THESIS

QUORUM: A SYSTEMS ENGINEERING FRAMEWORK FOR MIGRATION RISK ASSESSMENT

Submitted by

Stefan Haugen

Department of Systems Engineering

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer, 2026

Master's Committee:

Advisor: Aaron Brown

Rebecca Atadero
Steve Conrad
Erika Gallegos

ABSTRACT

QUORUM: A SYSTEMS ENGINEERING FRAMEWORK FOR MIGRATION RISK ASSESSMENT

This thesis presents QUORUM (Quantitative Understanding of Regional Outmigration Using Multivariate data), a modular, systems-engineered framework designed to assess development indicators with migration trends in a case example location of Central America. The work addresses a documented gap in research literature that no existing Water-Energy-Food (WEF) nexus model incorporates migration as a thematic focus, leaving policymakers and affected communities without the analytical tools needed to anticipate displacement pressures and plan adaptive responses.

This system is architected as three constituent modules. The Information Module handles heterogeneous data ingestion and harmonization. The Analytics Module performs statistical and machine learning analysis. The Design Module supports human-centered visualization and stakeholder communications. The system is developed using the V-Model of systems engineering, with functional requirements decomposed across each of the modules and governed by formal Interface Control Documents specifying data schemas, transformation rules, and acceptance criteria at each module boundary.

The Analytics Module was validated using a proof-of-concept implementation that integrates World Bank World Development Indicators with Meta mobility-derived migration data for seven Central American nations over a 2019 to 2022 study period. A four-stage analytical pipeline was applied containing descriptive migration characterization, lagged cross-correlation analysis, fixed-effects panel regression, and Random Forest modeling with Shapley Additive Explanations (SHAP) for feature importance attribution.

As a proof-of-concept demonstration, the analytics pipeline produced outputs broadly aligned with theoretical expectations, suggesting that publicly available environmental and socioeconomic indicators carry information relevant to migration pressures. Nine of 45 tested correlations reached statistical significance ($p < .05$), and SHAP analysis identified that metrics around a nation's carbon intensity and freshwater availability as the highest-ranked indicators within the exploratory analysis. Fixed-effects panel regression identified *Forest Cover (%)* as the most robust within-country migration predictor (beta = -11.53, $p = .029$).

The thesis contributes a reproducible, transparent, and grounded analytical architecture that explores how open-source data, when processed through a structured systems engineering pipeline, can surface preliminary quantitative relationships relevant to migration risk indicators with existing and available datasets. Limitations related to panel size (N = 14 country-year observations) and cross-country generalization are discussed, along with future directions for expanding temporal coverage, integrating higher-resolution monthly environmental data, and deploying bespoke Design Module elements for community-facing applications.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1
INTRODUCTION

1.1 Background and Rationale

1.1.1 The Planetary Context

Climate instability, resource insecurity, and human mobility are increasingly interlinked. As of 2023, Earth has exceeded six of nine recognized planetary boundaries (Richardson et al., 2023), signaling fundamental shifts in Earth system dynamics with direct implications for human welfare and security. Intensifying patterns of drought, crop failure, and water scarcity are now becoming consistent across the world's most fragile regions, with seasonal anomalies eroding agricultural livelihoods and forcing communities into cycles of displacement and vulnerability.

As the Intergovernmental Panel on Climate Change (IPCC, 2023) has documented, human migration is becoming a core indicator of systemic stress within coupled human-environment systems. Over 20 million people are displaced annually by weather-related disasters (IDMC, 2023). Migration, once primarily an adaptive strategy, has increasingly become a marker of systemic stress, underscoring the growing need for analytical tools that can interpret and anticipate these phenomena. An analytical framework that integrates development data with human movement patterns to anticipate and mitigate future crises is a critical tool that is desired by many investigative groups (García & Valencia, 2024; Huber et al., 2023). The challenge is compounded when considering that climate stress, migration, and resource security form not a linear chain of causation, but a feedback loop of interlocking and reactionary system shocks. Daher et al. (2021) identify migration as both a cause and a consequence of WEF system disruptions. The lack of water and food security forces communities to relocate, while the arrival of displaced populations places new and sudden demands on the water, energy, and food systems of the receiving areas. Daher et al.’s (2021) review cites the drought in Mexico as a historical driver of northward migration to the United States, while also documenting how refugee flows into Lebanon and Jordan have created water and food scarcity in those receiving states. The mismanagement of water and food systems in Central America has been identified as a leading driver of migration as agricultural livelihoods deteriorate. This finding has direct relevance to the Central American Dry Corridor that serves as the system’s empirical focus. Daher et al. (2021) conclude that multi-sectoral decision makers currently lack the tools to quantify how anticipated migration flows will stress these interdependent resource systems, and that no existing WEF model or framework incorporates migration as a thematic focus. This framework is designed thoughtfully to add useful insight into that void of understanding.

The mentioned fragmentation in analytical tools is mirrored by a fragmentation in operational response. Despite growing evidence that climate disasters and armed conflict compound displacement, humanitarian organizations continue to address displacement crises in isolation, with data reported separately based on climate-related and conflict-related triggers (IDMC & NRC, 2015; Peters et al., 2021). The Internal Displacement Monitoring Centre reported that 43 out of 45 countries and territories experiencing conflict displacement also recorded disaster displacement a year prior (IDMC, 2024), underscoring that these crises are rarely independent events. See et al. (2026) document how this operational disconnect creates blind spots in humanitarian response, with organizations classifying displacement by a single preceding “trigger” event, even when affected communities experience repeated, overlapping displacements from both conflict and climate hazards. The concept of the Humanitarian-Development-Peace (HDP) Nexus, or “triple nexus,” was developed in response to these converging challenges, calling for coherence between humanitarian, development, and peacebuilding efforts (Joireman & Haddad, 2023; OECD, 2024). The systems design is aligned with the premise of the triple nexus by fusing environmental, socioeconomic, and migration data within a single analytical pipeline. The framework provides the type of cross-domain situational awareness that siloed reporting systems poorly cover.

