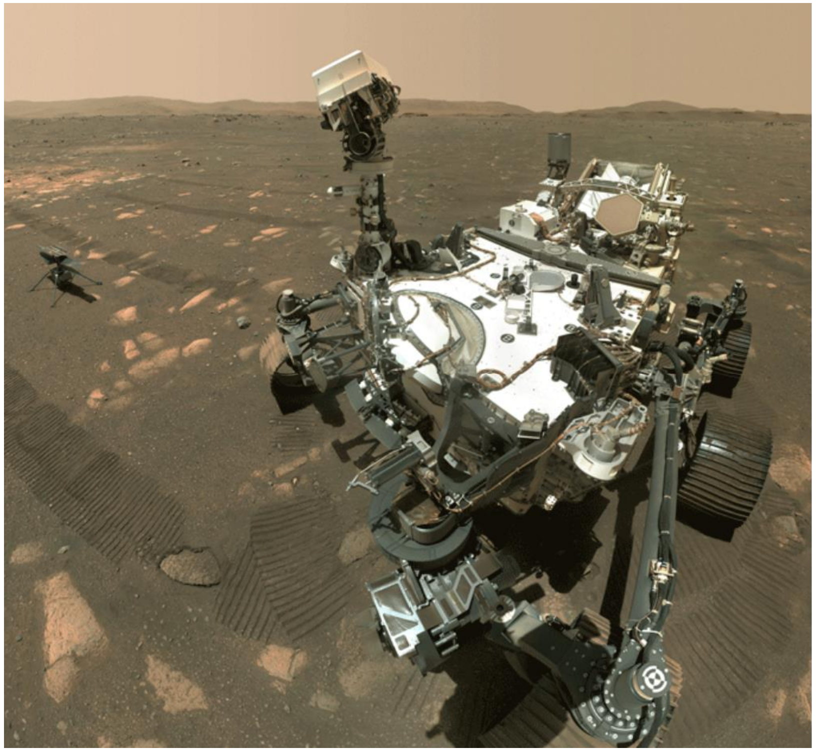
ADSURGO
ascend with analytics

DATAWorks 2024
Defense and Aerospace Test and Analysis Workshop



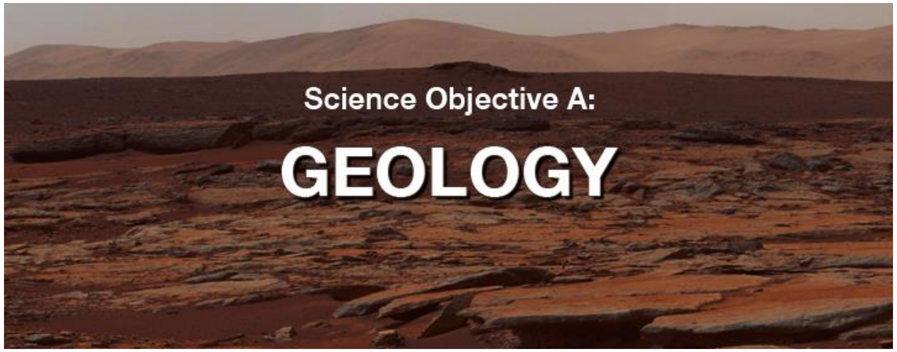
Design and Analysis of Nonlinear Models

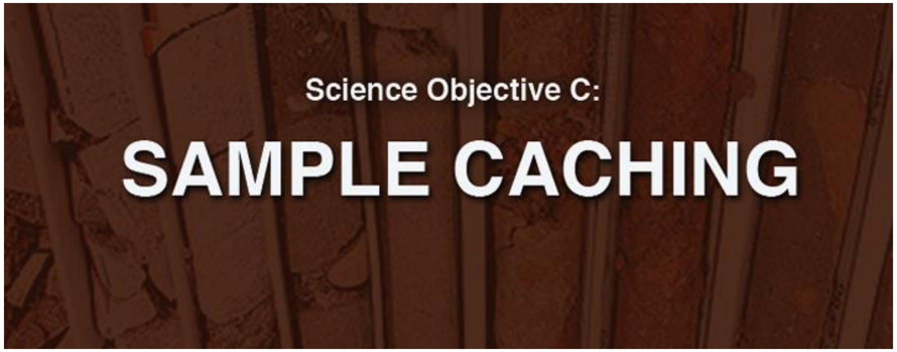
Plan

- Background on Mars Rover program
- Fitting data to nonlinear models: drill bit degradation
- Creating a design with a known nonlinear functional form: binary response for core quality
- Augmenting with a new factor known to have a nonlinear effect: spindle torque and space-filling designs

