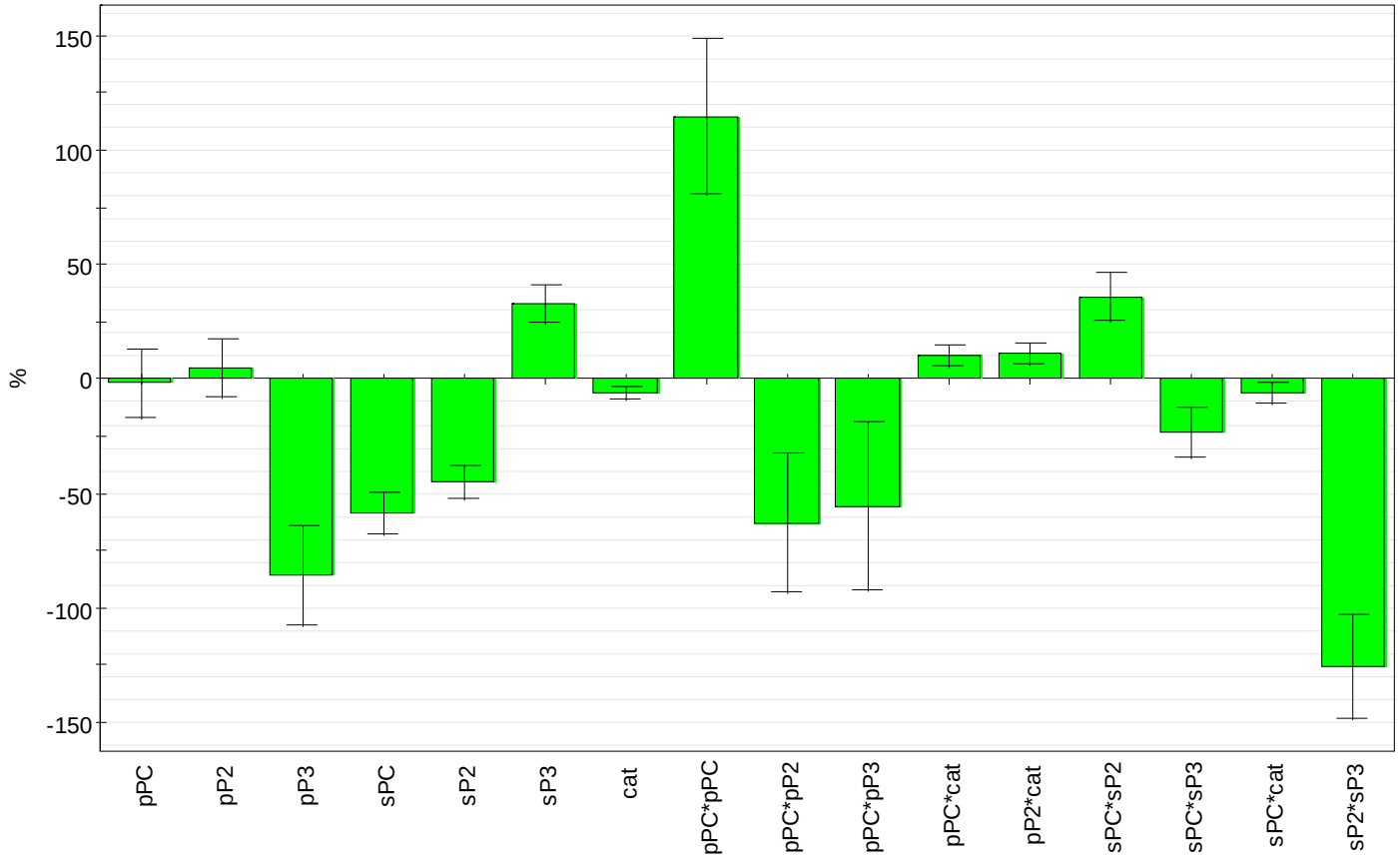
CatSci
Innovation in Catalysis

Replica Plot



CatSci
Innovation in Catalysis

Important Parameters



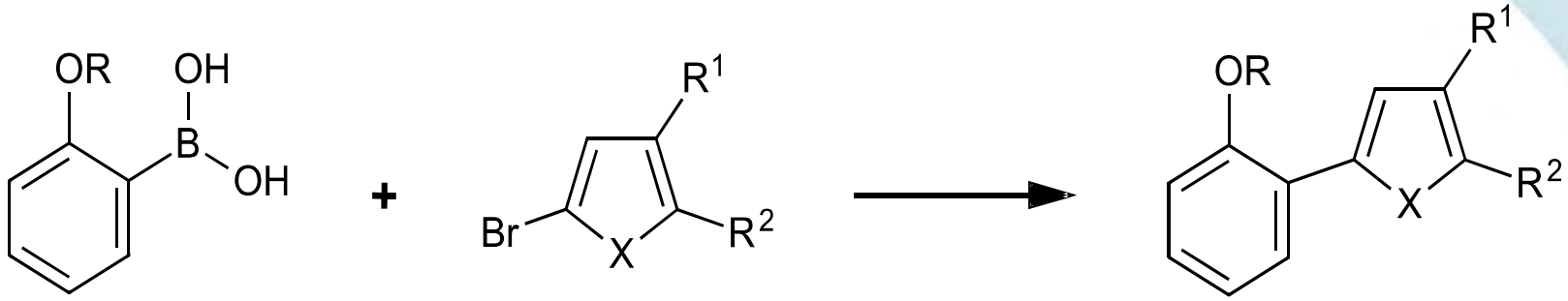
pPC	Phosphine PC1
pP2	Phosphine PC2
pP3	Phosphine PC3
sPC	Solvent PC1
sP2	Solvent PC2
sP3	Solvent PC3
Add	K_2CO_3,0, TFA
Cat	Ru, Ir

CatSci
Innovation in Catalysis

Further Case Studies in Brief

CatSci
Innovation in Catalysis

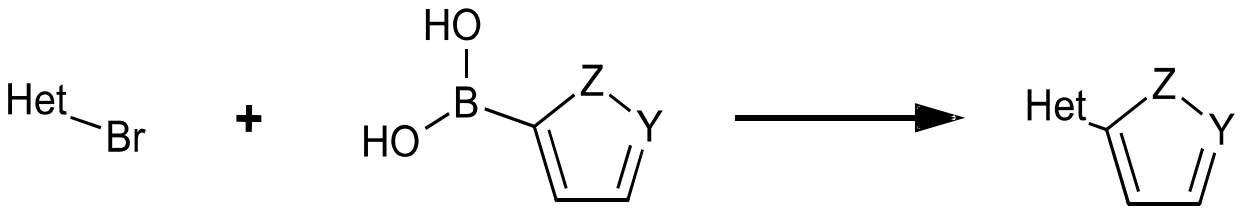
Case Study 1: Enabling Technology



- **Background**: a low-yielding, early phase Suzuki reaction with high catalyst loading
- **Aim**: find an active catalyst and investigate key parameters
- **Result**: a single design delivered a fit-for-purpose process *in only 2 weeks*