1.1.2 The Central American Dry Corridor: Empirical Focus

The Central American Dry Corridor is a tropical dry forest ecoregion extending from southern Mexico through Guatemala, El Salvador, Honduras, and Nicaragua. It serves as the regional focus for such a system as the region exemplifies the complex interplay between climate variability, agricultural vulnerability, and migration pressure that the framework is designed to address.

The Dry Corridor is characterized by pronounced wet and dry seasons, with rainfall concentrated in a bimodal pattern that agricultural communities have historically relied upon for subsistence farming. However, recent CLIMDEX analyses reveal an intensification of dry-season anomalies since 2000, causing displacement of previously sustainable livelihoods (Nakamura et al., 2024). This climatic variability, marked by both prolonged droughts and intense monsoons, has detrimentally affected rural populations and amplified existing socioeconomic challenges including poverty and inequality (Mejia et al., 2022).

Over the last decade, the Central American Dry Corridor has experienced severe droughts impacting over 1.3 million subsistence farmers and laborers, predominantly indigenous populations, leading to sizeable livelihood losses, acute food insecurity, and shifts in migratory behaviors (Nakamura et al., 2024). The region therefore provides both urgent motivation and a bounded empirical context for developing and validating the proposed analytical framework. Migration originating in these areas flows gravitationally toward the Mexico and U.S. border, the largest border crossing on Earth. A World Food Programme (WFP) publication of northern Central American migrants rejected at Mexican immigration with intentions of reaching the U.S. found that around half of that population had been working in the agricultural sector (Tower, 2021).

1.2 Key Definitions and Scope

This thesis adopts the following operational definitions to ensure clarity and consistency:

Migration flows "refer to the number of migrants entering or leaving a given country during a given period of time, usually one calendar year".

Climate-related migration refers to the movement of persons or groups who, predominantly for reasons of sudden or progressive environmental change due to climate variability, are obliged to leave their habitual place of residence, either temporarily or permanently, within or across an international border. This thesis distinguishes climate-related migration from conflict-driven displacement, though it acknowledges the frequent co-occurrence of these drivers via the World Bank Indicators.

Water-Energy-Food nexus denotes the interdependent systems of water, energy, and food resources in which changes to one sector produce cascading effects on the others. In the context of this thesis, the WEF nexus provides the conceptual architecture for understanding how environmental stress translates into livelihood disruption and, ultimately, migration pressure. (https://www.water-energy-food.org)

System of Systems (SoS) describes an arrangement of independently operable constituent systems that, when integrated, produce emergent capabilities exceeding those achievable by any single component. This system operates as an SoS in which the Information, Analytic, and Design Modules function autonomously with distinct outputs while contributing to the integrated risk assessment capability of the whole.

Explainability refers to the capacity of an analytical system to provide human-interpretable accounts of its outputs. This includes the identification of causal drivers, confidence levels, and the reasoning chain connecting input data to risk assessments. Explainability is operationalized through SHAP-based feature attribution, uncertainty quantification, and multi-level visualization designed for stakeholders with varying technical backgrounds.

Open-source data refers to datasets that are publicly accessible without proprietary licensing restrictions. In this thesis, the term encompasses World Bank World Development Indicators (WDI / **development data**), standardized evapotranspiration indices (SPEI) drought metrics, WFP food price monitors, and Meta's Data For Good migration datasets. The distinction between open-source and proprietary data is central to the thesis objective of building accessible analytical tools.

Complex displacement refers to situations in which communities are forced to flee amid the complex interaction of climate change, disasters, and conflict (See et al., 2026). The term captures the non-linear and compounding nature of displacement drivers, where climatic events, armed conflict, and underlying structural vulnerabilities are deeply entangled rather than operating as independent causal pathways. This framing adds to recent work that theorizes complexity as the cascade of risks across sectors and scales within coupled social-ecological systems (Simpson et al., 2021). The system's analytical architecture is designed to detect precisely these compounding interactions through its multi-source data ingestion and time-lagged cross-correlation approach.

Humanitarian-Development-Peace Nexus, denoted above as the "**triple nexus**," represents a policy and operational framework developed in response to contexts where humanitarian needs, climate change, and conflict coincide (Joireman & Haddad, 2023). The HDP Nexus emphasizes greater coherence between humanitarian, development, and peacebuilding efforts to address immediate needs, understand root causes, and support sustainable outcomes for displaced populations (OECD, 2024). Together with the durable solutions framework for internally displaced persons (IASC, 2010), the triple nexus reflects a recognition within the international community that responses to displacement must extend beyond short-term humanitarian assistance to encompass sustainable development perspectives and peacebuilding measures (Nguya & Siddiqui, 2020). The concept is relevant to the system because the framework's multi-sector analytical approach mirrors the integrated, cross-domain thinking that the HDP Nexus advocates for at the operational level.

1.3 Problem Statement and Research Gap

The capacity to anticipate and manage migration pressures remains deeply uneven across the Global South. While satellite networks, Internet of Things (IoT) sensor networks, proprietary and open-source datasets now generate extensive information relevant to climate vulnerability, those most affected by impending climate risks are rarely positioned to interpret the complexities hidden in data, or act upon the illuminated circumstances. Information infrastructures, ownership of data collection devices, data access permissions, and most critically the ability and means to translate information into community-based action have emerged as new dimensions of inequality.

Major advances in data processing, storage, and computational power have yielded substantial technologies accessible to anyone with a modern internet connection. This information age has also demonstrated that data requires meticulous context framing and often mathematical transformation and visualization before genuine human comprehension takes place.

