

Maximising Parametric Model Utilisation by Developing Excellent Cost Estimating Relationships

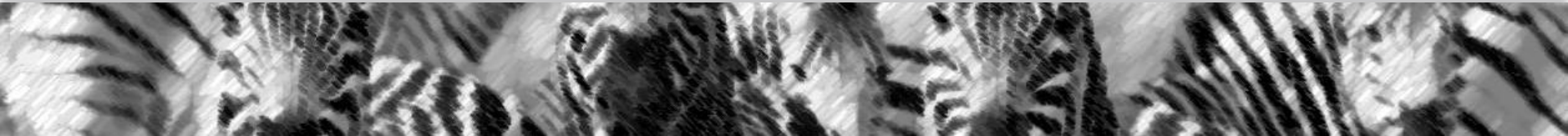
Dr Spencer Woodford, Burchelli Consulting Ltd

Society for Cost Analysis & Forecasting
8 February 2011



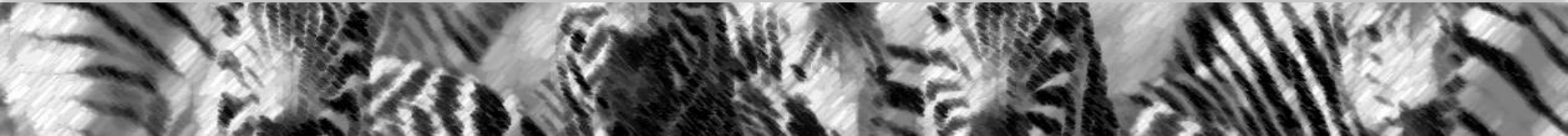
Contents

- Introduction
 - Parametric Cost Estimating Relationships
 - General Method
- Attributes of an 'Excellent' CER
 - Data
 - Model Structure
 - Algorithm Development
 - Accuracy
 - Documentation
 - Users
- Example – Aero Gas Turbine Development
- Conclusions

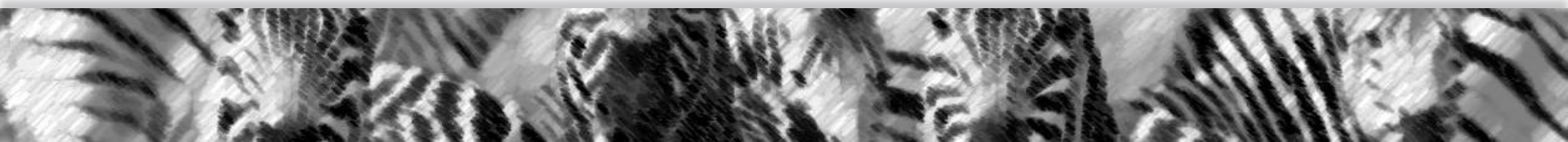
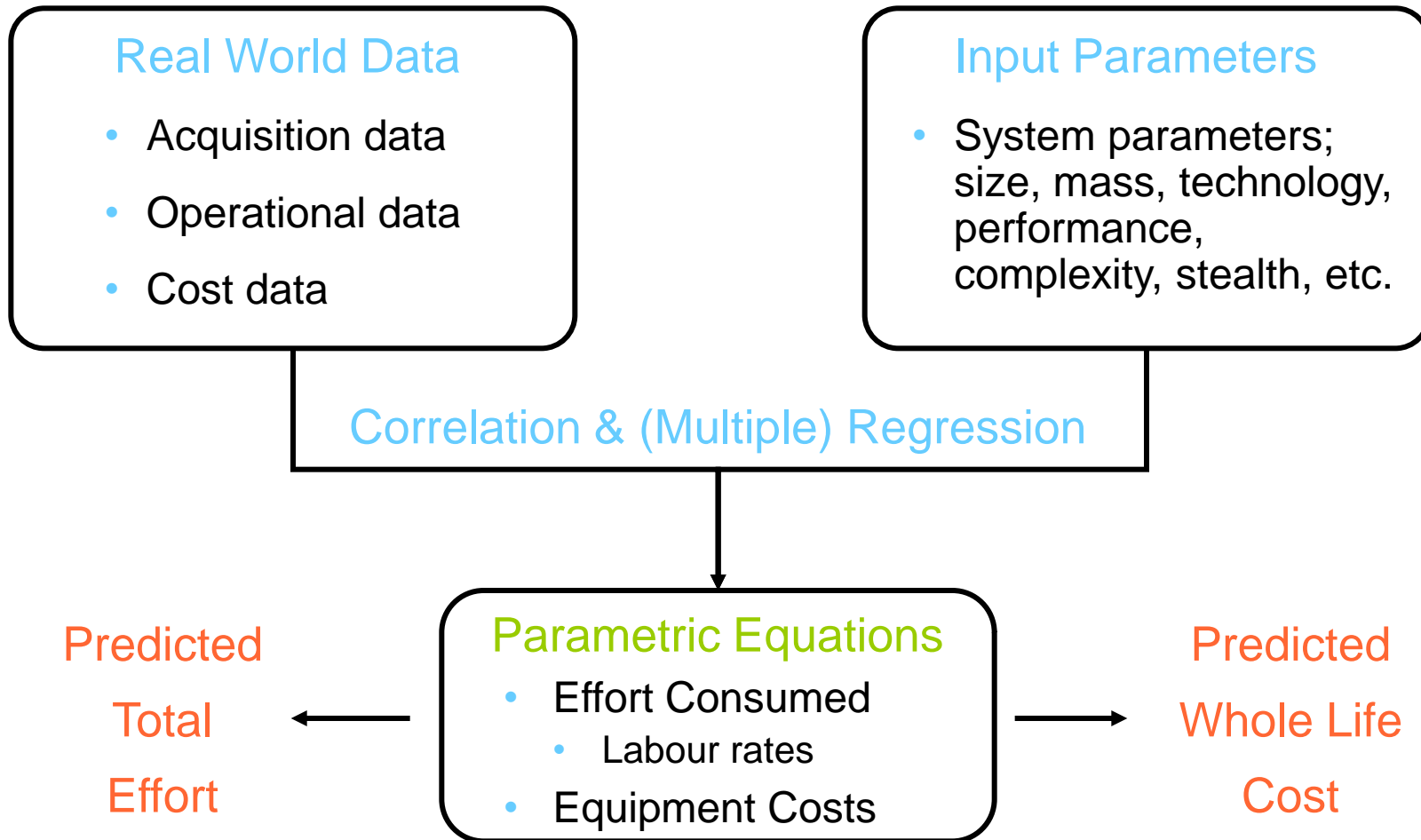


