

WRF #4742: Probability Management for Water Utilities

Copyright 2019

Water Research Foundation, Reproduced with permission

Research Objectives

1. Principles of Probability Management (PM)
2. Illustrate the use of PM tools in depicting and communicating uncertainty
3. SIPmath™ v2 standards

Funders

- Water Research Foundation
- Metropolitan Water District of Southern California
- Irvine Ranch Water District
- Tacoma Water
- Inland Empire Utility Agency
- Eastern Municipal Water District

Research Team

- **Dr. Thomas W. Chesnutt, PStat[®], CAP[®]** (Principal Investigator) of A&N Technical Services,
- **Dr. Michael Hollis, PStat[®]** of the Metropolitan Water District of Southern California,
- **Shayne Kavanagh** of the Government Finance Officers Association,
- **David L. Mitchell** of M. Cubed,
- **Dr. David M. Pikelney** of A&N Technical Services,
- **Dr. Jean-Daniel Rinaudo** (Outside Expert), of the French Geological Survey (BRGM), and
- **Marc Thibault** the lead author of SIPmath™ v2 standards for probabilitymanagement.org.

Room for improvement

Are studies reproducible?

The New York Times

TheUpshot

THE NEW HEALTH CARE

Science Needs a Solution for the Temptation of Positive Results

By Aaron E. Carroll

May 29, 2017



A few years back, scientists at the biotechnology company Amgen [set out to replicate](#) 53 landmark studies that argued for new approaches to treat cancers using both existing and new molecules. They were able to replicate the findings of the original research only 11 percent of the time.

Science has a reproducibility problem. And the ramifications are widespread.

Reliance on Averages

- ▶ Good decisions need to account for:
 - ▶ the entire shape of uncertainty (*distributions*) ; and
 - ▶ the consequences of risky outcomes (*loss function*).
- ▶ Averages obscure both, leading to mis-informed decisions.

WRF 4558 Uncertainty in Long-Term Water Demand Forecasts - Survey Responses

- Majority are 'cognizant of long term demand forecast uncertainties' and have taken some steps to address
- Larger systems (pop. served) means more attention devoted to long-term water demand forecasting
- Key future uncertainties: future growth in customers, future climate, the condition of the economy, and water efficiency
- ~50% of the utilities that address uncertainty do not address all variables
- There are deficiencies in available data and forecasts of available data
- Water supply & distribution system planning
 - bias for demand forecast scenarios reflect middle-to-high range of predictions,
 - under-predicting future demands outweigh the risks of over-predicting

ProbabilityManagement.org

- Non-profit
- Aim
 - Improve use of uncertainty in decision making
 - “communication, calculation, and credible estimates.”
- Data array aka Stochastic Information Packet
 - Define it
 - Way to contain it and use across software platforms
 - Meta data
 - Standard
- Free tools, purchase tools

Principles of Probability Management

(See probabilitymanagement.org)

- **Communicating Distributions as Data** –When estimating uncertain quantities, a “typical” or “average” value is used. Collapsing uncertainty to a scalar destroys information on variability.
- **Information** (measurement) can reduce total uncertainty and add credibility.
- **Interactive simulation** is a useful tactic for gaining simulated experience with how decisions affect risky outcomes. Visualization is an effective medium for communicating risk to decision-makers.
- **Coherence**—Distributions of causal forces are often related/dependent. The SIPmath Standard provides the **glue** logic for related uncertainties.

