Multivariate Design of Experiments for Engineering Dimensional Analysis

Chris Nachtsheim
Carlson School of Management and
School of Statistics
University of Minnesota

Daniel J Eck
Department of Statistics
University of Illinois

Thomas A Albrecht
Atlassian

FTC Webinar Series, November 3, 2021

Joint Work With:

Dennis Cook: University of Minnesota

Outline of Talk:

1. **Basic ideas of DA**
2. **Address dimensional analysis for multivariate responses**
3. **Examples**
4. **A glimpse at optimal design techniques for this problem**
5. **How we got involved in dimensional analysis**
6. **Conclusions**

Statistics vs. Engineers: A simple example

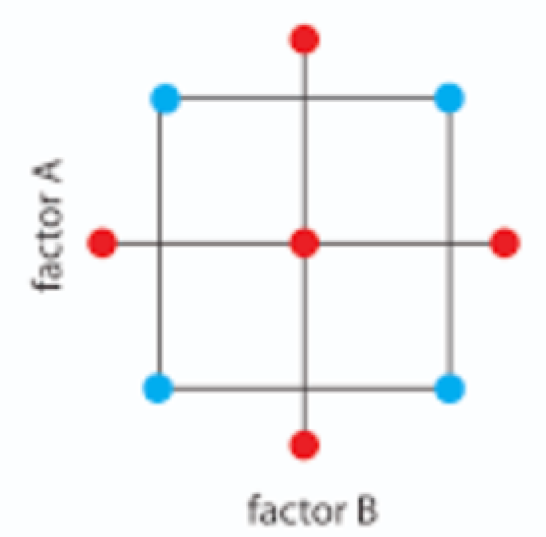
- **We wish to characterize the effects of velocity (X1) and time (X2) on the response variable distance, D.**
- **Now, we're ignorant, but an oracle knows that the true relationship is:**

$$D = V \times T + \varepsilon$$

- **How do we figure this out?**

Statistician's approach

- **DOE!!!** **We're concerned about curvatures and interactions, so run a central composite response surface design!**



- **After running the experiment, fit the second-order model:**

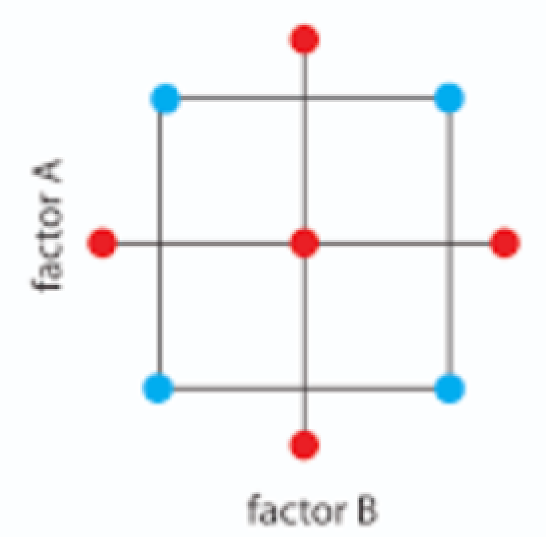
$$\mathrm{E}(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2$$

- **Drop out non-significant terms,**

$$\hat{Y} = 1.04 X_1 X_2 \approx V \times T$$

Statistician's approach

- **DOE! We're concerned about curvatures and interactions, so run a central composite response surface design!**



- **After running the experiment, fit the second-order model:**

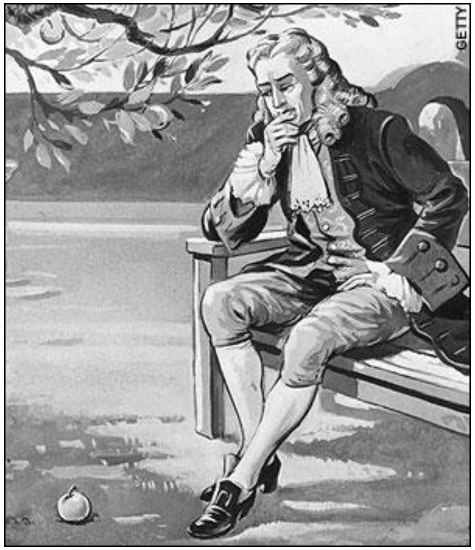
$$\mathrm{E}(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2$$

- **Drop out non-significant terms,**

$$\hat{Y} = 1.04 X_1 X_2 \approx V \times T$$

Engineer’s approach

- **The Engineer thinks more like Isaac Newton in that the mechanics of a physical systems are governed by physical laws**
  - ***A physical equation must be dimensionally homogenous***



$$D = h(V,T)$$

[D]=Miles

[h(V,T)]=Miles

7

Engineer's approach

- **Dimensions (Units) of the variables:**
  - **[D]=Miles**
  - **[V]=Miles/Hour**
  - **[T]=Hour**
- **One dimensionally homogeneous possibility:**

Dimensionally Homogenous → $D = cVT$

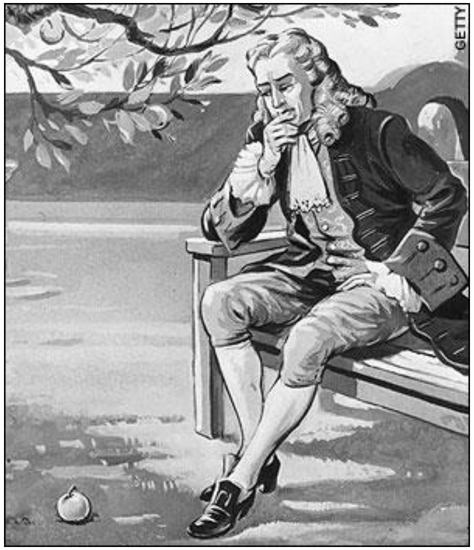
8

Engineer's approach

- **One Final Simplification: We make the equation dimensionless by dividing both sides by VT.**

Dimensionless Representation

$$\frac{D}{VT} = c$$



- **To estimate *c*, all we need to do is pick a velocity, V, pick a time, T, observe D and plug in (replicate as needed).**

Statistical vs. Engineering Approaches

	Statistician's Approach	Engineer's Approach
Factors to vary in experiment	Two (Time, Distance)	Zero (one fixed constant)
Model parameters	6 parameters	One fixed constant
Scalability	Model valid inside experimental range of factors	Will "scale" over any experimental range
Explanatory Power	Local, empirical model	Reveals a "universal law" between D,V, & T for physical systems