Introduction

- Parametric Cost Estimating Relationships (CERs):
 - are algorithms that forecast costs based upon the examination and validation of the relationships that exist between a project's technical, programmatic, and cost characteristics as well as the resources consumed during its development, manufacture, maintenance, and/or modification;
 - can be Simple (single parameter) or Complex (multiple parameters);
 - are inconsistently used for Defence projects;
 - suffer from a lack of support at higher levels due to poor understanding and lack of exposure to the levels of accuracy available from the tools.
- Similar algorithms are widely used in numerous industries to estimate outputs based on a number of technical and/or programmatic inputs, e.g.
 - size/mass estimation of aircraft components/systems;
 - reliability and maintainability metrics for complex systems;
 - manpower numbers required to operate and support complex systems, etc.



Introduction – General Method



Attributes of an 'Excellent' CER (1) – Data

- Quantity
 - Non-linear equations require a *minimum* of 11 data points for each parameter added to the equation
 - Never dismiss data for being 'too old' – technology (improvement) factors in the model could/**should** account for this
- Normalisation
 - All input data must be in consistent units, e.g. kg, m²
 - All costs must all be converted to the same currency in the same financial year (preferably using recognised & authoritative currency exchange rates)
- Visibility
 - A lack of **total** transparency of the data values, sources, missing/estimated input parameters, etc. will exacerbate the lack of confidence that others have in any model
 - 'Trust me' is not good enough to convince others that you have a good model



Attributes of an 'Excellent' CER (2) – Model Structure

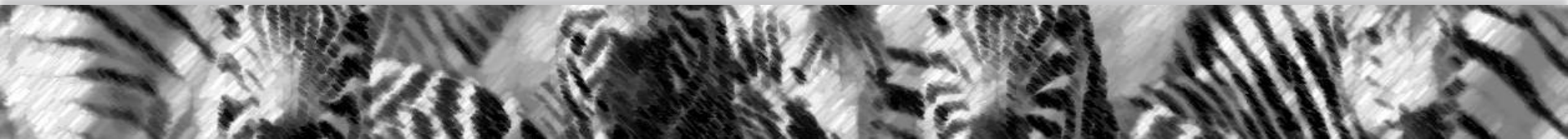
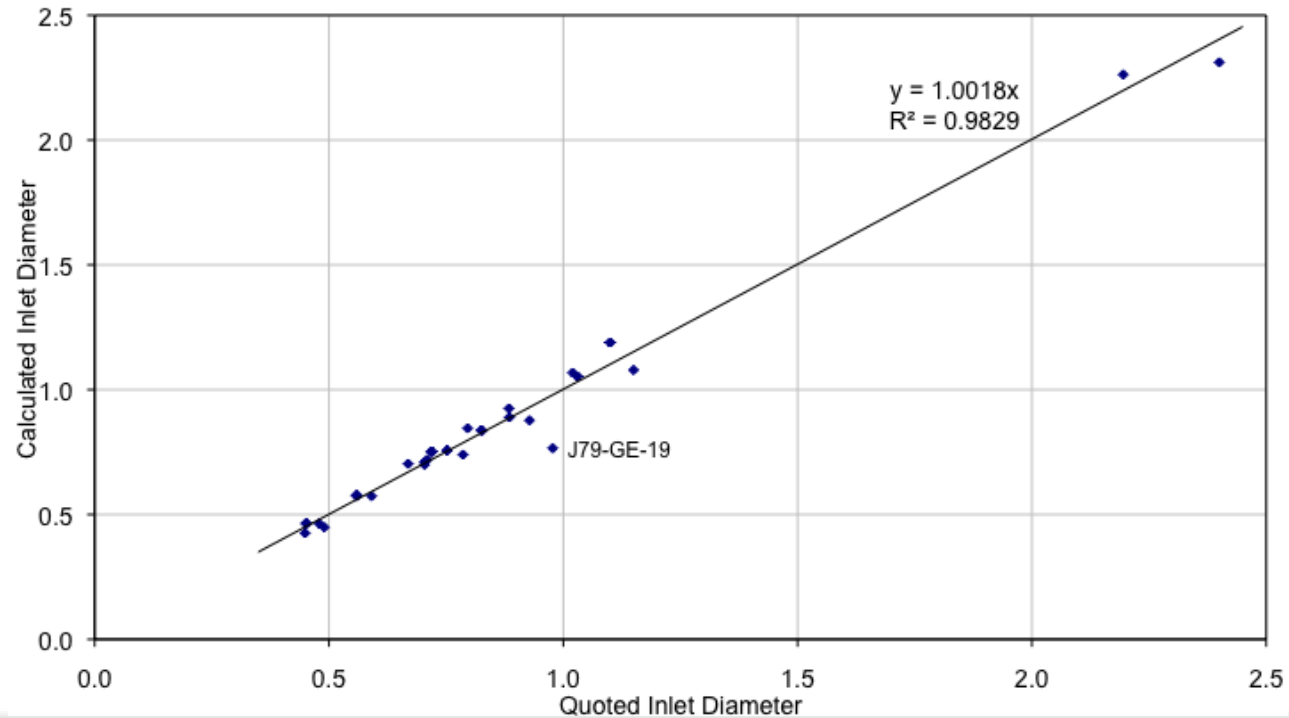
- There must be a direct (or strongly correlated) relationship between the inputs and the output of the model
 - If there is no good *engineering* link between parameters, they should be removed
 - Often, there is a *theoretical* link between input parameters & the output
 - This can be used to validate that the model coefficients are of the right order
- Complex systems will likely require more complex models to enable suitable fidelity to be captured
 - Consider splitting the overall output into several sub-CERs
 - Increases visibility and ability to correlate/validate individual elements