Probability Management's *SIPmath*TM 2.0 *Standard* defines:

SIP – Stochastic Information Packets

- *Uncertainties are communicated as data arrays*
- *Thus random draws from uncertain possibilities are stored as a column of realizations.*

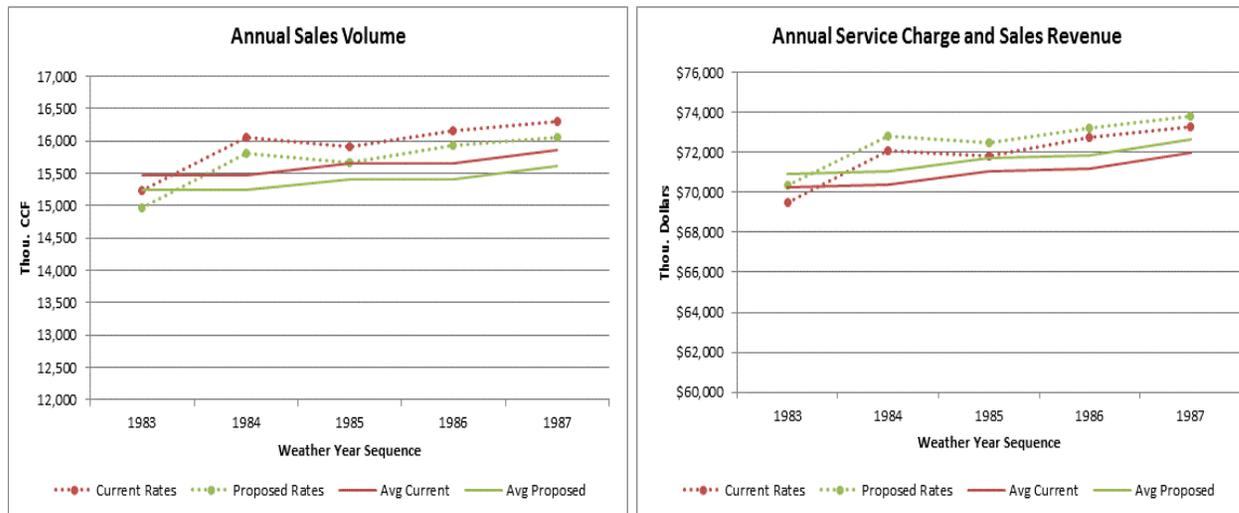
SLURP – Stochastic Library Unit Relations Preserved

- *A **coherent** set of SIPs that preserve statistical relationships between uncertainties*

SIPMath™ is Actionable

Sips can be used directly in calculations of uncertainty.
Cells in Excel can refer to SIPs instead of a single number.
No macros or add-ins need remain in the spreadsheet.

Volume and
Revenue
Volatility
demonstrated
visually



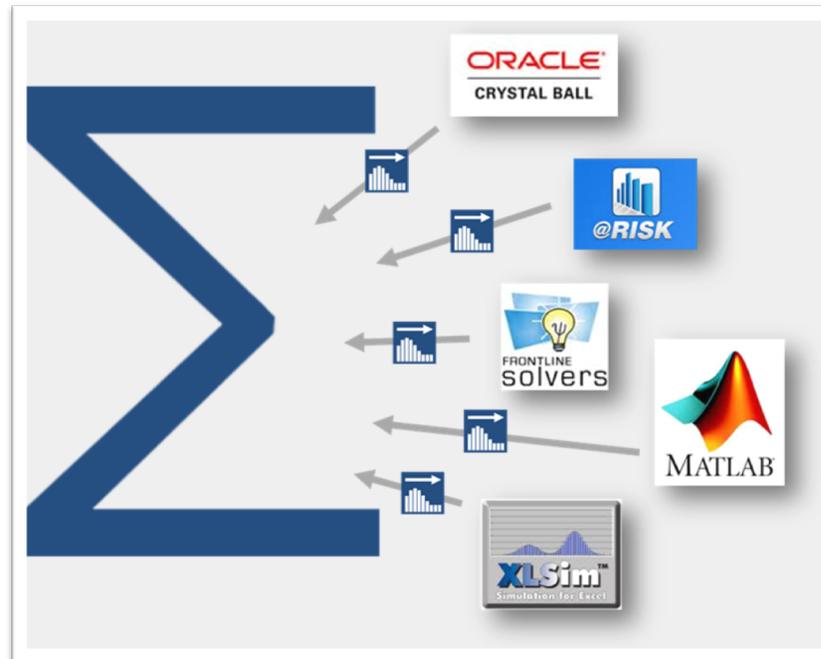
SIPmath compliant open-source model at <http://www.financingsustainablewater.org/tools/awe-sales-forecasting-and-rate-model>

SIPMath™ is Additive

Uncertainties can be summed, enabling enterprise risk management

SIPmath

- Platform independent
- Excel add in
 - free at probability management
 - any of these commercial simulation software.



