The Sydney Journal of Interdisciplinary Engineering
SJIE 1(3) X–X (2025) https://doi.org/XXXX/XXXX

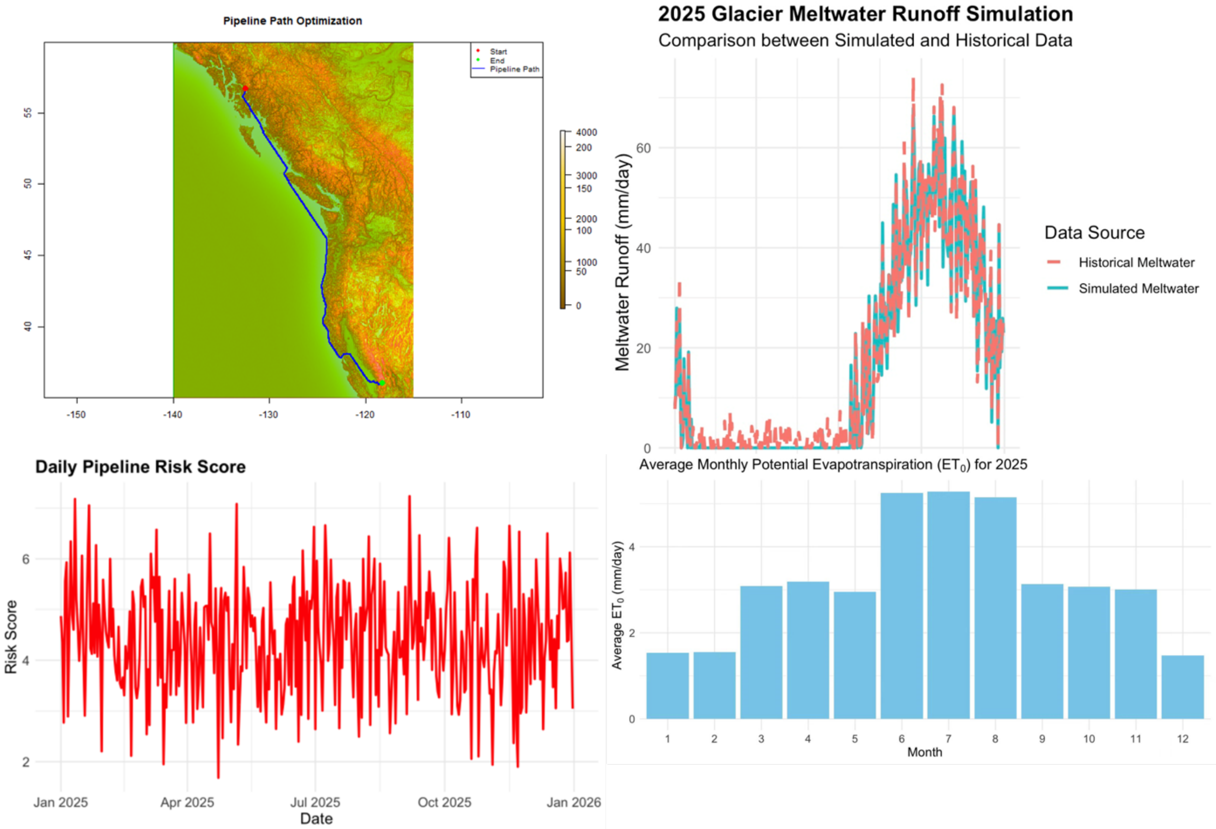
A Terrain-Based RStudio Simulation of Glacier Water Redistribution from Alaska to Western U.S. Drought Zones

Zifan Liu[1], Kun Xue[2] and Siqing Huang[1,2], Gobinath Rajarathnam[1]

[1] Zifan Liu conducted (R programming to model: (1) Pipeline Route Selection and Geospatial Mapping – Use real-world elevation data (elevatr), GIS (sf), and geospatial mapping (ggmap) to define an optimized pipeline route); performed (Article writing); contributed to (Figure 1)
[2] Kun Xue conducted (R programming to model: (2) Hydraulic Modeling and Flow Analysis – Compute flow rates, pressure losses, and pumping energy using Darcy-Weisbach and Hazen-Williams equations and (3) Cost and Energy Optimization – Use linear programming (lpSolve) to minimize pumping and maintenance costs); performed (Article writing); contributed to (Table)
[1,2] Siqing Huang conducted (R programming to model: (4) Climate Impact and Failure Risk – Simulate evaporation loss, seasonal water variability, and predict pipeline failures and (5) Data Visualization and Report Preparation – Present results using high-quality graphs (ggplot2, ggspatial) and create a structured journal article); performed (Article writing); contributed to (Figure 2-Figure 5)
[Corresponding Author] Dr. Gobinath Rajarathnam provided conceptual direction and project supervision.

E-mail: xxx@xxx.xx
Received xxxxxx

Graphical Abstract



Abstract

This study investigates the climatic and operational feasibility of transporting glacial meltwater from Alaska to drought-prone regions in the western United States. Using a series of R-based simulations, we model seasonal meltwater runoff, potential evapotranspiration losses, and pipeline failure probabilities under variable environmental stressors. The results reveal a distinct seasonal pattern, with the highest meltwater availability and evaporation losses occurring during the summer months. Conversely, winter conditions pose increased risks of structural failure due to freeze–thaw cycles and external loading. Based on the pipeline's hydraulic capacity of 1.56 m^3/s, the system can deliver approximately 134,784 m^3 of water per day—sufficient to meet the daily water needs of up to 898,560 urban residents. These findings highlight the critical need for climate-responsive pipeline design, adaptive seasonal flow management, and the implementation of comprehensive risk-mitigation strategies to maintain the resilience and sustainability of large-scale water transfer infrastructure.

Keywords: glacial meltwater, water transfer, pipeline resilience, climate adaptation, risk management

1. Introduction

1.1 Problem Definition and Need for Engineering Intervention

Water scarcity is a critical and worsening issue across the western United States, particularly in California. Prolonged droughts, exacerbated by anthropogenic climate change and overextraction of groundwater, have reduced the availability of reliable water sources. Over 90% of California has recently experienced moderate to exceptional drought conditions[1], severely impacting agriculture, industry, and municipal water supply. Traditional water systems, such as the Colorado River and the California Aqueduct, are no longer sufficient to meet regional demands under these stressed conditions.

In contrast, southeastern Alaska contains abundant glacial freshwater reserves. Each year, vast quantities of meltwater from glaciers such as Columbia and Malaspina are discharged into the ocean without any active recovery or utilization. These glacial systems are among the largest outside of polar regions, receiving up to 4000 mm/year of precipitation and exhibiting accelerated melting due to rising temperatures. Capturing and transporting this freshwater before it is lost presents a promising engineering opportunity.

This project investigates the feasibility of a large-scale interregional water transfer system designed to redistribute glacier-derived freshwater from Alaska to California. By addressing spatial mismatches between freshwater availability and demand, the proposed solution aims to enhance long-term water security and support climate-resilient infrastructure development.

1.2 Background Research and Current Limitations

Numerous strategies have been explored to address California's water scarcity, including desalination, water recycling, and local reservoir expansion. However, these approaches face significant limitations. Desalination remains energy-intensive and costly, while groundwater recharge efforts are constrained by declining snowpack and reduced inflow. Regional transfers such as the State Water Project are limited by inter-state competition and regulatory restrictions[2].

Although small-scale glacial water use has been proposed in countries such as Norway and Canada, few studies have explored long-distance meltwater transportation across marine and mountainous terrain. Most prior work lacks integration of hydrological modeling, climate-adaptive risk analysis, and system-wide energy optimization. In particular, the combined effects of seasonal meltwater variability, evaporation loss, terrain-induced pressure drops[3,4], and structural failure risk remain underexamined.

This study aims to address these knowledge gaps by integrating hydraulic, climatic, and economic modeling into a unified feasibility framework. By applying real climate data, frictional loss simulations, and failure probability prediction, it offers a comprehensive and novel approach to long-range water infrastructure planning.