Attributes of an 'Excellent' CER (3) – Model Algorithms (1)

- Simple models can make use of algorithms with single factors and coefficients, e.g.:
 - Aero Engine Inlet Diameter (m) = $A \times [\text{mass flow rate}]^b$
 - 'A' is a correlation *factor*; 'b' is a correlation *coefficient*: $A = 131.5$, $b = 0.5$

Note that 'b' aligns with the theoretical relationship between air mass flow rate and intake diameter for a fixed (i.e. 'ideal') compressor entry velocity

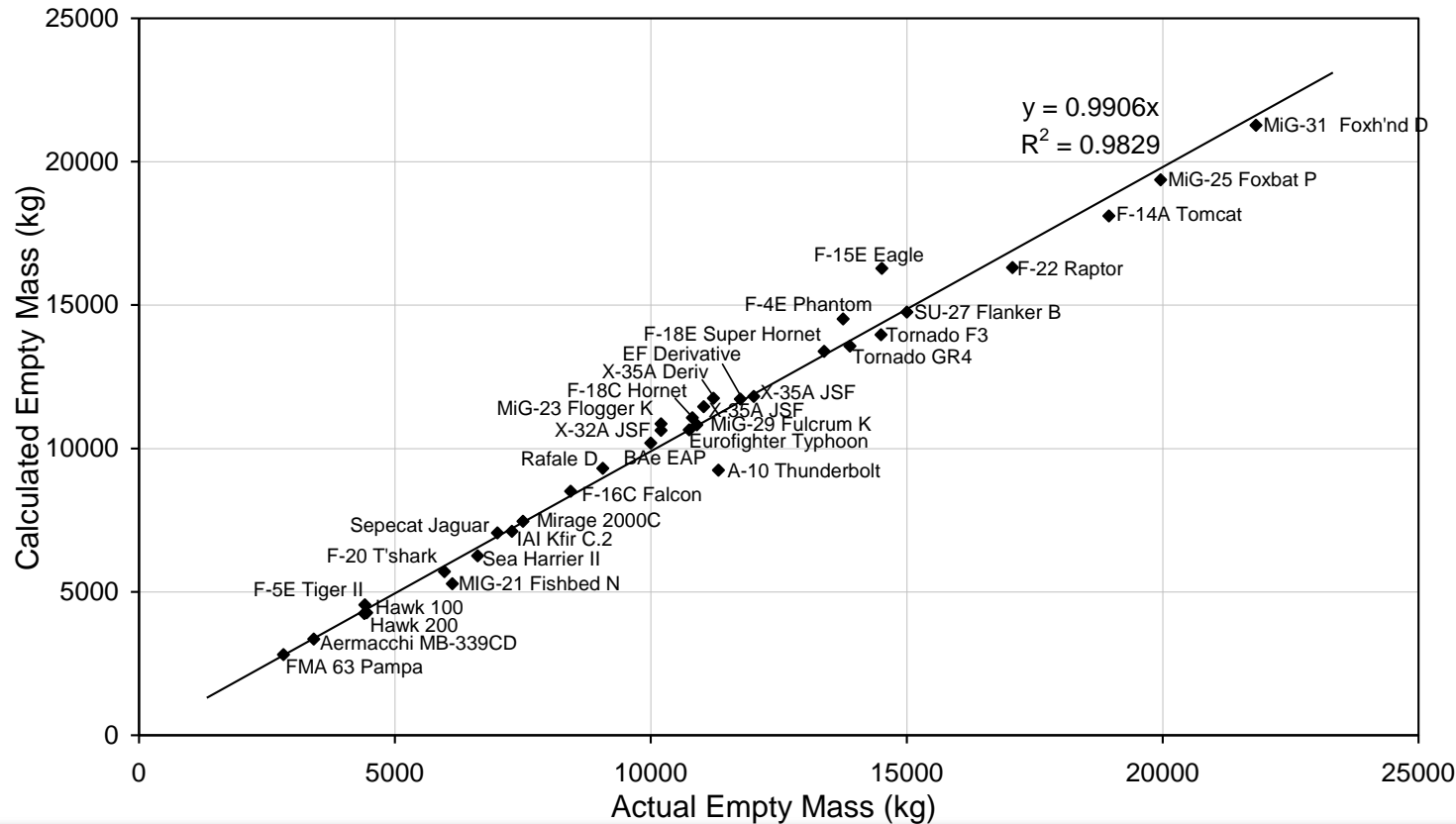


Attributes of an 'Excellent' CER (3) – Model Algorithms (2)

- Complex models may contain numerous inputs & coefficients, e.g.

$$\text{Wing Mass} = A \times \left[\frac{\text{Aspect Ratio}^{0.5} \text{Wing Area}^{1.5}}{\cos(\text{Quarter-chord Sweep})} \times \frac{(1+2 \text{ Taper Ratio})}{(3+3 \text{ Taper Ratio})} \times \frac{\text{Gross Mass}}{\text{Wing Area}} \times \text{Ultimate Load Factor}^{0.3} \times \left(\frac{\text{Design Diving Speed}}{\text{Thickness/Chord Ratio}} \right)^{0.5} \right]^{0.9}$$

