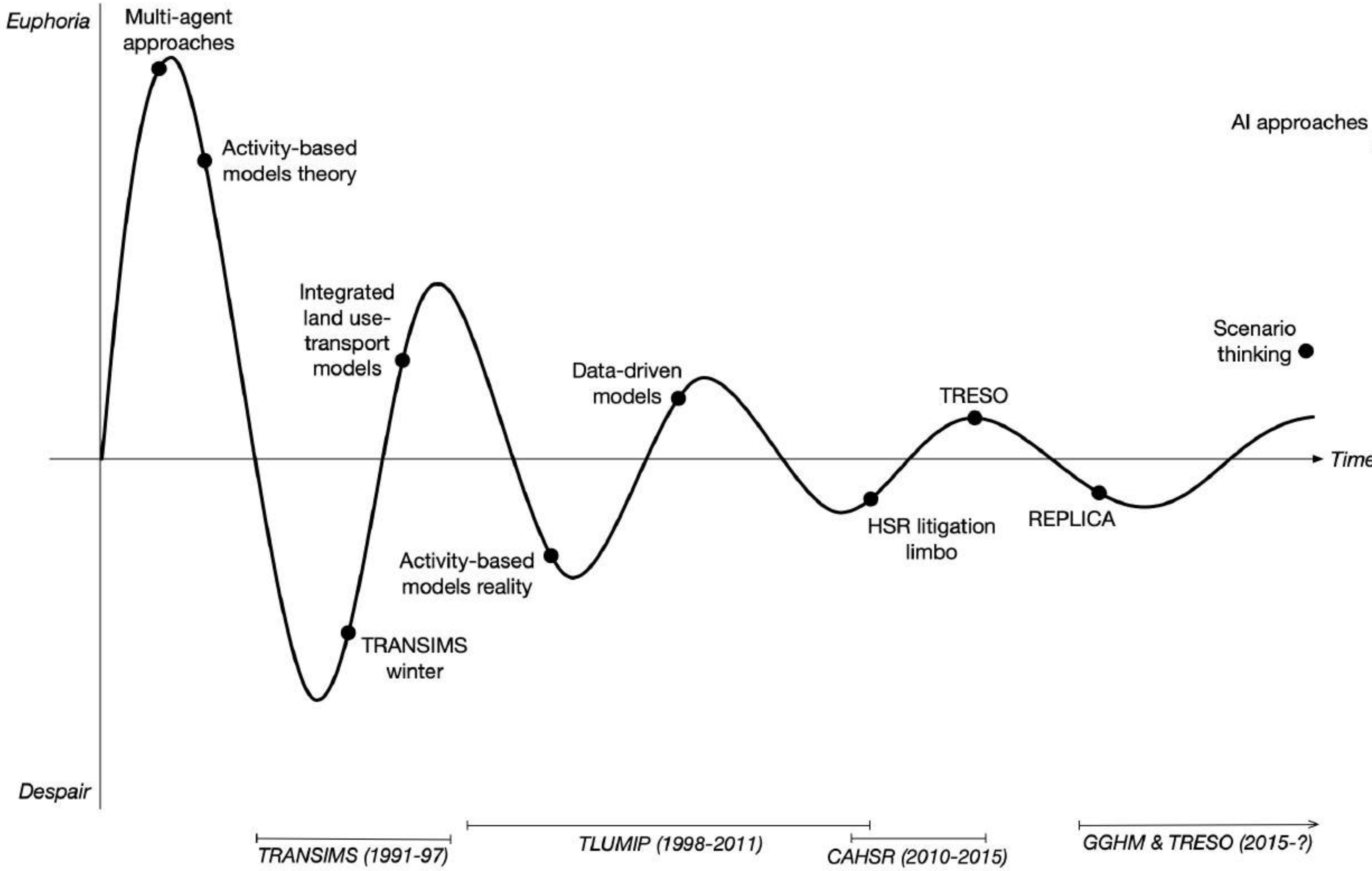
Rediscovering the lost art of travel forecasting

Rick.Donnelly@wsp.com | WSP | 3 December 2021

Preliminaries

1. Safety first
2. Standard disclaimers
3. Questions via chat

My career


Euphoria
Despair
Time
Multi-agent approaches
Activity-based models theory
Integrated land use-transport models
TRANSIMS winter
Activity-based models reality
Data-driven models
HSR litigation limbo
TRESO
REPLICA
Scenario thinking
AI approaches
TRANSIMS (1991-97)
TLUMIP (1998-2011)
CAHSR (2010-2015)
GGHM & TRESO (2015-?)

You do *what*?

"I create and simulate artificial societies from snapshots of real-world populations to study how policies and investments might effect social activity and travel behavior and its aggregate impacts upon the transport system and built environment in a safe, controlled laboratory setting."

Motivations

20th century

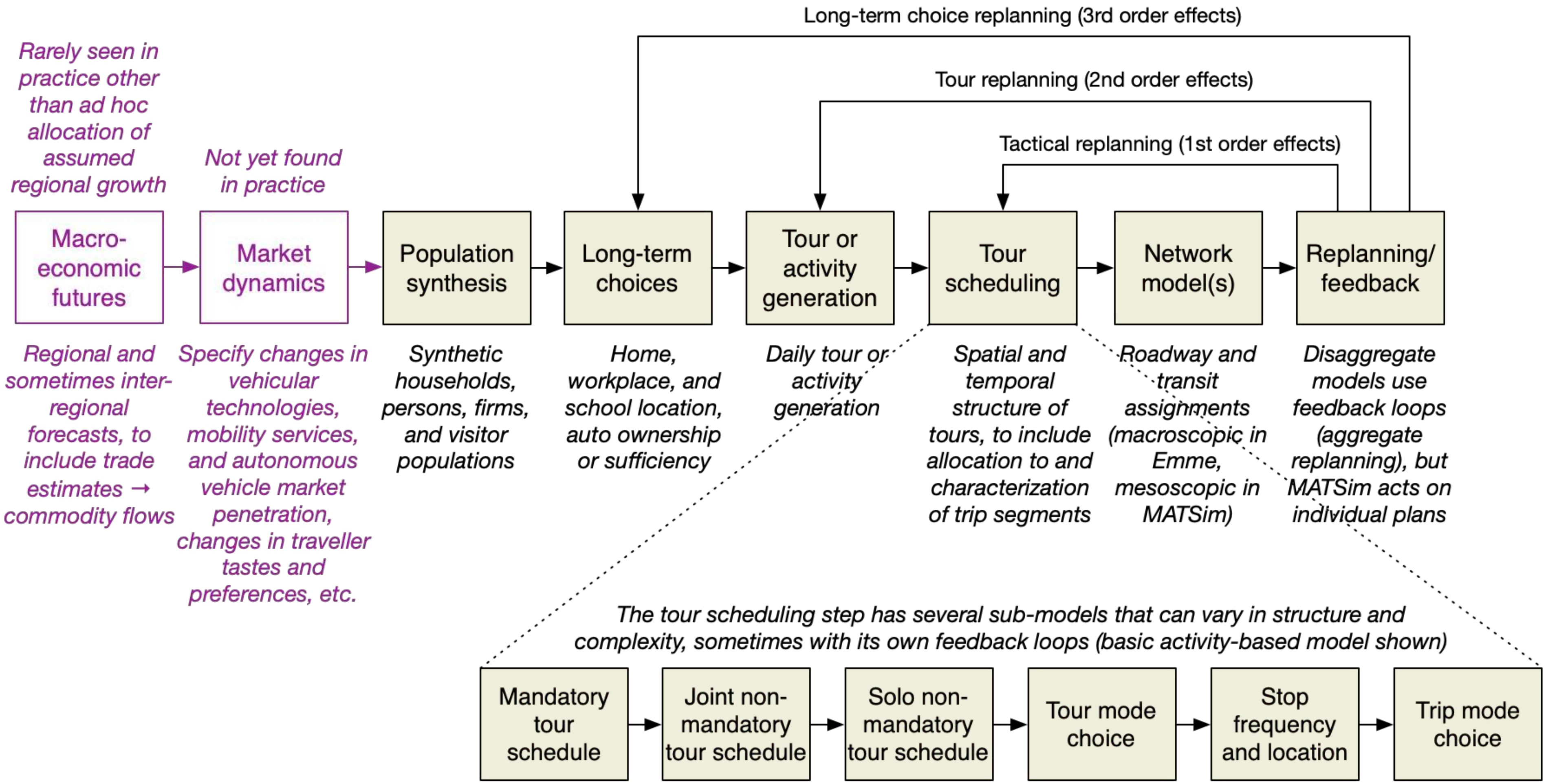
Transit demand and revenue
Major highway investments
Long-range regional planning

21st century

All of last century, plus:

Community connectivity
Links to economic and trade models
Commercial vehicle travel and impacts
Links to emissions models
Energy impacts
Travel demand management
Safety impacts
Modal redundancy studies
Network resilience measures
Economic impact analyses
Congestion management
Pricing studies
Managed lane studies
Cost-benefit analyses
Financial and social welfare metrics
Equity analyses
Active transport analyses
Health impacts
Fuel price increase impacts
Bottleneck analyses
Traffic engineering studies
Autonomous vehicles
Shared mobility services
Pandemic travel shifts
Changes in telecommuting
...

Ideal framework?