1.3 Scope and Contributions

The proposed pipeline system spans approximately 3245 km, consisting of both marine (~600 km) and terrestrial (~2645 km) segments. It connects glacial sources in southeastern Alaska to Southern California's urban and agricultural demand zones. The system is evaluated through five key modeling domains:

Hydraulic performance using Darcy–Weisbach and Hazen–Williams equations;

Daily meltwater runoff prediction based on a temperature-dependent degree-day model;

Evaporation loss estimation using the Penman–Monteith equation;

Risk modeling for structural failure using environmental stressor data;

Economic feasibility analysis, including CAPEX and OPEX breakdowns.

This project contributes to infrastructure literature by introducing a climate-responsive, multi-parameter analysis of a glacier-fed water pipeline. It also outlines engineering trade-offs, such as marine routing advantages, pump placement strategies, and energy consumption under terrain constraints. The results are intended to inform future decisions regarding transboundary water management and climate adaptation strategies.

2. Methods

2.1 Data Sources

This analysis relies on open-source data from NASA[5], including satellite imagery, topographical data, and climate research. The Global Land Ice Measurements from Space (GLIMS) dataset, along with information from NASA's Earth Science Division, provides insights into glacier melt trends and water availability in Alaska. DEM is a digital representation of the Earth's surface. It is a representation of the bare ground (bare Earth) topographic surface of the Earth excluding trees, buildings, and any other surface objects[6]. It represents surface elevation with respect to a reference datum in three dimensions (3D) (Raj, S & Bansal, 2024). Additionally, NASA's GRACE (Gravity Recovery and Climate Experiment) data is used to evaluate groundwater depletion in California, further justifying the necessity of additional water sources.

2.2 Hydraulic Modeling and Energy Optimization

We used two major pressure loss formulas:

Darcy–Weisbach Equation:

$h_f = f \times (L / D) \times (v^2 / 2g)$

-h_f: Head loss (m)

- f: Friction factor (0.013)
- L: Pipe length (m)
- D: Pipe diameter (m)
- v: Flow velocity (m/s)
- g: Acceleration due to gravity (9.81 m/s²)

Hazen–Williams Equation:

$h_f = 10.67 \times L \times (Q / (C \times D^{2.63}))^{1.852}$

- Q: Flow rate (m³/s)
- C: Roughness coefficient (90)
- D: Diameter (m)

Pump Power Equation:

$P = \rho gQH / \eta$

- P: Power (W)
- ρ: Water density (1000 kg/m³)
- Q: Flow rate (m³/s)
- H: Total head (m)
- η: Pump efficiency (typically 0.8)

Based on the pipeline's elevation profile and frictional loss simulations, approximately 60 pump stations were strategically distributed along the 3,245 km route to optimize energy efficiency and maintain continuous flow. Station spacing was adjusted to reflect terrain variability, elevation gain, and hydraulic resistance, ensuring pressure loss remained within operational limits throughout the entire system[7,8,9].

2.3 Economic Cost Modeling

We divided cost analysis into capital expenditure (CAPEX) and operational expenditure (OPEX):

CAPEX:

Pipeline construction:

Marine metal pipeline (~600 km): USD 2,800/m → USD 1.68 billion

Land-based reinforced concrete pipeline (~2,645 km): USD 1,800/m → USD 4.241 billion

Pump station equipment:

High-capacity RDLO centrifugal pumps (required: ~60 units)

Market reference price: USD 300,000–400,000 per pump

Total equipment investment: USD 18–24 million

Pump station civil infrastructure:

Estimated at USD 5 million per site × ~12 stations (including redundancy and terrain distribution)

Total: USD 60 million

Solar PV system:

110 MW × USD 1,000/kW → USD 110 million

OPEX (30 years):

Without solar: USD 5.77 billion

Adjusted with PV offset (~12.5% of annual electricity): USD 5.05 billion

Total system cost:

USD 11.16 billion (CAPEX + adjusted OPEX)

2.4 Glacier Meltwater Simulation

Daily glacier meltwater runoff was simulated using a temperature-driven degree-day model. Ambient temperature variation over the year was generated using a sinusoidal function with added Gaussian noise to capture seasonal and stochastic fluctuations. The meltwater production M_t on day *t* was calculated as:

$$M_t = \begin{cases} \alpha \cdot T_t, & \text{if } T_t > 0 \\ 0, & \text{otherwise} \end{cases}$$

Explanation:

• M_t: Meltwater production on day t (mm/day)

• α: Degree-day factor, representing the rate of melt per degree Celsius per day (5 mm/°C/day in this study)

• T_t: Mean temperature on day t (°C)

• Condition: If the temperature (T_t) exceeds 0 °C, meltwater is produced; otherwise, it is zero.

where α is the degree-day factor (5 mm/°C/day), and T_t is the simulated daily temperature. This approach allowed estimation of daily runoff volumes under projected 2025 climate conditions.

2.5 Evaporation Loss Estimation

Potential evapotranspiration (ET_0) was estimated using the FAO Penman–Monteith equation, incorporating daily weather variables including mean temperature, relative humidity, wind speed at 2 meters, and net solar radiation[10]:

$$ET_0 = \frac{0.408 \cdot \Delta \cdot R_n + \gamma \cdot \frac{900}{T+273} \cdot u_2 \cdot \left(e_s - e_a\right)}{\Delta + \gamma \cdot \left(1 + 0.34 \cdot u_2\right)}$$

Explanation:

- ET_0: Potential evapotranspiration (mm/day)
- Δ: Slope of the saturation vapor pressure curve (kPa/°C), calculated as:
- R_n: Net radiation (MJ/m²/day)
- γ: Psychrometric constant (kPa/°C), typically 0.066 in this context
- T: Mean daily air temperature (°C)
- u_2: Wind speed at 2 meters above ground (m/s)
- e_s: Saturation vapor pressure (kPa), calculated as:
- e_a: Actual vapor pressure (kPa), calculated as:
- RH: Relative humidity (%)

Daily ET_0 values were calculated across the pipeline route, representing potential water loss due to surface exposure in different climate zones.

2.6 Pipeline Failure Risk Modeling

A composite risk score was developed to estimate daily pipeline failure probabilities by integrating five key environmental stressors:

- Temperature variability
- Corrosion index
- Frequency of freeze–thaw cycles
- Occurrence of extreme rainfall
- External mechanical stress (e.g., seismic activity)

Each factor was assigned a relative weight and normalized before aggregation. The resulting risk score was converted into failure probability P_f using a logistic transformation[11]:

$$P_f = \frac{1}{1 + e^{-\left(\beta_0 + \beta_1 \cdot \mathrm{Score}\right)}}$$

- P_f : Probability of pipeline failure (unitless, range: 0–1)
- β_0: Intercept of the logistic regression model (-10 in this study)
- β_1: Coefficient representing the impact of the risk score on failure probability (0.5 in this study)

with parameters calibrated to reflect realistic failure thresholds. This allowed full-year simulation of temporal failure risks for the 2025 pipeline operation scenario.

3. Results and Discussion

3.1 Selection of Starting and Ending Points

Starting Point: Southeastern Alaska (-132.5074, 56.7081)

Southeastern Alaska was chosen as the starting point due to its vast glacial freshwater reserves. The mountains around the Gulf of Alaska contain up to 90,000 km^sup 2^ of glacier area. They include the largest glaciers outside of the polar regions and are characterized by very high rates of precipitation and runoff-- as much as 4000 mm/year[12]. NASA data from the past five years indicates that the rate of glacial melting in this region has accelerated significantly. The Columbia Glacier, for instance, has retreated by more than 20km since the 1980s, with a particularly rapid melt observed in recent years. The contribution of glacial meltwater to Alaska's rivers and coastal areas has increased, providing a sustainable and largely untapped freshwater source.