Mars 2020 Rover

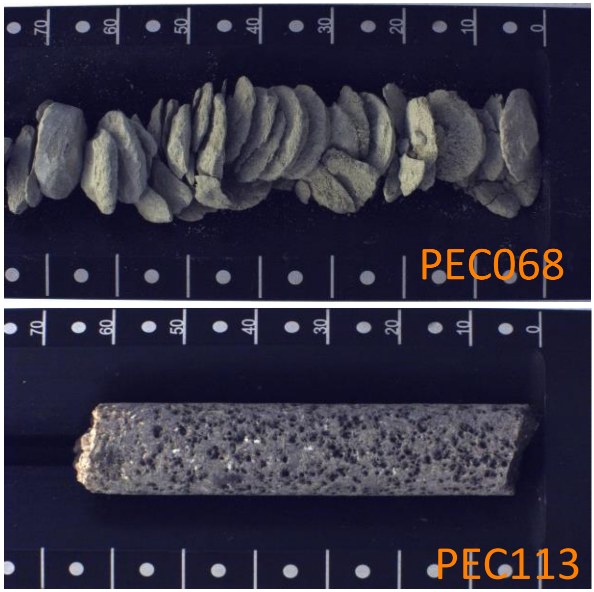
- The Mars 2020 Perseverance Rover has successfully drilled 20+ cores
- Nonlinearn DOE was key to success!





Subsystem Design

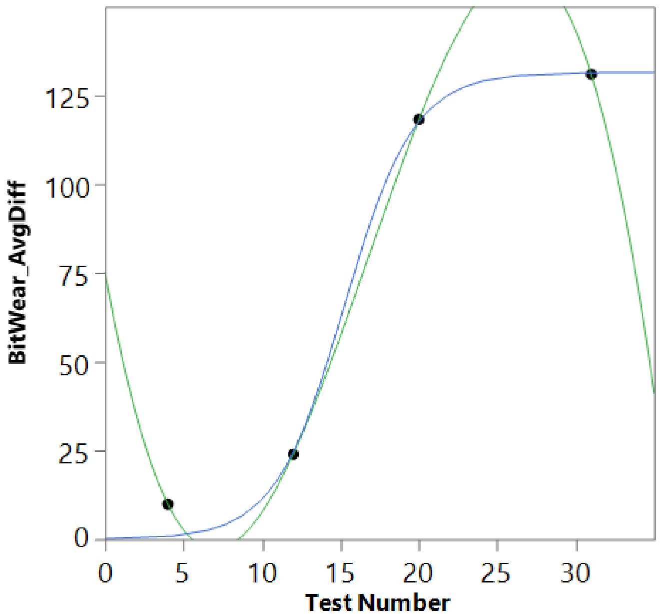
- We don't get to control Mars!
  - Subsystem design highly informed by testing
- Key requirements:
  - Collect ~40 cores of varying sample types, based on notional distribution of Earth analog rocks
  - Core quality - Best core has a few number of large pieces
  - Samples must be "hermetically" sealed
- Measurable responses
  - Core quality
    - Mass and number of pieces to pass through sieves of 2, 5, 10 and >10 mm
    - Sample volume
  - Seal leak rate
  - Drilling Performance
    - Avg cycle sideload
    - Avg cycle percussion current
    - Avg drilling torque
    - Avg percussion power
    - Avg rate of penetration



Nonlinear Modeling Intro
Modelling Coring Bit Wear

- **Problem:** How can we model the impact of bit wear on drill performance metrics? Is there a surrogate measure for bit wear?
- **Methodology:** Measure bit wear 4 times over a series of 30 tests (not an easy task), fit candidate nonlinear models, create new control variable as bit wear over time, run regression models against easily observable factors
- **Results:**
  - Bit wear approximated well by Logistic 3Parameter Sigmoid Curve
  - Makes sense from physics of failure/degradation models
  - Highly correlated with Time in USB rocks
  - Useful control variable for many of the responses

