From Operator Voices to Decision Models: Comparing Manual and AI-Assisted Problem Formulation for Rural Alaska Water Systems

Ana Lacau[1], Suzanne Pierce[2], Lauryn Spearing[3], Kasey M. Faust[1*]
[1]University of Texas at Austin; [2]Texas Advanced Computing Center; [3]Tufts University; *corresponding author: faustk@utexas.edu

1. PROBLEM STATEMENT

Rural Alaska water systems operate under extreme conditions, including permafrost, harsh climate, and geographic isolation [1, 2]. Many communities rely on haul systems that transport water in and wastewater out by vehicle. Unlike piped infrastructure, these systems require an additional workforce to drive trucks. Currently, we lack an understanding of how resource constraints, infrastructure limitations, and workforce issues interact to affect the performance of hauled water systems. Operators make daily decisions about resource allocation, maintenance, and emergency response without structured decision-support models—a critical gap, given that system failures can quickly escalate into public health emergencies [3, 4]. This research explores how to formulate operator challenges as structured decision problems. Here, we use semi-structured interviews with local water-sector experts from the Yukon-Kuskokwim Delta. We identify decision variables, constraints, and objective functions to reveal intervention points and create transferable decision-support frameworks. To accomplish this, we explore a comparative methodology: (1) qualitative hand-coding to extract problem parameters and constraints; and (2) LLM-assisted analysis to identify these elements at scale. We evaluate parameter identification accuracy, resource efficiency, and scalability to other resource-constrained water systems, establishing best practices for leveraging AI tools while identifying where human interpretation remains essential.

2. APPROACH

2.1 Data Collection

Semi-structured interviews with water system stakeholders were conducted in the Yukon-Kuskokwim (YK) Delta region between 2022 and 2023. For this analysis, we leverage interviews with regional water-sector experts in various roles, including water service operators, administrative leadership, and water truck drivers. Following principles of stakeholder theory [5], participants were selected based on direct interaction with water utilities. The interview protocol was designed to elicit comprehensive information about water distribution systems with particular attention to daily operations, stakeholder interactions, and operational challenges. All interviews were audio-recorded with informed consent, transcribed verbatim, and quality-checked before analysis.

2.2 Human-Led Qualitative Coding

Interview transcripts were analyzed using NVivo [6]. We developed a codebook based on components of decision-making problems, coding excerpts into seven categories: Objectives, Constraints, Trade-Offs, Decision Variables, Options, Solutions, and State Variables (see Table 1). These categories constitute a deductive coding framework, predicated on optimization theory [7]. However, the specific codes within each category emerged inductively from the interview data itself. This hybrid approach enabled us to systematically structure the data within a decision-making framework while remaining responsive to the unique challenges and nuances articulated by water system stakeholders [8].

Table 1: *Optimization Problem Components Coded in Qualitative Analysis*

Term	**Definition**	**Example**
Objectives	The goals or desired outcomes that operators seek to achieve.	Community education Providing safe water
Constraints	Limitations or boundaries that restrict actions.	Limited funding Workforce shortage

Trade-Offs	Situations where improving one objective necessarily compromises another.	Improving operator satisfaction compromises providing water to all customers.
Decision Variables	Controllable factors operators can adjust to influence system performance.	Operator work hours
Options	Discrete alternative courses of action that are available to operators when facing a decision point.	Residents choose between treated and natural water sources
Solutions	Strategies, workarounds, or approaches that operators or residents have implemented to address challenges; proven methods for navigating system constraints.	Remote worker program Managerial capacity
State Variables	Fixed or external factors that define the problem context but are not directly controllable by operators.	Permafrost Seasonal fluctuations

2.3 AI-augmented Qualitative to Quantitative Translation

Computational approaches create semantic bridges between qualitative stakeholder perspectives and formal scientific knowledge systems, enabling systematic translation between natural-language expressions and standardized scientific constructs. The approach employed three advanced NLP techniques: 1) topic modeling via Latent Dirichlet Allocation (LDA) to discover thematic structure [9], 2) dependency parsing to extract decision components with spaCy grammatical analysis, and 3) semantic similarity to link concepts via TF-IDF vectorization to scientific variables. The resulting decision components were mapped to standardized nomenclature across domains to establish science variable links to standard names, thereby supporting analytical workflow automation [10-12].