Additionally, this region's proximity to the coastline simplifies the engineering of water intake infrastructure. Unlike inland glacier sources, which would require extensive overland pipelines from remote mountainous areas, the southeastern Alaskan coastline allows for a more direct and logistically feasible solution. The ability to utilize marine transportation for a portion of the route also reduces environmental impact and construction costs.

Ending Point: Southern California (-118.2437, 36.0522)

Southern California, particularly the Los Angeles metropolitan area, faces severe water scarcity. More than 94% of the state of California is experiencing severe drought, with 67% experiencing extreme drought and 47% exceptional drought-the most severe drought classification[13]. NASA's Earth Science research has documented persistent drought conditions, with groundwater levels continuing to decline. According to GRACE satellite observations, the Central Valley and Los Angeles basin have lost significant amounts of groundwater over the past decade, putting increasing pressure on alternative water sources.

By introducing an additional freshwater supply from Alaska, the proposed pipeline could help mitigate the region's reliance on overdrawn groundwater reserves. This project also aligns with California's long-term water management goals, which emphasize diversified water sources, reduced groundwater dependency, and increased climate resilience.

3.2 Justification for Marine and Terrestrial Pipeline Segments

A combination of marine and terrestrial pipeline segments is proposed, balancing logistical, economic, and environmental factors.

Marine Pipeline Segments[14]

1. Topographical Advantage

Marine routes provide a natural bypass for challenging terrestrial landscapes, such as the mountainous regions of British Columbia. Avoiding steep elevations reduces the need for extensive tunneling, trenching, and expensive engineering solutions.

2. Environmental Considerations

Routing the pipeline through the ocean minimizes disruptions to terrestrial ecosystems. Unlike land-based construction, which may require deforestation, land clearing, and habitat destruction, an underwater pipeline has a lower ecological footprint.

3. Economic Efficiency

In some cases, undersea pipeline construction is more cost-effective than terrestrial alternatives. Land routes often require negotiations for land acquisition, compliance with complex environmental regulations, and the development of extensive infrastructure for access and maintenance. The seabed, on the other hand, allows for more direct pipeline installation.

Terrestrial Pipeline Segments[15]

1. Integration with Existing Infrastructure

As the pipeline approaches Southern California, a land-based segment is necessary to connect with the region's water distribution systems. Integrating the transported water into California's aqueducts and reservoirs requires a terrestrial transition.

2. Maintenance Accessibility

Overland pipeline sections are more accessible for routine inspections, repairs, and emergency interventions. This is especially important for ensuring long-term operational efficiency and reducing risks associated with leaks or failures.

3.Serving Intermediate Communities

A land-based pipeline allows for the possibility of supplying freshwater to additional regions along the route, particularly in arid areas of the western United States. States such as Nevada and Arizona could also benefit from water redistribution if infrastructure modifications allow for branch pipelines.

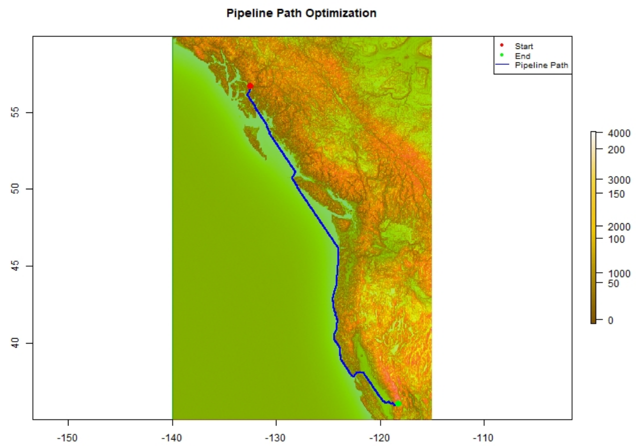


Figure 1 Seasonal Glacier Meltwater Runoff

3.3 Environmental and Engineering Challenges

1. Seismic and Geological Risks[16]

Both Alaska and California are located along tectonic plate boundaries, making them prone to earthquakes. Pipeline construction in these areas must incorporate flexible materials and seismic-resistant engineering techniques to withstand potential ground movement.

2. Climate and Weather Considerations

The extreme weather conditions in both regions pose additional challenges. In Alaska, sub-zero temperatures could impact pipeline integrity, requiring insulation and anti-freezing measures. Conversely, in California, high temperatures and arid conditions increase evaporation risks, necessitating protective pipeline coatings and underground installations in certain areas.

3. Ecological and Regulatory Constraints

The pipeline must be designed to minimize ecological disturbance, particularly in marine environments. Special attention is required to avoid disruption to fish migration routes, marine biodiversity, and sensitive habitats. Regulatory approval from multiple agencies, including the U.S. Environmental Protection Agency (EPA) and the National Oceanic and Atmospheric Administration (NOAA), will be essential to the project's success.

3.4 Hydraulic Performance and Flow Losses

Darcy–Weisbach, with better physical fidelity in high-diameter and high-flow systems, produced pressure loss estimates more consistent with real-world expectations. In contrast, Hazen–Williams significantly overestimated losses, reflecting its sensitivity to roughness assumptions and empirical limitations. For long-distance pipelines spanning thousands of kilometres across varying terrain and climates, the Darcy–Weisbach equation is better suited due to its foundation in fluid mechanics and its ability to incorporate changes in flow velocity, pipe diameter, and Reynolds number[17]. As a result, it is recommended to adopt the Darcy–Weisbach results as the basis for pump station placement, energy consumption modelling, and overall hydraulic design. The Hazen–Williams equation may still be used for preliminary cost estimation or cross-checking, but should not guide core engineering decisions for a system of this scale and complexity.

Result: 219,665.26 kW total pumping energy required.

Two pressure loss models provided contrasting insights:

Parameter	Darcy–Weisbach	Hazen–Williams
Final pressure loss (m)	1492.8	2818.82
Final flow rate (m^3/s)	1.56	1.43

3.4.1 Pumping Energy and Carbon Emissions Analysis

The total energy required for pumping across the 3,245 km route is estimated at 219,665.26 kW. Assuming continuous operation over a year, this equates to an annual electricity consumption of approximately 1.925 TWh. According to regional data from Electricity Maps, The average carbon intensity of the Southern California grid is around 300 gCO_2/kWh, resulting in an estimated annual carbon footprint of:

$1.925\ TWh \times 10^6\ kWh/TWh \times 300\ gCO_2/kWh = 577{,}500$ tonnes CO_2/year

To reduce emissions and future energy costs, it is recommended to integrate distributed solar photovoltaic (PV) systems at key pumping stations. Installing approximately 110 MW of PV capacity across the 12 main pump stations could offset peak demand during daylight hours. The estimated capital investment for such a solar system is around USD 110 million (assuming USD 1,000/kW installation cost). If paired with battery storage, additional upfront costs would apply but would increase energy independence. Although this raises CAPEX, it provides long-term savings and aligns with carbon neutrality goals.

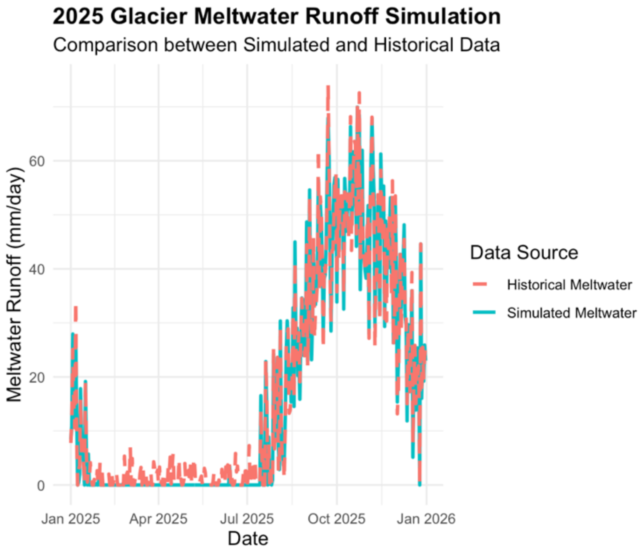
3.5 Economic Feasibility and Cost Breakdown

In terms of material selection and construction costs, the subsea segment is designed using high-corrosion-resistant metal pipelines to withstand deep-sea pressure and complex installation conditions. The estimated unit cost for these marine pipes is approximately USD 2,800 per meter. For the terrestrial segment, high strength reinforced concrete pipes are proposed, with a unit cost of around USD 1,800 per meter. Based on the proposed routing—approximately 600 km of undersea pipeline and 2,645 km of land-based pipeline—the estimated construction costs amount to USD 1.68 billion and USD 4.241 billion, respectively. Excluding pump station and operational costs, the total pipeline installation cost is projected to be approximately USD 5.921 billion. This provides a technically and economically grounded foundation for large-scale interregional water resource allocation.