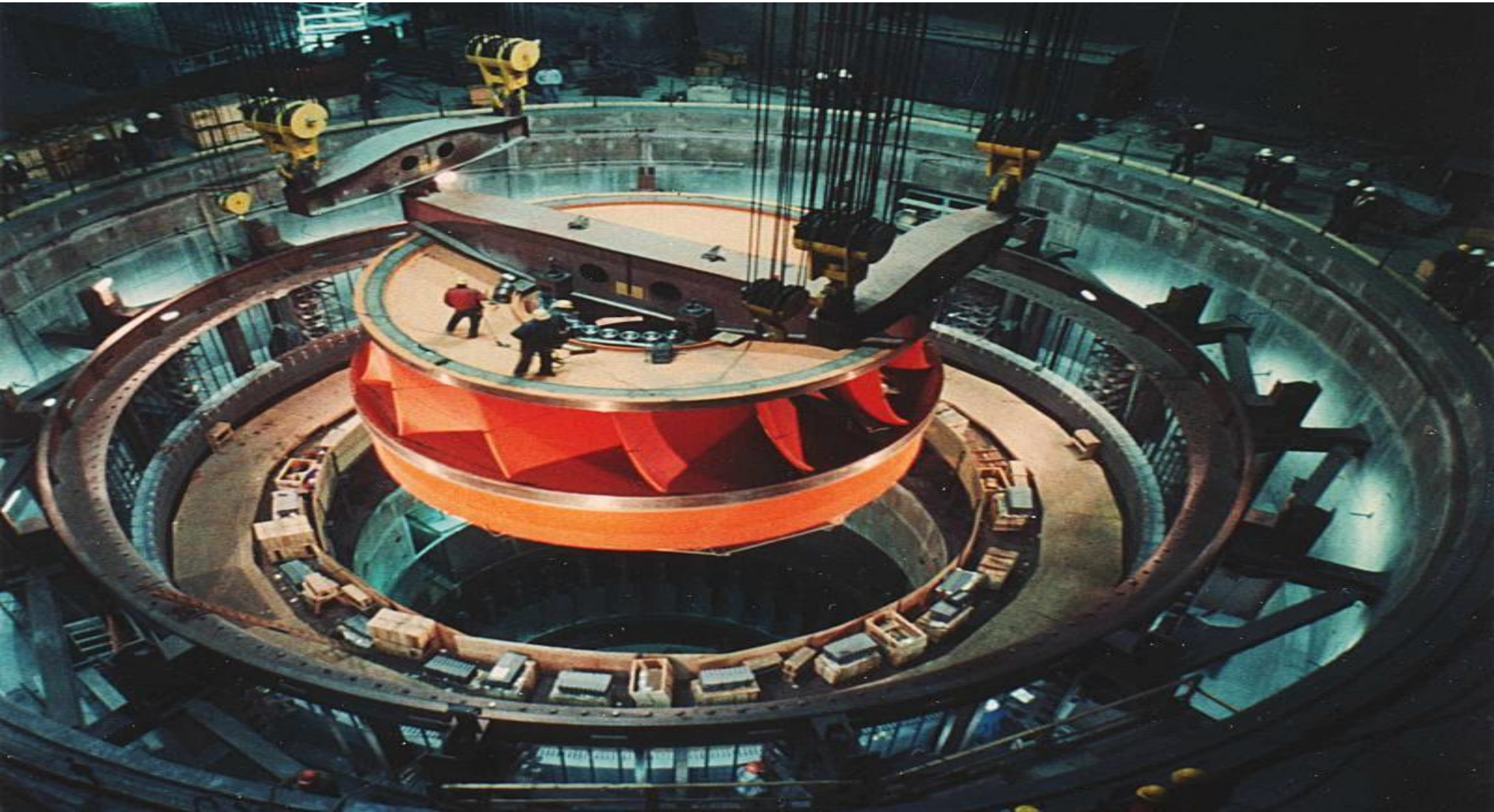
Dimensional Analysis (DA) Advantages

1. **Dimension reduction** ✔
   - **The number of factors is reduced by the number of measurement dimensions in the variables**
   - **Our example: Number of factors = 2. Number of dimensions is 2 (length, time). Resulting number of dimensions is zero!**
2. **Scalability** ✔
   - **Empirical models are valid within the ranges of the factors**
   - **Dimensionless models scale to any size**

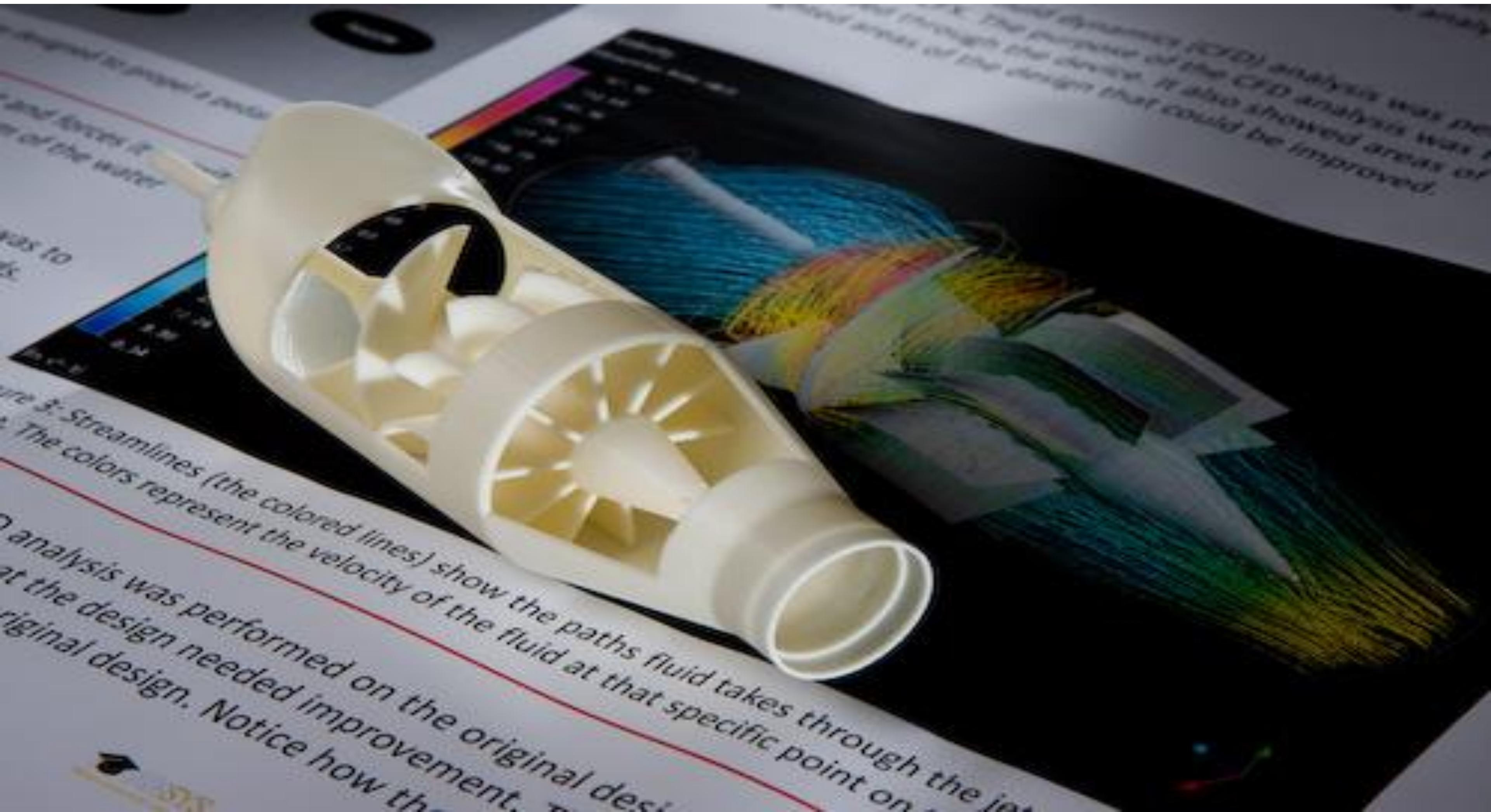
11

Why is scalability important?

Running experiments on a turbine of this size is impossible



Running experiments on a model of this size is easy. If all variables are dimensionless, we can extrapolate!