SIPMath™ is Auditable

The SIPMath™ standard requires provenance.

Saved SIPs can be replicated using same seed=auditability.

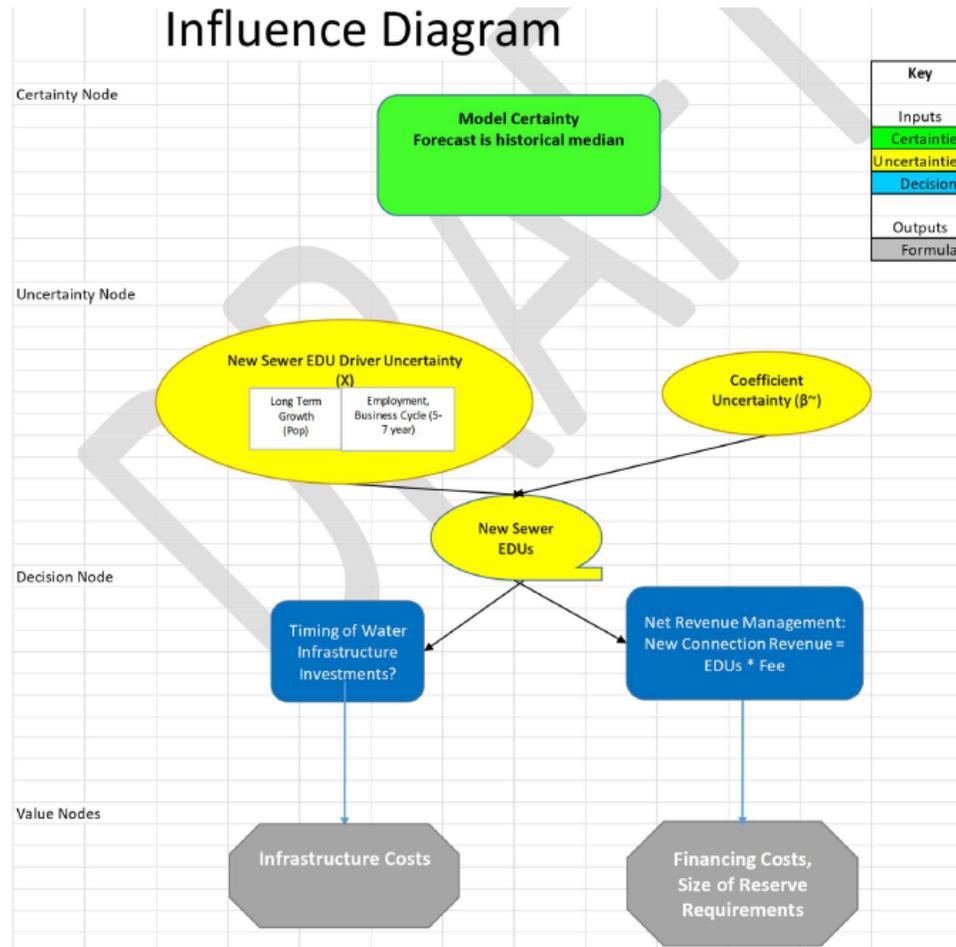
The diagram illustrates the XML structure for SLURP and SIP elements. It includes two callouts: one pointing to the 'provenance' attribute of the SLURP element, and another pointing to the 'provenance' attribute of the SIP element. A third callout points to the 'coherent' attribute of the SLURP element.

```
<SLURP name="exampleSLURP" count="1" coherent="true"
provenance="example SLURP provenance" >
  <SIP name="Domestic" count="1" type="CSV" csvr="1"
ver="1.0.0" provenance="Data from XYZ Co."
average="4.2" median="4.5">
    3.5,7.4,4.4,4.6,0.7,4.3,4.8,4.7,4.7,2.9
  </SIP>
  <SIP name="Foreign" count="10" type="CSV" csvr="1"
ver="1.0.0" provenance="Data from XYZ Co."
average="5.0" median="4.9">
    6.2,1.1,4.8,5.0,6.0,7.8,7.0,4.5,4.6,3.0
  </SIP>
</SLURP>
```