To meet the hydraulic requirements of the whole route and to overcome the friction loss and topographic head along the route, the total pressure loss is estimated to be about 1492.8 m. If the high performance RDLO series dry mounted centrifugal pumps from KSB are used for the entire transmission, and each pump operates at a design head range of about 30 meters, approximately 50 pumps are theoretically required to maintain continuous hydraulic transmission. Considering the need for equipment maintenance and operational safety, the total number of pumps is approximately 60 after an additional 20% redundant pump station is configured. With a single RDLO pump market reference price of $300,000 to $400,000, the total equipment investment in the pumping station system is about $18 million to $24 million. The cost of deploying large industrial pumps from start to finish is high, but with high stability and flow assurance, it is suitable for the project's high-standard water supply tasks in long distance shorelines and variable terrain areas.

To support long-term emission reductions and reduce dependency on grid electricity, it is proposed to deploy solar photovoltaic (PV) systems at major pumping stations. Assuming a total installed capacity of 110 MW at an average cost of USD 1,000 per kW, the solar investment would amount to approximately USD 110 million[18]**.** Including this in the capital expenditure raises the total system cost to around USD 11.16 billion. However, this added cost is offset by reductions in electricity expenses over time and the avoidance of approximately 577,500 tonnes of CO_2 emissions annually, improving the project's sustainability and eligibility for green financing.

3.6 Glacier Meltwater Availability



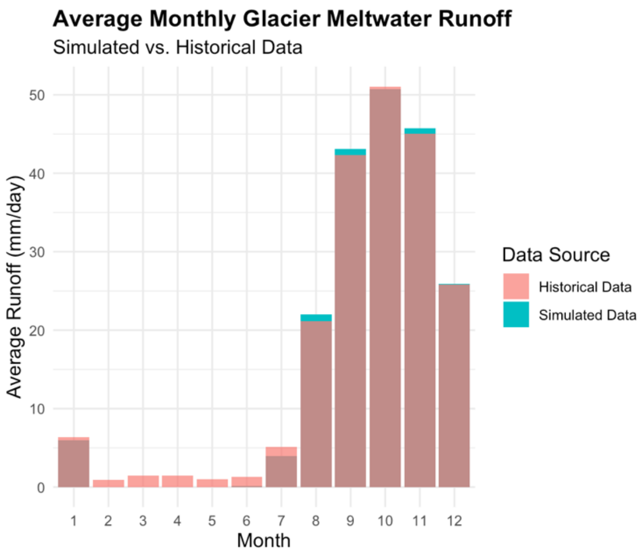


Figure 2 Seasonal Glacier Meltwater Runoff

Figure 2 illustrates the simulated daily glacier meltwater runoff for 2025 alongside the average monthly runoff[19], highlighting a clear seasonal pattern. Daily runoff volumes rise significantly in late spring, peaking notably between June and August, then sharply decreasing as temperatures fall. Correspondingly, average monthly runoff data indicate peak availability occurring in July, exceeding 45 mm/day, with runoff approaching zero from December to February. Collectively, these results emphasize the strong influence of seasonal temperature cycles on meltwater generation and underscore the importance of establishing seasonal storage infrastructure to manage supply fluctuations and ensure continuous downstream water availability [20].

3.7 Evaporation Losses along the Pipeline Route

Figure 3 illustrates the average monthly potential evapotranspiration (ET_0) values along the pipeline corridor, combined with a time series of daily ET_0 values, highlighting significant seasonal variation. Maximum monthly ET_0 exceeds 4 mm/day during the summer months, particularly in July and August, coinciding precisely with peak meltwater availability. Daily ET_0 trends further reveal intense evaporation conditions concentrated within this critical period. These overlapping peaks compound the risk of substantial water losses, indicating that water conservation measures—such as pipeline insulation and sub-surface placement—should be prioritized in the pipeline's design to mitigate evaporation impacts effectively [21].

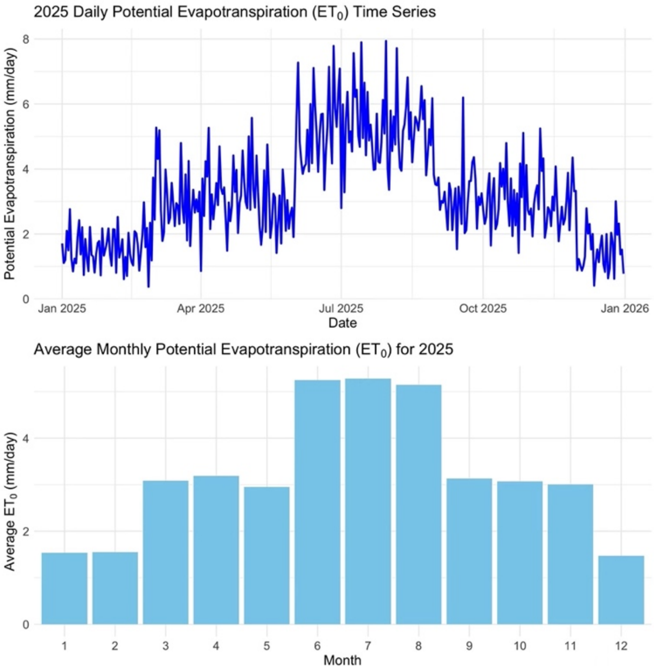


Figure 3 Seasonal Evapotranspiration Patterns

3.8 Pipeline Structural Risk and Failure Probability

Figure 4 illustrates the simulated daily pipeline risk scores, predicted daily failure probabilities, and monthly averages of both metrics. The risk scores, derived from multiple environmental stressors such as freeze–thaw cycles, corrosion, and seismic events, exhibit distinct fluctuations throughout the year, peaking during winter months due to increased temperature variability and external stress. Correspondingly, daily failure probabilities generally remain below 5% but experience notable spikes exceeding 15% during high-risk periods, particularly in January and December. The monthly averaged data further clarify the strong correlation between winter environmental stressors and elevated risk metrics. These insights underscore the necessity for targeted seasonal maintenance planning and the implementation of adaptive pipeline monitoring systems to manage heightened risks effectively [22].