Tour generation

Variable[a]	(1) Tour purpose	(2) Mandatory tour frequency	(3) Non-mandatory tour frequency	(4) At-work sub-tour frequency	(5) Stop frequency
Person type	X	X		X	X
Worker occupation				X	
Person age group	X	X			X
Gender	X	X	X		X
Household size	X		X		
Household composition					X
Children not at home			X		
Number of pre-school children				X	
Household income group	X	X	X	X	X
Dwelling type	X		X		
Usual workplace type	X				
Auto sufficiency w.r.t. workers	X		X		X
Accessibility[b] (number of variants used)	(2)	(2)	(4)	(2)	(1)
Distance to work/school		X			
Travel time to work/school		X			
Presence of mandatory tours			X		
Number of tours by purpose					X
Total household tours					X

a. Constants not shown

b. Several different specifications used. Values shown are number of different accessibility definitions used within each model.

Destination choice

	Variable	Maintenance tour[b]	Discretionary tour	Escorting tour	At-work sub-tour	Intermediate stop location
Impedance variables	Mode choice logsums	X	X		X	X
	Linear distance[a]	X	X	X		
	Squared distance[a]	X		X		X
	Natural log of distance[a]	X		X	X	X
Person and household attributes	Age group	X	X			X
	Gender	X				X
	Work status (full/part-time)	X	X			
	Income category					X
Tour attributes	Tour purpose					X
	Tour mode					X
	Direction (to/from anchor)					X
	Number of stop on tour					X
Size terms	Retail employment	X	X		X	X
	Service employment	X	X	X	X	X
	Government employment	X			X	X
	Military employment				X	X
	Hotel employment			X		
	Households		X	X	X	X
	School enrolments					X
	University/college enrolments					X

a. Logit modellers assert these terms should be considered a composite term rather than collinear variables

b. Used to denote habitual discretionary tours (e.g., grocery shopping)

Architectures and artifacts

Modeling systems

Macroeconomic
Population synthesis
Resident travel
Visitor model(s)
Commercial vehicles
Network assignment
Evaluation

Sketch planning models
Trip-based models
Activity-based models
Data-driven models
Probabilistic models
Generative models
Machine learning
System dynamic models
Random numbers
Group consensus

Markets

Households and persons
Residents making local trips
Residents making long-distance trips
Visitors
Firms and economic sectors
Commodities by mode of travel
Imports and exports
Long-distance trucks
Urban trucks
…

Traditional metrics

Aggregate network statistics (VKT, VHT, …)
Travel times and reliability
Per-capita change in VKT, non-auto travel, …
Wide economic benefits
Degree and extent of congestion
Public transport and pricing revenues
Environmental impacts
Network level of service
Cost-benefit analyses

Emerging metrics

Empty kilometres of travel (CVs)
Aggregate accessibility measures
Consumer surplus or user benefits
Network reliability and resilience
Social welfare statistics
Pricing revenues and equity
Risk and uncertainty analyses

Topologies

Global
State
Places (polygons)
Places (points)
Agents
Objects

Agents

Persons
Households
Vehicles
Roadways
Intersections
Mobility service providers
Public transport operators
Jurisdictions
Firms
Establishments
Buildings
Gateways

Agent properties

Preferences
Budgets
Choices
Activities
Tours
Trips
Routes

Objects

Buildings
Facilities
Vehicles
Signals
Sensors
Roadways
Junctions (intersections)
Transit lines

Household records

householdID	size	region	vehicles	income
1001	3	Parry Sound	2	102118
1002	1	Parry Sound	0	68115
1003	2	Parry Sound	1	71032

... (for each household in the synthetic population)

Generated using population synthesizer

Person records

householdID	personID	status	driver_lic	age	gender	home_loc	work_loc	sch_loc
1001	1	ft_service	Y	32	M	501	1803	NA
1001	2	pt_proftech	Y	34	F	501	2962	NA
1001	3	child	N	6	F	501	NA	503

... (for each person in synthetic population)

From long-term choice models

Person activity-travel records

householdID	personID	tourID	activity	duration	origin	destination	purpose	mode	dep_time	travel_time
1001	1	1	home	8.23	501	1803	HW	a	8.23	.41
1001	1	1	work	9.10	1803	501	WH	a	17.74	.59
1001	1	1	home	6.26						

... (for each person and activity in synthetic population)

From person records for mandatory activities and activity location model for non-mandatory

Tour generation

Mode choice model(s)

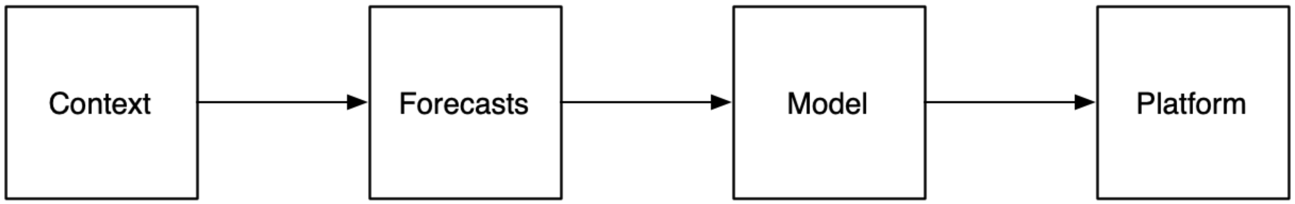
Tour scheduler (or temporal allocation)

Pre-assignment

Vehicle trip matrices by user classes by period

Results aggregated for traffic assignment

Davidson diagram



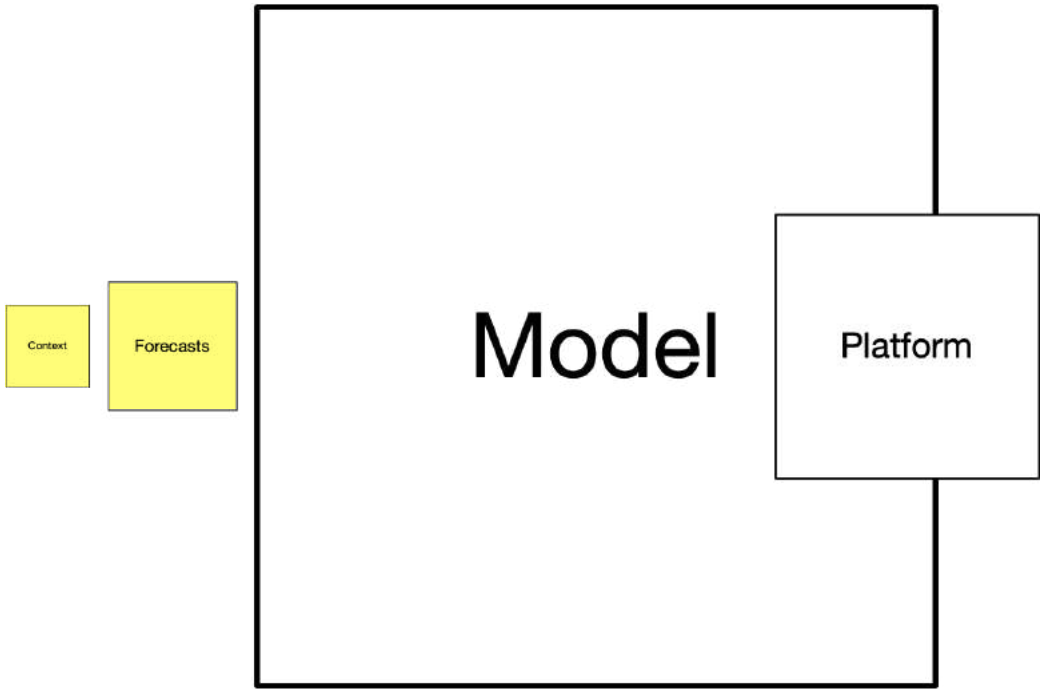
We build forecasting models to inform public policy and investment decisions. What story are we trying to tell, and who is the audience? What are relevant performance measures?

We are most often engaged to develop forecasts of future conditions. What range of assumptions and what properties of the modelled system are being tested?

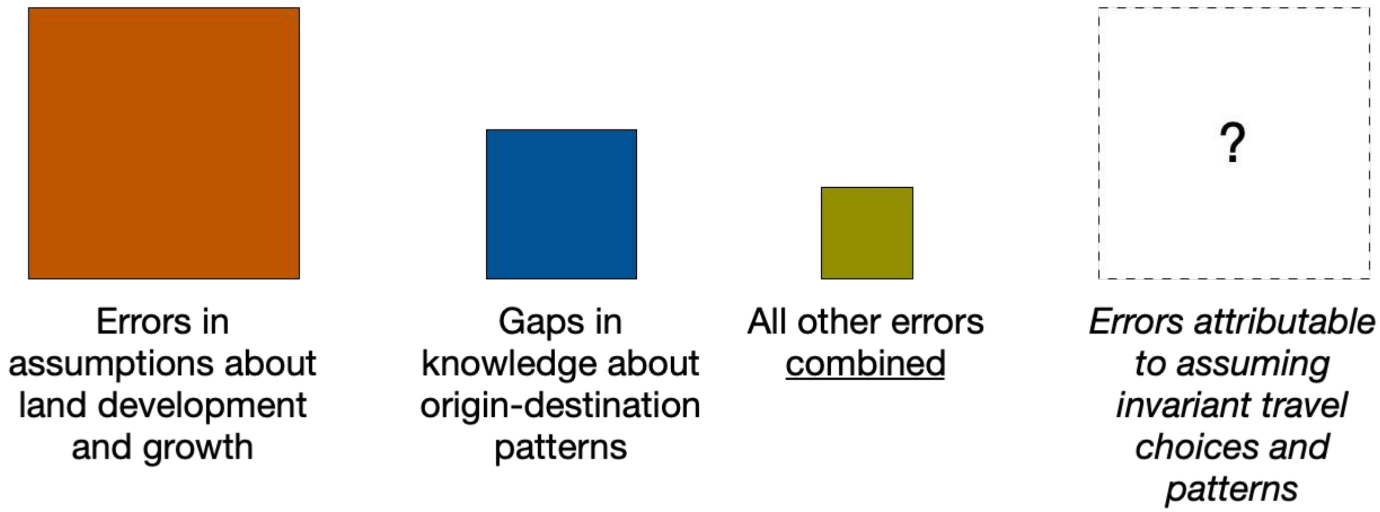
We can decide on the most appropriate modelling approaches and methods once we have the larger context defined.