The proliferation of affordable IoT devices, inexpensive processors, and solar-powered sensors offers opportunities for enhancing spatiotemporal data resolutions in any area. These elements should foster a capacity for participatory engagement among affected communities. However, realizing these benefits in climate-vulnerable regions like the Central American Dry Corridor demands overcoming critical barriers:

Genuine local engagement to ensure cultural relevance and community ownership of analytical outputs. Education resources miss key understanding gaps. Lack of consistency in engagement remains a barrier for traditional relief organizations.

Technical integration with legacy infrastructures and heterogeneous data sources across institutional boundaries. Access to portals with 'in-hand' interaction remains elusive, permitting personal migration agency only through a series of barriers.

Model transparency and accessibility for non-expert stakeholders who must act on system outputs. Culturally relevant communication, accessible colorways and informational organization lags behind the most modern interfaces of technical norms.

Ethical governance to mitigate data inequities and power imbalances inherent in external data collection. A watchful eye on data of vulnerable areas seems relegated to state intelligence agencies, mineral extraction or energy rather than serving the communities affected by these data.

Addressing these overlapping challenges necessitates a systems-based approach that prioritizes human-systems integration principles and processes complex datasets to serve end-users across diverse linguistic, skill, and educational backgrounds (Lynch et al., 2023). Vulnerable populations are typically left with after-the-fact relief frameworks rather than proactive resilience strategies. This disparity often stems from insufficient reliable data-driven insights on both physical and social processes in these regions, making it challenging to develop effective interventions and systematically underestimating the true impacts and costs of climate change (Johnson et al., 2024).

The absence of openly accessible and trustworthy analysis tools undermines both humanitarian response and community adaptation. The central challenge of this work is to design a transparent, equitable, and analytically robust system that transforms available yet often constrained datasets into actionable insights that are accessible and affordable to those who need them most.

1.4 Thesis Objectives and Central Hypothesis

This thesis advances a systems-engineering framework for a modular, explainable, and human-centered data-pipeline system called QUORUM (Quantitative Understanding Of Regional Outmigration Using Multivariate data) which integrates open environmental, socioeconomic and migration data to explain, and where statistically possible and appropriate, contribute to climate-related migration risk indicators.

Central Hypothesis: Indicators derived from defined open-source datasets can statistically validate and qualify emergent migration behaviors observed by a Meta mobility dataset in the Central American region.

An important framing distinction is necessary. QUORUM is a risk understanding framework, not a predictive model. The system is designed to identify and rank environmental and socioeconomic indicators that correlate with observed migration patterns, producing exploratory evidence of displacement-relevant relationships rather than operational forecasts. The proof-of-concept implementation presented in this thesis validates the architecture through correlation analysis, fixed-effects regression, and SHAP-based feature attribution, all of which characterize historical associations rather than generate forward-looking predictions. Operationalizing QUORUM as a forecasting tool would require substantially larger panel dimensions, prospective validation on out-of-sample time periods, and the stakeholder calibration embedded in the Design Module's functional requirements.

Theoretical basis: The theoretical foundation for this hypothesis rests on labor migration literature demonstrating that food price volatility directly affects household economic calculus, particularly for subsistence agricultural communities (Cai & Oppenheimer, 2013). When agricultural yields decline due to water stress, households face a decision threshold to either absorb the shock through savings depletion, seek alternative local employment, or migrate. The temporal lag between environmental stressor and migration response, typically 3-12 months in agricultural communities, reflects the system's sequential decision process and provides the basis for QUORUM's temporal architecture.

The proposed framework, which unites data science, humanitarian engineering, and human-centric design within a systems-engineering framework, aims to:

1. Ingest open-access environmental and socioeconomic data (precipitation anomalies, agricultural yield indices, energy reliability metrics, population dynamics) from APIs, scheduled releases, or database update triggers.

2. Compute analyses in operational sequence with conditional logic to interpret outputs and propagate uncertainty estimates.

3. Translate results into a human-centric, accessible graphical user interface with multi-level explainability.

4. Validate outputs and stakeholder requirements through structured information package verification protocols.

QUORUM builds on a history of decision support systems that link sustainability analysis with engineering design and stakeholder decision-making. Conrad et al. (2011) developed such a system for sustainable energy management that integrates technical modeling with stakeholder communication to inform resource planning. QUORUM applies similar foundational approaches to the migration and development domain, advancing the same principle that analytical outputs must be structured for the decision contexts in which they will be used.

1.5 Research Questions

This work addresses the following guiding questions:

RQ1: How can systems engineering principles be applied to design a transparent and adaptive risk assessment architecture for migration?

RQ2: Can widely accessible open-source datasets provide migration correlation with risk indicators with statistical power?

RQ3: Which data indicators illuminated migratory patterns in the Central American Dry Corridor for this case example?

1.6 Reason for a Systems-of-Systems Approach

1.6.1 Learning from the Global North: Hurricane Katrina

A historical event in the Global North illustrates the nonlinear interplay between human and environmental systems that motivates this project. In 2005, Hurricane Katrina inflicted catastrophic damage on New Orleans, Louisiana, (a city whose notorious geographic vulnerability had long been recognized) precipitating momentous population displacement and exposing profound socioeconomic inequalities in disaster response (Tarraga et al., 2020). Subsequent modeling revealed that the crisis exhibited the nonlinear escalation patterns characteristic of complex adaptive systems, where initial disruptions cascaded through interconnected social and physical infrastructures (Curtis et al., 2007).

