Causality and Uncertainty: A New Wave for Cost Estimation

Robert Stoddard

Mike Konrad

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213

Document Markings

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

DM19-0804

[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution. Please see Copyright notice for non-US Government use and distribution.

Agenda

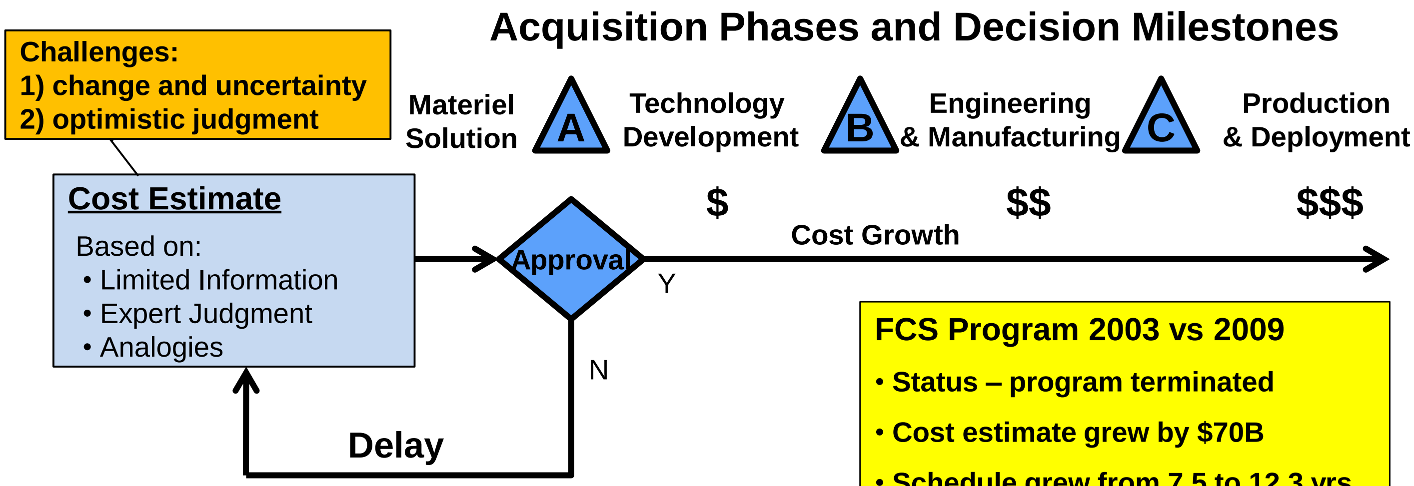
Original SEI Cost Research Motivation

The QUELCE Research Project Solution

SEI Causal Learning Research

Call to Action

Original SEI Cost Research Motivation



FCS Program 2003 vs 2009

- **Status – program terminated**
- **Cost estimate grew by $70B**
- **Schedule grew from 7.5 to 12.3 yrs**
- **Lines of code grew from 34M to 114M**

Source: GAO-10-406

Ground Combat Vehicle Delay Due to Reconciling Cost Estimates

- **4 months delay in obtaining approval to proceed**
- **30% discrepancy in cost estimates, Army v independent cost estimate**
- **Rework to conduct a new Analysis of Alternatives and to produce a new cost estimate**

Source: GAO-12-181T

Weapon Systems Acquisition Reform Act (WSARA) 2009 requirements for Milestone A approval.

The QUELCE Solution

Step 1: Identify Change Drivers and States

1. Identify Change Drivers & States → 2.Reduce complexity of Cause and Effect relationships via matrix techniques → 3. Assign Conditional Probabilities to BBN Model → 4. Calculate Cost Factor Distributions for Program Execution Scenarios → 5. Monte Carlo Simulation to Compute Cost Distribution

Change Driver	Nominal State	Alternative States				
Scope Definition	Stable	Users added	Additional (foreign) customer	Additional deliverable (e.g. training & manuals)	Production downsized	Scope Reduction (funding reduction)
Mission / CONOPS	As defined	New condition	New mission	New echelon	Program becomes Joint	
Capability Definition	Stable	Addition	Subtraction	Variance	Trade-offs [performance vs affordaility, etc.]	
Funding Schedule	Established	Funding delays tie up resources {e.g. operational test}	FFRDC ceiling issue	Funding change for end of year	Funding spread out	Obligated vs. allocated funds shifted
Advocacy Change	Stable	Joint service program loses particpant	Senator did not get re-elected	Change in senior pentagon staff	Advocate requires change in mission scope	Service owner different than CONOPS users
Closing Technical Gaps (CBA)	Selected Trade studies are sufficient	Technology does not achieve satisfactory performance	Technology is too expensive	Selected solution cannot achieve desired outcome	Technology not performing as expected	New technology not testing well
• • •	~~~~	~~~~	~~~~	~~~~	~~~~	~~~~

Domain-Specific Program Change Drivers Identified

The QUELCE Solution

Step 2: Reduce Cause and Effect Relationships via Design Structure Matrix Techniques

1. Identify Change Drivers & States → 2. Reduce complexity of Cause and Effect relationships via matrix techniques → 3. Assign Conditional Probabilities to BBN Model → 4. Calculate Cost Factor Distributions for Program Execution Scenarios → 5. Monte Carlo Simulation to Compute Cost Distribution