Once we understand the context, analytical requirements, and most appropriate model(s) we can decide upon the best data, software, and hardware solutions.

Distortion field



Forecasting errors in perspective

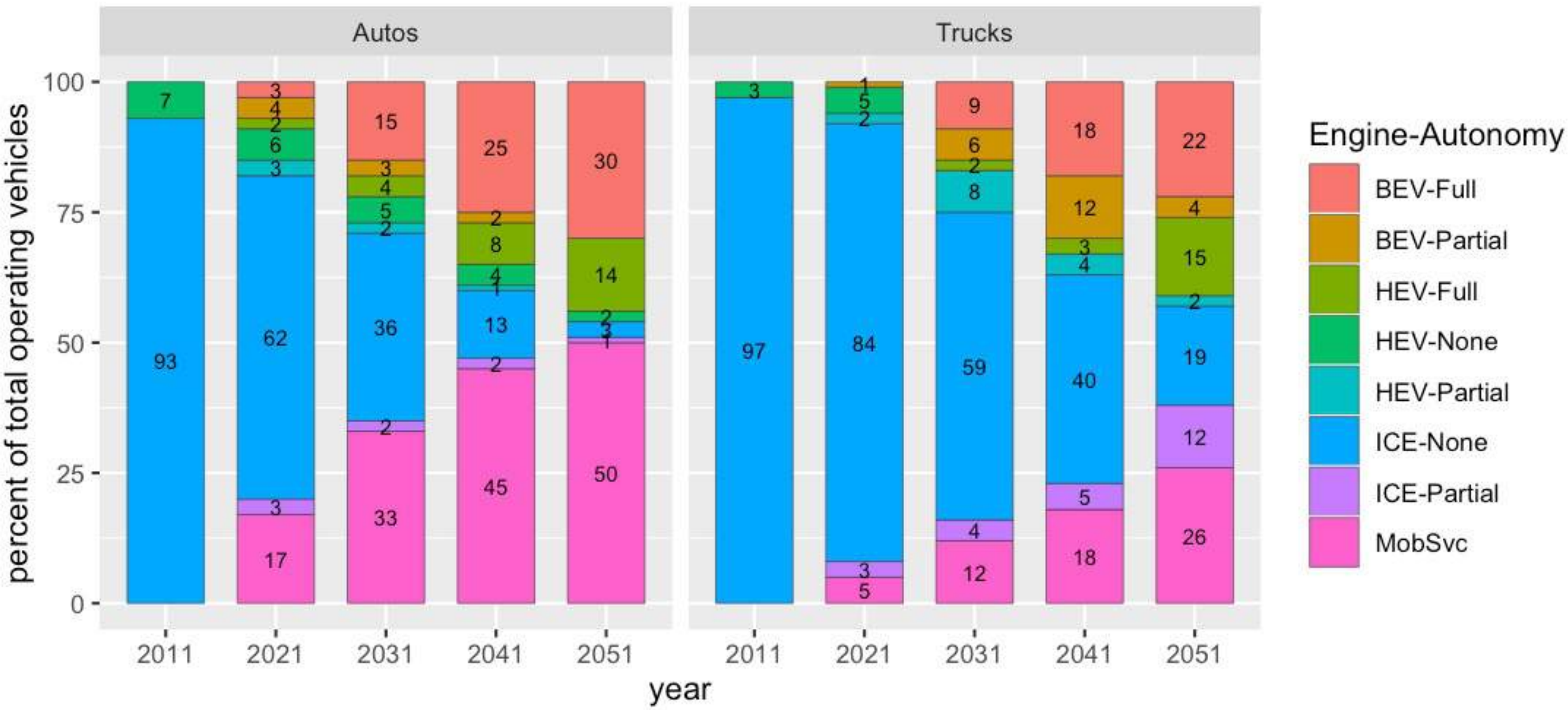


Insights:

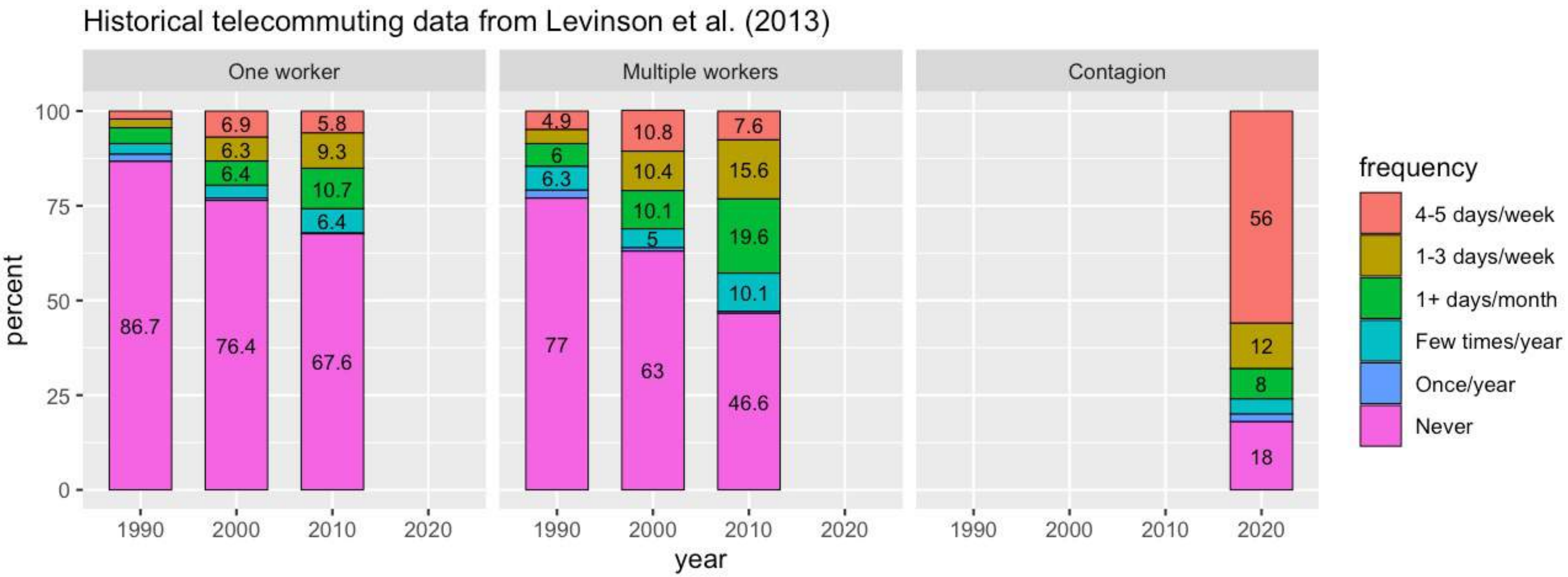
- B. Flyvbjerg et al. (2003), *Megaprojects and risk: An anatomy of ambition*, Cambridge University Press
- My experience to date

Forecasts vs reality

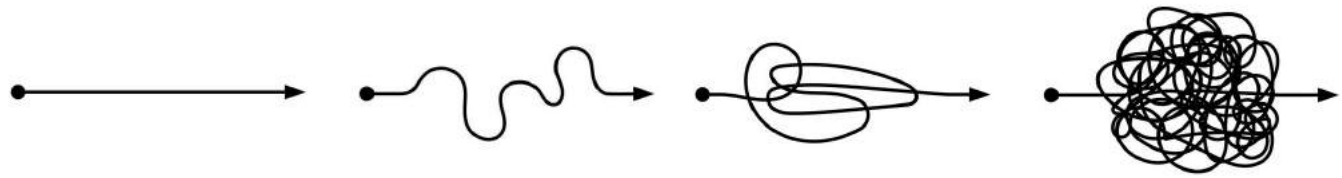
Changes in vehicle fleet



Telecommuting trends over time



What could go wrong?