While Katrina occurred in the Global North with relatively robust institutional response capacity, the underlying dynamics including nonlinear crisis escalation and feedback loops between environmental shock and social disruptions were misunderstood and mismanaged. These dynamics manifest with even greater intensity in under-resourced contexts in the Global South. The Central American Dry Corridor like many other climate vulnerable areas, shares Katrina's pattern of compounding stressors. Institutional infrastructure, early warning systems, and recovery resources are largely unavailable in the countries of CADC. This asymmetry stresses both the transferability of systems-thinking approaches across contexts and the urgency of developing accessible analytical tools for vulnerable regions where capacity may be limited.

1.6.2 V-Model system design

Non-linear human behaviors and natural-disaster dynamics create feedback loops among environmental shocks, political and economic pressures, conflict, uncertainty, and social networks. This complexity demands a structured approach to manage systemic interdependencies. The V-Model of systems engineering (Kossiakoff et al., 2020) was selected over alternative frameworks (spiral model, agile SE, waterfall) for three reasons:

First, the V-Model's explicit verification-validation pairing ensures that each requirement maps to a testable outcome. This is critical for a system that must earn stakeholder trust across diverse user communities with varying technical literacy and promotes integrity throughout the system.

Second, its structured decomposition from concept through subsystem design aligns with QUORUM's modular architecture, where independent modules (data ingestion, analytics, engagement) must integrate into a coherent whole.

Third, the V-Model's emphasis on requirements traceability supports the governance and accountability principles central to this thesis. Every design decision should be traceable to a stakeholder need. Every validation test can be traced to a requirement.

The left side of the V (decomposition) defines what QUORUM must do. The right side (integration and verification) verifies that it does so as designed. This structure ensures that conceptual intentions, technical performance, and human usability remain tightly linked throughout the system lifecycle.

The V-Model's structured verification and validation framework also allows a natural mode for identifying and analyzing systemic failure modes throughout the development lifecycle. Perreault and Gallegos (2025) demonstrate that the application of failure modes and effects analysis within V-Model based complex design can expose gaps between intended requirements and realized system behavior, a concern directly relevant to QUORUM's goal of producing trustworthy outputs for affected populations.

1.6.3 QUORUM as a System of Systems

At its core, QUORUM functions as a System of Systems (SoS), integrating three constituent systems:

Information Module: Climate sensing data, socioeconomic indicators, and humanitarian datasets ingested from different source locations with varying update frequencies, formats, and quality assurance levels.

Analytics Module: Machine learning models, statistical inference engines, and uncertainty quantification routines that transform raw data into insights and risk assessments with calibrated confidence intervals.

Design Module: Human-integration interfaces including visual dashboards, multilingual communication outputs, and educational interaction capabilities designed for users across technical literacy levels.

This design emphasizes traceability by linking every requirement to a verifiable outcome. This adaptability allows the framework to evolve with dataset availability and computational capability for any regional focus. It can run remotely, offline, and with minimal compute. This design ensures that the resiliency provided by this system is preserved as data sourcing and availability policies progress beyond recognition to historical data sources.

1.7 Concept of Operations (ConOps)

To ensure the QUORUM framework effectively translates complex environmental and socioeconomic data into actionable insights, the system's operational design is grounded in a task analysis of its intended end users. The system is architected to serve three primary stakeholder groups, each with distinct technical literacies and operational workflows. This Concept of Operations defines the expected interaction patterns that the Design Module must support.

1.7.1 Community Leaders and Local Decision-Makers

For users at the community level, the primary operational workflow is a risk inquiry sequence. A community leader navigating to their region is presented with a summary risk indicator featuring a plain-language severity description. Rather than raw numerical feature weights, the system presents the top contributing displacement factors as SHAP-attributed named indicators (for example, "declining freshwater availability"). This workflow must complete in fewer than five interactions, require no technical vocabulary, and support multilingual presentation to satisfy accessibility requirements (later defined by functional requirement 'FR-17').

1.7.2 Humanitarian Organization Analysts

For data analysts within humanitarian organizations, the primary workflow is a comparative monitoring sequence. These users select multiple countries or subnational regions, compare risk trajectories over configurable time windows, and analyze the SHAP beeswarm visualization to identify which indicators are driving divergent trends. This workflow supports data export in standardized formats compatible with coordination tools used by organizations such as UNHCR and IOM. The analyst must also be able to inspect the underlying data source for any displayed indicator, tracing it back through the ICD chain to its originating source, satisfying the auditability requirement.

1.7.3 Government Civil Protection Agencies

For national and regional government agencies, the primary task flow is a scenario exploration sequence. Users adjust input parameters such as projected drought severity, anticipated food price changes, and policy interventions, and observe how the model's risk assessment responds, enabling what-if analysis for resource pre-positioning and contingency planning. This workflow requires the highest level of technical engagement and presents confidence intervals alongside point predictions, making explicit what the model does and does not know. The scenario interface must prevent users from interpreting projections as deterministic forecasts, a cognitive bias that Nair et al. (2023) previously identify as a persistent failure mode in humanitarian forecasting systems.

1.7.4 The Human-in-the-Loop Feedback Cycle

Across all three task flows, the human-in-the-loop feedback mechanism operates as a cross-cutting capability. At any point in a workflow, allowing for intervention as needed, along with audit trails. A user can flag a system output as inconsistent with local knowledge acquired by an associated social network analysis, triggering a recorded annotation that enters the validation pipeline. This feedback loop closes the V-Model's right-side validation cycle by grounding statistical outputs in field-level experience, and it is the primary mechanism through which the Design Module satisfies QUORUM's CARE data sovereignty principles (Section 2.5.2). Detailed task flow diagrams, cognitive walkthrough protocols, and usability test scripts represent priority deliverables for the Design Module development. This requirement proves useful in the proof-of-concept environment to best interpret insights that arise as counter-intuitive or deeper connected than raw outputs suggest.