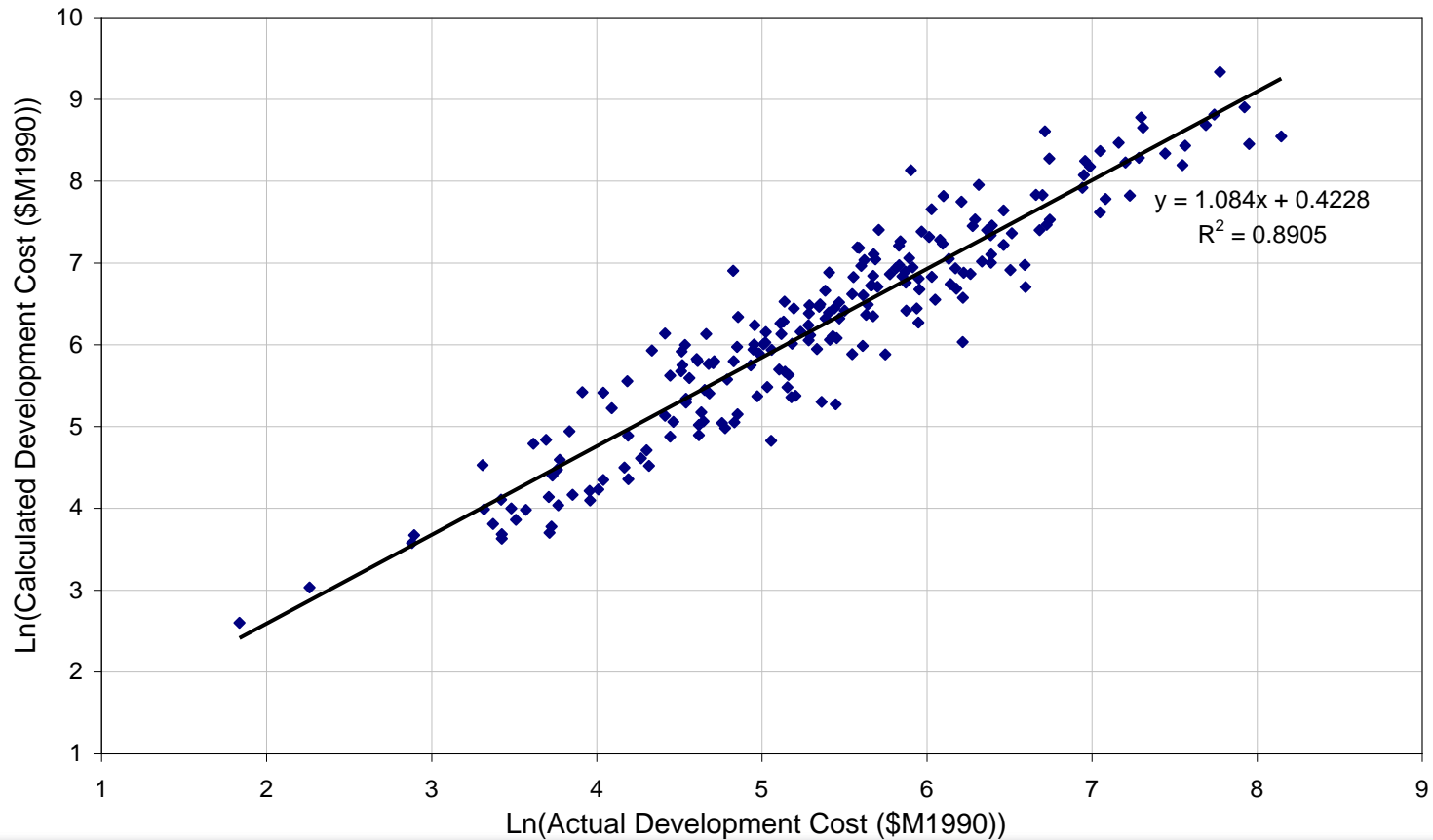
Empty Mass = Wing +
Fuselage + Empennage +
Undercarriage + Engine +
Cockpit + Fuel System +
APU + Flying Controls +
Hydraulics + Pneumatics +
Electrics + Env. Control +
Avionics + Sensors +
Trapped Fuel & Oil + Gun



Attributes of an 'Excellent' CER (3) – Model Algorithms (3)

- When conducting the regression analysis, choose the regression approach carefully

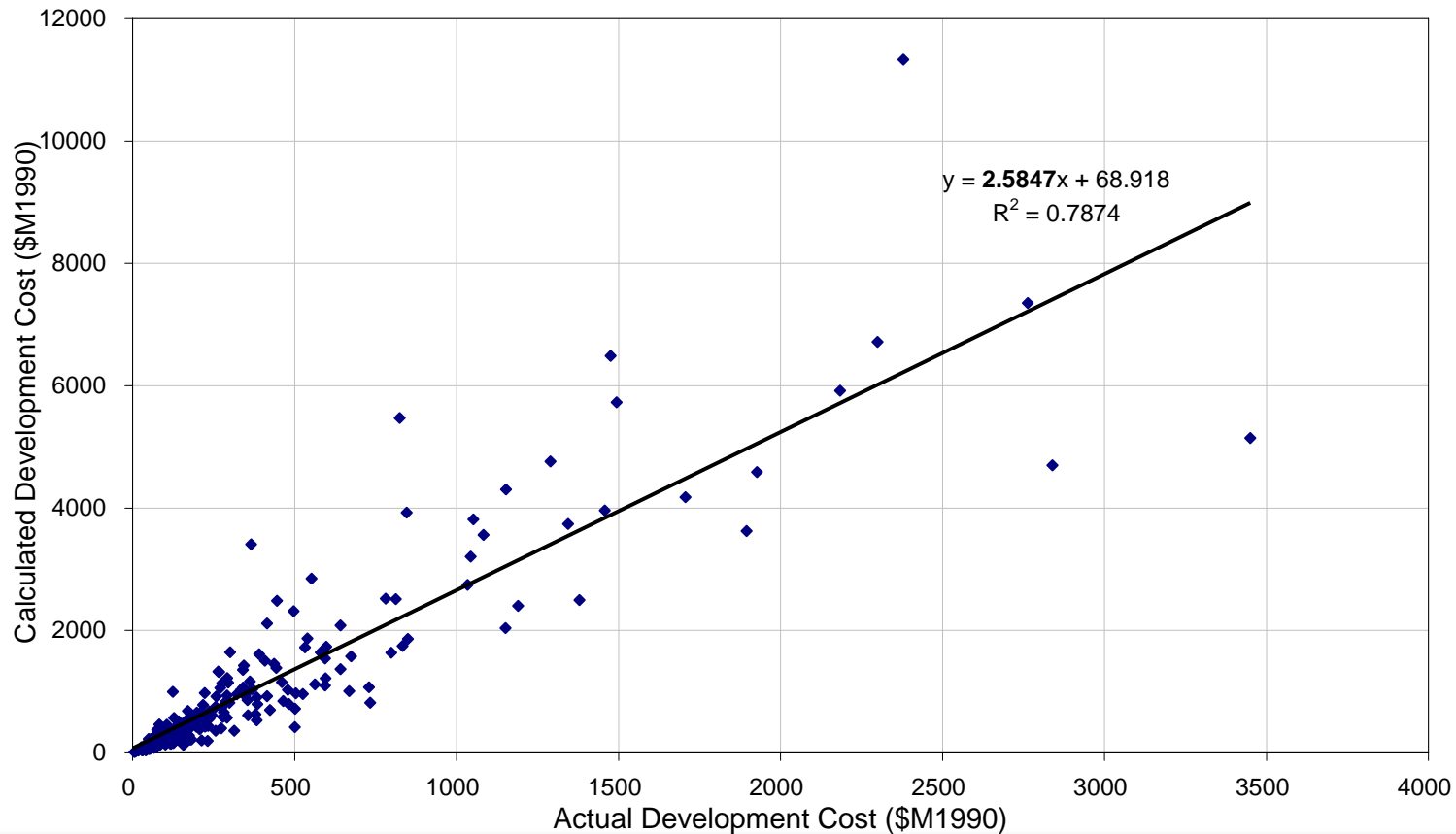
Logarithmic scales can produce **VERY** misleading results