Nonlinear Modeling Intro
Finding the Right Model



Prediction Model

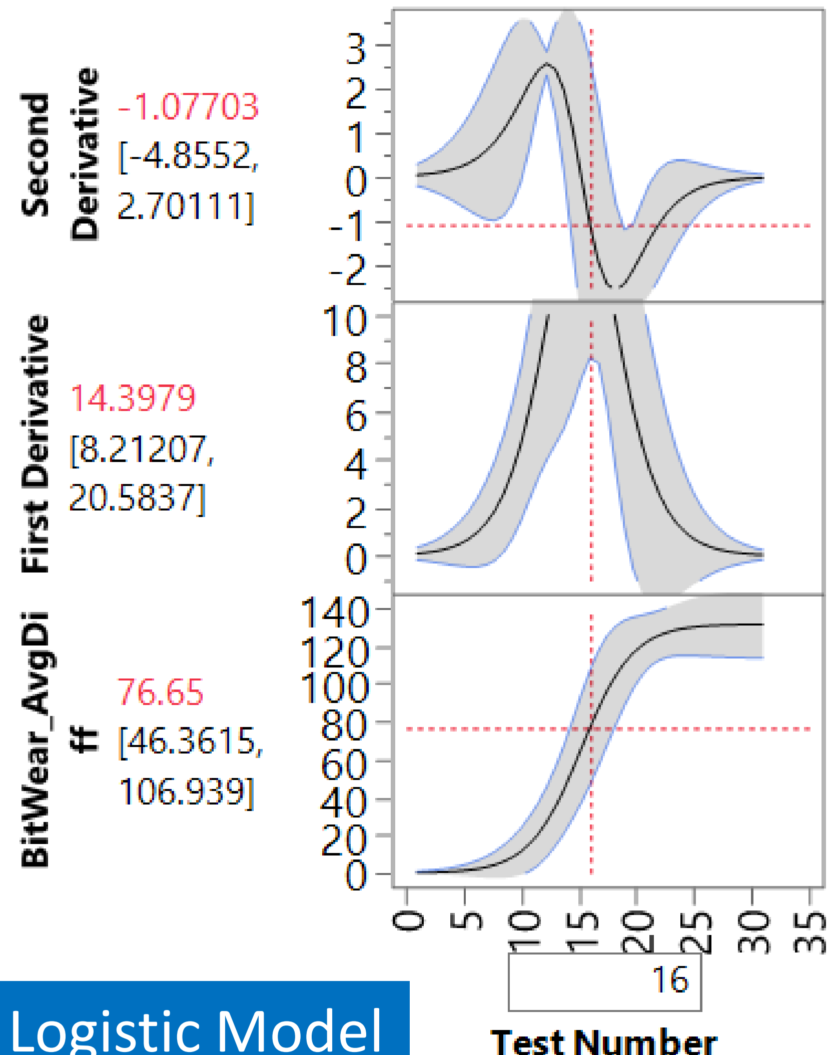
$$\frac{c}{\left(1+\text{Exp}\left(-a\cdot\left(\text{Test Number}-b\right)\right)\right)}$$