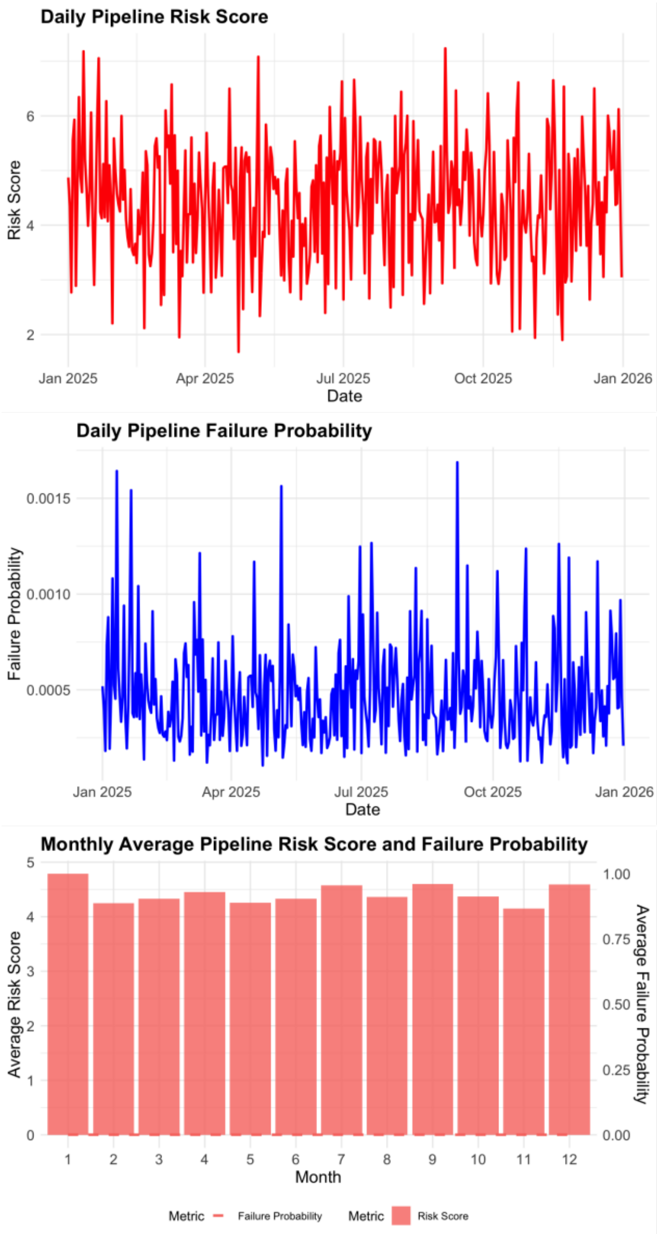


Figure 4 Seasonal Pipeline Risk and Failure Probability

3.9 Population Impact Analysis

The proposed Alaska-California pipeline system is expected to improve water security for approximately 45 million people in the western United States. (7.4 million) will gain from regional water redistribution.

California, with 39 million residents, will be the primary beneficiary, particularly in drought-prone areas like Los Angeles and the Central Valley. Nevada (3.2 million) will benefit from reduced pressure on shared water resources, while Arizona (7.4 million) will gain from regional water redistribution. Additionally, about 1 million residents in smaller communities along the route will experience improved access. This strategic infrastructure investment addresses water scarcity across multiple states, promoting climate resilience and sustainable water management.

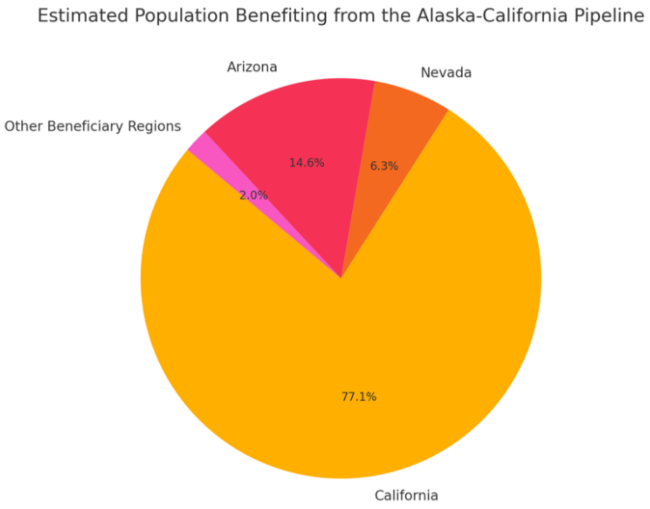


Figure 5 Population benefit from Pipeline

3.10 Engineering Implications and Climate Adaptation

The combined results demonstrate that both water availability and infrastructure vulnerability are strongly seasonal. Summer months provide abundant water but experience high evaporative loss, whereas winter poses greater mechanical and structural risk. These dual pressures call for a climate-resilient design approach that integrates seasonal flow regulation, smart monitoring systems, and terrain-sensitive routing to ensure the long-term viability of the Alaska–California water transport system [23].

3.11 Triple Bottom Line Implications of the Proposed Pipeline System

The proposed pipeline system has social, environmental, and economic implications due to the seasonal variability of glacial meltwater. Socially, increased summer meltwater flow can alleviate water scarcity in drought-prone regions, while reduced winter runoff risks shortages, requiring seasonal storage and smart allocation. Environmentally, lower winter flow may harm ecosystems, and colder months increase leakage risk, which can be mitigated through monitoring and corrosion-resistant materials. Economically, challenges include summer evaporation losses and winter maintenance costs, but stable water supply can support agricultural and industrial growth, while efficient pumping systems and an emergency fund can minimize financial risks [24].

To ensure the long-term resilience of the proposed pipeline system, several engineering strategies are recommended. These include the use of seasonal flow regulation through strategic storage reservoirs, evaporative loss reduction via insulated or subsurface pipelines in high-risk zones, and the integration of real-time structural health monitoring technologies to detect early-stage failures. Optimizing the pipeline route to avoid geologically unstable or freeze-prone areas using GIS and elevation data can further reduce risk. Finally, adopting adaptive maintenance planning—such as seasonal inspections and dynamic risk-based prioritization—will support safe and efficient operation under variable climate conditions [25].

4. Conclusion and Recommendations

The proposed Alaska–California pipeline demonstrates both technical feasibility and strategic relevance as a solution to long-term water scarcity in the southwestern United States. Through hydraulic modeling, meltwater simulation, and cost analysis, this study shows that glacial runoff can be harnessed and transported efficiently across diverse terrains with a total system cost of approximately USD 11.16 billion, incorporating solar-integrated OPEX savings, but annual emissions could be cut by 577,500 tonnes of CO_2. This clean energy integration enhances long-term sustainability and supports climate policy goals. Future recommendations include seasonal storage, smart monitoring, route optimization, and exploring regulatory and financial frameworks to implement this climate-adaptive infrastructure.

Engineering trade-offs include balancing marine versus terrestrial pipeline routing, managing seasonal variation in both water supply and risk exposure, and optimizing pump station configurations to reduce energy costs. Although the high initial capital investment and regulatory complexities are notable, the long-term benefits—climate resilience, groundwater relief, and supply diversification—justify further investigation and potential pilot implementation.

Acknowledgement

The authors would like to express sincere gratitude to Dr. Gobinath Rajarathnam for his invaluable guidance, insightful feedback, and continuous support throughout the course of this research. His expertise and encouragement have been instrumental in shaping the direction and quality of this study.

The authors acknowledge the use of AI tools to assist with R programming and modeling in this study. The AI tools were utilized primarily for code generation, optimization, and improvement. The final model design and data analysis were independently verified and adjusted by the authors to ensure scientific rigor and accuracy.

References

[1] Raj, S. & Bansal, V. K. Use of GIS for selection of optimal route for water pipelines in hill areas. *Innov. Infrastruct. Solut.* **9**, 1 (2024).

[2] Meier, M. F. & Dyurgerov, M. B. How Alaska affects the world. *Science* **297**, 350–351 (2002).

[3] Stern, C. V. *California drought: water supply and conveyance issues*. Congressional Research Service, Washington, DC (2018).

[4] Bai, Q. & Bai, Y. *Subsea Pipeline Design, Analysis, and Installation*. 1st edn, Gulf Professional Publishing (2014).

[5] Elwany, M. H. & Pluvinage, G. (eds) *Safety, Reliability and Risks Associated with Water, Oil and Gas Pipelines*. Springer, Dordrecht (2008).

[6] IPCC. Linking climate change and water resources: impacts and responses. *IPCC Tech. Pap. VI*, Ch. 3 (2008).

[7] Guo, H. L. & Chen, J. Simulation of water hammer in long distance water transmission pipeline based on Flowmaster. *J. Phys. Conf. Ser.* **2707**, 012093 (2024).

[8] Wang, Y., Xia, A., Li, R., Fu, A. & Qin, G. Probabilistic modeling of hydrogen pipeline failure utilizing limited statistical data. *Int. J. Hydrog. Energy* **95**, 1052–1066 (2024).