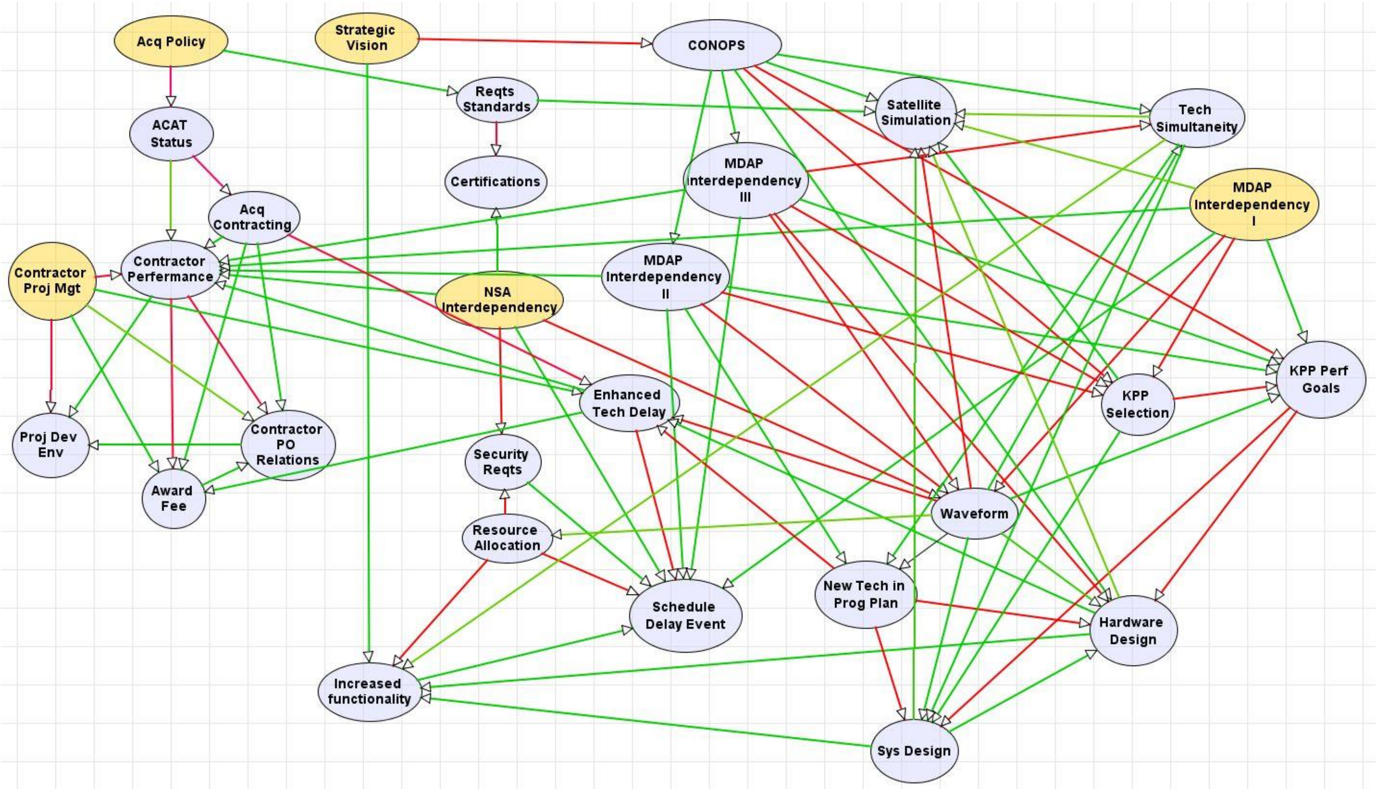
Change Drivers - Cause & Effects Matrix

Causes \ Effects	Mission / CONOPS	Change in Strategic Vision	Capability Definition	Advocacy Change	Closing Technical Gaps (CBA)	Building Technical Capability & Capacity (CBA)	Interoperability	Systems Design	Interdependency	Functional Measures	Scope Definition	Functional Solution Criteria (measure)	Funding Schedule	Acquisition Management	Program Mgt - Contractor Relations	Project Social / Dev Env	Prog Mgt Structure	Manning at program office
Mission / CONOPS			3								3							
Change in Strategic Vision			3	3		3			2				2					
Capability Definition								3					0	2	1	1	0	0
Advocacy Change											2			1			1	1
Closing Techn																		
Building Tech																		
Interoperabili																		
Systems Des																		

Capturing interrelationships among change drivers and reducing the complexity of the network