a = Growth Rate

b = Inflection Point

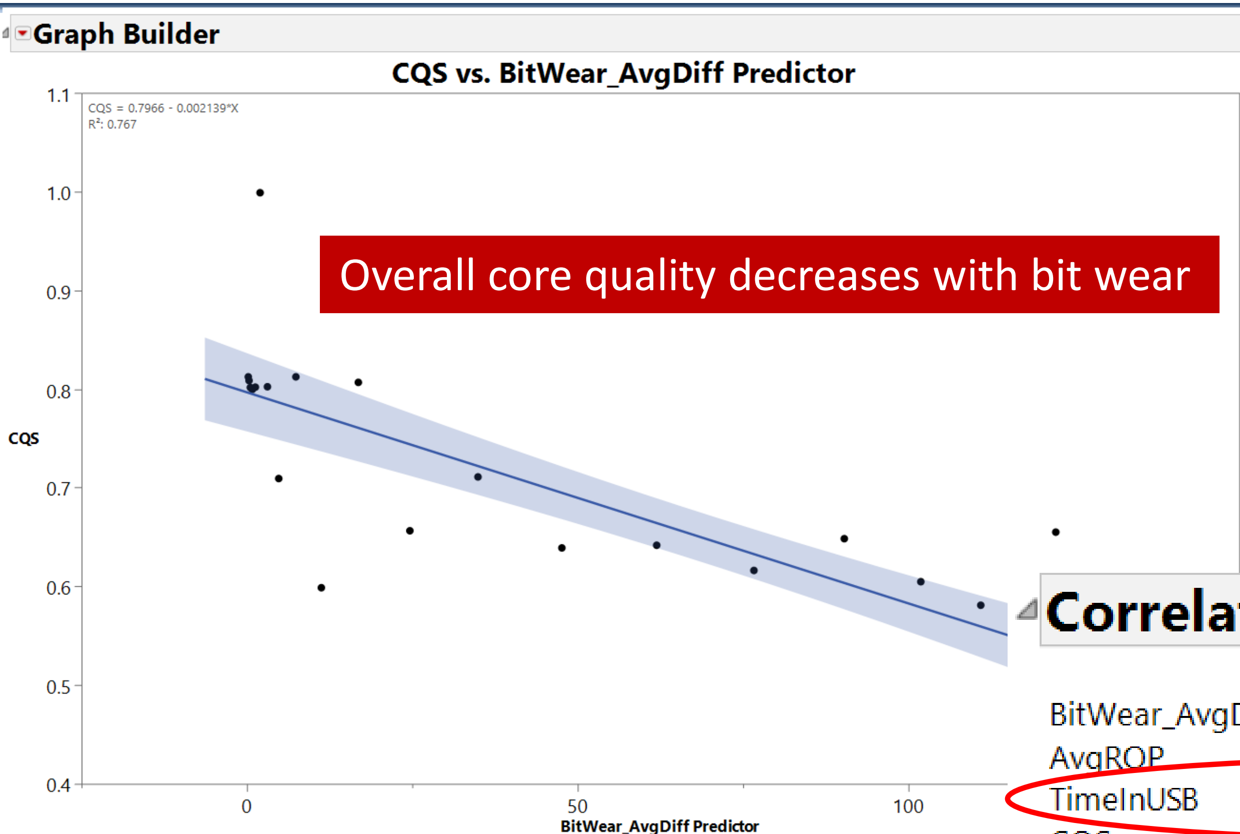
c = Asymptote

Prediction Profiler



3 Parameter Logistic Model

Nonlinear Modeling Intro
Finding a Bit Wear Surrogate

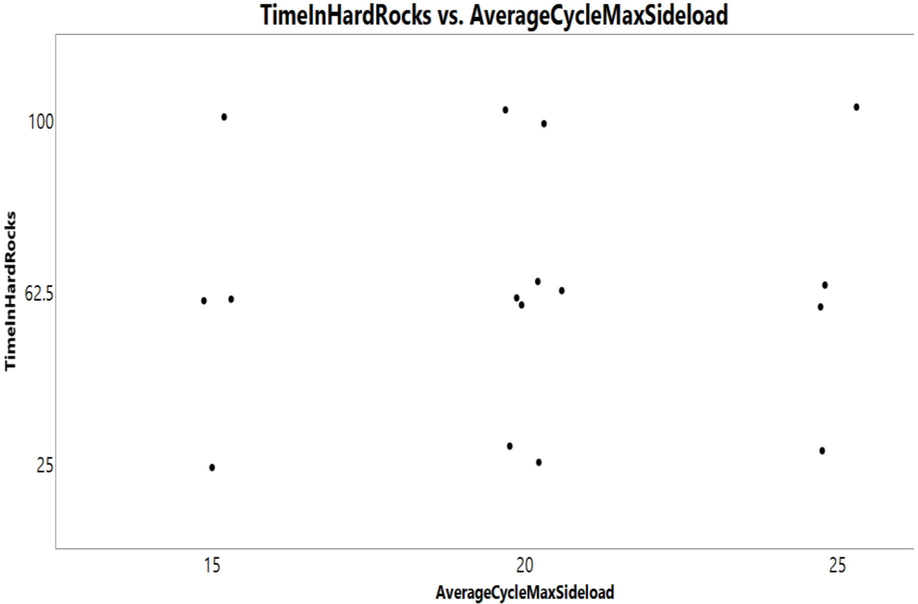


Correlations

	BitWear_AvgDiff Predictor
BitWear_AvgDiff Predictor	1.0000
AvgROP	-0.5094
TimeInUSB	0.9684
CQS	-0.8756
CoreTotalMass	-0.2533
CoreVolRatio	0.7149
TPEL_error	0.6313
CoreMassGT10mm	-0.3997
CoreNumPieces10_above	0.8828
CoreMassPerPiece_above_10mm	-0.8112
CoreVol_GT_10mm	-0.3997
CoreNumPieces	0.6037

Design for Nonlinear

- **Problem:** Ok got it that you can model nonlinear, but how do you actually set yourself up to have the right tests knowing you have a nonlinear relationship. Specifically, we have a binary response on core quality so we know logistic regression is our nonlinear relationship.
- **Methodology:** Limited options to take advantage of what you know to some level of the functional form. Bayesian D-Optimal Designs can help.
- **Results:**
  - Designs differ from Custom RSM
  - Better estimates where matters
  - General approach



Nonlinear Designs

- 2-level factorial designs are good for linear models and we can use centers to detect curvature and then augment with axials to determine quadratic effects
- Alternatively, use a response surface design or definitive screening design from the beginning if know higher order effects likely to be present
- Not all responses modeled adequately by 2^{nd} order Taylor Series approximation with normal error
- Engineering principals may point functional form to a different direction with complex curvature
- Possible to model nonlinearities, but how do you create a design that is set up to maximize the precision given the nonlinear behavior?