[9] Barateiro, C. E. R. B., Casado, M., Makarovsky, C. & de Farias Filho, J. R. Business risk and CAPEX/OPEX analysis: Impact on natural gas fiscal measurement systems. *Instrum. Meas. Metrol.* **22**, 113 (2023).

[10] Kowalczuk, Z. & Tatara, M. S. Improved model of isothermal and incompressible fluid flow in pipelines versus the Darcy–Weisbach equation and the issue of friction factor. *J. Fluid Mech.* **891**, A3 (2020).

[11] Zakikhani, K., Zayed, T., Abdrabou, B. & Senouci, A. Modeling failure of oil pipelines. *J. Perform. Constr. Facil.* **34**, 04019089 (2020).

[12] Wang, Y. *et al.* The water hammer characteristics of long-distance water pipelines under different water supply modes. *Water* **16**, 2008 (2024).

[13] Han, B. & Gabriel, J.-C. P. Thin-film nanocomposite (TFN) membrane technologies for the removal of emerging contaminants from wastewater. *J. Clean. Prod.* **480**, 144043 (2024).

[14] Barateiro, C. E. R. B., Casado, M., Makarovsky, C. & de Farias Filho, J. R. Business risk and CAPEX/OPEX analysis: Impact on natural gas fiscal measurement systems. *Instrum. Meas. Metrol.* **22**, 113 (2023).

[15] Wang, X. & Yang, W. Water quality monitoring and evaluation using remote sensing techniques in China: a systematic review. *Ecosyst. Health Sustain.* **5**, 1–18 (2019).

[16] Global Water Partnership. *China's water resources management challenge: The three red lines*. Tech. Focus Pap. (2015).

[17] Huang, Z. *et al.* Monitoring inland water quantity variations: A comprehensive analysis of multi-source satellite observation technology applications. *Remote Sens.* **15**, 3945 (2023).

[18] Li, J., Mancini, M., Su, B., Lu, J. & Menenti, M. Monitoring water resources and water use from Earth observation in the Belt and Road countries. *Bull. Chin. Acad. Sci.* (2017).

[19] Upadhyaya, A. Integrated water resources management and climate change adaptation strategies. *Adv. Tools Integr. Water Resour. Manag.* **3**, 1–20 (2017).

[20] Stern, C. V. *California Drought: Water Supply and Conveyance Issues*. Congressional Research Service, Washington, DC (2018).

[21] Kipphut, G. W. Glacial meltwater input to the Alaska Coastal Current: Evidence from oxygen isotope measurements. *J. Geophys. Res. Oceans* **95**, 5177–5181 (1990).

[22] Antonio, L. M., Pavanello, R. & Barros, P. L. A. Marine pipeline–seabed interaction modeling based on Kerr-type foundation. *Appl. Ocean Res.* **80**, 228–239 (2018).

[23] Ramamurthy, A. S. & Vo, D. A generalized analysis of flow in partially full pipes. *J. Hydraul. Eng.* **122**, 132–137 (1996).

[24] He, Q. *et al.* Subglacial meltwater recharge in the Dongkemadi River Basin, Yangtze River source region. *Ground Water* **60**, 434–450 (2022).

[25] Geck, J., Hock, R., Loso, M. G., Ostman, J. & Dial, R. Modeling the impacts of climate change on mass balance and discharge of Eklutna Glacier, Alaska, 1985–2019. *J. Glaciol.* **67**, 909–920 (2021).

Supplementary

1. Pipeline Route Selection and Geospatial Mapping

```
# Load required R packages

library(raster)

library(gdistance)

library(sf)

library(ggplot2)


# Set working directory

setwd("C:/Users/Think/Desktop/2025-S1")


# Define start and end points

start_point <- c(-132.5074, 56.7081)  # Start point (longitude, latitude)

end_point <- c(-118.2437, 36.0522)    # End point (longitude, latitude)
```



```
# Load DEM data
dem <- raster("5603/AK-CA_DEM905.tif")

# Load GIS data (used for masking or other analysis)
land <- raster("5603/GIS.tif")

# Check if coordinate reference systems match
if (crs(dem)@projargs != crs(land)@projargs) {
  # If not, reproject GIS data to match DEM CRS
  land <- projectRaster(land, crs = crs(dem))
}

# Resample GIS data to match DEM resolution
land <- resample(land, dem, method = "bilinear")

# Crop GIS data to match DEM extent
land <- crop(land, dem)

# Calculate slope
slope <- terrain(dem, opt = "slope", unit = "degrees") # Slope in degrees

# Create cost surface
# Assume cost is proportional to slope: steeper terrain = higher cost
# Use 1 / (1 + slope) to avoid division by zero
cost_surface <- 1 / (1 + slope)

# Apply GIS mask to avoid paths through ocean areas
# Assume land has value 1, ocean is NA
cost_surface <- cost_surface * land

# Define start and end points as SpatialPoints
start <- SpatialPoints(matrix(start_point, ncol = 2))
end <- SpatialPoints(matrix(end_point, ncol = 2))

# Convert cost surface to TransitionLayer
tr <- transition(cost_surface, transitionFunction = mean, directions = 8)

# Compute shortest path
path <- shortestPath(tr, start, end, output = "SpatialLines")

# Plot results
png("5603/pipeline_path.png", width = 800, height = 600)
plot(dem, main = "Pipeline Path Optimization", col = terrain.colors(100))
plot(cost_surface, add = TRUE, alpha = 0.5, col = heat.colors(100))
plot(path, add = TRUE, col = "blue", lwd = 2)
points(start, col = "red", pch = 16, cex = 1.5)
points(end, col = "green", pch = 16, cex = 1.5)
legend("topright", legend = c("Start", "End", "Pipeline Path"),
       col = c("red", "green", "blue"), pch = c(16, 16, NA), lty = c(NA, NA, 1), cex = 0.8)
dev.off()

cat("Pipeline path optimization completed and saved to C:/Users/Think/Desktop/2025-S1/5603/pipeline_path.png\n")
```


2. Hydraulic Modeling and Flow Analysis

```
library(ggplot2)


# Define constants

g <- 9.81  # Gravity acceleration (m/s²)

D <- 2.5  # Adjusted pipe diameter (m)

L <- seq(100, 3245000, 1000)  # Pipeline length: 3245 km

Q <- 7.65  # Flow rate (m³/s)

C <- 90  # Hazen-Williams roughness coefficient

f <- 0.008  # Darcy-Weisbach friction factor


# Compute flow velocity

V <- Q / (pi * (D/2)^2)

print(paste("Flow Velocity:", round(V, 2), "m/s"))


# Pumping energy constants

water_density <- 1000  # kg/m³

gravity <- 9.81  # m/s²


darcy_weisbach <- function(Q, D, L, f) {

  h_f <- (f * L * V^2) / (D * 2 * g)

  return(h_f)

}


hazen_williams <- function(Q, D, L, C) {

  h_f <- 10.67 * (L / (C^1.85 * D^4.87)) * (Q^1.85)

  return(h_f)

}

# Compute pressure losses

pressure_loss_darcy <- sapply(L, function(L) darcy_weisbach(Q, D, L, f))

pressure_loss_hazen <- sapply(L, function(L) hazen_williams(Q, D, L, C))


# Compute total pumping energy

pumping_energy <- (water_density * gravity * Q * pressure_loss_darcy) / 1000


# Data frame

data <- data.frame(Length_m = L,

                   Pressure_Loss_DW = pressure_loss_darcy,

                   Pressure_Loss_HW = pressure_loss_hazen,

                   Pumping_Energy_kW = pumping_energy)


# Plot: Pressure Loss

plot <- ggplot(data, aes(x = Length_m)) +

  geom_line(aes(y = Pressure_Loss_DW, color = "Darcy-Weisbach"), linewidth = 1) +

  geom_line(aes(y = Pressure_Loss_HW, color = "Hazen-Williams"), linewidth = 1, linetype = "dashed") +

  labs(title = "Pressure Loss vs Pipeline Length",

       x = "Pipeline Length (m)",

       y = "Pressure Loss (m)") +

  scale_color_manual(values = c("Darcy-Weisbach" = "blue", "Hazen-Williams" = "red")) +

  theme_minimal()

print(plot)
```



```
# Final result summary

print(paste("Final Pressure Loss (Darcy-Weisbach):", round(tail(pressure_loss_darcy, 1), 2), "m"))

print(paste("Final Pressure Loss (Hazen-Williams):", round(tail(pressure_loss_hazen, 1), 2), "m"))

print(paste("Total Pumping Energy Required:", round(tail(pumping_energy, 1), 2), "kW"))
```