The QUELCE Solution

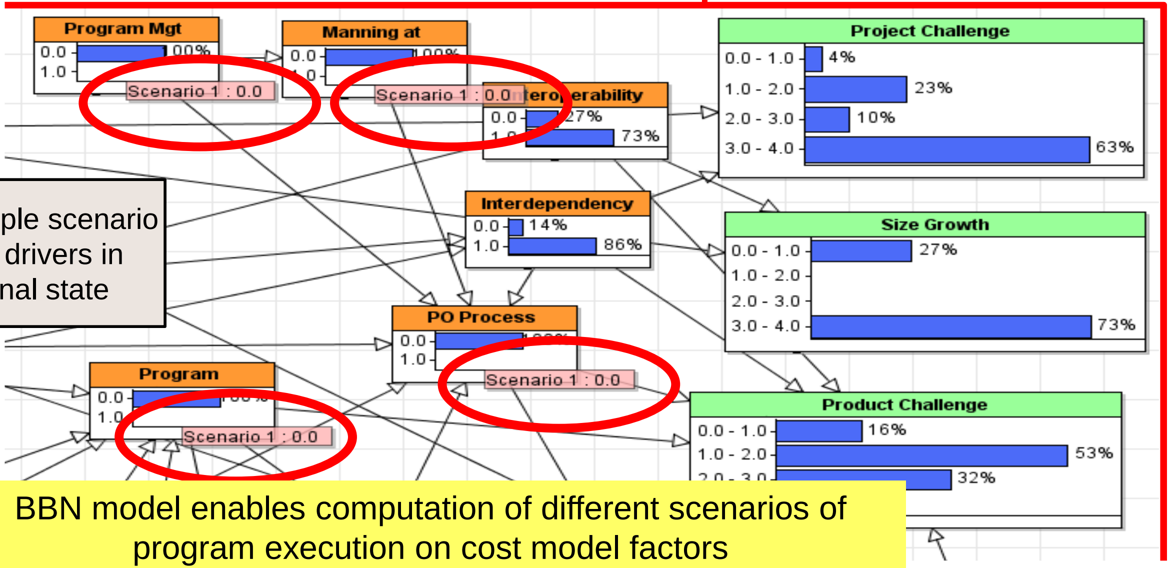
Step 3: Assign Conditional Probabilities to BBN Model



The QUELCE Solution

Step 4: Calculate Cost Factor Distributions for Program Execution Scenarios

1. Identify Change Drivers & States → **2. Reduce Cause** and Effect Relationships via Dependency Structure Matrix techniques → **3. Assign Conditional Probabilities to BBN Model** → **4. Calculate Cost Factor Distributions for Program Execution Scenarios** → **5. Monte Carlo Simulation to Compute Cost Distribution**



An example scenario with 4 drivers in nominal state

BBN model enables computation of different scenarios of program execution on cost model factors

The QUELCE Solution

Step 5a: Connecting BBNs to Cost Estimation Models

1. Identify Change Drivers & States → 2.Reduce complexity of Cause and Effect relationships via matrix techniques → 3. Assign Conditional Probabilities to BBN Model → 4. Calculate Cost Factor Distributions for Program Execution Scenarios → 5. Monte Carlo Simulation to Compute Cost Distribution

Understand and analyze cost model input factors

Group similar input factors based on empirical analysis in task 3.

Use empirical analysis from Repository as basis to map scale (XL … EH) of original cost model input factors to scale (1…5) of BBN output factors

COCOMO Parameter	
Scale Factors	PREC
	FLEX
	RESL
	TEAM
	PMAT
Effort Multipliers	PERS
	RCPX
	PDIF
	PREX
	FCIL
	RUSE
	SCED

Product Challenge factors (1=low…5=high)

COCOMO Parameter		XL	VL	L	N	H	VH	EH
Scale Factors	PREC			1	3	5		
	FLEX		1	2	3	5		
	RESL	1	2	3	4	5		
Effort Multipliers	RCPX			1	2	3	4	5
	PDIF			1	5			
	RUSE				1	3	5	

Project Challenge factors (1=low…5=high)

COCOMO Parameter		XL	VL	L	N	H	VH	EH
Scale Factors	TEAM	1	3	5				
	PMAT		1	2	3	4	5	
Effort Multipliers	PERS			1	3	5		
	PREX			1	2	3	4	5
	FCIL				1	3	5	
	SCED	1	3	5				

The QUELCE Solution

Step 5b: Monte Carlo Simulation to Compute Cost Distribution

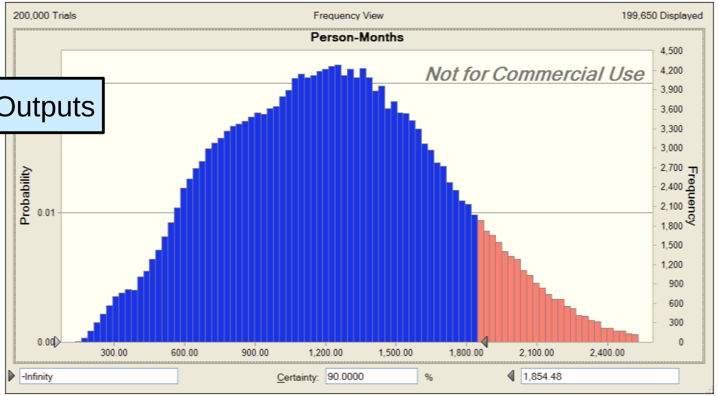
1. Identify Change Drivers & States →
2. Reduce complexity of Cause and Effect relationships via matrix techniques →
3. Assign Conditional Probabilities to BBN Model →
4. Calculate Cost Factor Distributions for Program Execution Scenarios →
5. Monte Carlo Simulation to Compute Cost Distribution

Monte Carlo simulation using program change factor distributions uses uncertainty on the input side to determine the cost estimate distribution