Provenance at the SIP and SLURP level

May be audited trial by trial

New Connection Revenue



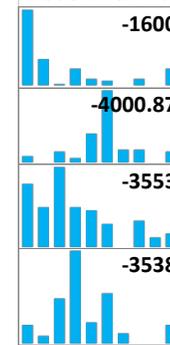
Predicting New Connection Revenue

NewSewerEDUs		%GrowthSewerEDUs	
	3,116	median	1.71%
	5,586	mean	3.14%
	4,985	std dev	3.33%



Four Models of the Annual Change in EDUs: Advantages and Disadvantages			
Model Key	Model Description	Advantages	Disadvantages
Model 0	Pick the historically most likely (median \approx 3,260 Delta_EDU)	Easiest to implement Easy to explain Overprediction can never be large	Does not predict booms or busts
Model 1	Pick year before Delta_EDU	Easy to implement/explain Captures some boom/bust Small errors (<2K) 60% of time	Most variable prediction error, on average Large errors (>2K EDU) one third of the time
Model 2	Lagged Percentage Growth of Employment	Relatively simple model Captures some boom/bust	Requires employment data and implementation
Model 3	year before Delta_EDU + Year before change in Emp growth + Difference of Emp growth from trend	Best predictor, on average	Not easy to implement or explain large errors still possible

ModelError=Actual-Predicted (Y-Y_hat)	Variability Std.Dev.	Median 50%
-1600 d_Model0Error	5,060	-
-4000.87 d_Model1Error	5,252	181
-3553 d_Model2Error	5,056	(471)
-3538 d_Model3Error	4,660	(691)