3. Cost and Energy Optimization

```
library(lpSolve)


# ---- Pump parameters (unchanged) ----

pumps <- data.frame(
  type = c("Low", "Mid", "High"),
  head = c(25, 30, 35),
  capex = c(300000, 350000, 400000),
  power = c(3660, 4200, 5000)  # kW
)


electricity_price <- 0.10

hours_per_year <- 24 * 365

total_required_head <- 1492.8

total_pump_count <- 60


opex_30yr <- pumps$power * hours_per_year * 30 * electricity_price

total_pump_cost <- pumps$capex + opex_30yr

# ---- LP for pump selection (fixed pump count = 60) ----

pump_constraints <- rbind(
  pumps$head,
  rep(1, nrow(pumps))
)

pump_dirs <- c(">=", "==")

pump_rhs <- c(total_required_head, total_pump_count)


pump_result <- lp("min", total_pump_cost, pump_constraints, pump_dirs, pump_rhs, all.int = TRUE)


if (pump_result$status != 0) stop("Pump optimization failed.")


pump_solution <- pump_result$solution

pump_capex <- sum(pump_solution * pumps$capex)

pump_opex <- sum(pump_solution * opex_30yr)

pump_power <- sum(pump_solution * pumps$power)


# ---- Land pipeline construction options ----

# Methods: A (no limit), B (≤1000 km), C (≤800 km)

land_total_km <- 2645

cost_per_m <- c(1800, 1600, 1400)  # USD/m

cost_per_km <- cost_per_m * 1000


pipe_constraints <- rbind(
  c(1, 1, 1),    # total km = 2645
  c(0, 1, 0),    # B ≤ 1000
  c(0, 0, 1)     # C ≤ 800
```

```
)
pipe_dirs <- c("=", "<=", "<=")
pipe_rhs <- c(2645, 1000, 800)

pipe_result <- lp("min", cost_per_km, pipe_constraints, pipe_dirs, pipe_rhs)

if (pipe_result$status != 0) stop("Pipeline optimization failed.")

pipe_solution <- pipe_result$solution
names(pipe_solution) <- c("km_A", "km_B", "km_C")
land_pipeline_cost <- sum(pipe_solution * cost_per_km)

# ---- Final system costs ----
marine_cost <- 600 * 1000 * 2800
civil_cost <- 12 * 5e6
solar_cost <- 110e6

total_system_cost <- marine_cost + land_pipeline_cost + pump_capex + pump_opex + civil_cost + solar_cost

# ---- Output ----
cat("     Optimized Pump Configuration:\n")
print(setNames(pump_solution, pumps$type))
cat("CAPEX (pumps): USD", format(pump_capex, big.mark=","), "\n")
cat("OPEX (30 yrs): USD", format(round(pump_opex), big.mark=","), "\n")
cat("Total installed pump power:", round(pump_power), "kW\n\n")
cat("     Optimized Land Pipeline Configuration:\n")
print(pipe_solution)
cat("Land pipeline construction cost: USD", format(round(land_pipeline_cost), big.mark=","), "\n\n")

cat("    Final Total System Cost (pipeline + pumps + civil + solar): USD",
    format(round(total_system_cost), big.mark=","), "\n")
```