	A	B	C	D
1	Effect			
2	Product Challenge	5		
3	Project Challenge	4		
4	Estimated Size (KSLOC)	50		
5	Product Challenge factors		5	
6	COCOMO Parameter		XL	VL
7	Scale Factors	PREC		4
8		Val		6.2
9		FLEX		5

BBN Outputs

Mapped COCOMO value



QUELCE Application and Challenge

Space program piloted QUELCE just after a recent cost estimate

- Anticipated 66 change drivers
- Realized 33 change drivers not considered in latest basis of estimate (BOE)
- Reported only 2/3 of change drivers in BOE were expected to have off-nominal performance
- SEI concluded at least 90% of historical cost growth events could have been identified and prevented by QUELCE

QUELCE workshops in past several years:

- Produced 200-400 change drivers
- Confirmed complexity explosion due to human judgement issue

Experts tend to attribute correlation as cause-effect!

SEI Causal Learning Research Addresses the Challenge

Beginning in 2016, SEI research focused on causal learning

- Causal learning appeared capable of distinguishing “spurious” correlation from causal-based correlation
- Causal learning did not require experimentation and thus, could operate on observational data such as historical cost research data
- Belief was that causal learning could help trim down the overwhelming list of software cost change driver relationships hypothesized by the experts

Causal learning was a novel leap from predominant use in medical research to use in software cost research

Causal learning has grown in use to six SEI research projects in past 3 years

SEI Research Journey from Uncertainty to Causality

-- PREDICTION --

(Correlation & Regression)

What are the subjective uncertainties to model in cost estimates?

WSARA 2009

QUELCE

Affordability

PRESCRIPTION

(Causal Search Collaboration **with CMU Dept of Philosophy**)

HOW TO CONTROL COST THROUGH …

… practitioner factors?

… team and leadership factors?

… organizational and enterprise factors?

… acquisition and technology factors?

Why Does Software Cost So Much? (LENS)

SCOPE LSI (3 Year LSI)

Pre-2017 | **2017** | **2018** | **2019** | **2020**

Motivation for Causal Learning

Controlling costs requires knowing which "independent factors" actually *cause cost* outcomes, so that we may change cost in a predictable manner.

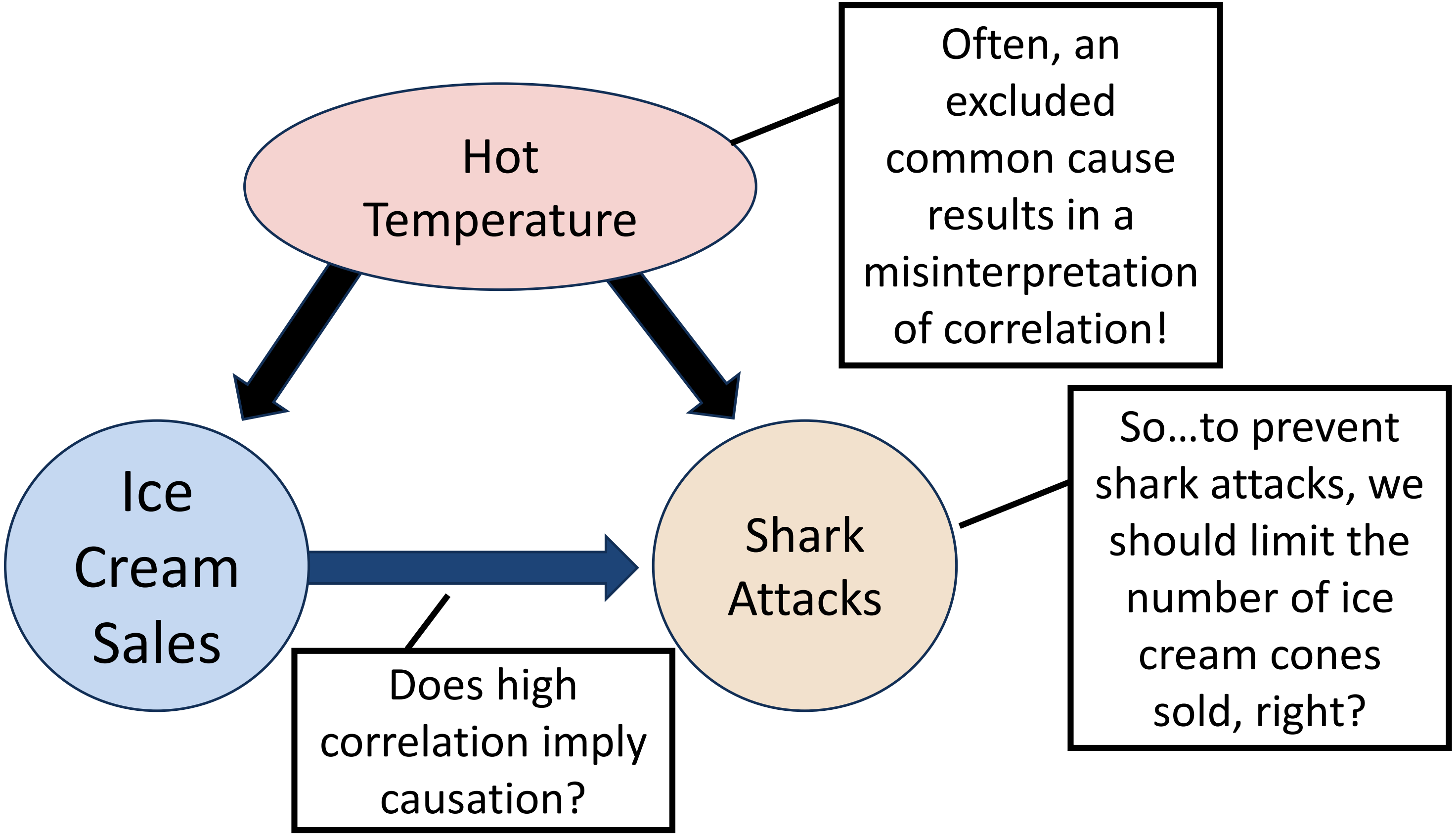
Just as correlation may be fooled by spurious association, so can regression