An even more compelling type of extrapolation…

15


CONCEPT VIGNETTES
A Collaboration between SUTD and MIT
YouTube

Mars Rover DA Process

$$v = f(d, m, g, \rho)$$

$v =$ velocity (dependent variable)	$[v] = L/T$
$d =$ diameter of parachute	$[d] = L$
$m =$ mass of the rover	$[m] = M$
$g =$ gravitational constant	$[g] = L/T^2$
$\rho =$ density of the atmosphere	$[\rho] = M/L^3$

Ipsen (1960) Stepwise Derivation of DA Model

- **Step-by-step approach that leads to dimensionless DA model**
- **At each step, one variable is used to eliminate a dimension (e.g., M, L, T) from the set of variables.**
- **The variable used is eliminated**
- **At termination, a reduced set of dimensionless variables is created**

Step 0: Initialize the variables by specifying dimensions

Step 0: Initialize	
Variable	Dimension
v	LT^{-1}
d	L
m	M
g	LT^{-2}
ρ	ML^{-3}

Step 1: Use d to eliminate length (L)

Step 0: Initialize		Step 1 Result: Remove L from Step 0 using d	
Variable	Dimension	Var.	Dim.
v	LT^{-1}	vd^{-1}	T^{-1}
d	L	———	———
m	M	m	M
g	LT^{-2}	gd^{-1}	T^{-2}
ρ	ML^{-3}	ρd^3	M

Step 2: Use m (rover mass) to eliminate mass (M)

Step 1 Result: Remove L from Step 0 using d		Step 2 Result: Remove M from Step 1 using m	
Var.	Dim.	Var.	Dim.
vd^{-1}	T^{-1}	vd^{-1}	T^{-1}
___	___	___	___
m	M	___	___
gd^{-1}	T^{-2}	gd^{-1}	T^{-2}
ρd^3	M	$\rho d^3 m^{-1}$	1

Step 3: Use gd^{-1} to eliminate time (T)

Step 2 Result: Remove M from Step 1 using m		Step 3 Result: Remove T from Step 2 using gd^{-1}	
Var.	Dim.	Var.	Dim.
vd^{-1}	T^{-1}	$v/\sqrt{dg}$	1
________	________	________	________
________	________	________	________
gd^{-1}	T^{-2}	________	________
$\rho d^3 m^{-1}$	1	$\rho d^3 m^{-1}$	1

Result

- **DA model:**

$$\frac{v}{\sqrt{dg}} = \phi\left(\frac{\rho d^3}{m}\right)$$

- **DA terminology: two dimensionless variables or “pi groups”**

$$\pi_0 = \phi(\pi_1)$$

Summary

Step 0: Initialize		Step 1 Result: Remove L from Step 0 using d		Step 2 Result: Remove M from Step 1 using m		Step 3 Result: Remove T from Step 2 using gd^{-1}	
Variable	Dimension	Var.	Dim.	Var.	Dim.	Var.	Dim.
V	LT^{-1}	Vd^{-1}	T^{-1}	Vd^{-1}	T^{-1}	$V/\sqrt{dg}$	1
d	L	—	—	—	—	—	—
m	M	m	M	—	—	—	—
g	LT^{-2}	gd^{-1}	T^{-2}	gd^{-1}	T^{-2}	—	—
ρ	ML^{-3}	ρd^3	M	$\rho d^3 m^{-1}$	1	$\rho d^3 m^{-1}$	1

$$\frac{v}{\sqrt{dg}} = \phi\left(\frac{\rho d^3}{m}\right)$$

So we have a one-factor DA experiment

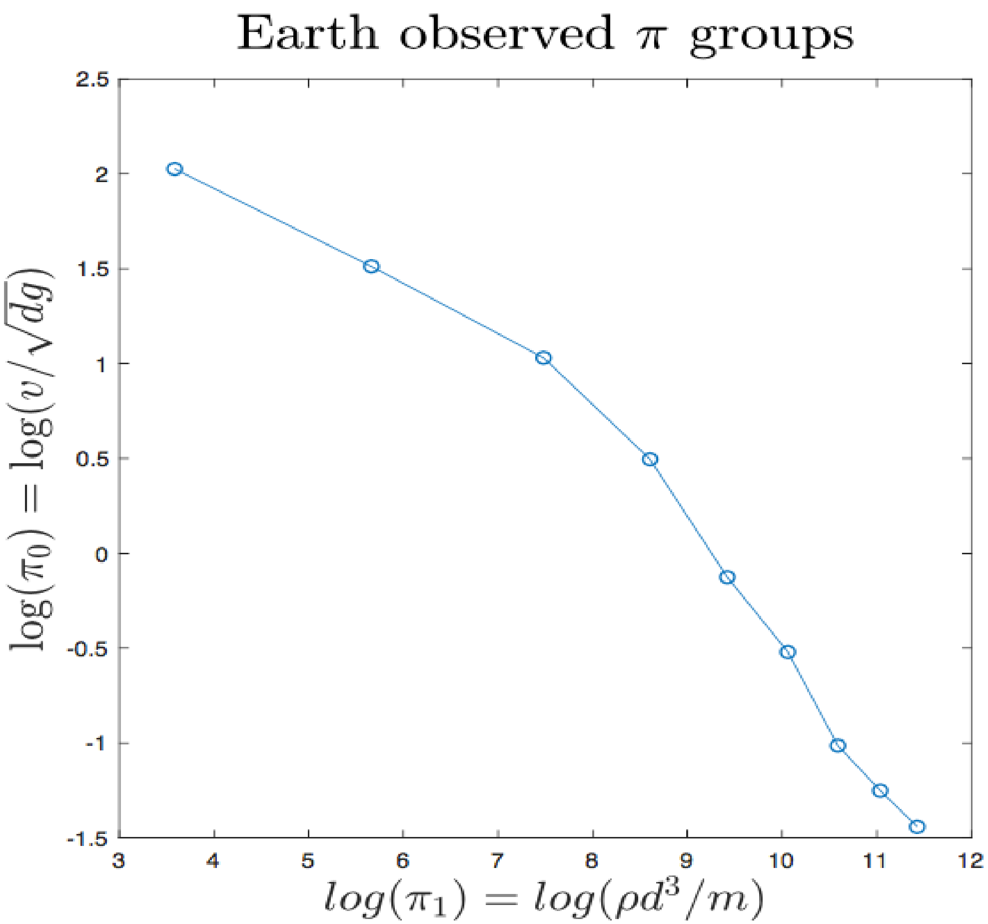
- **Model:** $$\pi_0 = \phi(\pi_1)$$

- **Objective: Estimate ϕ by varying the single factor π_1:**

$$\pi_1 = \frac{\rho d^3}{m}$$

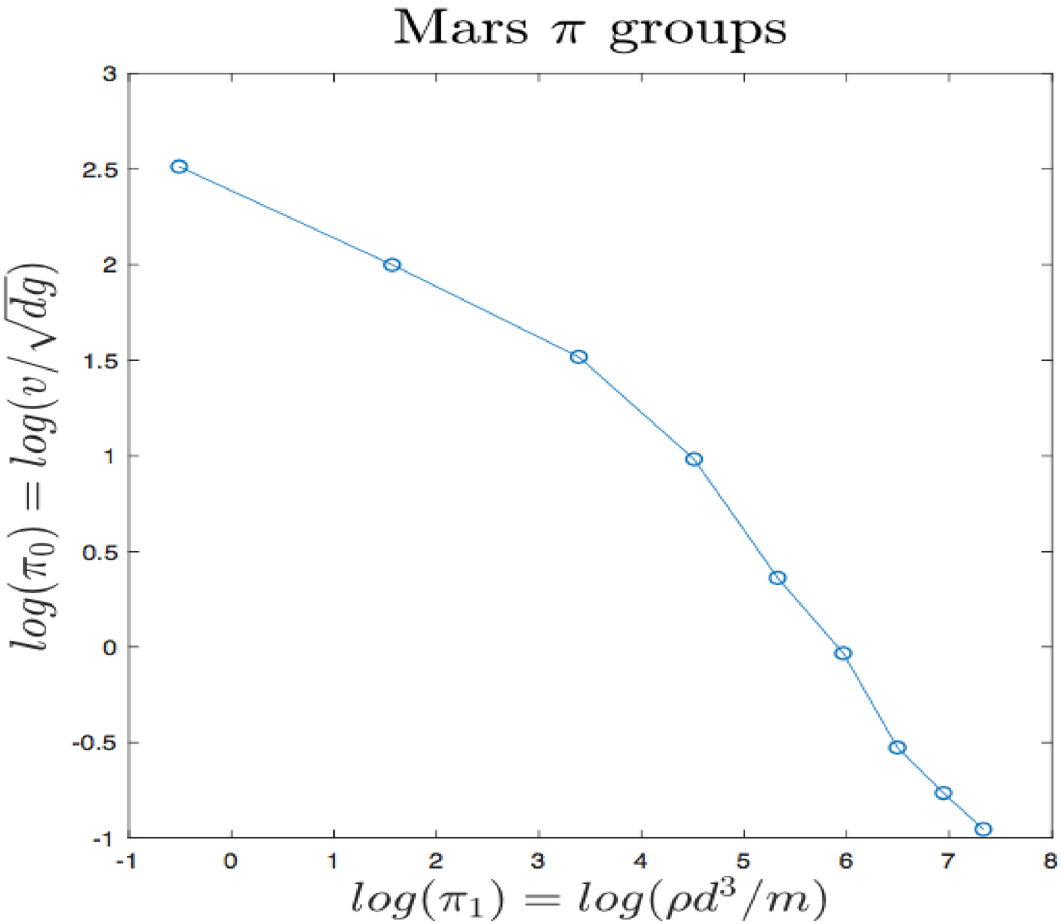
- **On Earth, cannot vary ρ = density of atmosphere.**
- **Could vary *m* = mass of rover but no need.**
- **Just vary d in order to vary π_1**

(Logged) Earth-bound results, n = 9



OK: But what about Mars?