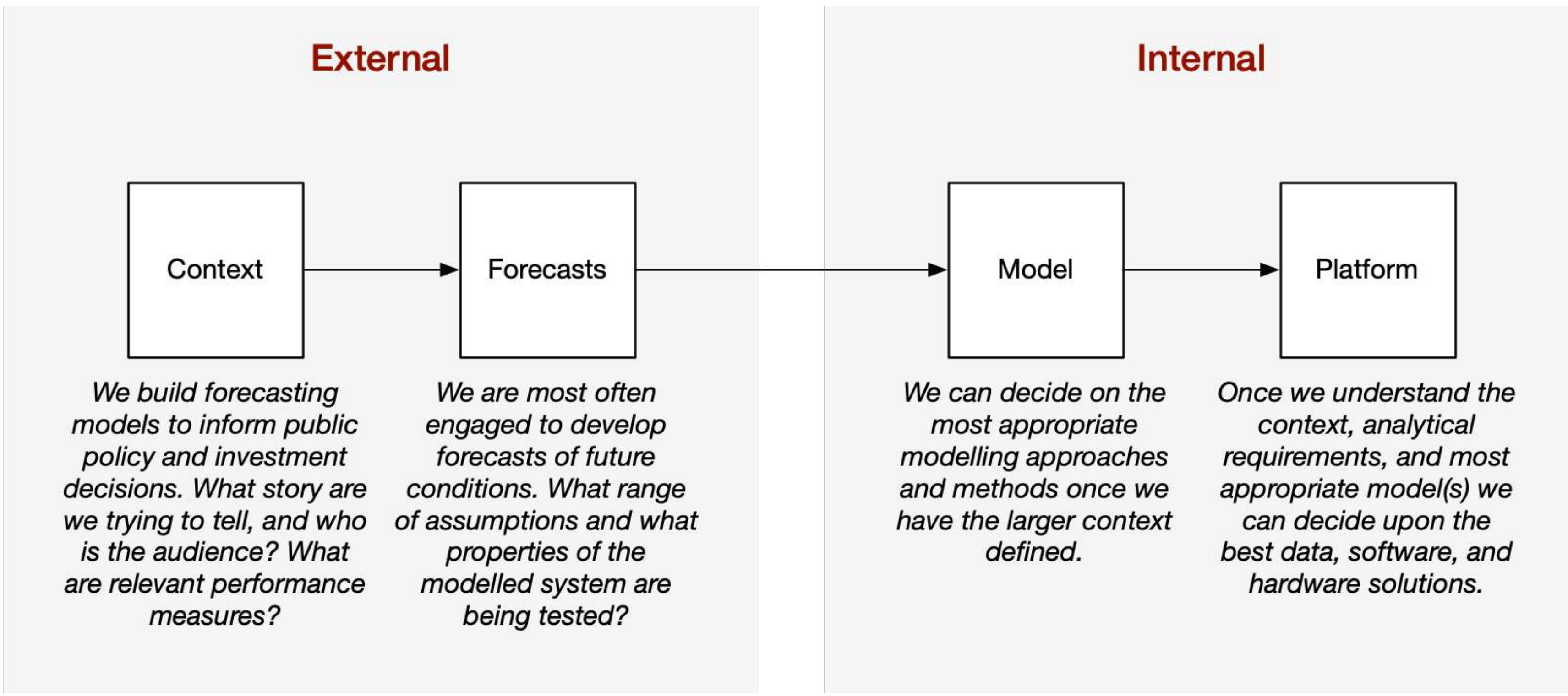
Internal

- Creeping complexity and complicatedness
- Overfit models
- Increasing computational burdens
- Noise vs signal
- Parameter storm
- Outdated assumptions
- Inaccurate forecasts
- Lack of resources

External

- Uncertainty
- High risk ← single future
- Issues evolve faster than models
- Accelerating social, behavior, and technological changes
- Ransomware infections
- Loss of confidence
- Irrelevance to policymaking
- Lack of timeliness

Davidson diagram

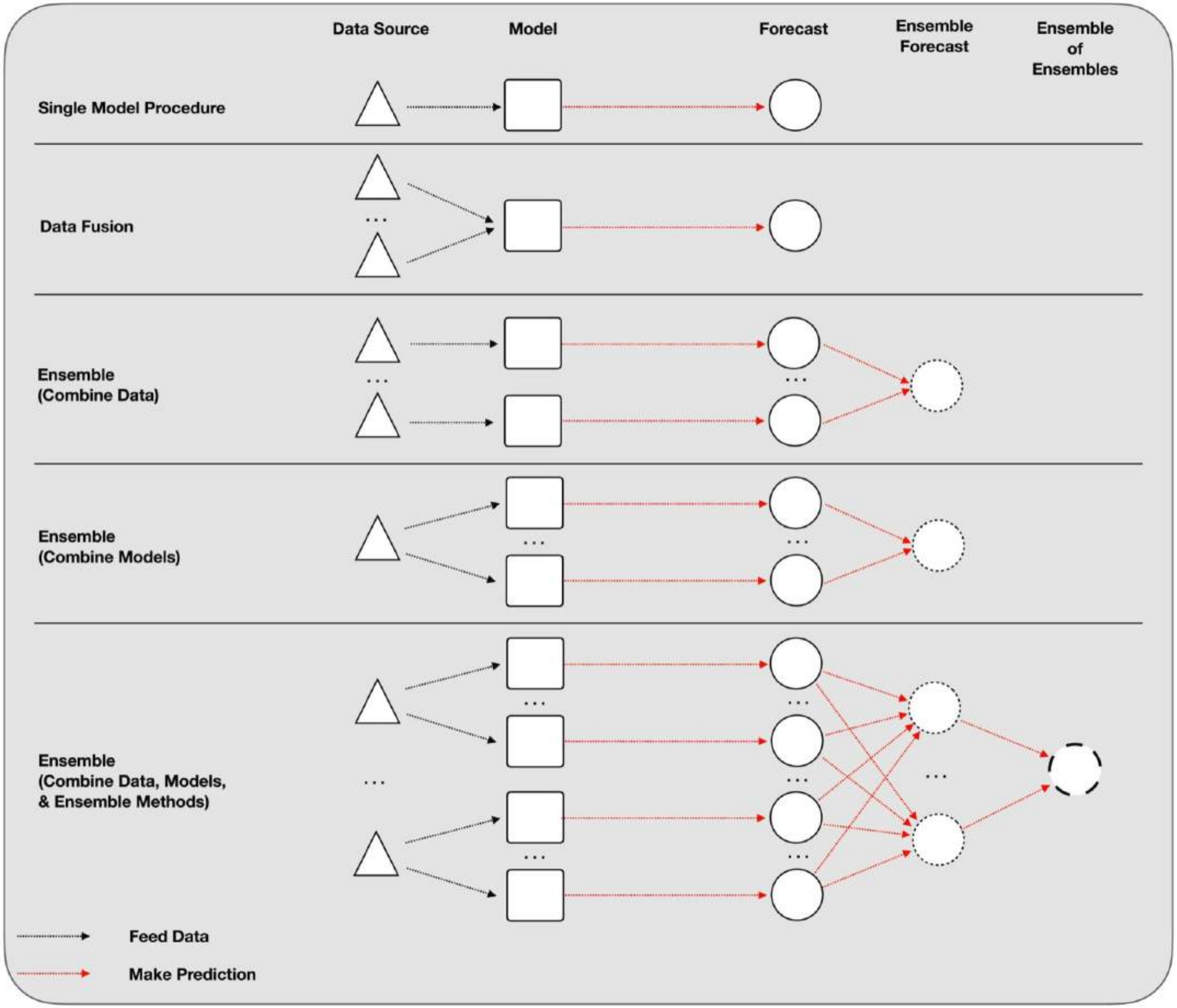


Change how I use them

Reduce cost and expand models

“Methods of combining data and models” (Figure 2.1)

From H. Wu (2021), Theory of ensemble forecasting – with applications to transport modeling, Unpublished PhD thesis, The University of Sydney



ML methods

Source: https://eulertech.wordpress.com/2017/10/03/machine-learning-algorithms-in-one-map/

AI context

Waves of the AI revolution?

- Internet AI
- Business AI
- Perception AI
- Autonomous AI

These four waves all feed off of different kinds of data…

Deep learning requires:

- Massive amounts of relevant data
- Strong algorithms
- Narrow domain
- Concrete goal

"If you're short any one of these, things fall apart."

Impactful AI requires:

- Abundant data
- Tenacious entrepreneurs
- Well-trained AI scientists
- Supportive policy environment

"…the center of gravity shifts from a handful of elite researchers to an army of tinkerers…"