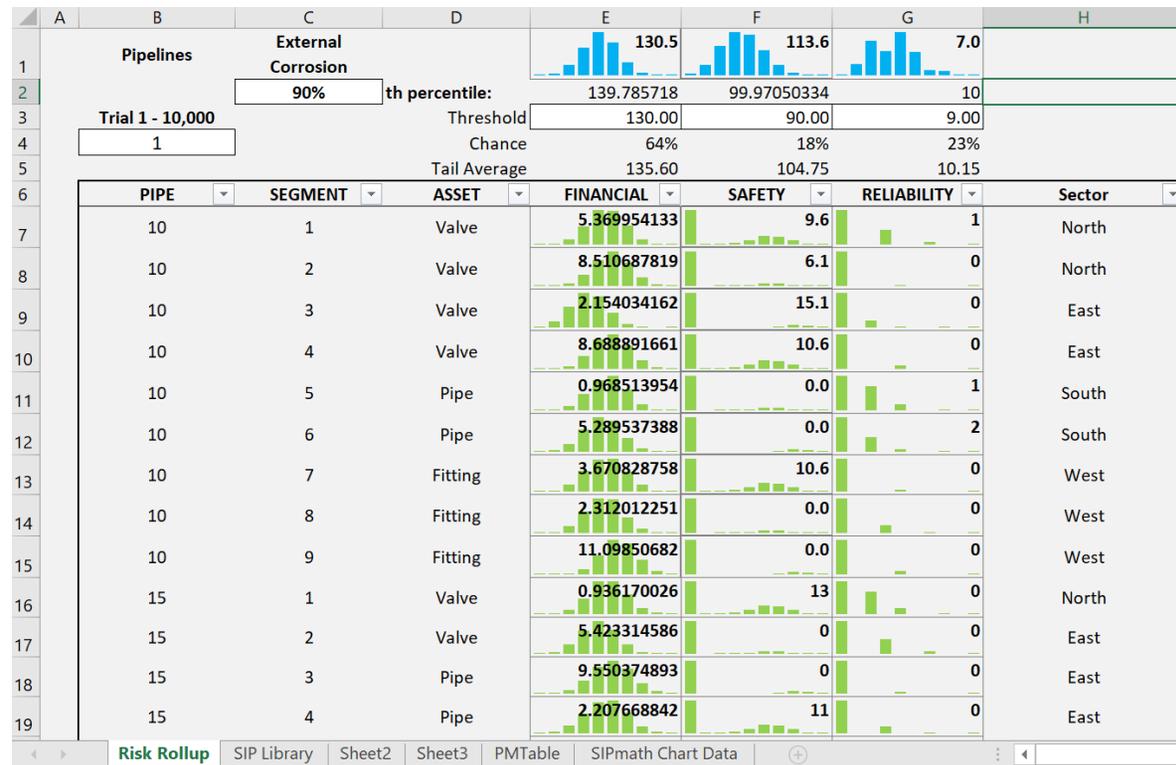
25%

60%

- # new connections each year is very uncertain on a one year time horizon.
- Explore the models

SIPmath Example for Diverse Risk Aggregation in a Gas Utility Network

- Aggregating operational risk across physical assets
- Risk Mitigation investment in a highly regulated environment require stakeholder understanding of Cost – Risk Tradeoffs across different types of risk: financial, safety, and reliability



Reference:

[Rolling up operational risk at PG&E](#)

by Jordan Alen, Christine Cowser Chapman, Melissa Kirmse, Farshad MirafTAB, & Sam L. Savage. OR/MS Today, December 2016, Volume 43, Number 6. The SIPmath paper airplane "Asset Level Model.xlsx" demonstrates how different types of risks—financial, safety, and reliability—can be combined to show cost tradeoffs.

Land Use Demand Forecasting

Probability management in residential land use demand forecasting models. This is a "paper airplane" spreadsheet model. Users are encouraged to fly it and gain some (simulated) pilot experience.

A grant work product of Water Research Foundation # A742: Probability Management for Water

Model Key

- Land Use Model 0: Water Use = Water Factor x Density x Land Use (This is a deterministic/certain model.)
- Model 1: Incorporates water factor and land use density sampling uncertainty into the model.
- Model 2: Adds weather uncertainty to Model 1.
- Model 3: Adds conservation trend to Model 2.
- Model 4: Adds price effects to Model 3.

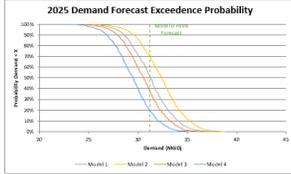
Cells taking inputs look like this:

Cells holding formulas look like this:

Cells holding uncertain values look like this:

Select forecast year to displace in chart --> 2025

* Asterisk indicates hypothetical data being used.



Probability Management in Residential Land Use Demand Forecasting Models

This is a "paper airplane" spreadsheet model. Users are encouraged to fly it and gain some (simulated) pilot experience.