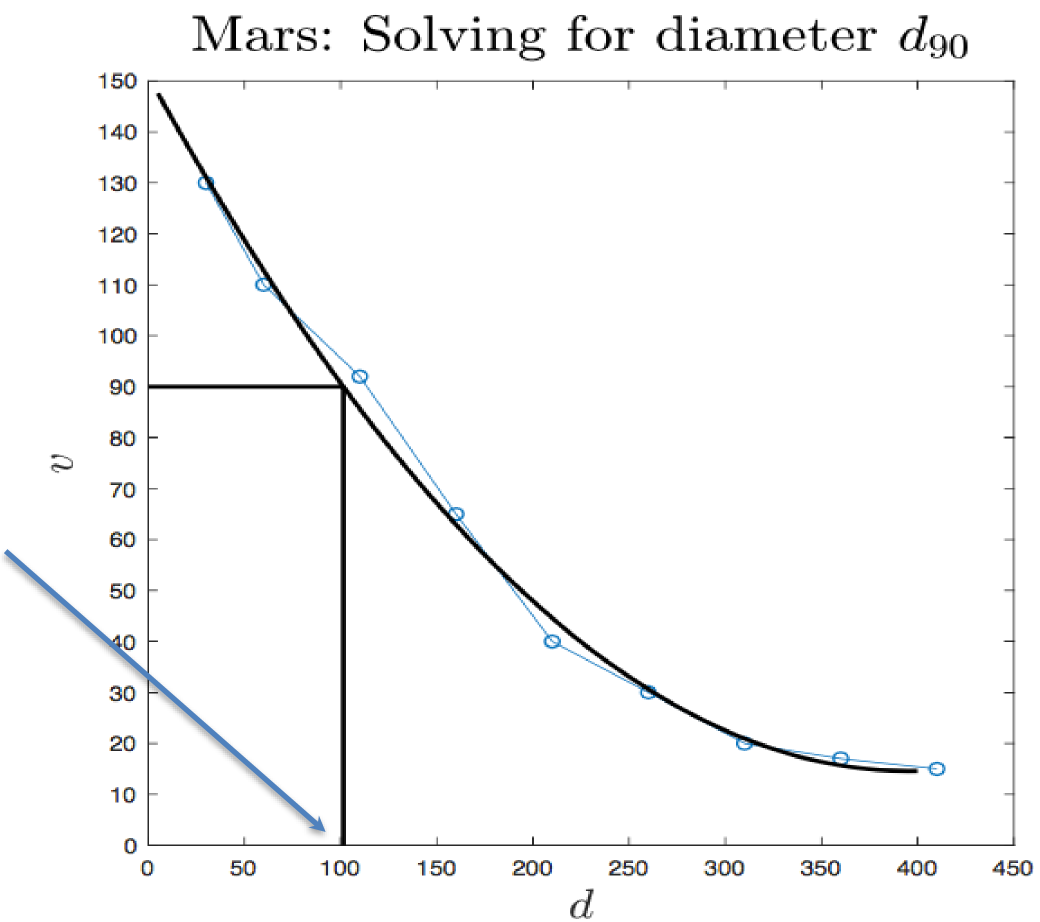
- **Just change the gravitational constant and the atmospheric density, and re-compute the π groups:**



Now solve for v, smooth, and back-solve for d_{90}

$$v = \pi_0 \sqrt{dg}$$

- **The diameter that slows the rover to 90 m/s is 101 meters**



Power of DA

- Example shows:
  - Not only can we extrapolate outside the range of the experimental region
  - We can understand the effects of changing physical constants without actually changing them
  - Run an experiment on Earth, extrapolate to Mars

29

How do we know this process works?

How do we know this process works?

Buckingham Π Theorem (1914)

If there is a physically meaningful equation involving a certain number *n* of physical variables, then the original equation can be rewritten in terms of a set of $p = n - k$ **dimensionless parameters $\pi_1, \pi_2, \ldots, \pi_p$ constructed from the original variables, where k is the number of physical dimensions involved.**

First proven by Joseph Bertrand (1878), in context of electrodynamics

31

Contributions of new paper

- **MV Buckingham Π Theorem**
- **New multivariate design criteria**
- **Illustrations: one simple, textbook example, one very complicated example**
- **Recommendations for optimal design algorithms in complicated situations**
- **(Maybe) a better introduction to DA for statisticians**

32

MV Buckingham Π Theorem

Theorem 1. *Assume the following:*

(i) A vector $Y \in \mathbb{R}^r$ *has a functional relationship with* p *predictors* $(x_1, \ldots, x_p)$*:*

$$Y = f(x_1, \ldots, x_p)$$

where f *is an unknown function of the predictors.*

(ii) The quantities $(Y_1, \ldots, Y_r, x_1, \ldots, x_p)$ *involve* k *fundamental dimensions labeled by* $L_1, \ldots, L_k$*. Then it is assumed that* $\boldsymbol{A} \subseteq span(\boldsymbol{B})$ *where* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *are, respectively, dimensional matrices for the responses and predictors*

(iii) Let Z *represent any of* $(Y_1, \ldots, Y_r, x_1, \ldots, x_p)$*. Then,* $[Z] = \prod_{i=1}^{k} L^{\alpha_i}$ *for some* $\alpha_i \in \mathbb{R}$*,* $i = 1, \ldots, k$ *which are the dimension exponents of* Z*.*

Then the formula $Y = f(x_1, \ldots, x_p)$ *can be rewritten as*

$$\tilde{\pi} = h(\pi_i, \ldots \pi_{p-rank(\boldsymbol{B})}),$$

where $\tilde{\pi} \in \mathbb{R}^r$ *is a vector of dimensionless quantities and* $\pi_i, \ldots \pi_{p-rank(\boldsymbol{B})}$ *are dimensionless predictors.*

Simple Example (White, Fluid Mechanics, 2008)

- **Pump design. Outputs are head pressure (gH) and brake horsepower (bhp)**
- **Predictors: Q, D, n, μ, ρ (ε not considered)**

Variable	Dimensions
gH	$[L^2T^{-2}]$
bhp	$[ML^2T^{-3}]$
Q	$[L^3T^{-1}]$
D	$[L]$
n	$[T^{-1}]$
ρ	$[ML^{-3}]$
μ	$[ML^{-1}T^{-1}]$
ϵ	$[L]$

$$\mathbf{A} = \begin{pmatrix} 2 & 2 \\ -2 & -3 \\ 0 & 1 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 3 & 1 & 0 & -3 & -1 & 1 \\ -1 & 0 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

$$\mathbf{A} \subseteq \operatorname{span}(\mathbf{B})$$

Simple Example (White, Fluid Mechanics, 2008)

- **DA model (two response π's; 2 factor π's):**

$$\begin{pmatrix} \frac{gH}{n^2 D^2} \\ \\ \\ \frac{bhp}{\rho n^3 D^5} \end{pmatrix} = g\left(\frac{Q}{nD^3}, \frac{\rho n D^2}{\mu}\right)$$