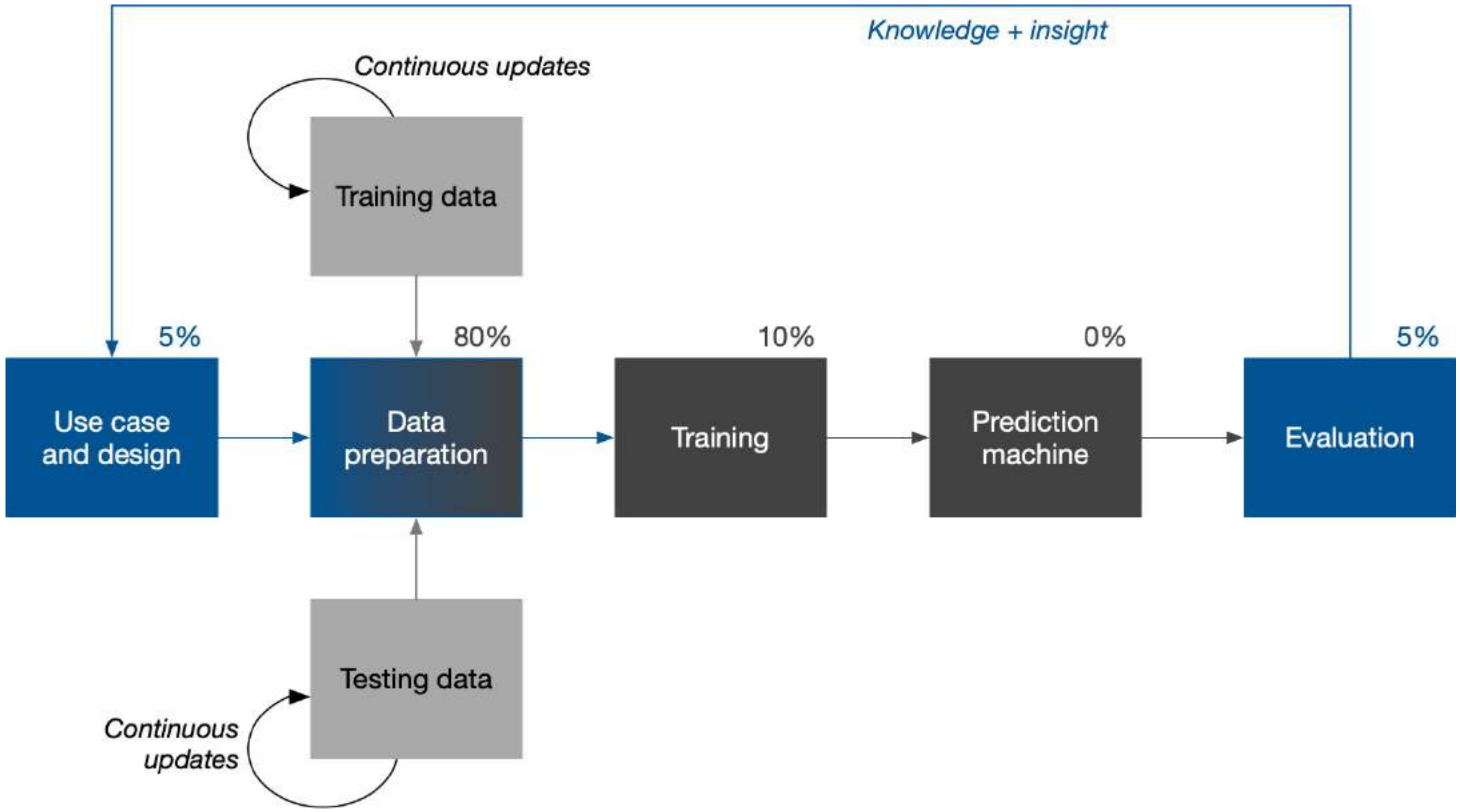
Optimization
Complex math
Deep interactions
Rule-based processes

+

Personal
Creative
Compassionate
Imaginative

Source: Kai-Fu Lee, *AI Superpowers: China, Silicon Valley, and the New World Order* (2018).

ML in a nutshell



Quick intercity mode choice example

Table 5.4: Long-distance travel surveys

Category	Attribute	NHTS	TSRC
Extents	Years included	2002	2012-14
	Total usable observations	45,118	167,481
Variables	Mode (of travel)	✓	✓
	Age group	✓	✓
	Gender	✓	✓
	Education	✓	✓
	Employment status	✓	✓
	Occupation	✓	
	Household income	✓	✓
	Travel party size	✓	✓
	Trip purpose	✓	✓
	Nights away	✓	✓
	Distance (one-way)	✓	✓
	Year	✓	✓
	Percent personally paid	✓	✓
	Annual frequency	✓	

Perfect case study in imbalanced data

Table 5.5: Number of observations by intercity mode of travel

	NHTS		TSRC	
Mode	Records	Percent	Records	Percent
Air	3,347	7.4	7,994	4.8
Auto	40,333	89.3	150,456	89.8
Bus	993	2.1	3,513	2.1
Other	77	0.2	3,427	2.0
Rail	392	0.9	1,268	0.8
Ship	36	0.1	823	0.5
Total	45,118	100.0	167,481	100.0

Starting position

Percent incorrect predictions

Mode	Logit model	Random guess
Air	15.1	95.6
Auto	4.2	10.3
Bus	69.7	98.2
Other	—	97.9
Rail	51.8	99.2
Ship	100.0	99.8

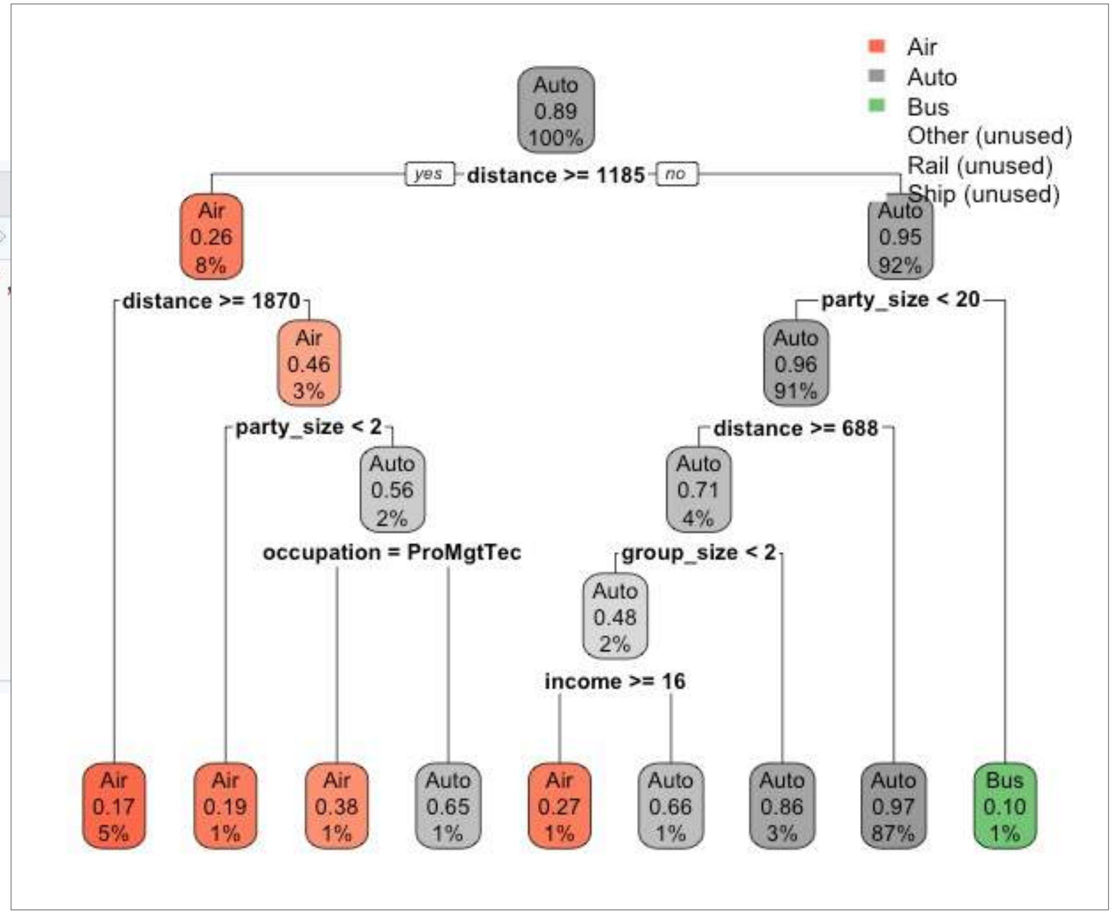
Console Terminal × Jobs ×

~/Library/Mobile Documents/com~apple~CloudDocs/ohplease/

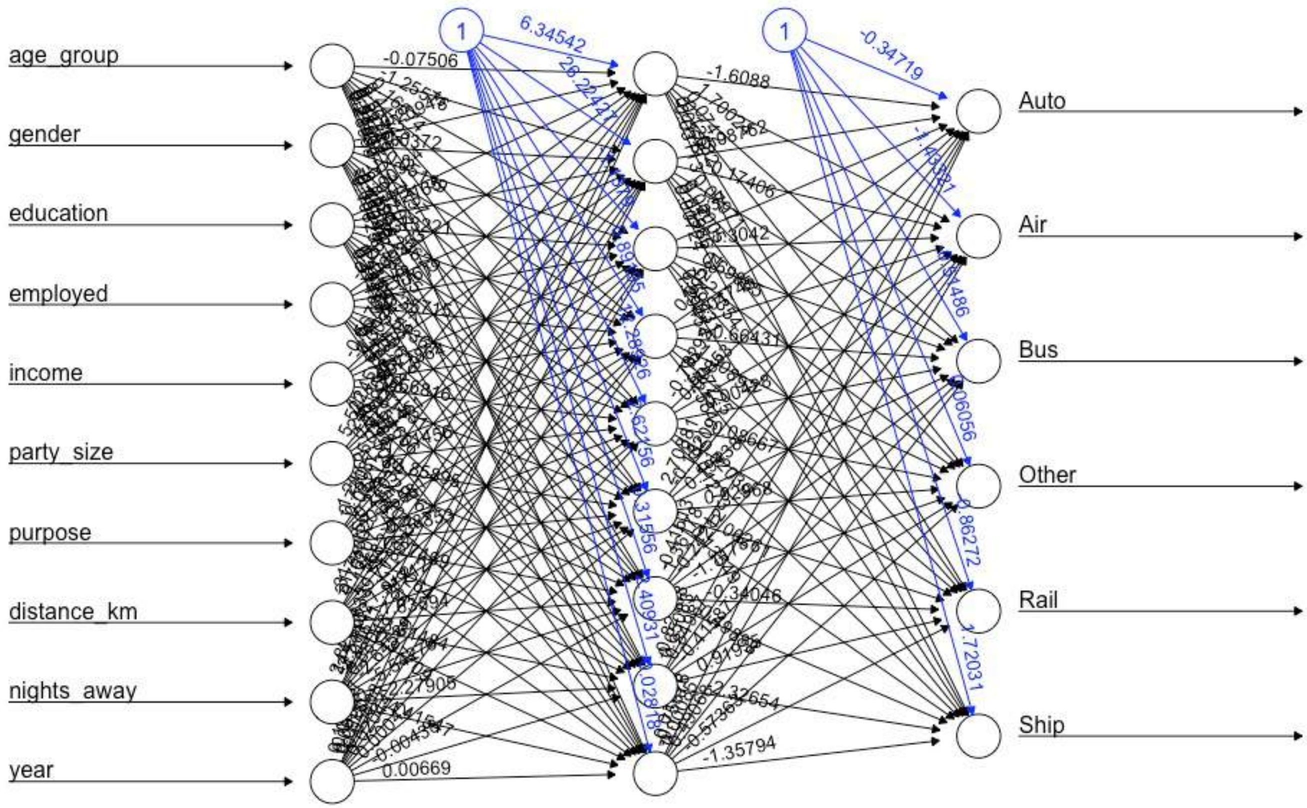
```
> combined$guess <- sample(observed_shares$mode, nrow(combined), r
+   prob = observed_shares$share)
> result <- xtabs(~mode + guess, data = combined, na.action = na.p
L)
> noquote(result)
       guess
mode       Air   Auto    Bus  Other   Rail   Ship
  Air      383   7209    155    161     49     37
  Auto    7282 134980   3141   3057   1212    784
  Bus      164   3150     81     74     27     17
  Other    181   3061     78     70     28      9
  Rail      69   1132     26     25     10      6
  Ship      53    732     20     12      3      3
> noquote(paste("Random guess accuracy = ", accuracy(result)))
[1] Random guess accuracy =  0.81
> modal_accuracy(combined, combined$guess)
   mode Correct Incorrect pctIncorrect
1   Air     383      7611         95.2
2  Auto  134980     15476         10.3
3   Bus      81      3432         97.7
4 Other      70      3357         98.0
5  Rail      10      1258         99.2
6  Ship       3       820         99.6
> 
```