CatScI
Innovation in Catalysis

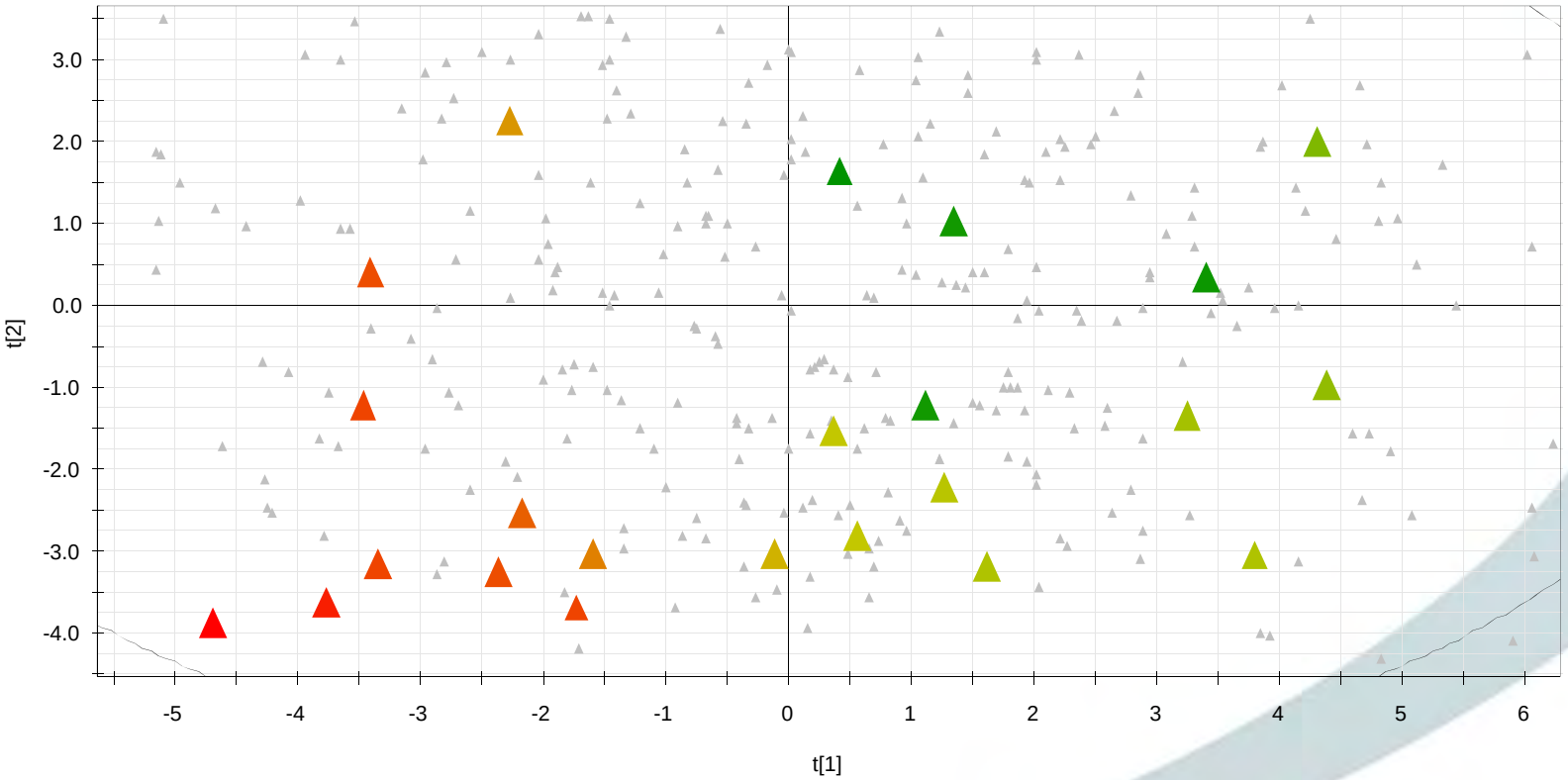
Case Study 2: Solvent Selection



- **Background**: a challenging non-robust Suzuki reaction which necessitated a high catalyst loading
- **Aim**: develop a robust reaction and reduce the catalyst loading
- **Result**: reduced catalyst loading combined with a higher yield saved $250,000 per 60 kg delivery