- **Only Q, n, and D are varied, μ and ρ are held constant.**

There are two design spaces

- **Design space in the original variables, Q, n, and d:**

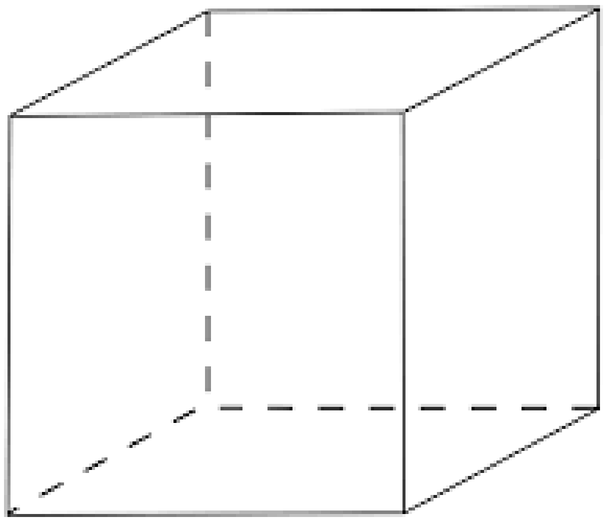
$$\chi = \{\,(Q, n, D) : 4 \le Q \le 30,\;\; 710 \le n \le 1170,\;\; 28 \le D \le 42\,\}.$$

- **Design space in the two pi groups:**

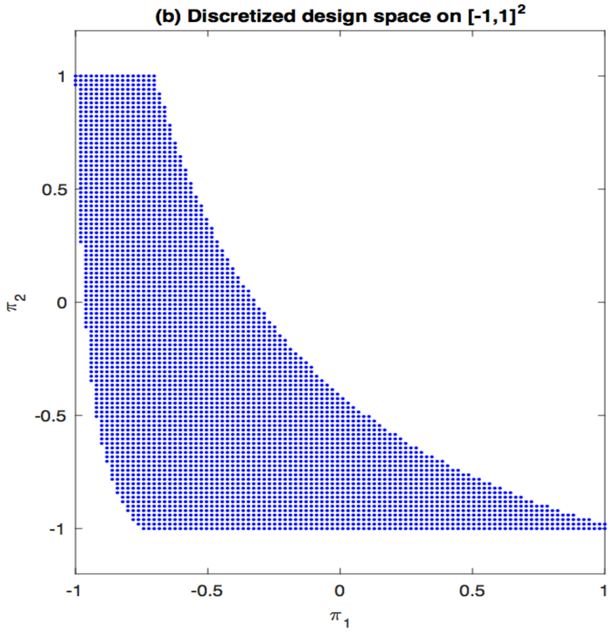
$$\chi_\pi = \{\,(\pi_1, \pi_2) : \pi_1 = Q/(nD^3), \pi_2 = nD^2 \text{ where } (Q, n, D) \in \chi\,\}$$

Pictures of Design Spaces

χ



χ_π



Design Recommendations

- **Previous work we recommended use of:**
  - **Nonparametric designs**
    - **Uniform designs**
    - **Minimax distance designs (etc.,)**
  - **Parametric designs**
    - **D-optimal designs for 3rd-order (or higher-order) models**
    - **Suggested I-optimal, $\bar{L}$-optimal* designs as alternatives**

*Cook, Nachtsheim "Model Robust Linear-Optimal Designs, *Tech*, 1983

MV I-optimality

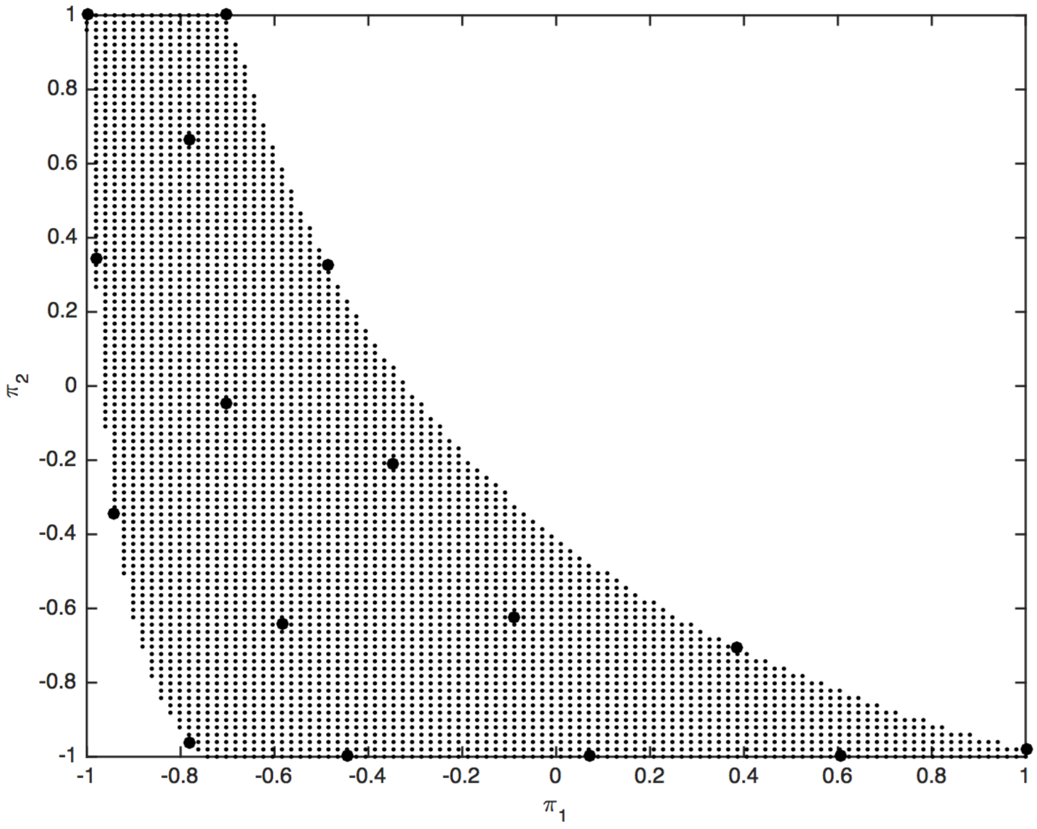
- **The MV I-optimality criterion:**

$$V_{MV}(\xi_n) = r^{-1} v_{\chi}^{-1} \sum_{i=1}^{r} w_i^{-1} \mathrm{Trace}[\mathbf{D}_i \mathbf{M}_i]$$

- **The MV criterion is simply the model-robust, L-optimality criterion of Cook and Nachtsheim (1982, *Technometrics*), where the weights are given by the variances, w_i^{-1}.**

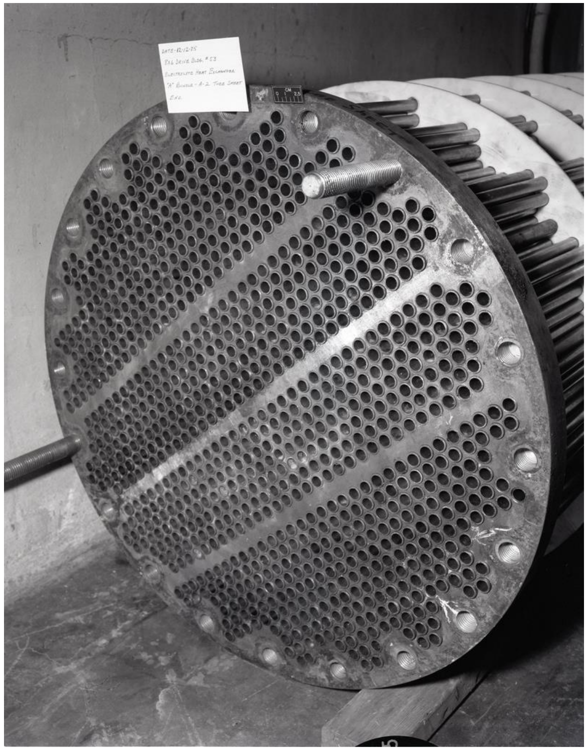
39

I-Optimal Design (n = 16)