Simple decision tree

```
Console    Terminal ×    Jobs ×
~/Library/Mobile Documents/com~apple~CloudDocs/ohplease/
> noquote(paste("All modes: decision tree accuracy = ",
[1] All modes: decision tree accuracy =  0.94
> source("./feature_accuracy.R")
> noquote(feature_accuracy(df_test, mode, pred))
   mode .groups Correct Incorrect pctIncorrect
1   Air    drop     517       180         25.8
2  Auto    drop    7936       138          1.7
3   Bus    drop      54       107         66.5
4 Other    drop       0        10        100.0
5  Rail    drop       0        72        100.0
6  Ship    drop       0        10        100.0
> 
```



Maybe neural net



Results

(a) Full dataset

Percent incorrect predictions

Mode	Logit model	Random guess	Decision tree	Bagged tree	Neural net (h=24)	SEDO bagged tree	SEDO + random
Air	15.1	95.6	41.7	29.2	10.1	9.5	9.5
Auto	4.2	10.3	0.5	1.4	7.4	0.8	0.8
Bus	79.7	98.2	100.0	91.1	24.8	26.9	21.8
Other	—	97.9	100.0	94.9	77.5	36.5	29.3
Rail	81.8	99.2	100.0	93.4	40.0	77.4	37.9
Ship	100.0	99.8	100.0	96.8	63.3	64.4	60.3

(b) Best over/under sampling outcome

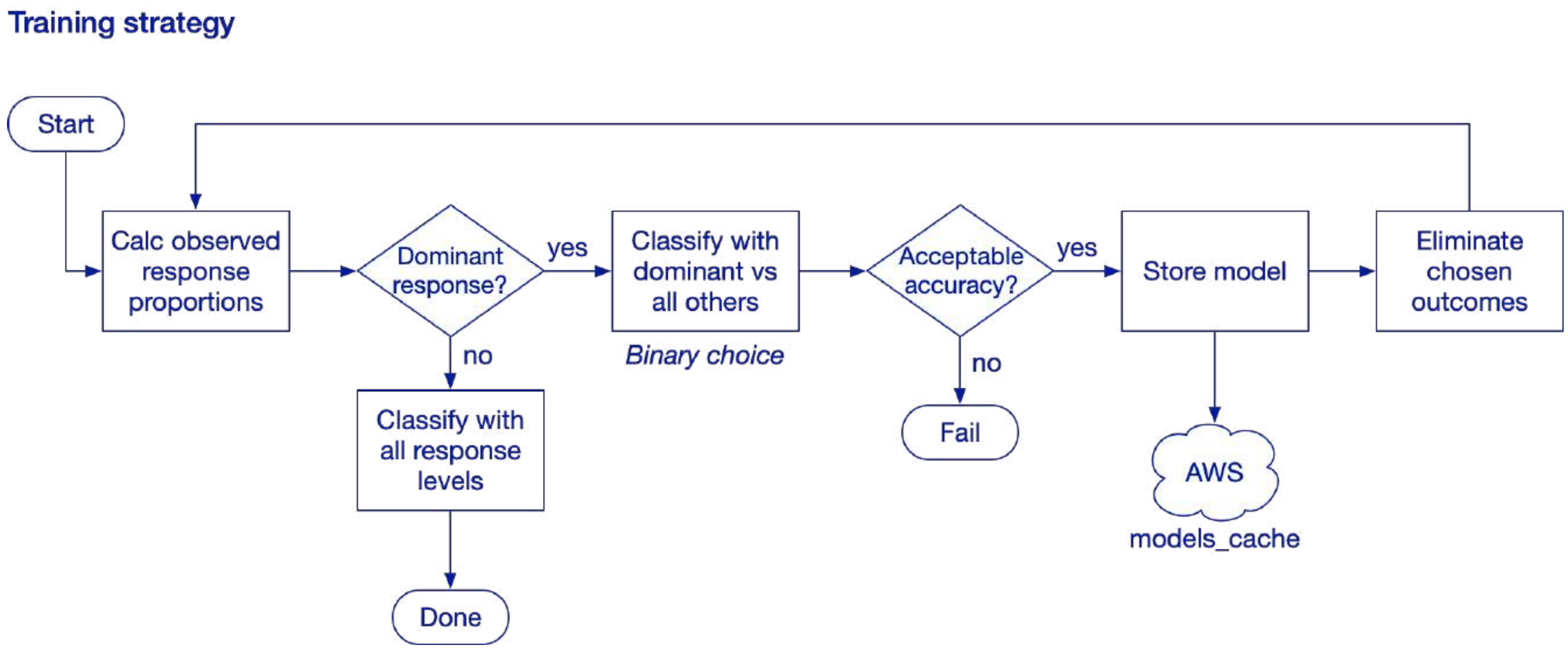
Percent incorrect predictions

Mode	Decision tree	Bagged tree	Neural net (h=24)
Air	8.3	6.9	7.0
Auto	5.2	2.5	3.6
Bus	100.0	51.9	22.3
Other	100.0	74.9	47.1
Rail	100.0	87.6	39.6
Ship	100.0	80.7	73.3

(c) Ranking

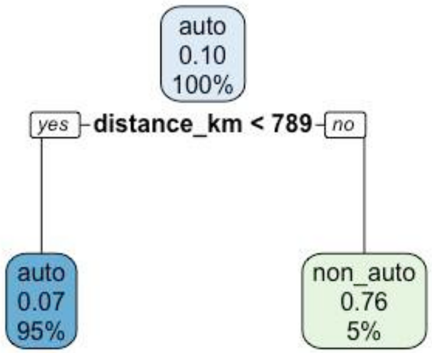
Approach	max(pct incorrect)
SEDO + random	60.3
O/U neural net (h=24)	73.3
SEDO bagged tree	74.5
Neural net (h=24)	77.5
O/U bagged tree	87.6
Bagged tree	96.8
Random guess	99.8
Logit model	100.0
Decision tree	100.0
O/U decision tree	100.0

Successive elimination of dominant outcomes (SEDO)

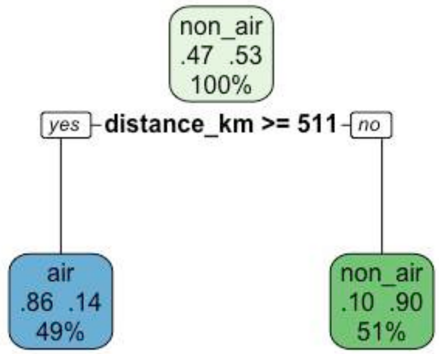


SEDO (3 levels)