1.8 Significance and Contributions

This work advances both theory and practice in humanitarian systems engineering through the following contributions:

Methodological contribution: Development of a verifiable, systems-based platform integrating environmental and social datasets through a modular pipeline architecture with explicit interface control.

Technical contribution: Demonstration of a multi-level explainability framework for risk assessment that provides appropriate detail for stakeholders ranging from technical analysts to policymakers to community organizers.

Governance contribution: A data governance architecture aligned with FAIR (Findable, Accessible, Interoperable, Reusable) and CARE (Collective Benefit, Authority to Control, Responsibility, Ethics) principles that operationalizes ethical data science in humanitarian contexts.

Human-systems contribution: Integration of Human-Systems Integration (HSI) and Behavior Change Communication (BCC) strategies to ensure outputs are both technically valid and completely digestible, comprehensible, and actionable for affected communities.

Practical contribution: A reproducible methodological template for interdisciplinary collaboration among engineers, social scientists, and policymakers working on climate adaptation challenges.

The humanitarian engineering foundations of this work draw on collaborative program models that prioritize community partnership and cooperative-design. Brown and De Regil Sánchez (2021) describe the structure and outcomes of a virtual international humanitarian engineering collaborative program, demonstrating that effective humanitarian technology development requires sustained and intentional engagement between engineering teams and the communities they serve. This principle informs QUORUM's Design Module architecture, which places affected communities as fellow interpreters of system outputs rather than passive, top-down technology consumers.

In uniting data, design, and attention to community agency, this system exemplifies how transparent, explainable technical workflows can support communities in interpreting and acting upon the information that shapes their futures. It is important to distinguish QUORUM's primary contribution from the analytical results it produces. The modular three-subsystem design, formal documentation workflows, and defined data exchanges are the contribution of this thesis. The results presented in Chapter 4 are a proof-of-concept demonstration that the pipeline executes correctly and produces outputs consistent with theory. These results support the conclusion that the architecture performs as designed, with outputs generated as intended within the modular system. Migration and development indicator results discussed in Chapter 5 are presented to qualitatively validate current literature findings.

Outside of scope: This thesis focuses on developing and validating the core QUORUM architecture with proof-of-concept implementation for the Information Module and Analytics Modules for a test case of the Central American Dry Corridor (and all seven countries of Central America). Full-scale deployment, extensive user testing with affected communities, and expansion to additional geographic contexts represent natural extensions for future research. The Design Module hinges on community engagement / interaction and therefore has been omitted from the scope of this work. The functional requirements in Chapter 3 cover all three modules, as well as systemic requirements.

Chapter 2
LITERATURE REVIEW

2.1 Introduction

The literature at the intersection of machine learning, climate change, and human migration reveals both the technical progress of modern analytical approaches and persistent systemic gaps in humanitarian approaches to modern agency. This review synthesizes findings across domains including the foundations of climate-migration dynamics, modeling approaches, limitations of current methods, data governance considerations, and human-systems integration requirements. Each domain informs specific design decisions in the QUORUM framework.

2.2 Environmental Drivers and Nonlinear Responses

Studies demonstrate that environmental shocks, particularly those that disrupt water systems and agricultural productivity, are consistently identified as primary migration drivers, with extremes at both ends of the hydrological spectrum posing distinct challenges (National Academy of Sciences, 2024). Agent-based models have simulated labor migration in response to drought-induced yield loss, revealing complex household decision dynamics (Cai & Oppenheimer, 2013). The increasing frequency and intensity of extreme weather events including flooding, heat stress, and sea-level rise are ominously impacting livability and food security, thereby exacerbating migratory pressures (Johnson et al., 2024).

However, climate vulnerability alone is not a deterministic predictor of migration probability (Cattaneo et al., 2019). Severe drought may paradoxically immobilize populations by depleting the resources needed to move. Meanwhile, milder but chronic stressors can accelerate urbanization (Johnson et al., 2024; Cai et al., 2016). This “immobilization under extreme stress” pattern is a recurring finding in the literature, highlighting the need for models that capture thresholds and nonlinear dynamics rather than assuming linear relationships between environmental stress and migration response (Willekens et al., 2024).

2.2.1 The Water-Energy-Food Nexus

The interdependencies and access to water, energy, and food are well established as the non-negotiables in human civilization. The relationship follows dependence based cycles in which water is needed to produce energy, energy is required to extract, distribute, and treat water, and food production depends on both (Shannak et al., 2018). Despite this recognition, the institutions tasked with governing these resources have historically operated in isolation. Managing water, energy, and food through separate agencies with distinct missions, staff, data streams, and analytical tools Shannak et al. (2018) found that this fragmentation has produced noticeable inconsistencies in WEF modeling. These have included variations in methodological approach, data type, and policy interpretation that undermine the goal of integrated and systemic planning. Their comparative review of nine frameworks and models spanning the Water Evaluation and Planning System, the Environmental Livelihood Security model, Climate Land-use Energy and Water Strategies (CLEWs), Multi-Scale Integrated Analysis of Societal and Ecosystem Metabolism, and others reveal that while each tool offers robust capabilities within its scope, none fully integrates water, energy, and food dynamics across spatial and temporal scales simultaneously. The CLEWs model, for instance, adds a climate dimension to WEF analysis but still lacks the analytical tools to trace how shifting cropping patterns affect energy systems over time and space. This fragmentation is a technical problem and has direct policy consequences for regions where decisions about agricultural investment, water allocation, and energy infrastructure are deeply interwoven.

2.2.2 Multi-Driver Interactions and Feedback Loops

Climate impacts interact with political uncertainty and economic opportunity structures, creating trade-offs that households must navigate. Notably, fewer than 15% of reviewed migration studies explicitly model these multi-driver interactions (Willekens et al., 2024), emphasizing the importance of feedback-aware frameworks. Understanding the contextual factors and mechanisms influencing migration is crucial for developing effective policy interventions aimed at protecting vulnerable populations (Hoffmann et al., 2021).