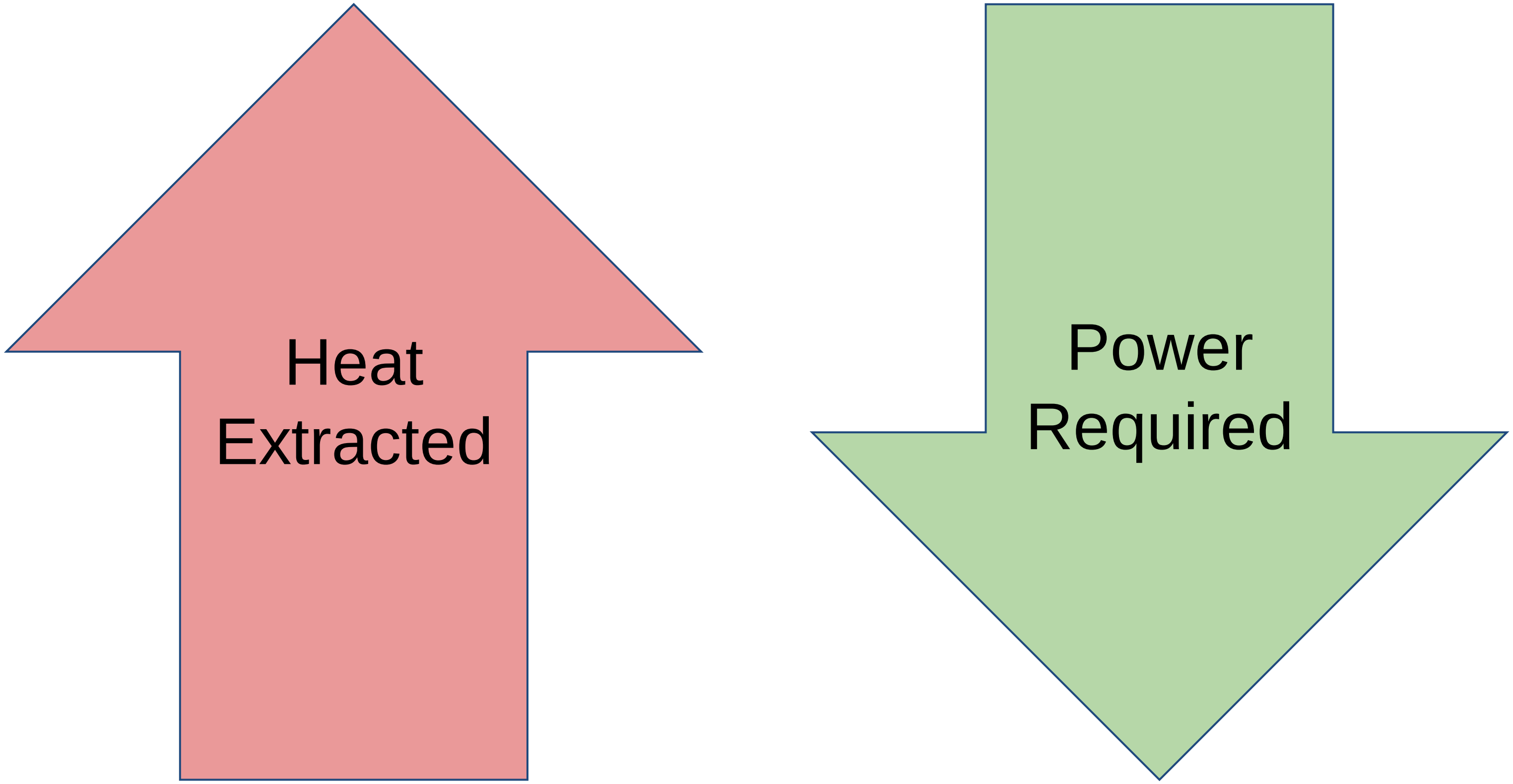


40

Designing Efficient Heat Exchangers

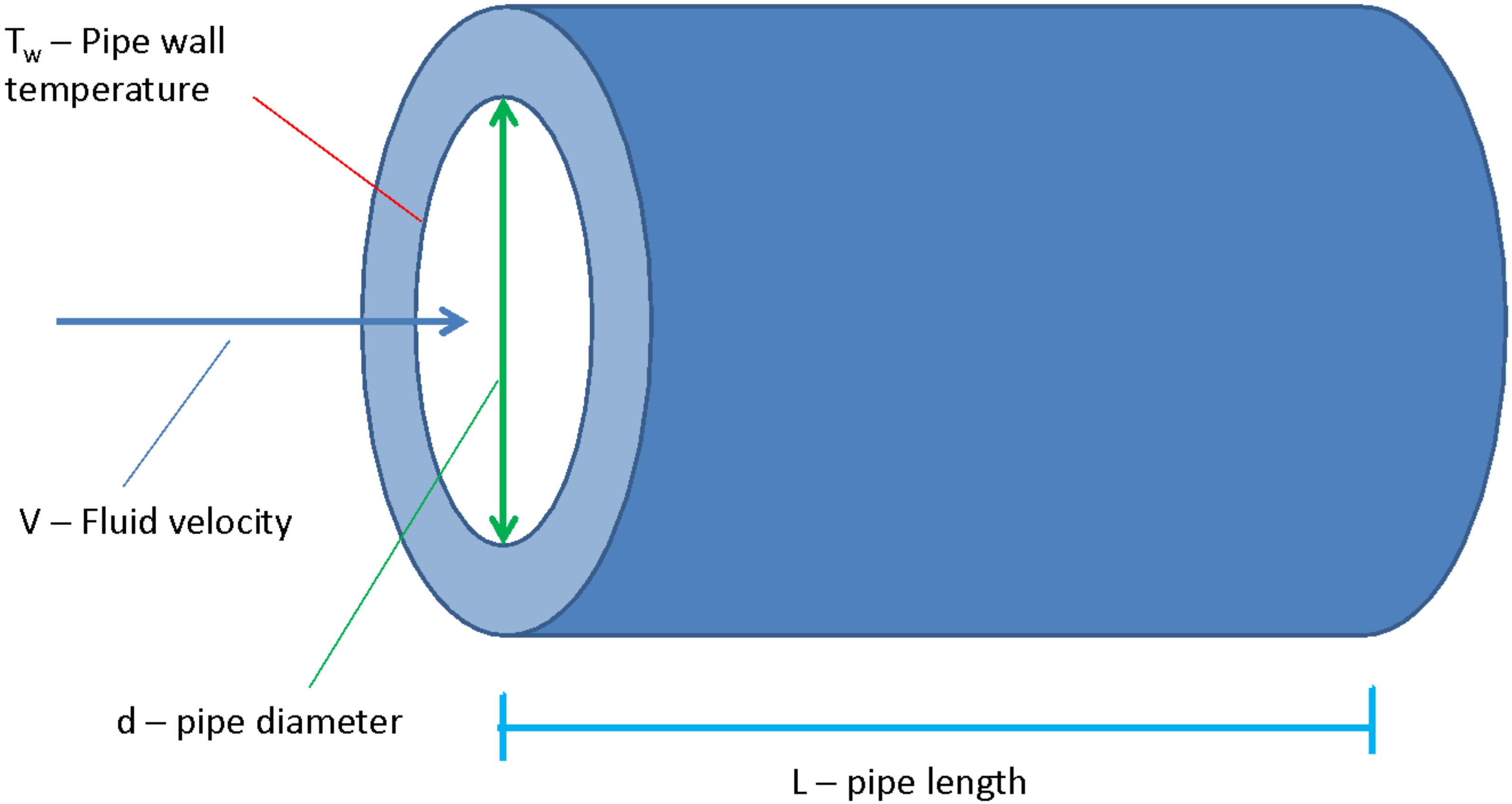


Why is this Multivariate?