CatSci
Innovation in Catalysis

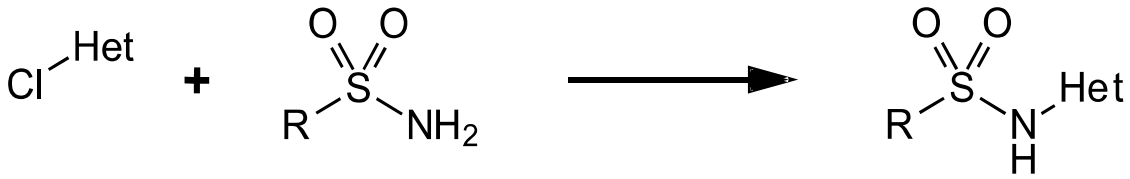
Case Study 2: Solvent Selection



Our rational approach makes sense of solvent space

CatScI
Innovation in Catalysis

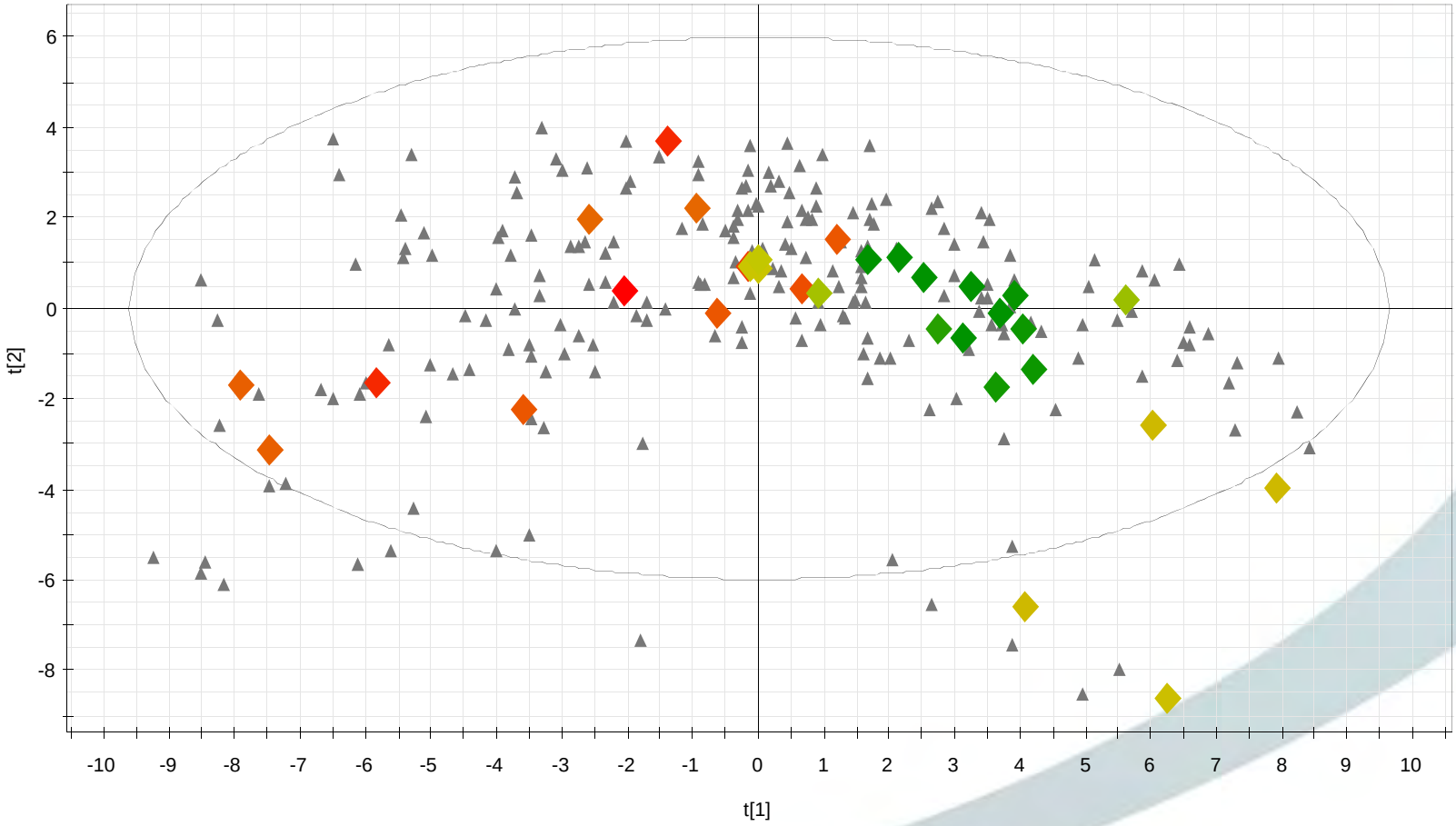
Case Study 4: Predictive Catalysis



- **Background**: a late phase Buchwald-Hartwig sulfamidation requiring an expensive ligand
- **Aim**: investigate alternative ligands, Pd sources and solvents to explore thoroughly the reaction space
- **Result**: a cheaper IP-free ligand was found saving €190,000 per 80 kg delivery without compromising other factors