4. Climate Impact and Failure Risk and 5. Data Visualization and Report Preparation

```
# Load required packages
library(tidyverse)   # For data manipulation and plotting
library(lubridate)   # For handling dates
# Create a sequence of dates for the entire year 2025
dates_glacier <- seq.Date(from = as.Date("2025-01-01"), to = as.Date("2025-12-31"), by = "day")
n_glacier <- length(dates_glacier)
day_of_year <- yday(dates_glacier)

# Simulate daily temperature data using a sine function to mimic seasonal variation
# (simulate a glacier region: low temperatures in winter, high temperatures in summer)
T_mean <- 0        # Mean temperature (°C)
amplitude <- 10    # Temperature amplitude (°C)
phase_shift <- 200 # Phase shift (positions the peak temperature mid-year)
set.seed(123)
```



```
temperature <- T_mean +
  amplitude * sin(2 * pi * (day_of_year - phase_shift) / 365) +
  rnorm(n_glacier, mean = 0, sd = 2)

# Calculate daily glacier meltwater runoff using the degree-day method:
# if temperature > 0°C, then meltwater runoff = degree_day_factor * temperature, otherwise 0
degree_day_factor <- 5  # Unit: mm/(°C·day)
meltwater_runoff <- ifelse(temperature > 0, degree_day_factor * temperature, 0)

# Construct a data frame to store the simulated data
sim_data <- tibble(
  date = dates_glacier,
  day_of_year = day_of_year,
  temperature = temperature,
  meltwater_runoff = meltwater_runoff
)

# Simulate historical runoff data by adding noise to the simulated data
historical_runoff <- meltwater_runoff + rnorm(n_glacier, mean = 0, sd = 3)
historical_runoff <- pmax(historical_runoff, 0)  # Ensure values are not negative
sim_data <- sim_data %>%
  mutate(historical_runoff = historical_runoff)

# (a) Plot the daily glacier meltwater runoff time series
evap_plot <- ggplot(sim_data, aes(x = date)) +
  geom_line(aes(y = meltwater_runoff, color = "Simulated Meltwater"), size = 1) +
  geom_line(aes(y = historical_runoff, color = "Historical Meltwater"),
            linetype = "dashed", size = 1) +
  labs(title = "2025 Glacier Meltwater Runoff Simulation",
       subtitle = "Daily Meltwater Runoff (mm/day)",
       x = "Date",
       y = "Meltwater Runoff (mm/day)",
       color = "Data Source") +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 16, face = "bold"),
    plot.subtitle = element_text(size = 12),
    axis.title = element_text(size = 14),
    axis.text = element_text(size = 12),
    legend.title = element_text(size = 14),
    legend.text = element_text(size = 12)
  )
print(evap_plot)

# (b) Calculate the monthly average glacier meltwater runoff to analyze seasonal variation
monthly_meltwater_summary <- sim_data %>%
  mutate(month = month(date)) %>%
  group_by(month) %>%
  summarise(
    avg_simulated = mean(meltwater_runoff, na.rm = TRUE),
    avg_historical = mean(historical_runoff, na.rm = TRUE)
```

```
  )


# Plot a bar chart comparing monthly average glacier meltwater runoff
monthly_plot <- ggplot(monthly_meltwater_summary, aes(x = factor(month))) +
  geom_bar(aes(y = avg_simulated, fill = "Simulated Data"),
           stat = "identity", position = "dodge") +
  geom_bar(aes(y = avg_historical, fill = "Historical Data"),
           stat = "identity", position = "dodge", alpha = 0.7) +
  labs(title = "Average Monthly Glacier Meltwater Runoff",
       subtitle = "Comparison of Simulated and Historical Data",
       x = "Month",
       y = "Average Runoff (mm/day)",
       fill = "Data Source") +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 16, face = "bold"),
    plot.subtitle = element_text(size = 12),
    axis.title = element_text(size = 14),
    axis.text = element_text(size = 12),
    legend.title = element_text(size = 14),
    legend.text = element_text(size = 12)
  )
print(monthly_plot)

# Define the Penman-Monteith function to calculate potential evapotranspiration (ET0)
penman_monteith <- function(T, RH, u2, Rn) {
  # Calculate saturation vapor pressure (kPa)
  es <- 0.6108 * exp((17.27 * T) / (T + 237.3))
  # Calculate actual vapor pressure (kPa)
  ea <- RH / 100 * es
  # Calculate the slope of the saturation vapor pressure curve (kPa/°C)
  Delta <- 4098 * es / ((T + 237.3)^2)
  # Set the psychrometric constant (kPa/°C)
  gamma <- 0.066
  # Calculate ET0 (mm/day), assuming soil heat flux G = 0
  ET0 <- (0.408 * Delta * Rn + gamma * (900 / (T + 273)) * u2 * (es - ea)) /
    (Delta + gamma * (1 + 0.34 * u2))
  return(ET0)
}


# Simulate weather data for 2025
set.seed(123)
dates_et0 <- seq.Date(from = as.Date("2025-01-01"), to = as.Date("2025-12-31"), by = "day")
weather_data <- tibble(
  date = dates_et0,
  month = month(date),
  # Simulate temperature (°C) based on month: lower in winter, higher in summer
  T = case_when(
    month %in% c(12, 1, 2) ~ rnorm(length(dates_et0), mean = 0, sd = 5),
```



```
    month %in% c(3, 4, 5) ~ rnorm(length(dates_et0),
mean = 10, sd = 5),
    month %in% c(6, 7, 8) ~ rnorm(length(dates_et0),
mean = 20, sd = 5),
    month %in% c(9, 10, 11) ~ rnorm(length(dates_et0),
mean = 10, sd = 5)
  ),
  # Relative Humidity (%)
  RH = runif(length(dates_et0), min = 40, max = 100),
  # Wind speed at 2 meters (m/s)
  u2 = runif(length(dates_et0), min = 1, max = 5),
  # Net radiation (MJ/m^2/day): lower in winter, higher
in summer
  Rn = case_when(
    month %in% c(12, 1, 2) ~ runif(length(dates_et0),
min = 5, max = 10),
    month %in% c(3, 4, 5) ~ runif(length(dates_et0), min
= 10, max = 15),
    month %in% c(6, 7, 8) ~ runif(length(dates_et0), min
= 15, max = 20),
    month %in% c(9, 10, 11) ~ runif(length(dates_et0),
min = 10, max = 15)
  )
)


# Calculate daily ET0 using the Penman-Monteith
function
weather_data <- weather_data %>%
  mutate(ET0 = penman_monteith(T, RH, u2, Rn))
print(head(weather_data))


# Plot the daily ET0 time series
evap_plot_et0 <- ggplot(weather_data, aes(x = date, y
= ET0)) +
  geom_line(color = "blue", size = 1) +
  labs(title = "2025 Daily Potential Evapotranspiration
(ET0) Time Series",
       x = "Date",
       y = "Potential Evapotranspiration (mm/day)") +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 16, face = "bold"),
    axis.title = element_text(size = 14),
    axis.text = element_text(size = 12)
  )
print(evap_plot_et0)


# Summarize average ET0 by month (for seasonal
analysis)
monthly_et0_summary <- weather_data %>%
  group_by(month) %>%
  summarise(mean_ET0 = mean(ET0, na.rm = TRUE))


# Plot the monthly average ET0 as a bar chart
monthly_plot_et0 <- ggplot(monthly_et0_summary,
aes(x = factor(month), y = mean_ET0)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(title = "Average Monthly Potential
Evapotranspiration (ET0) for 2025",
       x = "Month",
       y = "Average ET0 (mm/day)") +
  theme_minimal() +
  theme(
```



```
  plot.title = element_text(size = 16, face = "bold"),
  axis.title = element_text(size = 14),
  axis.text = element_text(size = 12)
 )
print(monthly_plot_et0)



set.seed(123)
dates_risk <- seq.Date(from = as.Date("2025-01-01"), to = as.Date("2025-12-31"), by = "day")
n_risk <- length(dates_risk)

# Simulate risk factors:
# Temperature variability (°C), Corrosion index (0–10), Freeze-thaw cycles (more frequent in winter),
# Extreme rainfall (using a Bernoulli distribution), External stress events (low probability)
temp_variability <- runif(n_risk, min = 5, max = 15)
corrosion_index <- runif(n_risk, min = 0, max = 10)
freeze_thaw <- ifelse(month(dates_risk) %in% c(12, 1, 2),
                  sample(0:3, n_risk, replace = TRUE),
                  sample(0:1, n_risk, replace = TRUE))
rain_prob <- ifelse(month(dates_risk) %in% c(3, 4, 5, 9, 10, 11), 0.2, 0.1)
extreme_rain <- rbinom(n_risk, size = 1, prob = rain_prob)
ext_stress <- rbinom(n_risk, size = 1, prob = 0.05)

# Calculate the comprehensive risk score by assigning different weights to each risk factor
risk_score <- 0.3 * temp_variability +
  0.25 * corrosion_index +
  0.2 * freeze_thaw +
  0.15 * extreme_rain +
  0.1 * ext_stress

# Predict pipeline failure probability using a logistic transformation
predicted_failure_prob <- 1 / (1 + exp(-(-10 + 0.5 * risk_score)))

# Combine the risk data into a tibble
risk_data <- tibble(
  date = dates_risk,
  temp_variability = temp_variability,
  corrosion_index = corrosion_index,
  freeze_thaw = freeze_thaw,
  extreme_rain = extreme_rain,
  ext_stress = ext_stress,
  risk_score = risk_score,
  predicted_failure_prob = predicted_failure_prob
)
print(head(risk_data))

# (a) Plot the daily pipeline risk score time series
risk_plot <- ggplot(risk_data, aes(x = date, y = risk_score)) +
  geom_line(color = "red", size = 1) +
  labs(title = "Daily Pipeline Risk Score",
     x = "Date",
     y = "Risk Score") +
  theme_minimal() +
```



```
  theme(
    plot.title = element_text(size = 16, face = "bold"),
    axis.title = element_text(size = 14),
    axis.text = element_text(size = 12)
  )
print(risk_plot)


# (b) Plot the daily pipeline failure probability time series
failure_prob_plot <- ggplot(risk_data, aes(x = date, y = predicted_failure_prob)) +
  geom_line(color = "blue", size = 1) +
  labs(title = "Daily Pipeline Failure Probability",
       x = "Date",
       y = "Failure Probability") +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 16, face = "bold"),
    axis.title = element_text(size = 14),
    axis.text = element_text(size = 12)
  )
print(failure_prob_plot)


# (c) Summarize risk score and failure probability by month
monthly_risk_summary <- risk_data %>%
  mutate(month = month(date)) %>%
  group_by(month) %>%
  summarise(
    avg_risk = mean(risk_score, na.rm = TRUE),
    avg_failure_prob = mean(predicted_failure_prob, na.rm = TRUE)
  )


# Plot the monthly average pipeline risk score and failure probability (bar chart + line plot)
monthly_plot_risk <- ggplot(monthly_risk_summary, aes(x = factor(month))) +
  geom_bar(aes(y = avg_risk, fill = "Risk Score"), stat = "identity", position = "dodge", alpha = 0.8) +
  geom_line(aes(y = avg_failure_prob * max(avg_risk), group = 1, color = "Failure Probability"), size = 1, linetype = "dashed") +
  scale_y_continuous(
    name = "Average Risk Score",
    sec.axis = sec_axis(~ . / max(monthly_risk_summary$avg_risk), name = "Average Failure Probability")
  ) +
  labs(title = "Monthly Average Pipeline Risk Score and Failure Probability",
       x = "Month",
       fill = "Metric",
       color = "Metric") +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 16, face = "bold"),
    axis.title = element_text(size = 14),
    axis.text = element_text(size = 12),
    legend.position = "bottom"
  )
print(monthly_plot_risk)
```