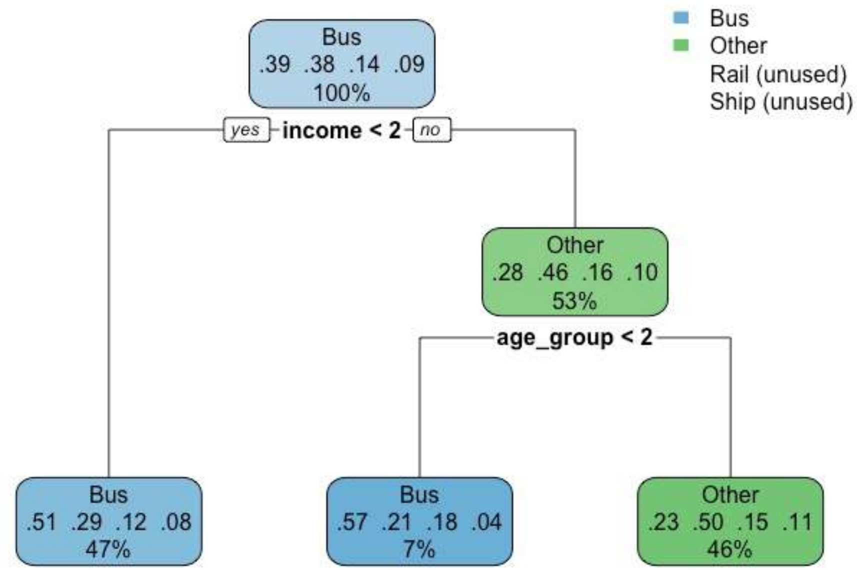
L1
(auto vs. non-auto)



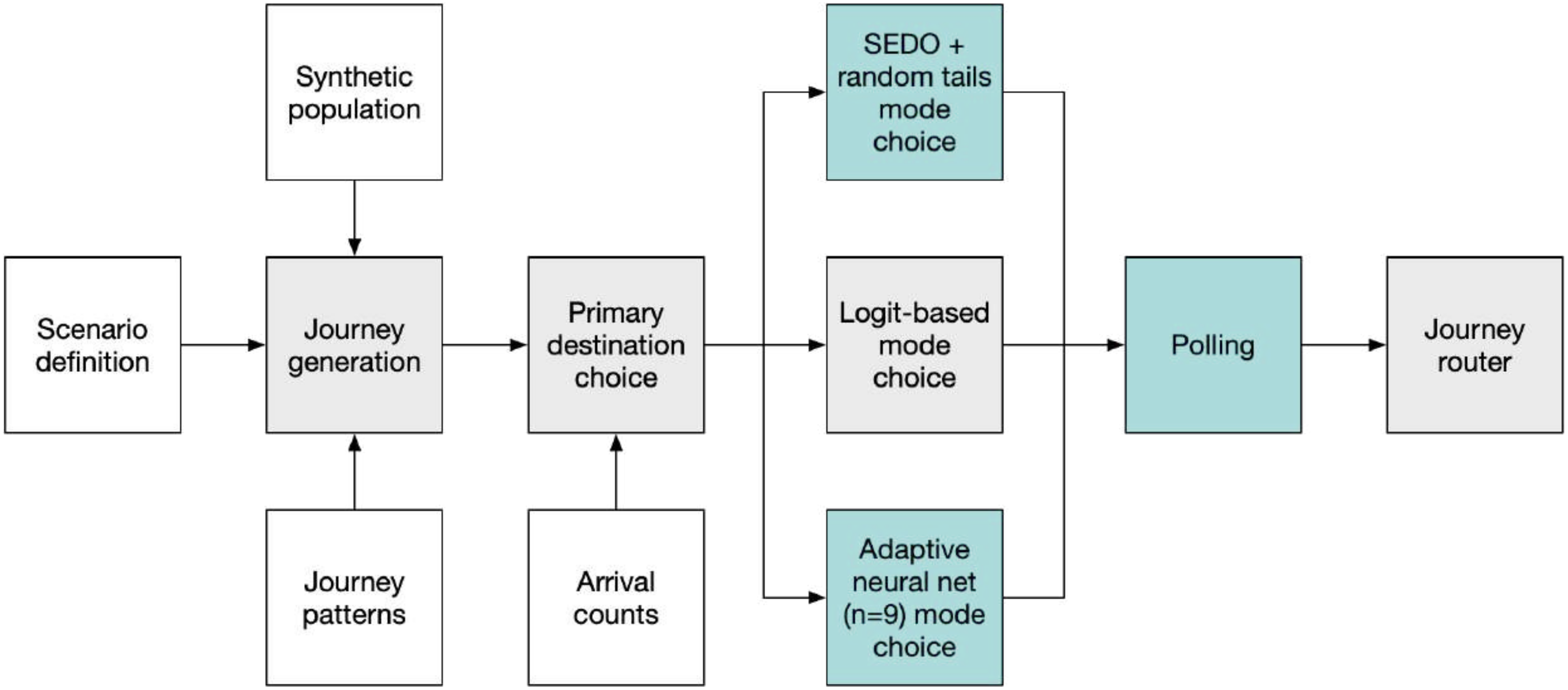
L2
(air vs. non-air)



L3
(remaining modes)



Final modeling system



My view

How can we build evidence-based planning models that overcome:

ML limitations

- Data limitations (quality, quantity, stationarity, …)
- Data silos
- Stochastic
- Lack of interpretability
- P-hacking
- AI solutionism
- Ethical concerns

Human limitations

- Biases and prejudices
- Agendas
- Replication mindset
- Mistakes
- Misinterpreting results
- Linear thinking
- Difficulty comprehending multi-dimensional interactions

Scenario thinking example

Contagions

1. Return to 2019
2. Rolling sheltering and isolation
3. Relative calm between cyclical pandemics
4. Rolling pandemics the new normal
5. A universal vaccine or cure emerges