While manual qualitative coding provides a rigorous interpretive foundation, it poses scalability challenges for decision-support applications that require rapid synthesis across diverse stakeholder groups, geographies, or problem domains. Here, we apply topic modeling to identify thematic structures across interviews, extract Scientific Variable Objects (SVOs) to identify measurable quantities in stakeholder interviews, and map qualitative expressions to formal scientific domains via semantic similarity analysis. Comparing multiple topic model configurations (baseline, enhanced stopword filtering, and strict parameter settings) enables us to select the configuration that best preserves stakeholder meaning while supporting computational workflows.

3. KEY FINDINGS

3.1 Human-Led Findings

Initial results are shown in Table 2; coding is ongoing at the time of submission.

Table 2: *Distribution of Decision Components Identified Through Qualitative Coding*

Code	Frequency of Interviewees	Total Excerpts
Constraints	**2**	**21**
Financial Constraints	1	7
Household Responsibilities Impact Revenue	1	1
Meeting Funding Qualifications	1	1
Limited Resources	1	3
Customer Non-Payment	1	2
Workforce Constraints	2	12
Extensive Training	1	1
Workforce Shortage	2	6
Costly Operator Training	1	1
Subsistence-Related Seasonal Absence	1	2
Non-contextual National Operator Certification Exams	1	1
Recruiting	1	1
Project and Supply Constraints	1	3

Intermittent Supply Issues	1	2
Long Project Timelines	1	1
Decision Variables	**2**	**11**
Workforce Decision Variables	1	4
Long Hours	1	1
Overtime Compensation	1	1
Funding CDL Training	1	1
Wages	1	2
Managerial Decision Variables	1	4
Communication	1	3
Technology Integration	1	1
Water Treatment Decision Variables	1	3
Operator Skepticism	1	1
Fluoridation	1	1
Water Quality Sampling	1	1
Objectives	**2**	**16**
Community Awareness Objectives	**1**	**3**
Promote Safe Water and Hygiene Practices	1	2
Community Education	1	1
Financial Objectives	**1**	**1**
Profit	1	1
Operational Objectives	**2**	**8**
Providing Onsite Assistance to Operators	1	2
Providing Operator Training	1	2
Providing Water to Customers	1	1
Providing Water When There is a Fire	1	1
Running a Sustainable System	1	1
Improve Operator Satisfaction	1	1
Public Health Objectives	**1**	**4**
Ensure Water Safety	1	3
Water Fluoridation	1	1
Options	**2**	**5**
Operator Work Schedule	1	2
Full-time Work (benefits + lower pay)	1	1
Part-time Work (no benefits + higher pay)	1	1
Water Source	**1**	**2**
Groundwater	1	1
Surface Water	1	1
Sustainable Practices	**1**	**1**
Graywater Recycling	1	1
State Variables	**2**	**12**
Arctic Conditions	**2**	**10**
Permafrost	1	2
Erosion	1	1
Temperature Issues	1	1
Running a Reliable System in Extreme Conditions	1	2
Pipe Freeze	2	2
Seasonal Fluctuation in Water Source Quality	1	2
Workforce Culture	**2**	**2**
Male-dominated	2	2
Trade-Offs	0	0
Solutions	**1**	**7**
Financial Solutions	1	2

Funding	1	2
Public Health Solutions	1	2
Regular Sampling	1	1
Public Notifications	1	1
Workforce Solutions	1	3
Remote Worker Program	1	1
Stable Workforce	1	1
Managerial Capacity	1	1

Constraints affecting technical performance were most frequently cited, with workforce challenges dominating, followed by financial limitations and sub-Arctic operational conditions. Objectives focused on operational and public health goals, while decision variables spanned workforce management, communication, and water treatment. Solutions were identified less frequently, and trade-offs were never explicitly discussed in the interviews.

Several leverage points emerged where constraints, objectives, and interventions align. The most prominent is workforce capacity. Workforce shortage constraints affect training objectives, with decision variables including compensation and employer-funded certification, and solutions like remote worker programs. A second leverage point links water safety objectives to quality sampling and monitoring solutions, particularly given seasonal environmental fluctuations. Notably, although financial and sustainability objectives were identified, the decision variables considered were limited to external funding, indicating a gap in operator agency.

3.2 AI-Led Findings

Analyses were conducted using 9 interviews (including the two hand-coded interviews). The additional interviews provided multi-stakeholder perspectives and a broader context for the analysis. Figure 1 shows the distribution of SVOs extracted from stakeholder interviews. The center represents all variables (n= 49 unique); the middle ring shows science domains color-coded by discipline; the outer ring displays the 10 most frequently mentioned variables per domain. Segment size proportional to mention frequency.

