Engineering Model Independence
A Strategy to Encourage Independence Among Models

Zachary Pirtle*

Department of Engineering Management and Systems Engineering

The George Washington University

Collaborators: Zoe Szajnfarber, Jay Odenbaugh

*This work does not necessarily reflect the views of NASA or the United States Government, and opinions expressed belong to the author.

Related challenges

- Going beyond models – doing empirical and theoretical work that is independent from models – is a holistic challenge
- Understanding how models differ from themselves can be quite difficult
- **What does it mean for models to be independent from one another?**

Independence among Climate Models

- Pirtle et al 2010 was part of early efforts to push for understanding how models are different from one another (Parker)
- Drawing on Richard Levins' work on model pluralism ("truth lies at the intersection of independent lies"), my 2019 systems engineering dissertation created a heuristic framework for individuating models from one another

ELSEVIER

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/envsci

**What does it mean when climate models agree? A
assessing independence among general circulation**

Zachary Pirtle [a], *Ryan Meyer* [a,b,*], *Andrew Hamilton* [b]

[a] Consortium for Science, Policy, and Outcomes, Arizona State University, PO Box 875603, Tempe, AZ 85287-5603, U

[b] School of Life Sciences, Arizona State University, PO Box 874501, Tempe, AZ 85287-4501, USA

ARTICLE INFO

Published on line 18 May 2010

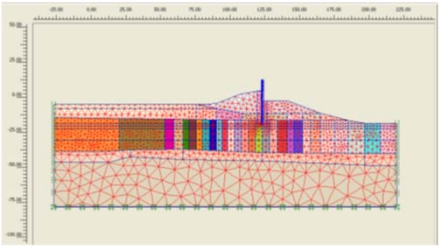
Keywords:
Models
Prediction
Climate science
Science policy
Robustness

ABSTRACT

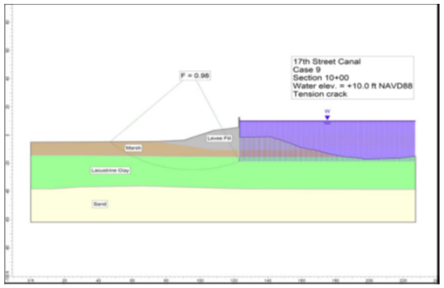
Climate modelers often use agreement among multiple general circu
as a source of confidence in the accuracy of model projections. Howe
model agreement depends on how independent the models are fr
climate science literature does not address this. GCMs are independe
dent on one another, in different ways and degrees. Addressing
independence is crucial in explaining why agreement between mode
dence that their results have basis in reality.