Attributes of an 'Excellent' CER (3) – Model Algorithms (4)

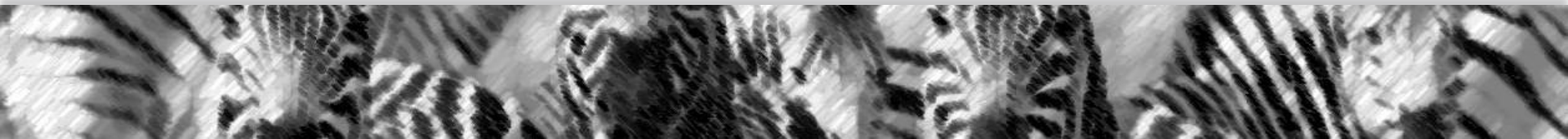
- When conducting the regression analysis, choose the regression approach carefully

Gradient is more than **twice** that expected & intercept is poor



Attributes of an 'Excellent' CER (3) – Model Algorithms (5)

- When conducting the regression analysis, choose the 'objective function' (of the internal optimisation) carefully
- R^2 is a very commonly used indicator of 'goodness of fit'
 - R^2 is the 'Coefficient of Determination' and is defined mathematically as:
$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (f_i - \bar{y})^2}$$
 where \bar{y} is the mean of the observed data.
 - Therefore, using R^2 as part of the regression will 'skew' the observed error to reduce with increasing values of x
 - This is not a problem for some programmes, but across a range of results, it may be preferable to have a consistent percentage error for all outputs
 - Many programmes will be measured on a '% error basis', not on total error
- Consider using absolute percentage error instead of R^2
 - Generates consistent *percentage* errors across all values



Attributes of an 'Excellent' CER (4) – Accuracy

- Displaying/exposing model accuracy is a 'two-edged sword'
 - Telling a customer that a model is 'inaccurate' can prevent its use (or significantly reduce their willingness to use it)
 - It is vitally important that we explain that no model can be 100% accurate
 - Must expose the level of modelling error in the modelling algorithm
 - Use multiple approaches to detail the inaccuracy and how it has been captured
- Example contained in later slide



Attributes of an 'Excellent' CER (5) – Model Documentation

- Model documentation must exist to explain the data, sources, normalisation, model structure, rationale, and applicability
 - Lack of documentation will guarantee that the model will not be used
- No-one likes producing model documentation, but it is a necessary evil
 - Must be complete, accurate, and synchronised with model developments
- Should be treated as a 'log-book' for the model
 - Can be a powerful source of evidence for use with Customers, CAAS, Scrutiny, etc.

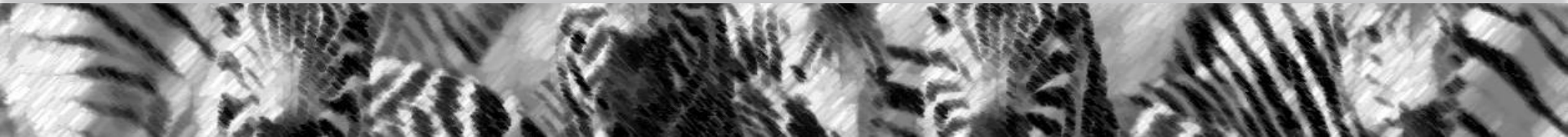


Attributes of an 'Excellent' CER (6) – Expert Users

- Parametric models can very quickly generate highly inaccurate outputs in the hands of an inexperienced/poorly trained user
 - All parametric models have a limited applicability
 - Understanding the model and its applicability is key to generating realistic, accurate outputs
 - The user must understand:
 - the data, sources, normalisation, and validation
 - the model structure and the factors and coefficients
 - the optimisation approach and whether the model was generated with R^2 , $|\%|$ error, or some other objective function
 - The units and economic conditions of the model output, and how to translate that value in to the value required to answer the question
- A 'poor user' will almost guarantee poor results

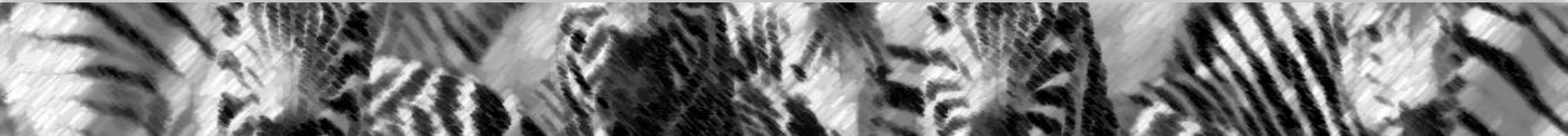


Example – Generation of an (almost) Excellent Aero Gas Turbine Development Cost Estimating Relationship