Future of work

1. Return to 2019
2. Increased telework and hybrid office-remote work
3. Sustained shift towards remote work

Automation + AI

1. AI winter
2. Second Machine Age scenario with higher unemploy-ment
3. Automation trends plateau
4. AI dominance

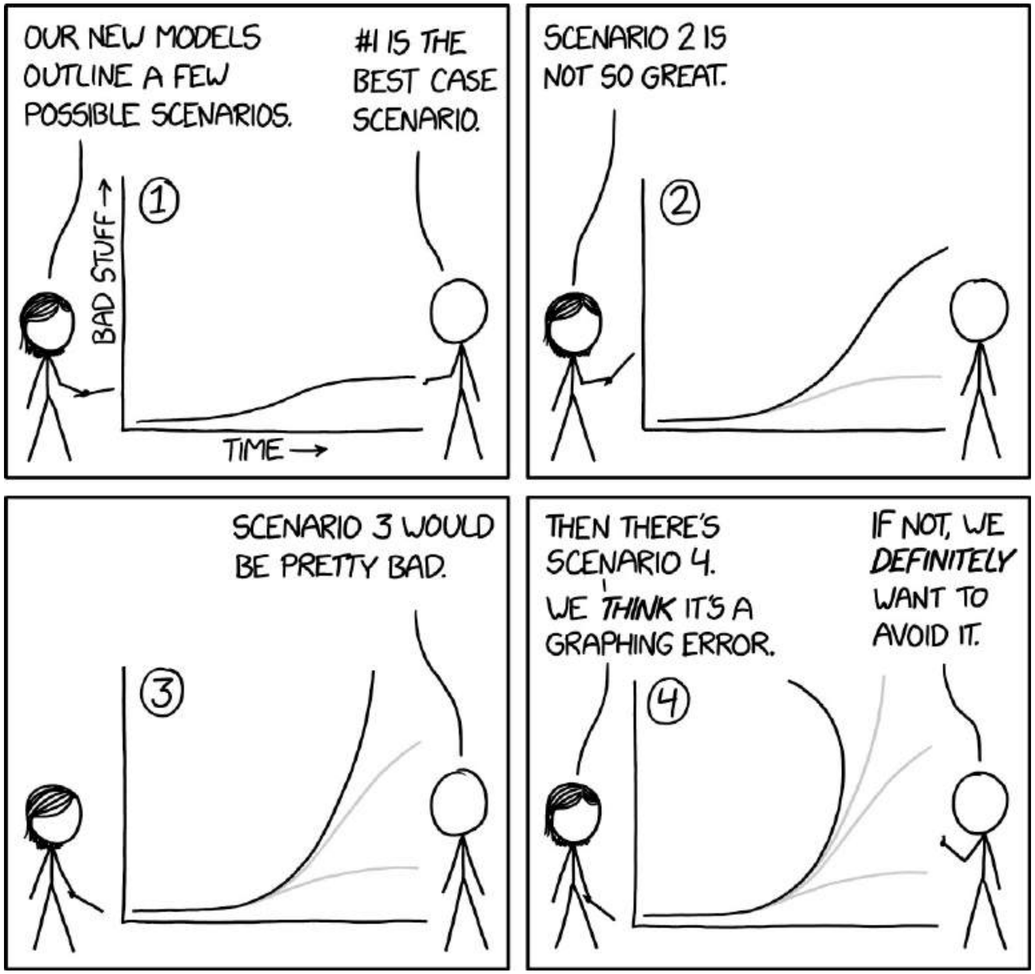
Autonomous vehicles

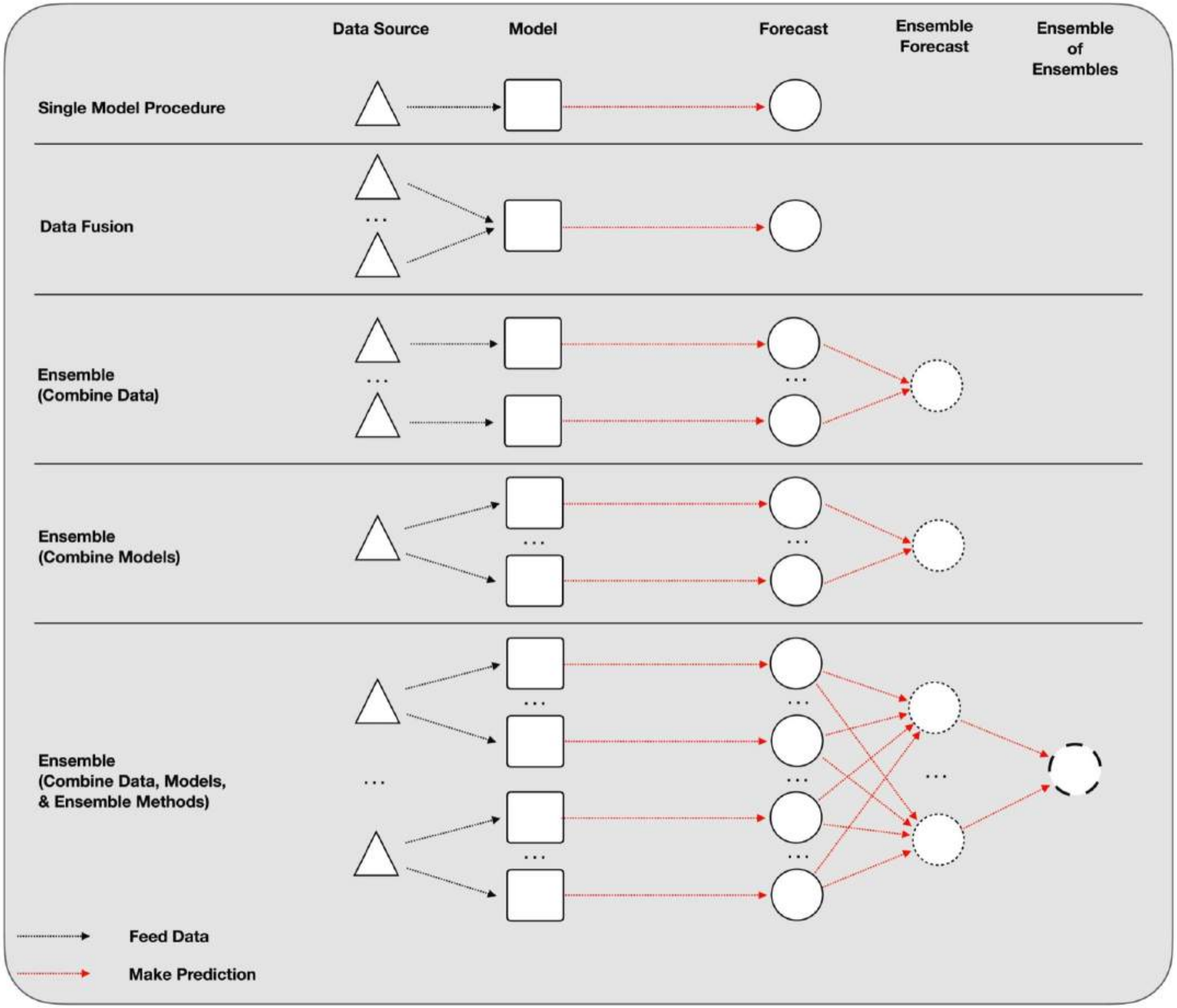
1. Bureaucratic and regulatory inertia
2. AVs remain niche products
3. Widespread adoption of AVs
4. Level 5 automation dominates travel

Military presence

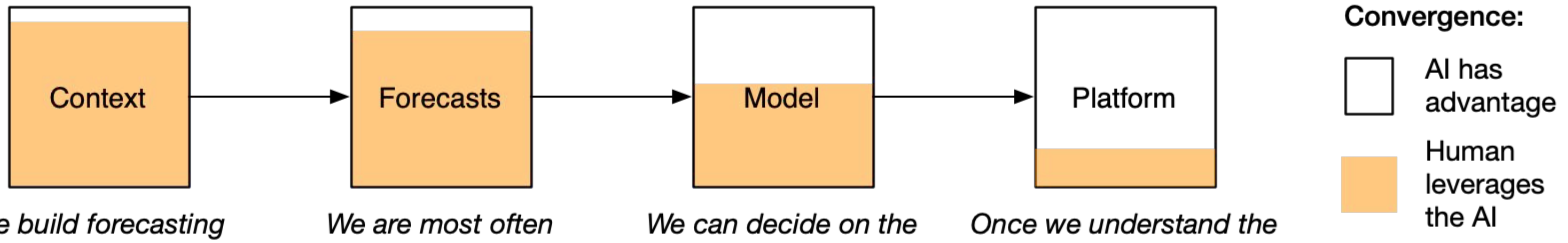
1. Remain at current levels
2. Digital warfare focus reduces traditional forces
3. Stronger Pacific presence to deter Chinese expansion
4. Drones replace human warriors

Scenario thinking example




Data Source
Model
Forecast
Ensemble Forecast
Ensemble of Ensembles
Single Model Procedure
Data Fusion
Ensemble (Combine Data)
Ensemble (Combine Models)
Ensemble (Combine Data, Models, & Ensemble Methods)
Feed Data
Make Prediction

Davidson diagram redux



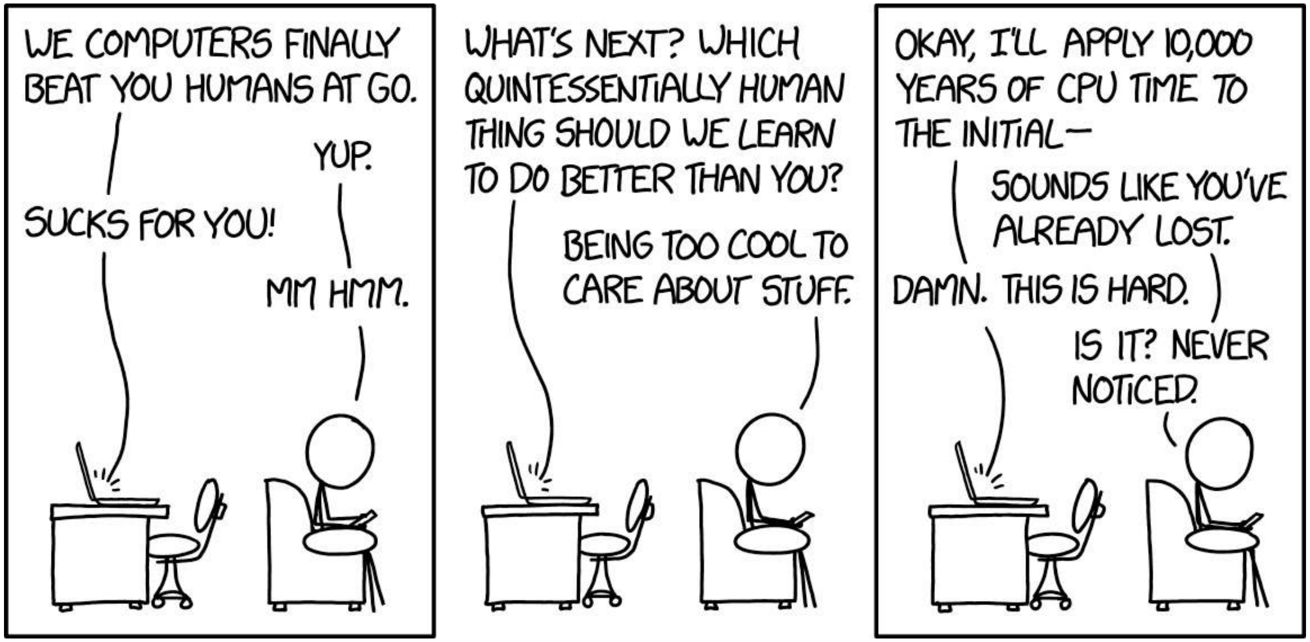
We build forecasting models to inform public policy and investment decisions. What story are we trying to tell, and who is the audience? What are relevant performance measures?

We are most often engaged to develop forecasts of future conditions. What range of assumptions and what properties of the modelled system are being tested?

We can decide on the most appropriate modelling approaches and methods once we have the larger context defined.

Once we understand the context, analytical requirements, and most appropriate model(s) we can decide upon the best data, software, and hardware solutions.

Questions?



WE COMPUTERS FINALLY BEAT YOU HUMANS AT GO.
YUP.
SUCKS FOR YOU!
MM HMM.
WHAT'S NEXT? WHICH QUINTESSENTIALLY HUMAN THING SHOULD WE LEARN TO DO BETTER THAN YOU?
BEING TOO COOL TO CARE ABOUT STUFF.
OKAY, I'LL APPLY 10,000 YEARS OF CPU TIME TO THE INITIAL—
SOUNDS LIKE YOU'VE ALREADY LOST.
DAMN. THIS IS HARD.
IS IT? NEVER NOTICED.

Highly recommended

- H. Wu (2021), Theory of ensemble forecasting – with applications to transport modeling, Unpublished PhD thesis, The University of Sydney.
  https://ses.library.usyd.edu.au/handle/2123/26252
- W. Li & K. M. Kockelman (2021), "How does machine learning compare to conventional econometrics for transport data sets? A test of ML versus MLE", *Growth and Change,* in press.
  https://doi.org/10.1111/grow.12587
- D. Kahneman, O. Sibony & C.R. Sunstein (2021), *Noise: A Flaw in Human Judgment*, Little Brown Spark, London.
- "Machine learning" online Coursera course by Andrew Ng.
  https://www.coursera.org/learn/machine-learning
- A. Ben-Zvi (2020), Scenarios for the COVID-19 future.
  https://breakwaterstrategy.com/scenarios-for-the-covid-19-future/