Model Key

Land Use Model 0	Water Use = Water Factor x Density x Land Use (This is a deterministic/certain model.)
Model 1	Incorporates water factor and land use density sampling uncertainty into the model
Model 2	Adds weather uncertainty to Model 1
Model 3	Adds conservation trend to Model 2
Model 4	Adds price effects to Model 3

Cells taking inputs look like this:

Cells holding formulas look like this:

Cells holding uncertain values look like this:

Inputs Needs for Model 0

Code	Land Use Label	Water Factors (gpcd)	Land Use Density (durs)	Land Use Forecast (mgd)					
		2019	2020	2025	2030	2035	2040*		
102	Low Density - In-use	308	3.0	1,936	2,224	2,422	2,705	2,984	3,233
102	Medium Density - In-use	202	7.5	5,262	5,660	6,195	6,706	7,246	7,845
102	Medium-High Density - In-use	80	17.0	2,420	2,430	2,830	3,300	3,852	4,322
102	High Density - In-use	125	32.5	1,064	1,201	1,571	1,906	2,304	2,664

Model 0 Water Use Forecast (mgd)

Code	2019	2020	2025	2030	2035	2040
102	116	131	221	251	281	310
102	629	1,071	760	1,052	1,352	1,652
102	64	63	75	73	81	87
102	450	453	644	821	1,026	1,231
Total	259	288	311	329	324	356

Additional Inputs Needs for Model 1

Code	Land Use Label	Water Factors (gpcd)	Dist	Name	Mean	Std Err	Code	Dist	Name
102	Low Density - In-use	308	1.95	100	0.82	0.10	102	LogN	Wt_HD_102
102	Medium Density - In-use	202	0.64	200	0.23	0.02	102	LogN	Wt_HD_102
102	Medium-High Density - In-use	80	0.70	200	0.43	0.02	102	LogN	Wt_HD_102
102	High Density - In-use	125	0.70	200	0.43	0.02	102	LogN	Wt_HD_102

Model 1 Water Use Forecast (mgd)

Code	2019	2020	2025	2030	2035	2040
102	116	131	221	251	281	310
102	629	1,071	760	1,052	1,352	1,652
102	64	63	75	73	81	87
102	450	453	644	821	1,026	1,231
Total	259	288	311	329	324	356

Additional Inputs Needed for Model 2

Code	Land Use Label	Water Factor	Ratio of Max to Min Month Demand*	Monthly Demand Index	Weather	Dist	Name										
				Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec		
102	Low Density - In-use	312	2.2	0.70	0.68	0.79	0.85	1.00	1.05	1.40	1.50	1.50	1.00	0.85	0.72	0.62	Wt_HD_102
102	Medium Density - In-use	323	1.7	0.76	0.76	0.79	0.89	1.00	1.05	1.21	1.26	1.12	0.95	0.83	0.69	0.62	Wt_HD_102
102	Medium-High Density - In-use	95	1.8	0.79	0.77	0.79	0.89	1.02	1.14	1.20	1.24	1.13	0.97	0.85	0.62	0.62	Wt_HD_102
102	High Density - In-use	112	1.2	0.85	0.81	0.86	0.98	1.06	1.06	1.19	1.19	1.06	0.89	0.86	0.81	0.62	Wt_HD_102

Weather Seed: 500

Note: Model 2 also requires historical precipitation and temperature data for service area to be entered on the Weather Data worksheet.

Model 2 Water Use Forecast (mgd)

Code	2019	2020	2025	2030	2035	2040
102	116	131	221	251	281	310
102	629	1,071	760	1,052	1,352	1,652
102	64	63	75	73	81	87
102	450	453	644	821	1,026	1,231
Total	259	288	311	329	324	356

Additional Inputs Needed for Model 3

Model 3 requires estimates of non-Persons Per Household (PPH) by land use code and a proportion of future water savings from plumbing codes and programs

Code	Land Use Label	PPH*	Minimum	2020	2025	2030	2035	2040
102	Low Density - In-use	1.0	2.27	3.85	4.90	5.60	6.07	
102	Medium Density - In-use	3.0	2.27	3.85	4.90	5.60	6.07	
102	Medium-High Density - In-use	2.0	2.27	3.85	4.90	5.60	6.07	
102	High Density - In-use	2.0	2.27	3.85	4.90	5.60	6.07	

Most Likely

Code	2020	2025	2030	2035	2040
102	2.84	4.81	6.12	7.00	7.59
Total	2.84	4.81	6.12	7.00	7.59