Data Sources and Processing

- Historical database obtained from US source
 - Contains 32 technical and programme parameter values for each engine, plus normalised development cost (1990 e.c.)
 - 307 data points made up of 10 engine types
 - Removed piston, rotary, auxiliary power units, automotive & rocket engines, leaving 236 data points
 - 17 mis-classified or highly outlying engines removed
 - 219 aero gas turbine engines remaining
 - 115 turbofan, 34 turbojet, 26 turboshaft, 44 turboprop
- 6 parameters difficult to estimate prior to start of development, therefore not used
 - Mass, Part count, Programme duration, Number of test hours, Number of stages, Number of rotating spools



Parametric CER Derivation

- Focused on the development of a model of the form:

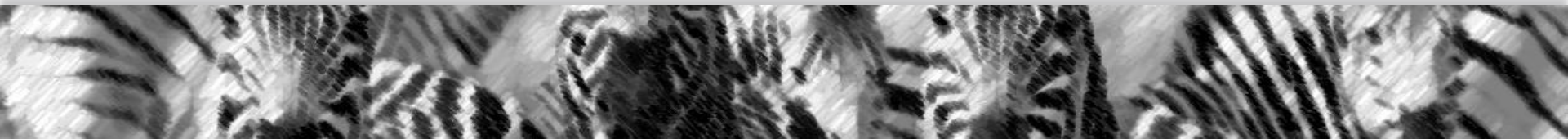
$$\text{Cost} = \alpha \times (\text{Parameter A})^a \times (\text{Parameter B})^b \times \dots \times y^{(\text{Parameter } \beta)} \times z^{(\text{Parameter } \gamma)}$$

- α = Constant \times Type \times Problem
- β, γ = dummy variable parameters
- a, b, y, z coefficients calculated using Generalised Reduced Gradient non-linear optimisation code (GRG2)
- Possible variables identified using a combination of engineering knowledge and statistical evidence
- Iterative process to identify significant variables to be included within the model & derive coefficient values
 - Relevance of variables checked using statistical checks, e.g. p-test, t-test...



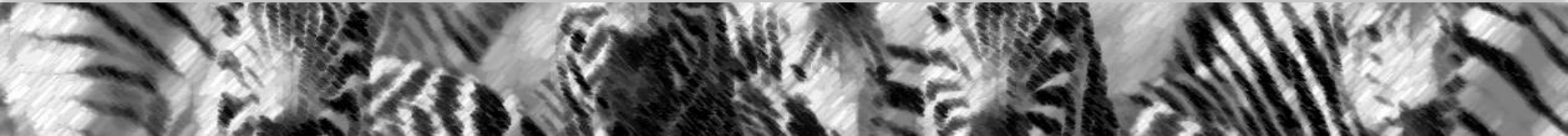
Common 'Errors' with Engine CERs

- Technical Parameter Separation (Engineering)
 - Thrust as a Technical Parameter
 - Thrust is a function of both engine size and technical advance. As such, it is not a good technical parameter for this model, despite statistical evidence
- Cross-correlation (Statistics)
 - Bypass Ratio (BPR) and Mass Flow Rate (WA)
 - Large civil aircraft have high WA and BPR. Small turbojets have low WA and BPR. Small engines are unlikely to have high BPR, as this results in a small, inefficient engine core
 - Turbine Entry Temp. (TET) and Overall Pressure Ratio (OPR)
 - Engines require high OPR to make use of efficiencies generated through high TET. Both are measures of technological advance, and should not appear in the same equation together

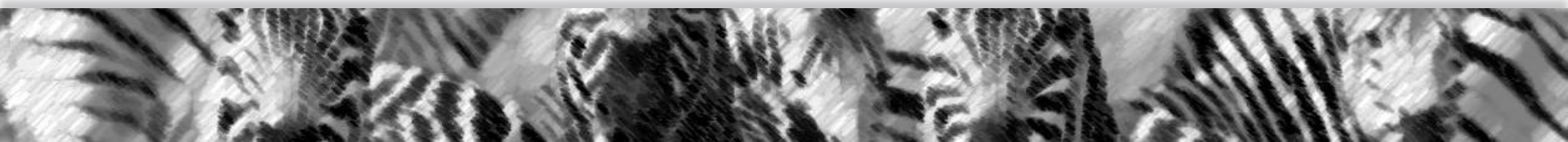
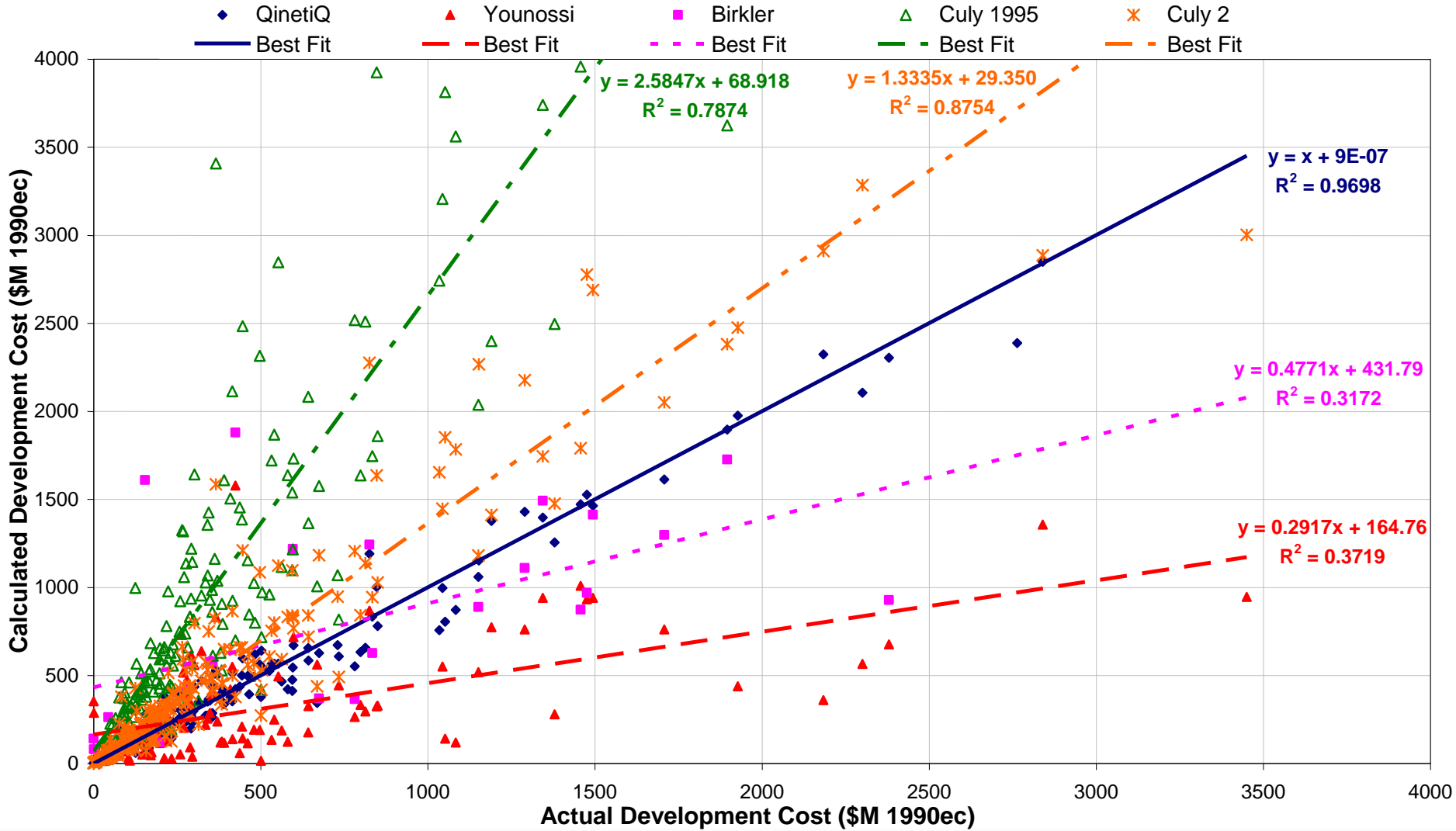