We must move beyond correlation to causation, if we want to make use of cause and effect relationships

We can now evaluate causation without expensive and difficult experiments

Establishing causation with observational data remains a vital need and a key technical challenge, but is becoming more feasible and practical.

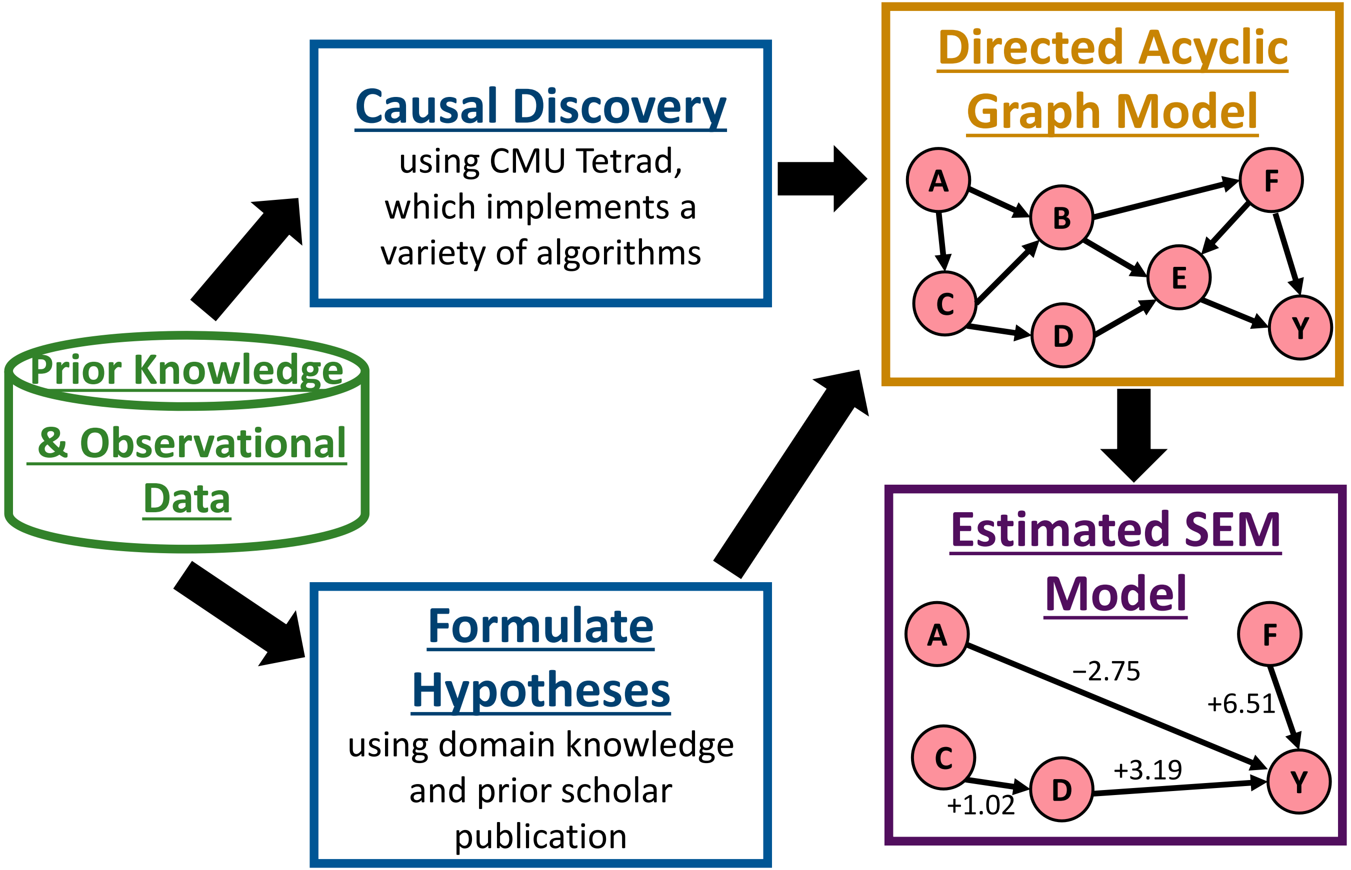
Primary Reason for Spurious Association



Different Uses for Correlation versus Causation

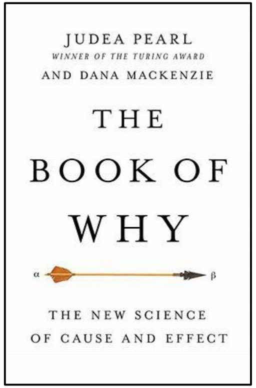
Correlation	Causation
Classifying & identifying	Influencing & acting
Informational value of different evidence	Using evidence to guide policy or actions
Prediction & reasoning given observations	Prediction & reasoning given interventions
Probable explanations for some event or issue	Ways to produce or prevent an event or problem