Maximum

Code	2020	2025	2030	2035	2040
102	3.85	5.29	6.74	7.70	8.35
Total	3.85	5.29	6.74	7.70	8.35

Seed: 500

Dist: LogN

Name: consrv_0 consrv_1 consrv_2 consrv_3 consrv_4 consrv_5

IFH-GPCD savings projections for Orange County prepared for DVRM (M.Cubed, 2016)

Model 3 Water Use Forecast (mgd)

Code	2019	2020	2025	2030	2035	2040
102	116	131	221	251	281	310
102	629	1,071	760	1,052	1,352	1,652
102	64	63	75	73	81	87
102	450	453	644	821	1,026	1,231
Total	259	288	311	329	324	356

Additional Inputs Needed for Model 4

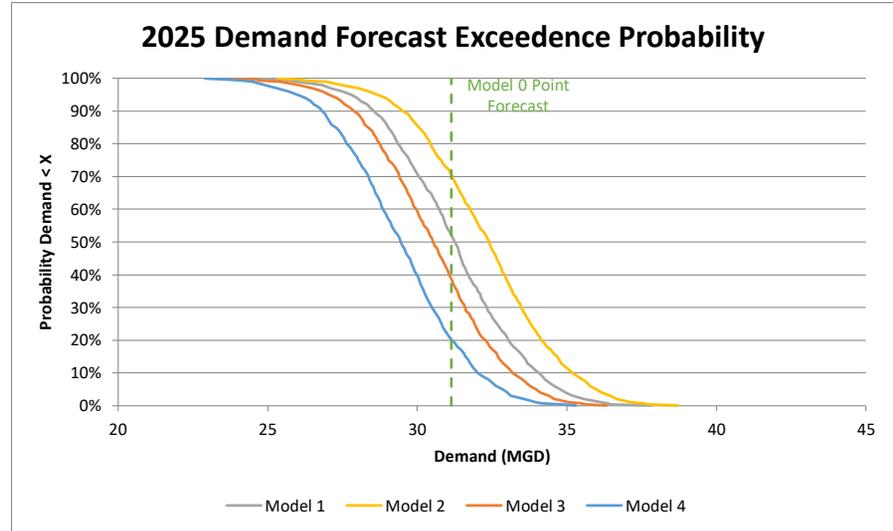
Model 4 requires price elasticity estimates and a forecast of the rate of increase in the real (inflation-adjusted) volumetric water rate

Code	Land Use Label	Price Elasticity Estimate	Projected Relative Increase in Real Volumetric Rate for Residential Service					
			2019	2020	2025	2030	2035	2040
102	Low Density - In-use	-0.25	0.00	0.00	0.00	0.00	0.00	0.00
102	Medium Density - In-use	-0.25	0.00	0.00	0.00	0.00	0.00	0.00
102	Medium-High Density - In-use	-0.16	0.05	0.00	0.00	0.00	0.00	0.00
102	High Density - In-use	-0.10	0.03	0.00	0.00	0.00	0.00	0.00

Note: This could be based on historical rate of growth in volumetric rate or 8 could use projections prepared by the finance department. The demonstration model is assuming a 2% annual average rate of increase.

Model 4 Water Use Forecast (mgd)

Code	2019	2020	2025	2030	2035	2040
102	116	131	221	251	281	310
102	629	1,071	760	1,052	1,352	1,652
102	64	63	75	73	81	87
102	450	453	644	821	1,026	1,231
Total	259	288	311	329	324	356



Conclusion

- By reducing the cost of creating, validating, and communicating probabilistic analysis... Probability Management gives permission to face uncertainty and inform decisions.
- Ignoring uncertainty does not make it go away. It only leads to ulcers and sleepless nights.

Resources:

- [Probabilitymanagement.org](http://probabilitymanagement.org)
- Open source risk based forecast book, <http://www.gfoa.org/forecastbook>
- AWE, www.financingsustainablewater.org