42

Schematic

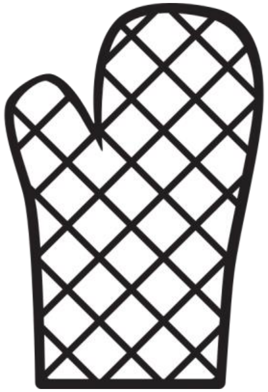


T_w – Pipe wall temperature
V – Fluid velocity
d – pipe diameter
L – pipe length

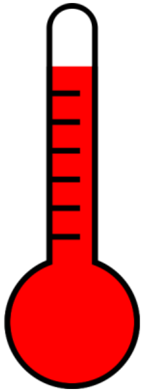
Fluid Properties

Viscosity

Density



Thermal Conductivity



Temperature

A more complex example

- **Design of a heat exchanger with bivariate response**
  - **Response 1 is the pressure loss (ΔP)**
  - **Response 2 is heat transfer rate (Q)**
- **There are 9 independent variables (next slides)**
- **There are four fundamental dimensions: L, M, T, and t**
- **Thus there will be 9 - 4 = 5 independent pi groups**

$$\begin{pmatrix} \Delta P \\ Q \end{pmatrix} = f(D, L, V, T_W, T_f, \mu, \rho, g, K)$$

Variable	**Symbol**	**SI Units**	**Base Quantity**
Pressure change	ΔP	Pa	M/Lt^2
Heat transfer rate	Q	W	ML^2/t^3
[redacted]			
Inner diameter of pipe	d	M	L
Length of pipe	L	M	L
Mean velocity of fluid	V	m/s	L/t
Temperature of inner pipe wall	T_w	K	T
Mean temperature of fluid at the start of the tube	T_f	K	T
Viscosity of fluid	µ	$N\text{-}s/m^2$	M/Lt
Density of fluid	p	kg/m^3	M/L^3
Acceleration due to gravity	g	m/s^2	L/t^2
Thermal conductivity of fluid	K	W/m-K	ML/Tt^3

46

Derivation of DA Model

Step 0: Initialize		Step 1 Result: Remove M from Step 0 using ρ		Step 2 Result: Remove L from Step 1 using d		Step 3 Result: Remove t from Step 2 using V/d		Step 4 Result: Remove T from Step 3 using T_f	
Variable	Dimension	Var.	Dim.	Var.	Dim.	Var.	Dim.	Var.	Dim.
ΔP	$ML^{-1}t^{-2}$	$\Delta P/\rho$	$t^{-2}L^2$	$\Delta P/(\rho d^2)$	t^{-2}	$\Delta P/(\rho V^2)$	1	$\Delta P/(\rho V^2)$	1
Q	mL^2t^{-3}	Q/ρ	$t^{-3}L^5$	$Q/(\rho d^5)$	t^{-3}	$Q/(\rho d^2 V^3)$	1	$Q/(\rho d^2 V^3)$	1
d	L	d	L	—		—		—	
L	L	L	L	L/d	1	L/d	1	L/d	1
V	Lt^{-1}	V	Lt^{-1}	V/d	t^{-1}	—		—	
T_W	T	T_W	T	T_W	T	T_W	T	T_W/T_f	1
T_f	T	T_f	T	T_f	T	T_f	T	—	
μ	$ML^{-1}t^{-1}$	μ/ρ	L^2t^{-1}	$\mu/(\rho d^2)$	t^{-1}	$\mu/(\rho d V)$	1	$\mu/(\rho d V)$	1
ρ	ML^{-3}	—		—		—		—	
g	Lt^{-2}	g	Lt^{-2}	g/d	t^{-2}	gd/v^2	1	gd/v^2	1
K	$MLt^{-1}t^{-3}$	K/ρ	$L^4T^{-1}t^{-3}$	$K/(\rho d^4)$	$T^{-1}t^{-3}$	$K/(\rho d V^3)$	T^{-1}	$KT_f/(\rho d V^3)$	1

$$\begin{pmatrix} \frac{\Delta P}{\rho V^2} \\ \frac{Q}{\rho d^2 V^3} \end{pmatrix} = \phi\left(\frac{L}{d}, \frac{T_W}{T_f}, \frac{\mu}{\rho d V}, \frac{gd}{v^2}, \frac{KT_f}{\rho D v^3}\right)$$

47

Derivation of DA Model

Euler's number

Reynold's number

$$\left(\frac{\frac{\Delta P}{\rho V^2}}{\frac{Q}{\rho d^2 V^3}}\right) = \phi\left(\frac{L}{d}, \frac{T_W}{T_f}, \frac{\mu}{\rho d V}, \frac{gd}{v^2}, \frac{K T_f}{\rho D v^3}\right)$$

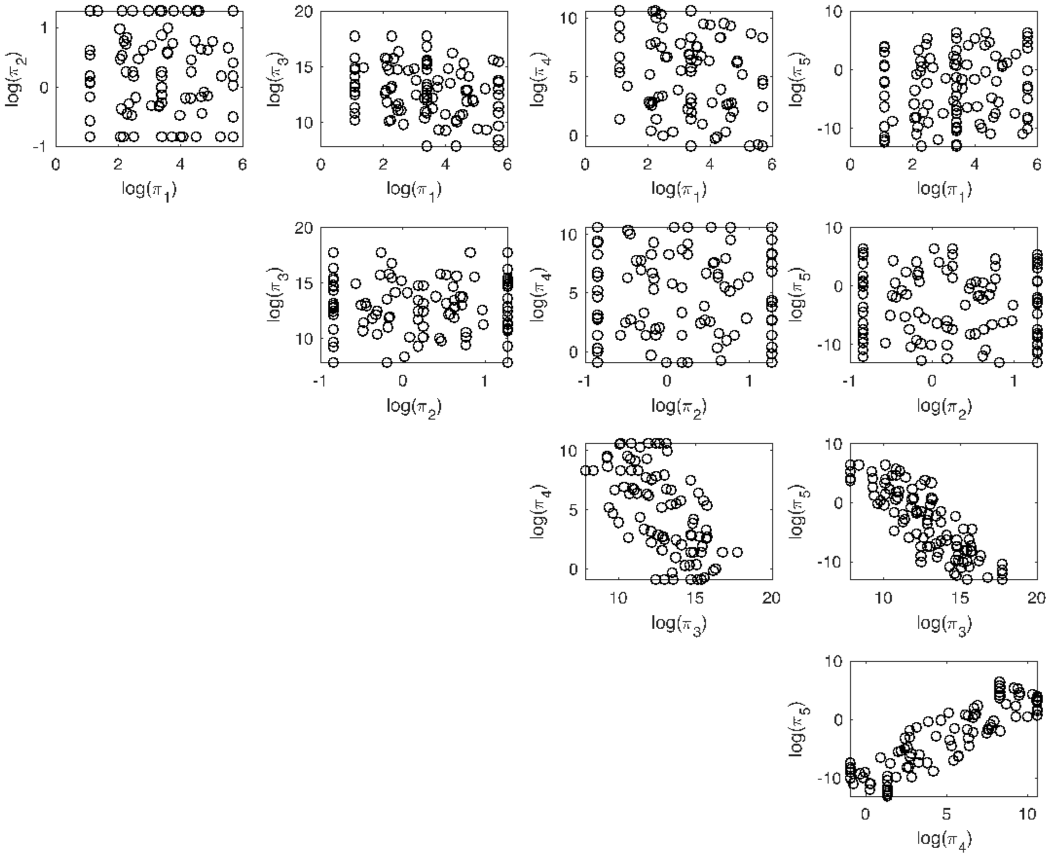
Design Generation Findings

- **Parametric design**
  - Generate the designs using the 8-dimensional original variables space
  - Use coordinate exchange with continuous optimizer for each coordinate
- **Nonparametric design**
  - Generate 100K uniformly distributed points in the 5D Pi design space for use as candidate set
  - Use candidate set with Fast Flexible Filling* to create designs