CatSci
Innovation in Catalysis

Case Study 4: Predictive Catalysis

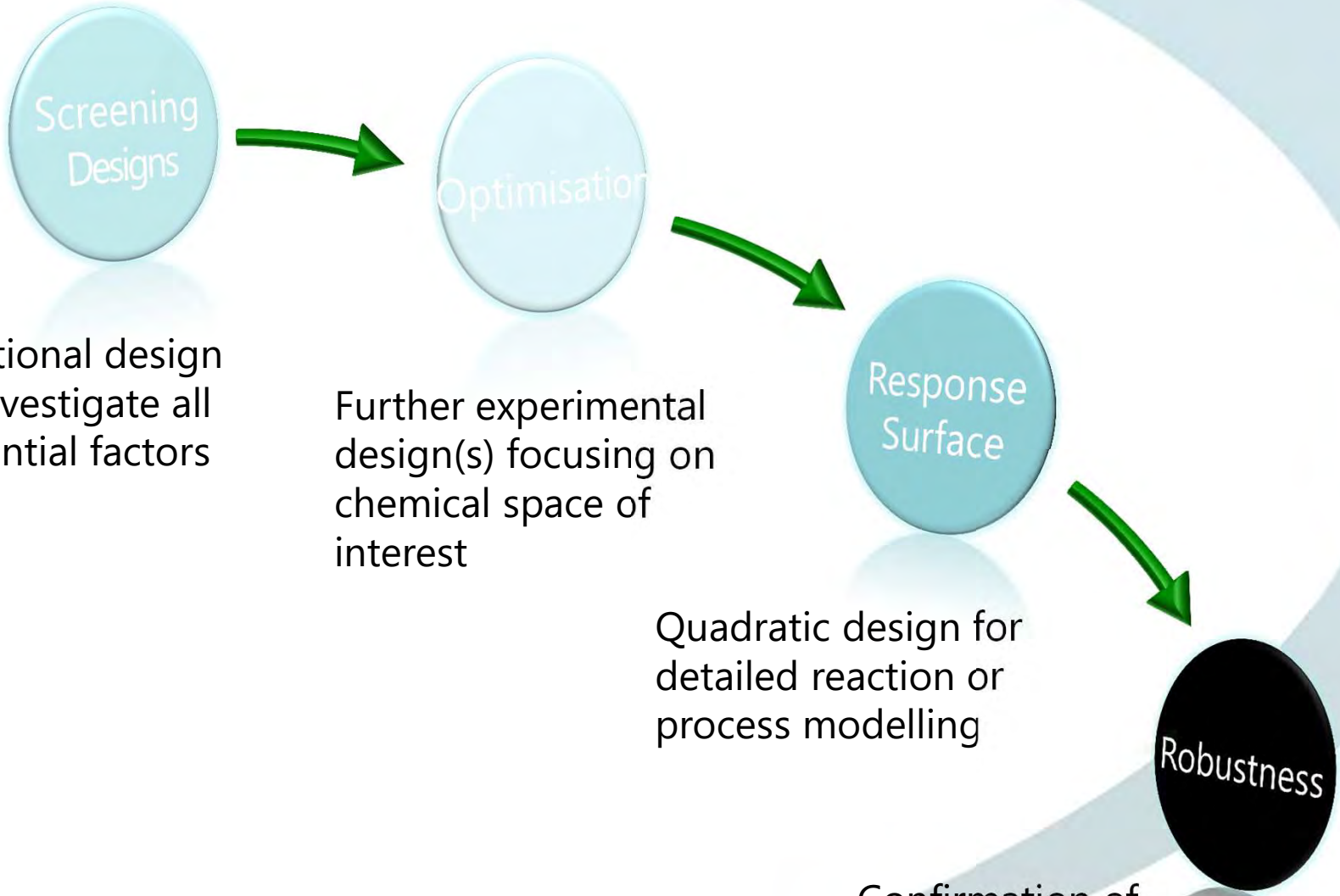


CatScI's unique ligand mapping capability exemplified

CatScI
Innovation in Catalysis

Conclusions

CatSci
Innovation in Catalysis



CatScI
Innovation in Catalysis

Summary

- DoE is a powerful technique for maximising data from a minimal number of reactions
  - can be applied to many chemistries
  - especially valuable for TM-catalysed reactions
- PCA provides a method to quantify discrete factors as continuous factors
- DoE training available from CatScI
- Let us optimise your catalysis!

Acknowledgements

- Paul Murray (CSO)

- Ross Burn
- Ed Turp
- Simon Tyler

- www.catsci.com

Org. Process Res. Dev., **2013**, *17*, 40-46

CatSci
Innovation in Catalysis