Results – Model Overview

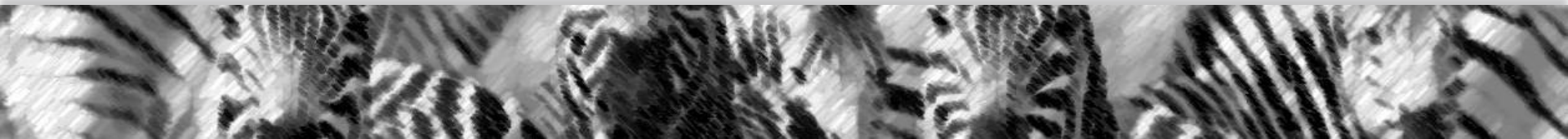
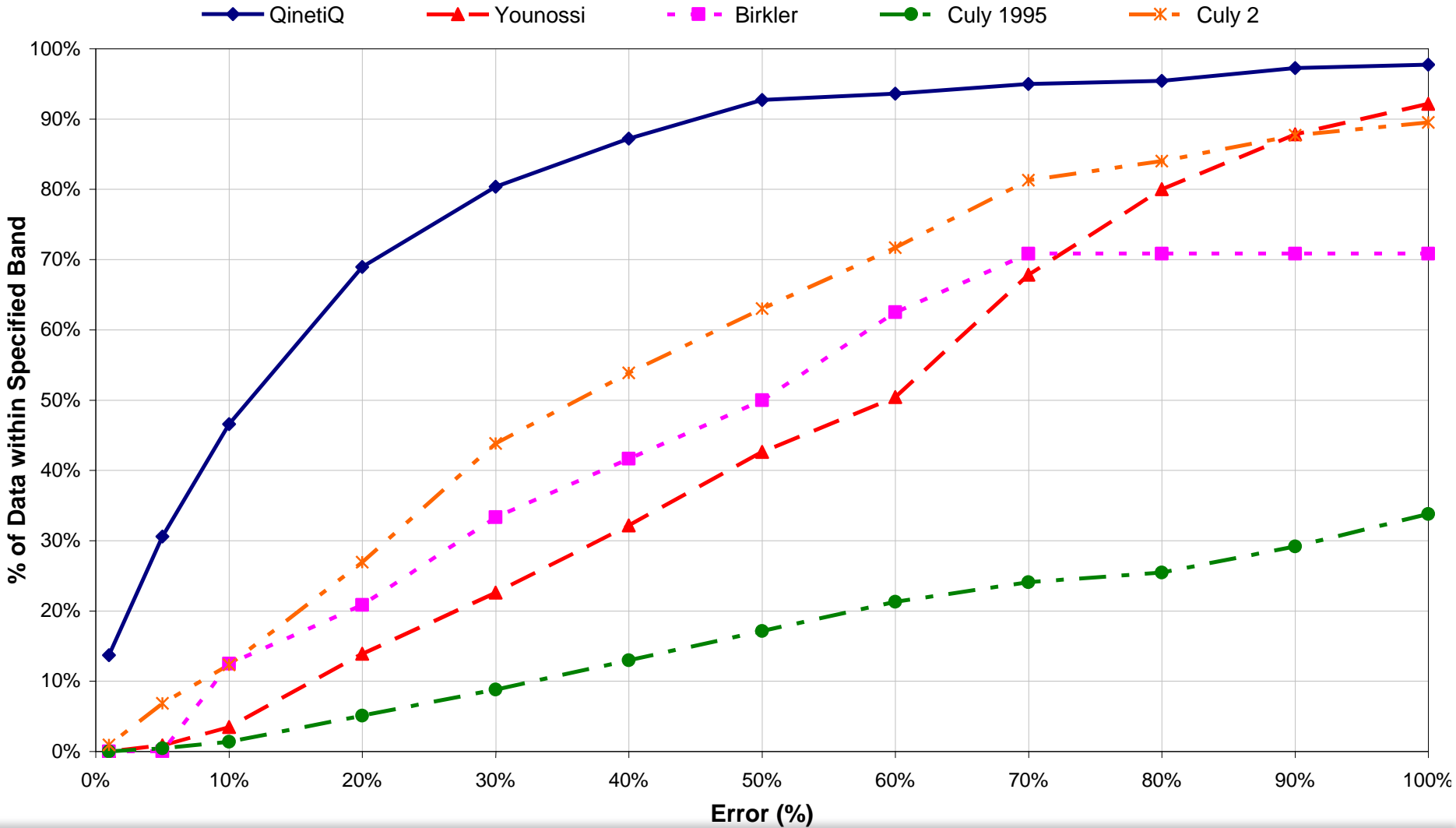
- Best fitting model was found to contain:
 - Engine Type, Problem, WA, OPR, Joint Venture, No: test engines, Mach
- Includes factors for:
 - Complex Programmes, Military Engines, Country of Origin, “Skunk Works” Programmes
- ‘Type’ factor gives relative costs of Turboprop and Turboshaft engines, and Turbofan & Turbojet engines, with and without reheat
- ‘Problem’ factor gives cost variation according to novelty of proposed engine design
 - 11 levels from ‘All-New Design’ to ‘De-Rate’
- All factors are ‘trade-able’ – which is most beneficial?