- To optimize the design points, you need to know the functional form and a good estimate of the parameters (so why are we testing if we know all this already???)

Nonlinear Designs
JMP Implementation

- Consider using a Bayesian approach where the parameter values are restricted to some reasonable range via the prior distribution
- Bayesian D-optimal criteria is expectation of Log |Information| with respect to parameter vectors from the prior

$$\phi(\mathbf{D}) = \int \log |\mathbf{D}'\mathbf{D}| f(\boldsymbol{\beta})\, d\boldsymbol{\beta}$$

Where D is matrix of partials of the expectation function evaluated wrt each model parameter evaluated at each design point

- Gotwalt, Jones, and Steinberg (*Technometrics*, 2009) solve this and implement in JMP
- Goal is to place points at high prediction variance points in the design space—not the vertices like linear models though!

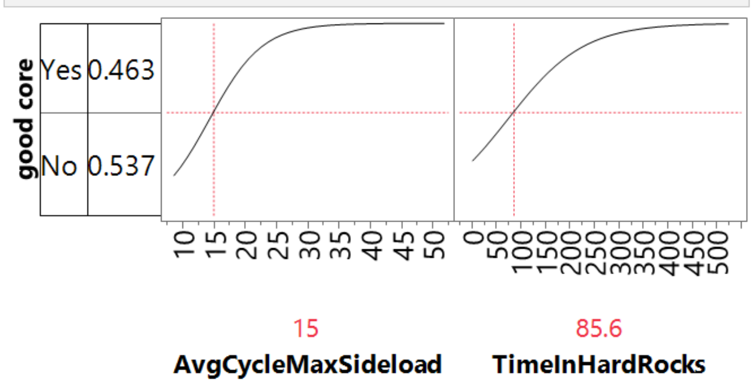
Nonlinear Designs Binomial Example

- Core Quality is a function of many variables and ultimately a very noisy and nonlinear process
- For simplicity, we model whether or not core was good as a function of Maximum Sideload and Time in HardRocks in a logistic regression model for USB rock

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	-4.5081043	1.0104159	19.91	<.0001*
AvgCycleMaxSideload	0.23676095	0.0552635	18.35	<.0001*
TimeInHardRocks	0.01289752	0.0050375	6.56	0.0105*

Prediction Profiler



Confusion Matrix

Training

Actual	Predicted Count	
good core	No	Yes
No	69	17
Yes	15	34

Nonlinear Designs
Binomial Example

- We want to design a set of tests accounting for these two factors given a binary response
- We know the functional form of logistic regression

  1 / (1 + Exp(-(b0 + b1 * :AvgCycleMaxSideload + b2 * :TimeInHardRocks)))