A persistent challenge in this domain is accurately forecasting displacement dynamics and identifying root causes, given the inherent complexity and fluid nature of real-world contexts where multiple stressors co-evolve (Nair et al., 2023).

Recent qualitative evidence reinforces these quantitative findings about multi-driver complexity. See et al. (2026) conducted 32 semi-structured interviews with humanitarian practitioners in the Philippines and found that practitioners acknowledged the overlapping nature of climate and conflict displacement in practice, yet their labeling and classification language still fell short of capturing the full reality of what the authors term "complex displacement". Temporal classification reveals that climate-induced displacement often affects more people for shorter durations, while conflict-induced displacement tends to affect fewer people but persist far longer. This distinction is one that annual WEF indicators may struggle to capture. These practitioner-level findings illustrate why QUORUM's multi-source, time-lagged analytical approach is necessary. It shows that a framework that ingests only single-trigger displacement data, or collapses temporal dynamics into annual panels, risks replicating the very blind spots that practitioners on the ground have identified.

2.3 Modeling Approaches in Migration Research

2.3.1 Machine Learning Approaches

Methodological toolkits now span agent-based models and AI-enhanced hazard engines (Best, 2019; Best et al., 2022), but most studies still investigate single causal pathways rather than the full interaction among climate, social, and economic drivers. This siloed approach limits the complex nature of migration as well as policy relevance. Studies report that migration tends to be a tool of higher socioeconomic demographics in mostly flood-prone areas, leaving the most vulnerable populations trapped in deteriorating conditions (Duijndam et al., 2025; Reimann et al., 2023).

The application of artificial intelligence to WEF nexus analysis represents a rapidly advancing frontier. A growing body of WEF systems research identifies a suite of AI tools applicable to nexus management, including machine learning for pattern recognition, natural language processing, computer vision, optimization algorithms, and predictive analytics. This literature argues that AI's capacity to identify critical trends within operational data is particularly valuable for untangling the complexities of complex resource challenges involving diverse and nonlinear variables. They propose that the integration of AI and Internet of Things (IoT) sensor data could enable real-time adaptive decision-making, allowing system managers to dynamically respond to changing environmental conditions and anticipate potential disruptions before they escalate. For QUORUM, this framing directly supports the Analytics Module's use of machine learning (including Random Forest and regression approaches) for time-lagged cross-correlation between WEF stressor events and migration anomalies. Additionally the system is designed precisely to detect the leading-indicator signatures of displacement dynamics within open-source data streams, including and especially interfacing with sensor networks.

2.3.2 System Dynamics and Hybrid Methods

A recent and instructive precedent for integrated WEF nexus modeling comes from Correa-Cano et al. (2022), who developed a modular Water-Energy-Food-Environment (WEFE) modeling toolkit applied to agricultural export development in a severely water-stressed region of Peru region facing dynamics comparable in character, if not scale, to those of the Central American Dry Corridor. Their toolkit combines used modeling components including irrigation simulation, partial equilibrium economic modeling, and life cycle assessment to simultaneously quantify water and energy productivity at the field level, socioeconomic outcomes (employment, income, food prices) at the regional level, and environmental impacts at both local and global scales. A key innovation is the described "soft-linked" architecture, in which modules are integrated through data exchange rather than hard-coded coupling, allowing each component to remain technically rigorous within its domain while still contributing to a integrated, cross-sectoral picture. Correa-Cano et al. (2022) found that existing nexus tools had failed to simultaneously capture both biophysical and socioeconomic dimensions, a gap their toolkit was specifically designed to close. For QUORUM, this precedent validates the modular, multi-source architecture of the Information and Analytics modules, and affirms that combining environmental data streams with socioeconomic indicators is both technically feasible and analytically essential for producing actionable insights in resource-stressed contexts. Additionally, the system is adaptable to any input as long as the interface document control parameters are met between modules.

2.4 Limitations of Current Models

Even in the age of information, well-designed and modular toolkits face the challenge of translating complex multi-sector outputs into insights accessible to diverse stakeholders. Correa-Cano et al. (2022) acknowledge this directly in their limitations noting that despite its methodological rigor, the complexity of the results demands careful synthesis and communication to ensure that findings reach relevant decision-makers in accessible ways. They call for stronger collaboration between researchers, policymakers, and stakeholders to enable data sharing and turn analytical outputs into actionable guidance. This is a constraint that QUORUM treats not as a limitation but as a design requirement using the Human-Systems Integration and Behavior Change Communication strategies built into QUORUM's Design Module architecture ensure that risk assessments are technically valid in addition to being comprehensible and actionable for the communities most affected by climate-related displacement. The ultimate takeaway from Correa-Cano et al. (2022) is that a model's social utility depends on what it calculates in addition to whether the people who need it can understand and act on it.

2.4.1 Interpretability and Black-Box Concerns

Few frameworks have the capacity to integrate real-time, open-source environmental and socio-economic data streams or embed governance and ethical safeguards into their computational pipelines (Cattaneo et al., 2019). Literature review helped to reveal that many machine-learning approaches lack interpretability and can avoid acknowledgment of errors. This ultimately limits stakeholder trust or can increase the spread of misinformation when outputs are acted upon without understanding. The structural research gap identified by Daher et al. (2021), that no WEF model or assessment framework had incorporated migration as a thematic focus (as introduced in Section 1.1), is particularly consequential given that the United Nations 2030 Agenda for Sustainable Development explicitly recognizes migration as an adaptive strategy (SDG Target 10.7) alongside water (SDG 6), energy (SDG 7), and food (SDG 2). This gap directly motivates the design of QUORUM.