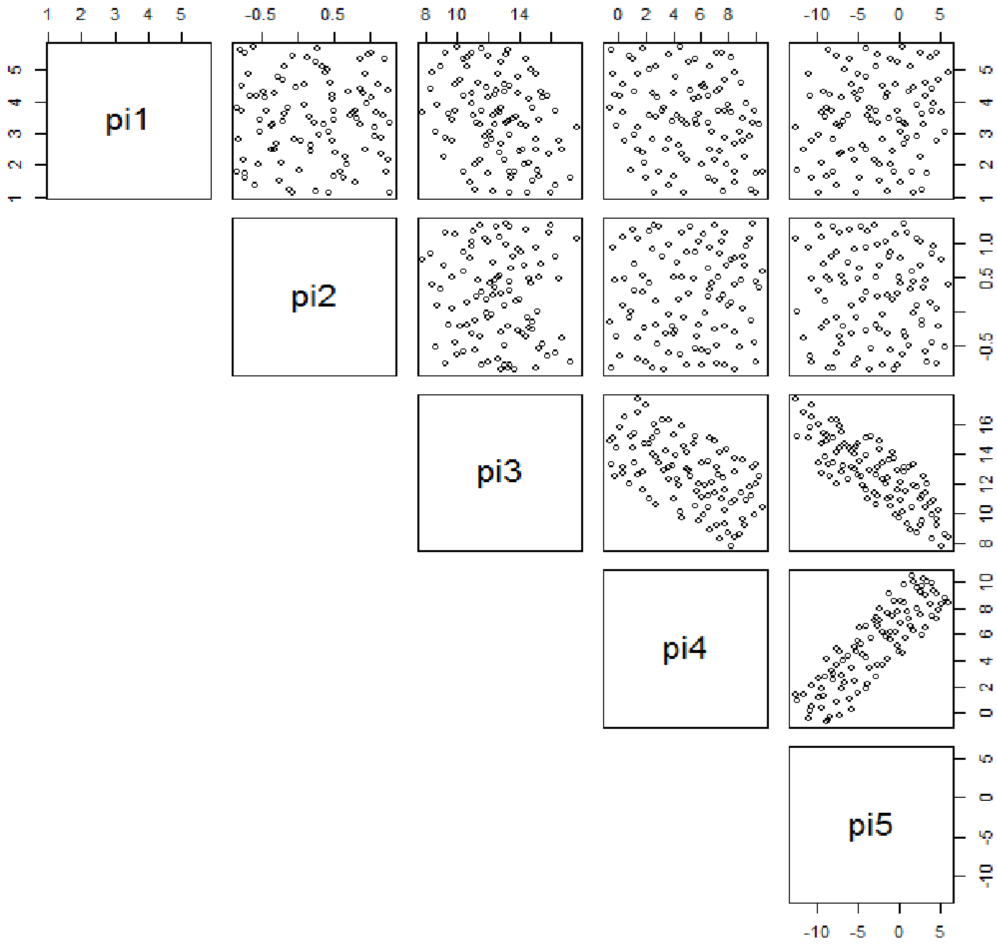
*Ryan Lekivetz, Brad Jones, QREI, 2014

49

Parametric Design: 2D Projections, n = 100



Nonparametric Design: 2D Projections, n = 100



51

Comments/Conclusions

- **DA is a powerful tool for modeling physical systems**
  - **Parsimony, Scalability, Dimension Reduction**
- **Multivariate responses not uncommon**
- **Gave generalization of Buckingham for multivariate responses**
- **Gave new criterion of multivariate parametric design**
- **Gave recommendations for design construction**

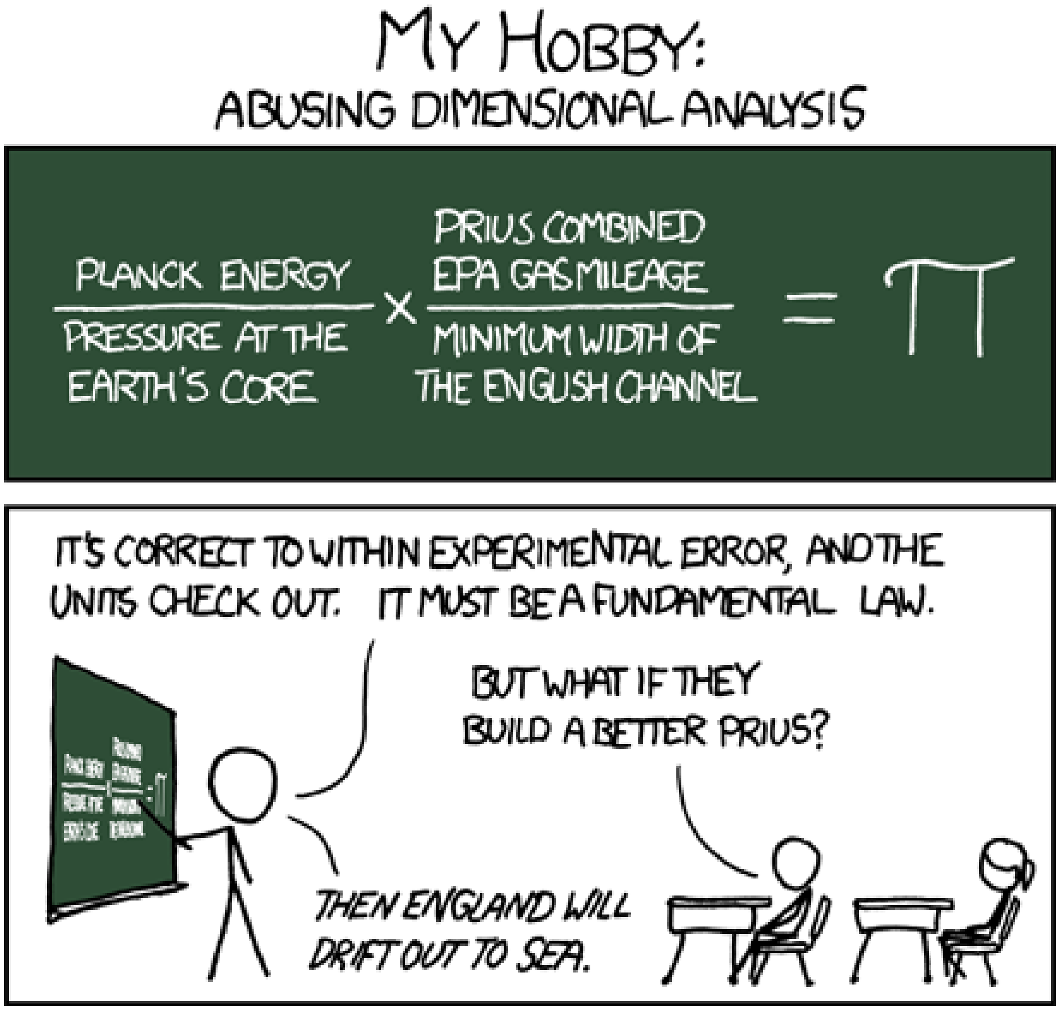
Epilogue (Chris):

How did we stumble into the "design for DA" problem?

How did we statisticians get involved in DA?

- **(c. 1980)   Chris's early consulting– giving bad advice**
- **1988 talk on DOE at Carlson school to Mgt Sci department**
- **2011, Tom Albrecht, Boston Scientific**
- **2013, M. Albrecht, Nachtsheim, Cook, T. Albrecht, Experimental Design for Engineering Dimensional Analysis, *Technometrics***
- **2014, Dennis and Chris DOE PhD seminar: Dan Eck proves the MV Buckingham Pi Theorem**
- **2021, our paper**

Thank you


MY HOBBY:
ABUSING DIMENSIONAL ANALYSIS
PLANCK ENERGY / PRESSURE AT THE EARTH'S CORE × PRIUS COMBINED EPA GAS MILEAGE / MINIMUM WIDTH OF THE ENGLISH CHANNEL = π
IT'S CORRECT TO WITHIN EXPERIMENTAL ERROR, AND THE UNITS CHECK OUT. IT MUST BE A FUNDAMENTAL LAW.
BUT WHAT IF THEY BUILD A BETTER PRIUS?
THEN ENGLAND WILL DRIFT OUT TO SEA.