Case Example on Levee failure models

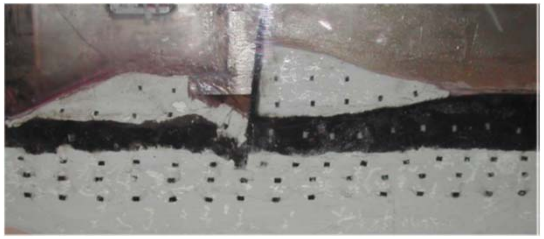
Three models to assess Levee failure



L1: Finite Element Analysis model



L2: Limit Equilibrium Analysis model



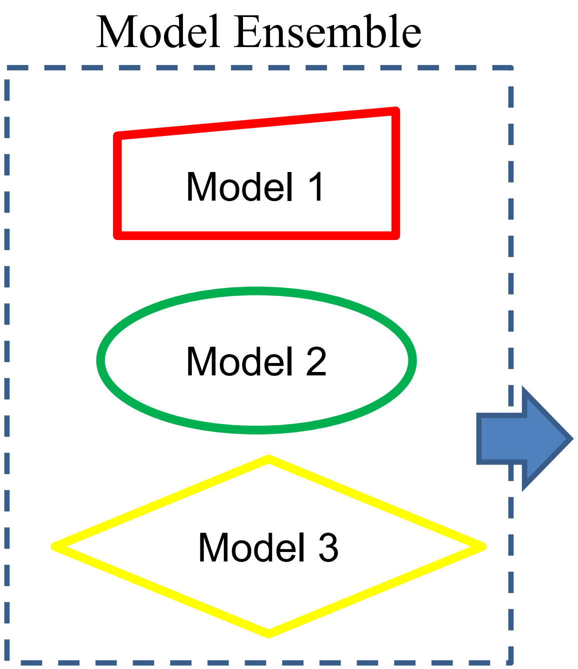
L3: Centrifuge model

In this example, all three models agreed on the cause of the levee failure, with this agreement being deemed more significant due to the independence among the models being utilized.

To develop dimensions of independence, iteratively compared each model using an open coding approach, identifying key differences

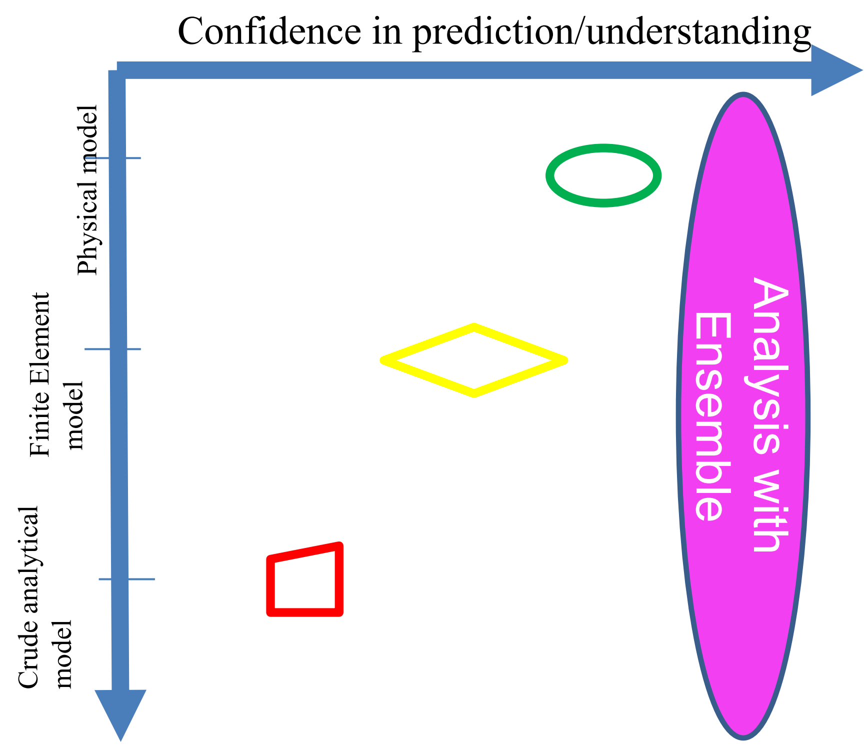
Example: Idealization in LEA model removes many details of possible failure, uses broad areas of strength, whereas FEA decomposes the world into more grid cells
Resulting independence dimensions: Model structure, idealization/causal logic

Model Independence is differences among a group of models



Who? Model teams, integrated analyst

Knowledge/ Prediction about aspect of a system



Each model is an analysis type being pursued to observe the same (or part of the same) system

- Amongst the broader 'ensemble,' a model is independent in the ways and degrees it is different from the other projects

For model agreement to be beneficial, the models must be independent from one another

Sufficient independence among the models?