Results & Comparison with Published Models



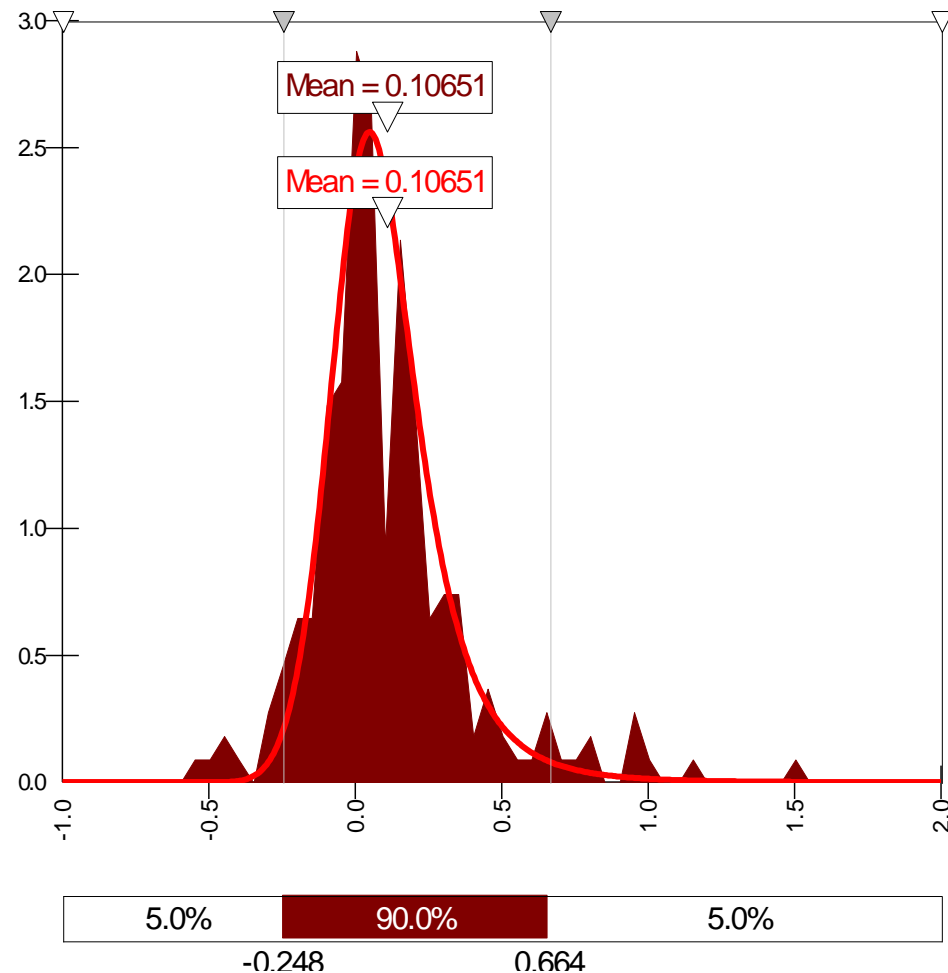
Accuracy & Comparison with Published Models



Accuracy – Capture & Addition to Model Algorithm

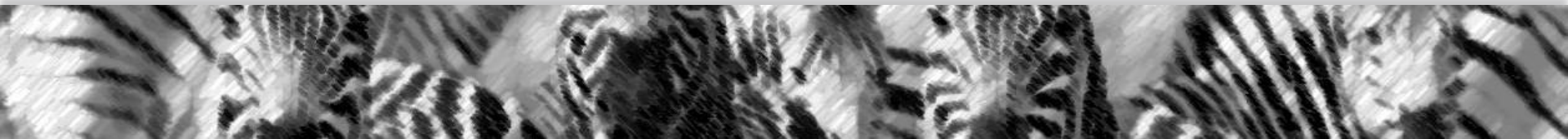
- Model ‘inaccuracy’ captured using ‘best-fit distribution’ tools within COTS software product
 - Gives clear link between ‘accuracy’ (e.g. R^2) and confidence levels
- Model algorithm updated to include ‘inaccuracy’ as part of the stochastic modelling process
 - Allows inclusion of model ‘inaccuracy’ as well as input ‘uncertainty’

Engine Development Error Data & Best-Fit Distribution



Conclusions

- Parametric CERs are used inconsistently across MoD projects
- The generation of Excellent CERS requires:
 - a large database of high-quality data points;
 - accurate and consistent data ‘cleansing’ and normalisation;
 - inclusion of parameters that demonstrate theoretical/engineering links;
 - an appropriate level of complexity to reflect the engineering reality;
 - adoption of a powerful regression technique that does not hide error;
 - selection of an ‘objective function’ that is relevant for the question/model;
 - capture and usage of model ‘inaccuracy’ to enhance the model output;
 - comprehensive documentation that explains all of the above (and more!);
 - trained, expert users to ensure that the model is applied correctly.
- Generating Excellent CERs *may* increase their use within MoD projects
 - Doing so would increase quality and confidence in resulting cost forecasts



Burchelli Consulting Limited

Few things in life are as simple as Black and White