2.4.2 Siloed Approaches and Single-Hazard Focus

Across the reviewed literature, most models focus on single hazards or regions, neglecting integrated Water-Energy-Food nexus investigations that capture the compound stressors communities actually face. A persistent structural weakness across WEF frameworks is their tendency toward single-sector optimization. Shannak et al. (2018) found that even models somewhat designed for nexus thinking tend to privilege one resource, usually water, as the primary driver, leaving the understanding of complex development drivers incomplete. Their review further highlights that policymakers working at sub-national and regional scales are particularly underserved in their toolsets. While macro-level aggregate models exist, the spatial and temporal granularity needed to capture how resource availability changes across geography and seasons is mostly missing. This is a critical gap for a system like QUORUM, which must operate at the subnational level in the Central American Dry Corridor, where seasonal precipitation anomalies, localized crop failure, and community-level energy access can be informed by IoT devices and interact in ways no single-sector model can adequately represent. Shannak et al.'s (2018) call for "integrated planning and system thinking rather than optimizing the use of one resource over another" is precisely the design mandate that QUORUM's multi-source data ingestion and WEF indicator integration module is built to fulfill.

This siloing extends beyond analytical models to the humanitarian operational landscape itself. See et al. (2026) demonstrate that the humanitarian response system in the Philippines remains structured around distinct organizational mandates, with UNHCR leading on conflict displacement while IOM and IFRC lead on disaster displacement, even when the same communities experience both types of crisis in rapid succession. Their qualitative research identifies factors that shape organizational decision-making about where and how to respond with formal government requests for assistance, existing operational presence (a pattern related to what the humanitarian literature calls “tarmac bias”), security constraints in conflict-affected zones, and the logistics of aid distribution to centralized versus dispersed populations. Together, these factors create a pattern in which protractedly displaced conflict populations are systematically overlooked when new climate disasters strike the same areas. The Marawi Siege of 2017 and the subsequent landfall of Typhoon Tembin just two months later illustrate this compounding dynamic. The humanitarian responses to the typhoon were narrowly structured around the new trigger event, effectively erasing the lingering effects of the preceding conflict for communities still in displacement (See et al., 2026; Weerasinghe, 2021). For an analytical system like QUORUM, this finding has direct design implications, the framework must be capable of representing compounding displacement histories rather than treating each observation period’s migration signal as an isolated event. QUORUM’s lagged cross-correlation approach, which explicitly models delayed responses between environmental stressors and migration outcomes, and random forest analyses both offer a mechanism for capturing these accumulations.

2.4.3 Uncertainty Quantification

A persistent limitation across climate-migration modeling is the insufficient treatment of uncertainty. Willekens et al. (2024) note that fewer than 20% of reviewed migration models provide calibrated uncertainty estimates alongside point predictions, a deficiency that undermines both scientific credibility and policy utility. When decision-makers act on migration forecasts without understanding the confidence bounds of those forecasts, the resulting interventions risk being either insufficient or misallocated.

Uncertainty in climate-migration systems arises from multiple compounding sources. Data uncertainty stems from experimental design flaw, measurement error, reporting lags, and inconsistent spatial or temporal coverage across datasets collected by different agencies with different methodologies. Model uncertainty reflects the choice of analytical method, hyperparameter (feature) selection, and the degree to which a model's structural assumptions capture the true data-generating process. Scenario uncertainty reflects the fundamental unpredictability of future climate trajectories, policy interventions, and socioeconomic evolution. Each of these uncertainty sources propagates through the analytical pipeline, and their interactions can amplify total prediction error in nonlinear ways.

Nair et al. (2023) demonstrate that even well-performing humanitarian forecasting systems can produce misleading outputs when uncertainty is left unquantified, particularly in contexts where decision-makers may treat any numerical output as precise. Their work on explainable forecasting for humanitarian needs assessment highlights that communicating what a model does not know is as operationally important as communicating what it predicts.

QUORUM addresses this gap through multiple design mechanisms. At the Analytics Module level, the system computes confidence intervals for regression coefficients, employs leave-one-group-out cross-validation to test generalization limits, and uses SHAP values to attribute prediction variance to specific features rather than presenting opaque aggregate scores. At the system level, the Interface Control Documents require explicit documentation of data provenance, missingness rates, and coverage flags at every module boundary, ensuring that uncertainty information travels with the data through the full pipeline rather than being discarded at intermediate stages.

2.5 Data Governance and Ethical Considerations

2.5.1 Digital Colonialism and Data Asymmetries

Advanced technologies have expanded analytical capacities but also deepened asymmetries in data access across populations of internet-users (Chi et al., 2025). Control over these capabilities is concentrated among a limited set of institutional actors, a concentration that risks reinforcing digital colonialism, where data from vulnerable regions fuels models without commensurate benefits returning to those communities (Best, 2019; Couldry & Mejias, 2019). Information itself has become a tradable commodity (Yar et al., 2024), and those with access wield power over those left behind (Gaikwad et al., 2021). While performing data-sourcing for this project, the CIA Factbook, a foundational information database was removed from public online access, forcing many researchers to use an archived upload of the 2020 version hosted using the programming library GitHub. This speaks to the increasingly prohibitive access of data systems, a prime example of how anyone looking for information to aid in their circumstances, could have access revoked with little or no options available to once foundational sources of data.

While these datasets are so valuable, they are also susceptible to noisy data trends, availability issues (more on this in ‘alternatives analysis’) and the data collection can also be inconsistent. High quality satellite equipment and analysts supporting this investigation are expensive and are usually owned by private corporations or government agencies with limited budgets. Corporations and organizations often don’t spend resources on areas that either lack resources or are deemed too insufficiently valuable to warrant further attention. Systematic underrepresentation of certain geographies and demographics can only worsen these asymmetries.