Pursued deep case studies to develop a framework of how real-world models are independent from one another

Multiple Dimensions by which models can be independent

Categories
Causal Logic/ Idealization
Comprehensiveness
How Model Represents the System
Manageability
Social Context
What the Model Does
Model result agreement

Multiple Dimensions by which models can be independent

Categories	Dimensions of Model Independence
Causal Logic/ Idealization	
Comprehensiveness	
How Model Represents the System	Structure/scope assumed by the model
	Parameters
	Data included in the model.
Manageability	
Social	
What the Model Does	
Model result agreement	

Applying the Framework

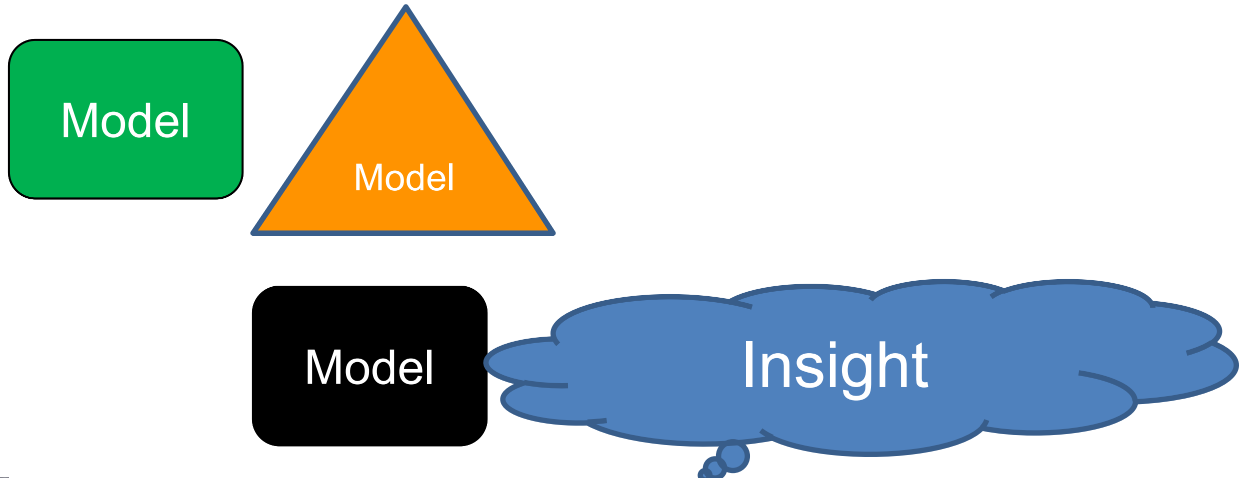
(or, how Model differences shape discovery as well as justification)

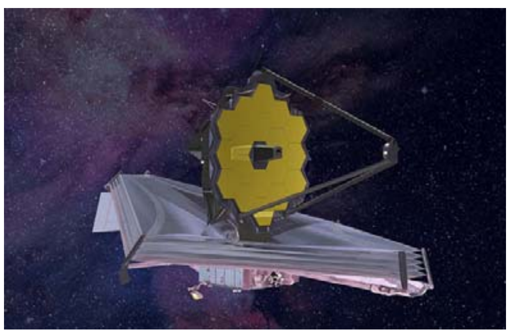
GW

Multiple Cost and Schedule Models to assess NASA projects

- NASA requires multiple model assessments of the expected cost and schedule of new satellites and models
- Formal commitments for cost and schedule for rockets are based on having the results of the models be assessed by managers

Multiple Cost and Schedule Models to assess NASA projects





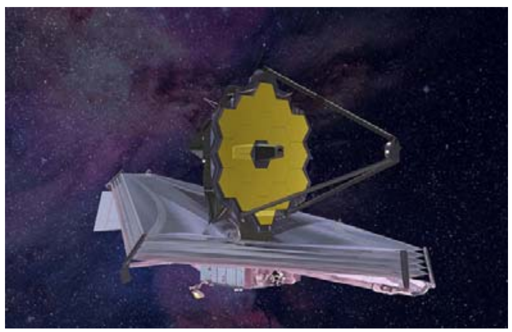
(Hacking, "Do we See Through Microscopes")