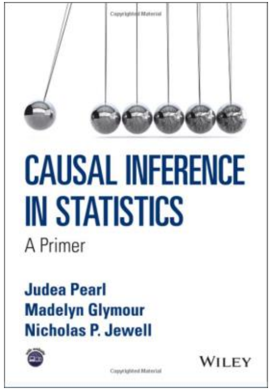
Landscape of Causal Learning

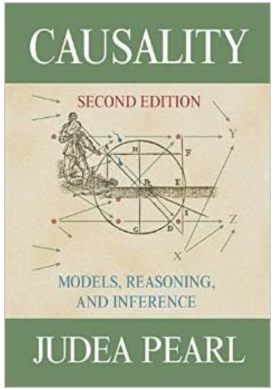


Causal Learning as a Discipline

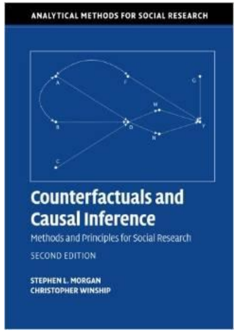
Judea Pearl

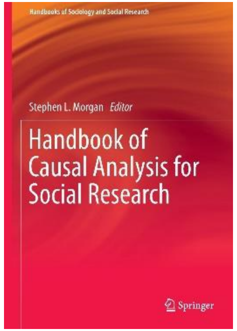


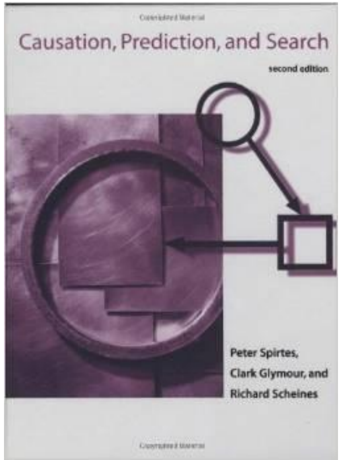




Stephen Morgan







Richard Scheines

David Danks

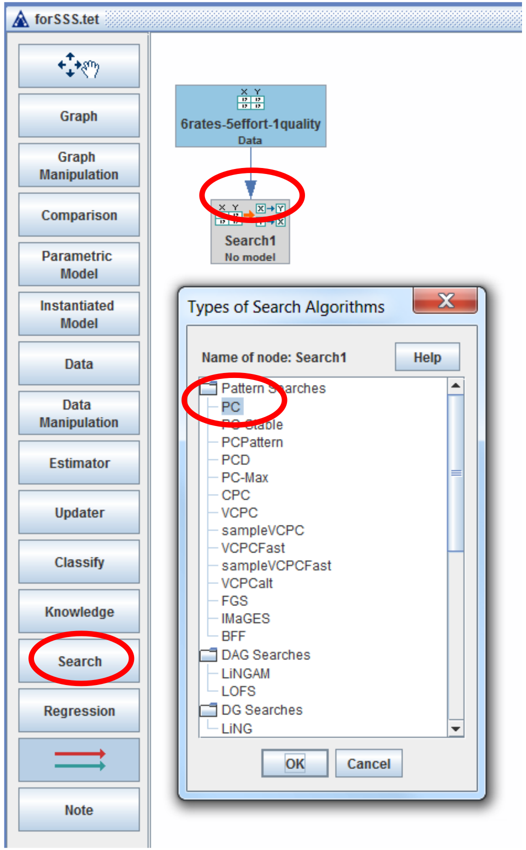
Clark Glymour

Peter Spirtes

Joe Ramsey

Kun Zhang

Causal Search Algorithms and Tooling

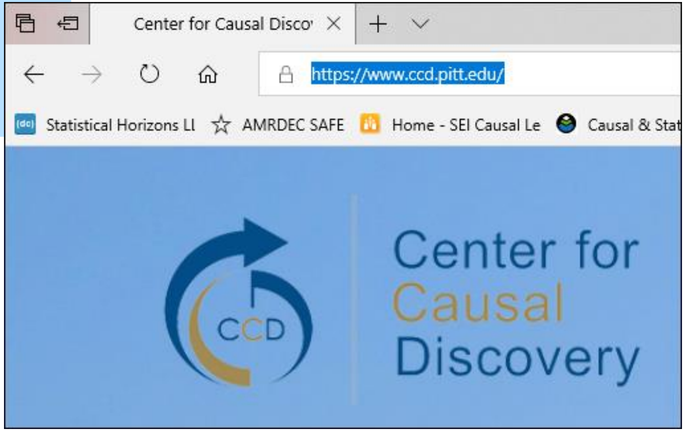


Step 4: Initiate a causal search by inserting a Search box with a connection from the Data box.

Select from over a dozen different algorithms for causal discovery.

The PC algorithm, named for Peter Spirtes and Clark Glymour (CMU), its creators, is one of the most widely used and known causal search algorithms.