- We also have historical data to suggest parameter estimates to inform our prior distribution
  - Create new data table with columns for two input factors Sideload and Time as well as a ***continuous*** response
  - Create column that has logistic regression model formula specifying parameters for b0, b1, and b2 {b0 = -5, b1 = 0.25, b2 = 0.015}
  - From this table run Nonlinear DOE to input the range of factors and the prior information
  - Specify number of runs and create design
  - Plot factors to see geometry

Nonlinear Designs
Binomial Example

Default Diffuse Prior

Nonlinear Design

Factors

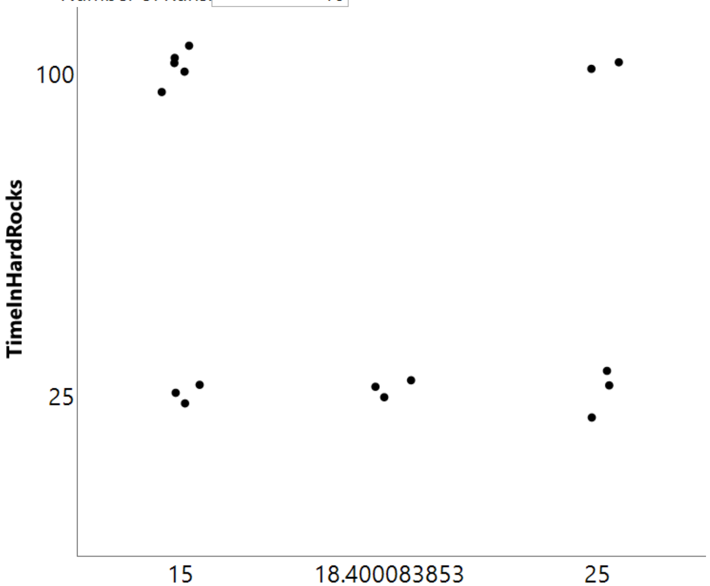
Name	Role	Values	
AvgCycleMaxSideload	Continuous	15	25
TimeInHardRocks	Continuous	25	100
optional item			

Parameters

Name	Distribution	Values	
b0	Normal	-2.5	-7.5
b1	Normal	0.125	0.375
b2	Normal	0.0075	0.0225
optional item			

Design Generation

Number of Runs: 16



Informative Prior

Nonlinear Design

Factors

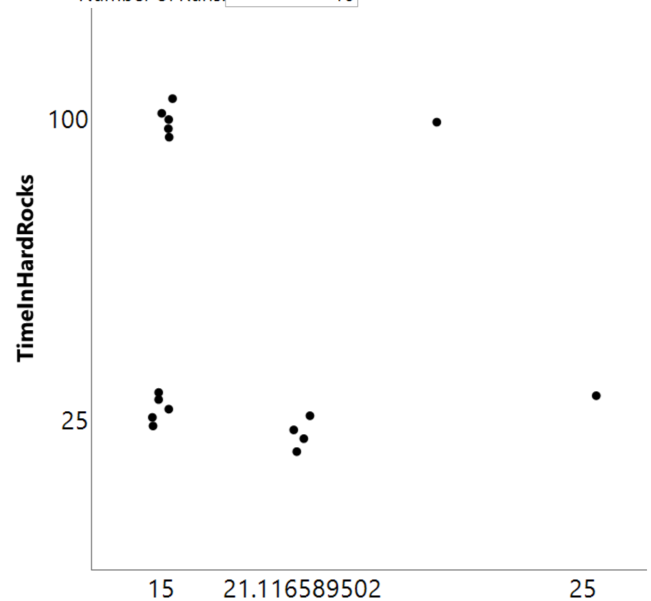
Name	Role	Values	
AvgCycleMaxSideload	Continuous	15	25
TimeInHardRocks	Continuous	25	100
optional item			

Parameters

Name	Distribution	Values	
b0	Normal	-5.5	-3.5
b1	Normal	0.2	0.3
b2	Normal	0.012	0.014
optional item			

Design Generation

Number of Runs: 16



Nonlinear Designs
Augmenting Binomial Example

- Augment existing diffuse prior with **6 more runs (diffuse, uniform)** and another **6 (informative, uniform)**