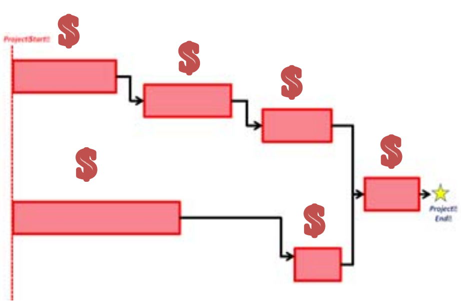
Model results = insight, actionability, predictions, confidence in agreement claims

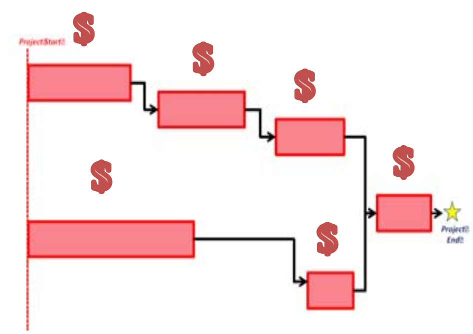
Multiple Cost and Schedule Models to assess NASA projects

- NASA requires multiple model assessments of the expected cost and schedule of new satellites and models
- Formal commitments for cost and schedule for rockets are based on having the results of the models be assessed by managers

(Hacking, “Do we See Through Microscopes”)

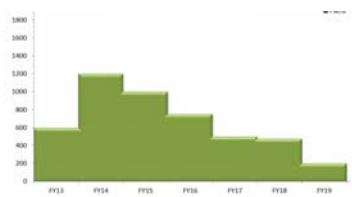




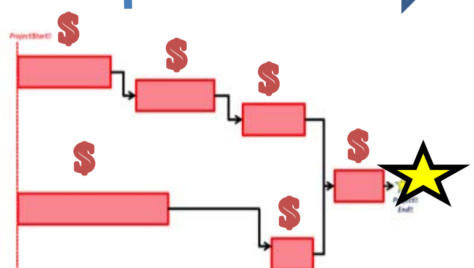


Explain Different Results Based on Independence

Independence



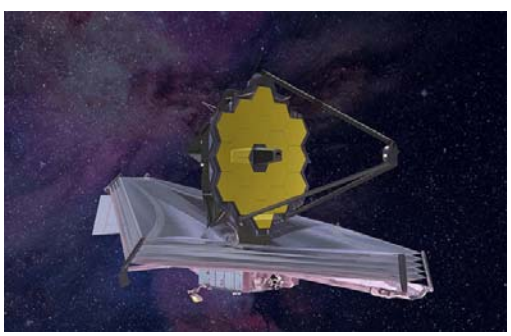
Model 1



Model 2

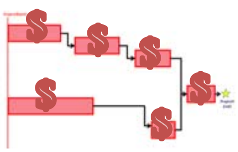
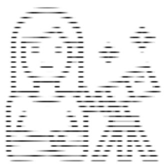
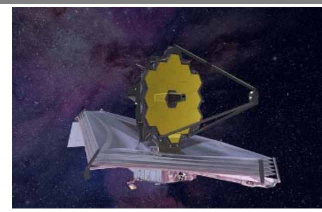
	Model 1	Model 2
Prog Challenge #1: Unknown Scope	Forces assessment	Neutral
Prog Challenge #2: Budget cap	Forces assessment	Blindspot
Prog. Challenge #3: Schedule logic	Blindspot	Forces assessment
…other prog. challenges	…	…

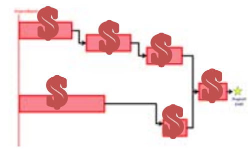
Different Insight?