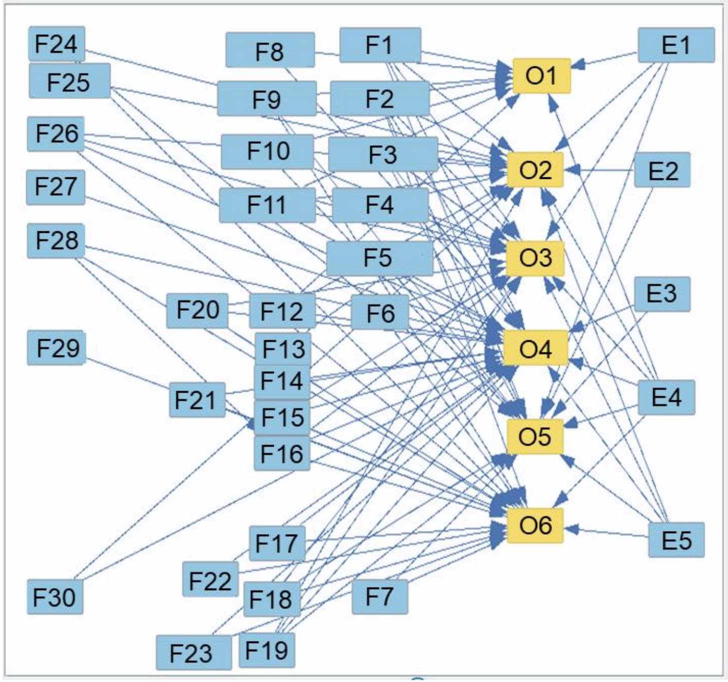
https://www.ccd.pitt.edu/

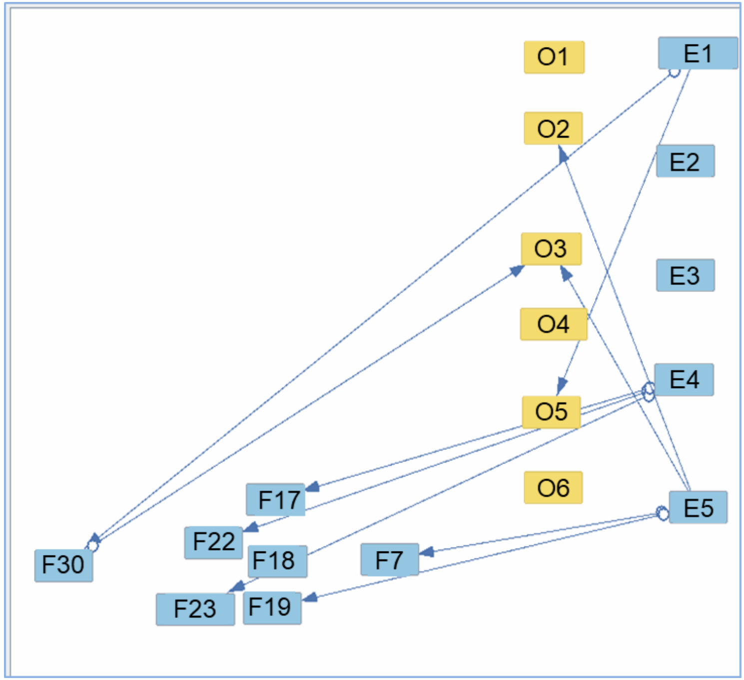


Causal Search Outputs

Comparison of causal graphs unhealthy vs. healthy for a cyber physical system (CPS)

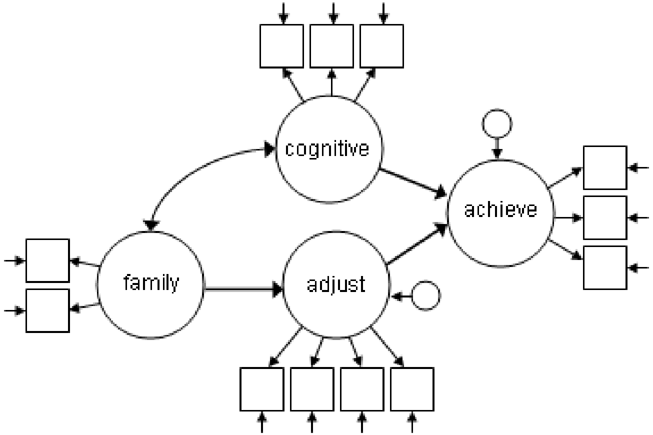
- In an unhealthy CPS, outcomes have many internal causes, including wear and tear from earlier missions
- In a healthy CPS, outcomes are almost entirely driven by the environment





Key: O1 Outcome
E2 Environment variable
F3 Internal Performance factor

Causal Estimation Techniques



Structural Equation Modeling

Multivariate modeling: may involve measured and latent factors; with a simultaneous set of regression equations; where factors can be independent and dependent within the same model

Instrumental Variables

Finding a third factor correlated to your independent factor but not directly related to your dependent factor; useful when unmeasured confounders might exist between your independent and dependent factors

Propensity Scoring

An approach to matching and trimming within a data set to evaluate cause effect relationships

Causal Algebra Do-Calculus

A new causal algebra by Pearl et al to compute causal effects using three new rules of do-calculus in addition to traditional probability manipulation rules