This analysis successfully created a semantic bridge between stakeholder narratives and scientific knowledge systems, enabling bidirectional translation between qualitative community input and quantitative scientific frameworks. Analysis revealed thematic clusters spanning hydrological processes, infrastructure systems, climate adaptation, and community resilience. Automated SVO extraction identified specific measurable quantities (water levels, flood frequency, permafrost depth). While this does not yet present decision-making formulation findings, the resulting structured representations – objectives, constraints, trade-offs – will be developed in the next phase of analysis, maintaining explicit provenance to the source stakeholder language through topic-to-domain mapping and SVO-theme connection matrices. This will enable direct comparison between human-led and AI-led approaches, supporting transparent, defensible decision processes that ground models in authentic community priorities while also identifying critical data needs for model refinement.

Figure 1: *Scientific Variables Distribution by Science Domain.*

4. IMPLICATIONS

This research offers both practical and methodological contributions with implications for utility management in resource-constrained settings and for qualitative research methods more broadly. By framing operator challenges as decision-making problems, this work provides a foundation for decision-support tools tailored to the constraints of remote Alaska. Identifying decision variables, constraints, and trade-offs enables managers and policymakers to systematically evaluate interventions and identify high-leverage points for improvement. Understanding which constraints are most binding (e.g., workforce capacity, equipment availability, budget) allows funding agencies to target resources effectively. Mapping trade-offs helps operators balance competing priorities, such as immediate service needs and long-term infrastructure investments.

The comparative analysis of hand-coding versus LLM-assisted analysis provides critical insights into AI's role in qualitative research. By evaluating accuracy, efficiency, and scalability, this work establishes evidence for when and how AI can augment traditional methods. If LLM analysis demonstrates comparable accuracy with substantial time savings, it enables scaling qualitative research to larger datasets—capturing more diverse perspectives without proportional resource demands. Conversely, identifying where human interpretation remains highlights the value of domain expertise in translating lived experience into formal structures. This suggests hybrid approaches where AI handles initial pattern identification while human researchers focus on nuanced interpretation and validation.

5. REFERENCES CITED

[1] "Infrastructure Challenges in the Alaskan Arctic | The Belfer Center for Science and International Affairs." Accessed: Apr. 09, 2025. [Online]. Available: https://www.belfercenter.org/publication/infrastructure-challenges-alaskan-arctic

[2] "Alaska Infrastructure Report Card | ASCE," ASCE's 2025 Infrastructure Report Card |. Accessed: Apr. 09, 2025. [Online]. Available: https://infrastructurereportcard.org/state-item/alaska/

[3] L. A. Spearing, P. Mehendale, L. Albertson, J. A. Kaminsky, and K. M. Faust, "What impacts water services in rural Alaska? Identifying vulnerabilities at the intersection of technical, natural, human, and financial systems," *J. Clean. Prod.*, vol. 379, p. 134596, Dec. 2022, doi: 10.1016/j.jclepro.2022.134596.

[4] M. LaPatin, N. Ritsch, D. E. Armanios, L. Albertson, L. E. Katz, and K. M. Faust, "A Stakeholder-Systems Analysis of Water Provision in Rural Alaska," *ACS EST Water*, vol. 4, no. 2, pp. 575–590, Feb. 2024, doi: 10.1021/acsestwater.3c00612.

[5] "Stakeholder Theory: Reviewing a Theory That Moves Us." Accessed: Jan. 16, 2026. [Online]. Available: https://journals.sagepub.com/doi/epdf/10.1177/0149206308324322

[6] *NVIVO*. (2023). QSR International Pty Ltd.

[7] S. S. Rao, *Engineering Optimization: Theory and Practice*. John Wiley & Sons, 2019.

[8] J. Saldaña, *The coding manual for qualitative researchers*, 2nd ed. Los Angeles: SAGE, 2013.

[9] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Jan. 2003.

[10] Y. Gil *et al.*, "Intelligent systems for geosciences: an essential research agenda," *Commun. ACM*, vol. 62, no. 1, pp. 76–84, Dec. 2018, doi: 10.1145/3192335.

[11] M. D. Wilkinson *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship," *Sci. Data*, vol. 3, no. 1, p. 160018, Mar. 2016, doi: 10.1038/sdata.2016.18.

[12] Eaton, B., Gregory, J., Drach, B., Taylor, K., Hankin, S., Blower, J., Caron, J., Signell, R., Bentley, P., Rappa, G., Höck, H., Pamment, A., Juckes, M., & Raspaud, M., *NetCDF Climate and Forecast (CF) Metadata Conventions*. (2021). [Online]. Available: http://cfconventions.org/