Model results = insight, actionability, predictions

Framework Applied to NASA Independent Assessment (Chapter 9.5, using GAO 2018)



Dimensions of Model Independence	Differences between SRB and Project Model Assessments
Causal Logic/ Idealization	No independence between models
Comprehensiveness	No independence between models
Structure/scope assumed by the model	No independence between models Independent reviewers tweak and adjust schedule, breaking out higher resolution in key areas. Usually, 80%+ of inputs are in common
Parameters	**Moderate amount of independence** Independent reviewers often make most of their changes to uncertainty parameters and by addition of risks. These can lead to significant date changes in model outputs
Data included in the model.	No independence between models
Comprehensibility	No independence between models
Disciplinary background of Assessor	No independence between models
Function	No independence between models
Goals of Assessor Team	Context dependent. Can be significant, but generally no independence between models
Accuracy/Agreement	No independence between models determinable

Causal logic/idealization strongly shapes implicit idea of how engineers even manage and control cost/schedule

<u>To identify challenges</u>: structure, parameters, inputs

<u>For mitigating challenges</u>: causal logic/idealization and parameters are key

Conclusion

- Model independence is multi-dimensional
- Rigorous definitions of independence must be context specific
- My goal has been to have a heuristic framework that can help a practitioner disambiguate across multiple models

Techné: Research in Philosophy and Technology ISSN: 1091-8264
Online First: June 29, 2018 DOI: 10.5840/techne201862283

Engineering Model Independence: A Strategy to Encourage Independence Among Models

Zachary Pirtle, Jay Odenbaugh, Andrew Hamilton, and Zoe Szajnfarber

Abstract: According to population biologist Richard Levins, every discipline has a "strategy of model building," which involves implicit assumptions about epistemic goals and the types of abstractions and modeling approaches used. We will offer suggestions about how to model complex systems based upon a strategy focusing on

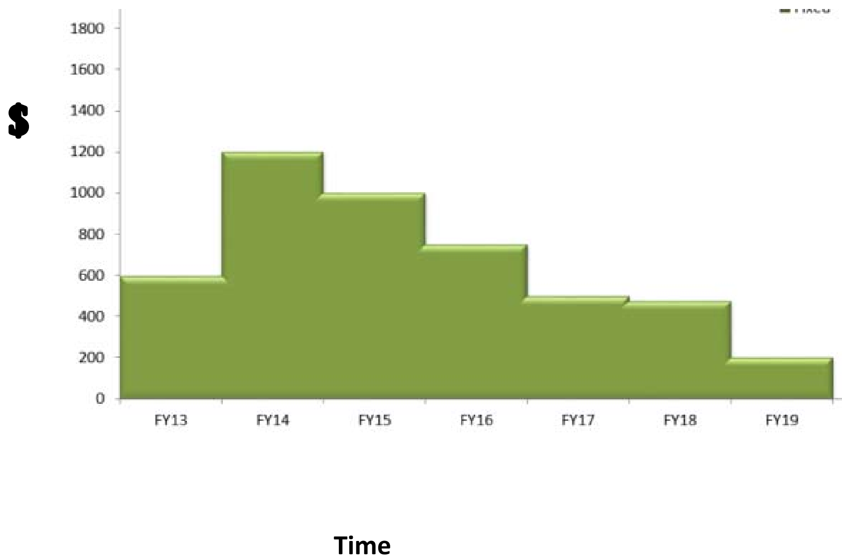
- For policy relevant problems, need to explain issues of model independence to managers and policy decision-makers