Example SEI Causal Learning Cost Research Results

Controlling Size: Only 2 of 4 code size measures appear causal on effort and quality

Controlling Complexity: Only 1 of 3 factors appears causal on performance and quality

Controlling Architecture Violations: Only 1 of 4 violation factors appears causal on quality

Controlling Team Performance: Only 1 of 20+ factors appears causal on quality and cost

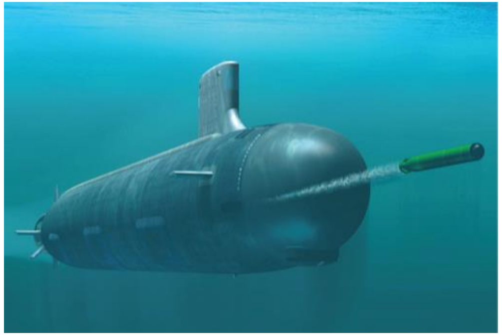
Causal search may provide useful feedback:

- ***Presence of causal links***
- ***Absence of causal links***

Other SEI Causal Learning Research: Simulation and Test

Problem
Lack of accredited simulators

Technical Challenge
Experts unsure of the expected result for a given simulated scenario

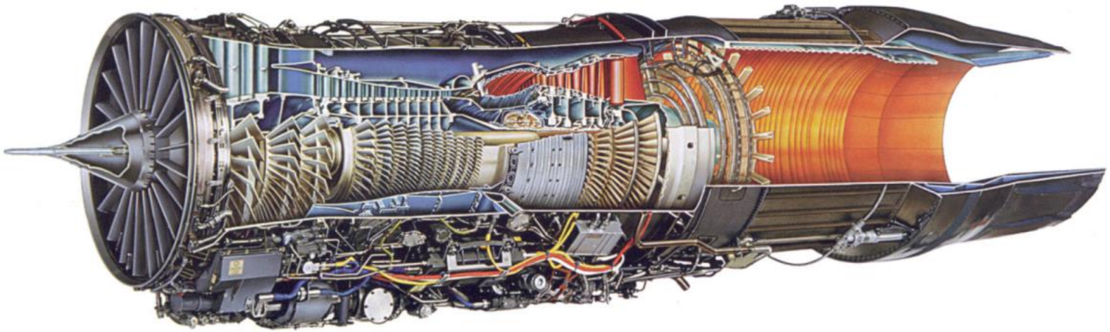


Research Questions
1. Scale up metamorphic testing to test very complex DoD systems?
2. Machine learning to identify metamorphic relations for testing?
3. Causal learning to drive metamorphic relations testing?

Other SEI Causal Learning Research: Sustainment

<u>Problem</u>
Unscheduled maintenance creates unacceptable costs

<u>Technical Challenge</u>
Traditional statistical approaches helpful, but insufficient



<u>Research Questions</u>

1. Machine learning of engine sensor and control data improve scheduled maintenance?
2. Causal learning integrated with machine learning add value?

Future SEI Causal Learning Research: Examples

Affordable

- Acquisition practice improved using causal models
- Cost estimates and budget execution using causal models
- Simpler but more effective ROI models based on causal factors (e.g. Model Based Engineering, Architecture practice, Technical Debt)

Trustworthy

- Causal factors threatening cyber defenses
- Causal factors limiting resilience
- CL combined with ML tools for more affordable and trustworthy SW technologies
(e.g. DOD initiative in Digital Engineering)
- Expected behavior from autonomous systems   (e.g. “*Explainable AI*”; Jensen, UMass)

Capable

- Causal drivers of workforce performance
- SW architecture strategies and tactics driving system performance
- More efficient experimentation of technical solutions
- Increased realism of complex system simulation
- Autonomous systems controlling consequences
- Machine learning with human-like intelligence (e.g. *“Strong AI”; Pearl,*
“The Book of Why”)

Timely

- Causal structures from DevOps information stream to control process and lifecycle
- Agile causal systems situationally prescribe practices aligned with goals
- Project risks controlled through causal structures of project parameters

A Vision for Software Cost Research

Prescriptive Cost Guidance

- Guide cost and pricing negotiations
- Identify interventions during program execution
- Formulate causal-based lessons learned

Explainable AI Cost Estimation

- Increase transparency of future AI-based cost estimation and management
- Increase trust in such solutions and reduce bias

Cost Model Transferability

- Determine when a cost model may be safely used in a new context or situation
- Identify in advance when cost models may not be trusted

Call to Action

Demand causal knowledge to guide interventions

Engage with SEI causal researchers studying software cost

Motivate data collection and sharing for more repeatable and reproducible causal studies

Build causal learning competency in your organization

Contact Information

Presenter / Point(s) of Contact

Robert Stoddard Dr. Mike Konrad

Email:

rws@cmu.edu
mdk@sei.cmu.edu

Telephone:

+1 412.268.1121 (Bob)

+1 412.268.5813 (Mike)

Other SEI Cost Research Team Members

Chris Miller

Bill Nichols

Sarah Sheard

Dave Zubrow

Other CMU Cost Research Contributors

Madelyn Glymour

Joe Ramsey

Kun Zhang

USC Cost Research Contributors

Jim Alstad

Barry Boehm

Anandi Hira