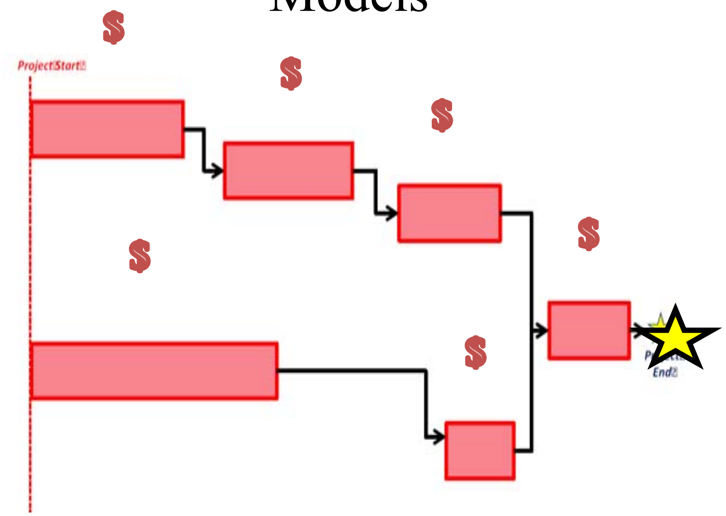
Backup

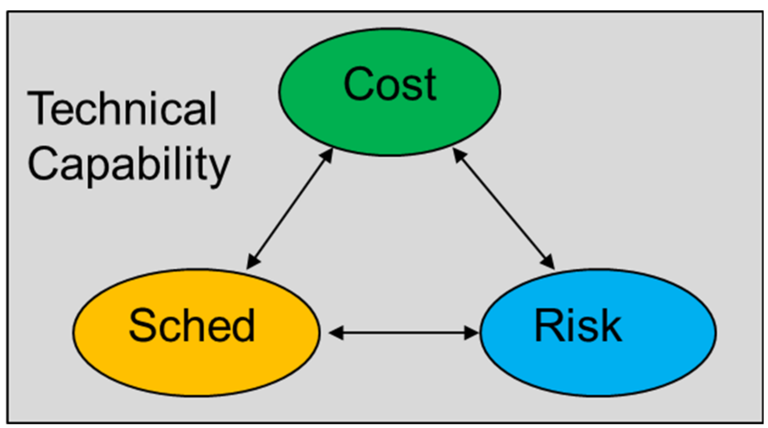
Multiple Cost and Schedule Models to assess NASA projects

Cash Flow (CF) Models

Joint Cost and Schedule Confidence Level (JCL) Models







CF to JCL Independence

Inputs, Assessments, Model/Modeler Characteristics

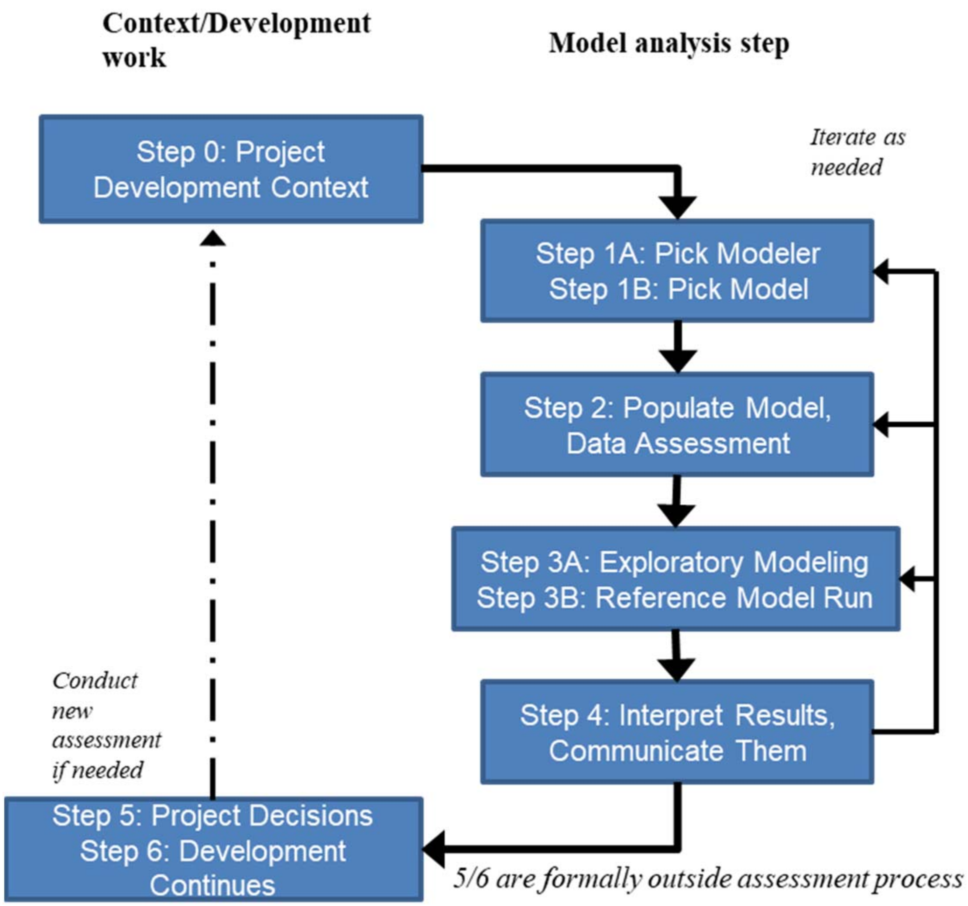
Categories	Dimensions of Model Independence	Scale type	JCL/CF relevant metric (italics if disparate)	Qual score	Differences Between Cash Flow and JCL models
Causal Logic/ Idealization	Causal Logic/ Idealization	Disp	*[extent of shared logic/ idealization]*	Med	Key equation for CF balances annual cost, JCL traces schedule logic to completion. However, both use the same overall theory for programmatic analysis and project management, including assumed tradeoffs between cost, schedule and risk. Technical work is idealized as cost for CF, as schedule for JCL .In PP&C terms, these are almost different ontologies.
How System is Represented	H1: Structure/scope assumed by the model	Cont.	Size. Same content, degree of fidelity in describing it	Med	JCL generally has hundreds of lines of schedule activities and associated logic, CF uses just one/few, uses annual cost totals. Both purport to include all project content inside of the model, albeit at different levels of resolution.
	H2: Parameters	Disp.	*[different parameters/key inputs to system]*	Med	CF has unique parameters on inefficiency penalty, fixed cost inputs. JCL has parameters for correlation, and also includes uncertainty as inputs.
	H3: Data included in the model.		Same content, degree of fidelity in describing it	Low	Both use the full set of Project Baseline Data on top level costs/schedule/risks, albeit much more lower level data nuances are included in JCL model as cash flow includes data at an aggregate summary level.
Social	S1: Epistemic culture, Organizational structure	Disp.	*[different training]*	Low	Little difference in disciplinary background, as analysts using the model usually come from similar backgrounds, and can be used at multiple places in an organization (project, enterprise) levels.
	S2: Project Team/Institutional Goals	Disp.	*[different goals]*	Low	Different analysts can have different goals (tell truth, advocacy, fame), but the way in which the model is used is not fundamentally shaped by this. Different goals is most likely to manifest in uncertainty assumptions by each model.
Function	W1: Core function, what the model does	Cont. + Disp.	Predicted launch date *[different function kinds]*	Low	Both CF and JCL make predictions of a project's launch date and cost, and can assign probabilistic values to those predictions.
Realism	R2: Comprehensiveness;	Cont.	Magnitude of detail (representation) about a system	Med	Both models include full program content at an aggregate level, but JCL resolves much more of it in the model, including schedule logic connections. There is much greater resolution and information incuded in the JCL model.
	R3: Model result agreement/ accuracy	Cont.	% deviation of predicted results vs actual results	Low	Both models assumed to have reasonable predictive power. No serious benchmarking to quantify accuracy has been done of either model.
Manageability	M1: Comprehensibility of analysis	Cont.	Manager/analyst perception that they understand totality of analysis	High	CF can be understood in an hour or so of review, JCL inputs much more complicated. The causal explanation of how the models work and how they come to a given result is much easier to understand in CF.

Zachary Pirtle, PhD

CF to JCL Independence

Categories	Dimensions of Model Independence	Scale type	JCL/CF relevant metric	Qual score	Differences Between Cash Flow and JCL models
Realism	R1: Idealizations	Disp.	Extent of shared idealization approaches	High	Technical work is idealized as cost for CF, as schedule for JCL .In PP&C terms, these are almost different ontologies.
	R2: Comprehensiveness;	Cont.	Magnitude of detail (representation) about a system	Med	Both models include full program content at an aggregate level, but JCL resolves much more of it in the model, including schedule logic connections. There is much greater resolution and information incuded in the JCL model.
	R3: Model result agreement/ accuracy	Cont.	% deviation of predicted results vs actual results	Low	Both models assumed to have reasonable predictive power. No serious benchmarking to quantify accuracy has been done of either model.
Manageability	M1: Comprehensibility of analysis	Cont.	Manager/analyst perception that they understand totality of analysis	High	CF can be understood in an hour or so of review, JCL inputs much more complicated. The causal explanation of how the models work and how they come to a given result is much easier to understand in CF.

How models *see*



Mix of Strengths and Weaknesses, driven by independence dimensions

Strong
Neutral
Weak

	Prog. Challenge	Cash Flow (normal low detail)		JCL (normal high detail)	
		Identify	Mitigate	Identify	Mitigate
1	**Unknown scope**	Force assessment • Model structure • Parameters	Directly replans • Causal Logic/ Idealization	Neutral • Model Structure	Failure to simulate • Causal Logic/ Idealization • Manageability
2	**Execution budget challenges**	Force assessment • Model stru • Parameters	Directly replans	Blindspot	Failure to simulate • Causal Logic/ Idealization Parameters Manageability
3	**Schedule logic**	Blindspot • Causal Log Idealizatio • Comprehen • Data Input			irect replan Causal Logic/ Idealization Comprehensiveness Model Structure
4	**Poor perform.**	Neutral • Causal Logic/ Idealization • Data Input	• Causal Logic/ Idealization • Data Input	• Causal Logic/ Idealization • Data Input	eutral • Causal Logic/ Idealization • Data Input
5	**Technical Issues**	Force assessment • Data inputs	Neutral • Data inputs	Force assessment • Data inputs	Direct replan • Causal Logic/ Idealization

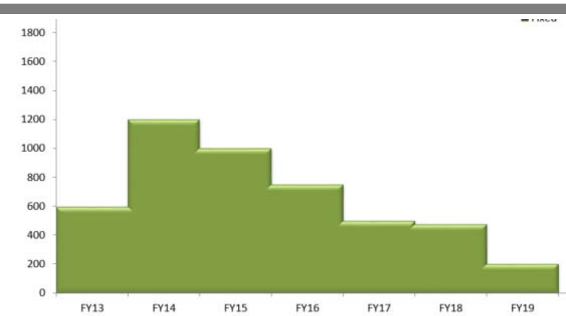
Some shared results due to similarities between models

Six overall dimensions shape results - Not all due to structure and causal logic/idealization

Independent: Significant Difference in Causal Logic/Idealization

CF

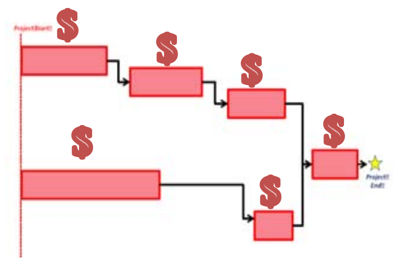
(schedule tied to cost) $



Time

JCL

(cost-loaded schedule)



Time

<u>Significant Difference in Causal Logic/idealization</u>:

Immeasurable; ‘different chunks of physics’ (Hacking 1983),

Hard to measure/define a greater difference:

Akin to Mass and Volume:

Gravity applies to mass,

friction applies to surface area~volume.

JCL idealizes technical work primarily as schedule;
CF idealizes as costs.