2.5.2 FAIR and CARE Principles

Effective migration insights require transparent, and reproducible datasets, yet control over these systems often remains external to affected communities. The commodification of migration and environmental data perpetuates power asymmetries. Implementing FAIR (Findable, Accessible, Interoperable, Reusable) alongside CARE (Collective Benefit, Authority to Control, Responsibility, Ethics) principles transforms open-data frameworks into modes of data sovereignty. This approach ensures that data collection and analysis serve the interests of affected communities, fostering trust and enabling localized, culturally appropriate responses to migratory pressures (Hoffmann et al., 2021).

The FAIR principles, originally articulated for scientific data management, establish four requirements that QUORUM's Information Module is designed to satisfy. Findability requires that every dataset carry persistent, globally unique metadata identifiers so that both human users and automated systems can locate it. In QUORUM, this is operationalized through systems engineering verification tools. Accessibility requires that data be retrievable through standardized, open protocols. QUORUM satisfies this by drawing exclusively from publicly accessible repositories (World Bank API, SPEI NetCDF downloads, WFP data portal) and documenting the exact retrieval procedure for each source in the pipeline codebase. Interoperability requires that data use formal, shared vocabularies and representations so that it can be combined across sources. QUORUM addresses this through its canonical tabular schema with standardized ISO3 country identifiers, harmonized temporal keys (year_month), and consistent variable naming conventions enforced at the ICD-02 boundary. Reusability requires that data carry rich provenance metadata and be released under clear usage licenses. QUORUM's documentation generation function produces metadata lineage, transformation history, and coverage flags for every data element, ensuring that any downstream user can assess the fitness-for-purpose of the analytical inputs.

2.6 Human-Systems Integration and Trust

2.6.1 Behavioral Dimensions of Migration Decisions

Migration is fundamentally a human decision process shaped by cognitive biases, social learning, and bounded rationality (Willekens et al., 2024). Decisions are often mediated through social networks, collective perceptions of risk, and cultural norms, which means that ignoring behavioral feedback loops can lead to inaccurate predictions and ineffective engagement strategies (Willekens et al., 2024).

Incorporating herd effects, social contagion, and prospect-theory-based decision frameworks into migration modeling improves both realism and policy relevance. The importance of understanding community communication dynamics and early humanitarian engagement using quality information well ahead of crisis onset can appreciably impact displacement outcomes and support for managed retreat where appropriate (Lenton et al., 2023).

2.6.2 Human-in-the-Loop Requirements

A robust human-in-the-loop framework is essential for validating model outputs and ensuring that the insights generated are actionable and relevant to the lived experiences of affected populations, thereby mitigating risks associated with over-reliance on automated systems (McAuliffe & Triandafyllidou, 2023). Distrust or avoidance of a system immediately disqualifies its usefulness regardless of technical performance. This pattern recurs frequently, illustrated by examples such as USAID-funded water sanitation that eroded from existence due to funding cuts leaving communities reliant on infrastructure that was never designed for long-term sustainability. Such examples show that technical performance alone cannot sustain a system's operational value. Without embedded mechanisms for ongoing community validation, stakeholder feedback, and adaptive maintenance, even well-funded interventions degrade. QUORUM's human-in-the-loop architecture is designed to prevent this outcome by requiring that system outputs be continuously calibrated against the lived experience of community members in the field, ensuring that the systems remain grounded in local realities rather than drifting toward purely algorithmic abstraction.

2.6.3 Linking HSI Principles to System Design

Human-systems integration as it applies to QUORUM ensures that any computational model remains socially grounded, ethically aligned, and culturally responsible. For QUORUM, this translates to specific and intentional system design requirements. Interfaces must accommodate users across technical literacy levels, outputs must be interpretable within local cultural contexts, as well as dashboard style interface for information and education. Any feedback mechanism feature design must allow communities to validate or challenge system assessments based on local knowledge. Information derived from social network analysis, and the contexts of the communities at risk, must be present at the beginning of system design.

Research on technology adoption in complex sociotechnical systems provides further grounding for QUORUM's human-centered design approach. Johnson, Miller, and Conrad (2022) identify that public perception, integration factors, and trust are critical determinants of technology adoption in Urban Air Mobility systems, findings that generalize to any human facing analytical system. Their work highlights that technical performance alone is insufficient for system acceptance, and that perceived transparency, stakeholder engagement, and integration with existing decision-making workflows are equally important. These adoption factors map directly onto QUORUM's Design Module requirements for explainability (FR-16), cultural awareness (FR-14), and stakeholder communication (FR-15).

Future extensions of this work could incorporate participatory methods such as multilingual workshops, culturally contextual visualizations, and co-designed interfaces developed with affected communities to strengthen understanding and trust (Besinga & Theophilus, 2025; Monteiro de Carvalho, 2026). The demand for community-centered, participatory approaches in WEF nexus should break from being an aspirational principle as it is increasingly recognized to being a technical necessity. WEF systems design research increasingly argues that technical and environmental efficiency must be broadened to incorporate social and economic dimensions, including equity, justice, and distributional impacts of resource access. Community engagement and participatory modeling are not only to enhance the legitimacy of solutions but also unlock localized knowledge that top-down technical systems routinely miss. This principle maps directly onto QUORUM's HSI design philosophy, which incorporates BCC strategies, multilingual interfaces, and educational mechanisms as described in Section 1.8.

2.7 Research Gaps and Thesis Aim

This thesis aims to bridge these gaps through a modular, explainable, and human-centric system architecture that integrates environmental and social data across the systems-engineering lifecycle, with accessibility and stakeholder communication as central design principles.

The literature reviewed in this chapter reveals five converging gaps that QUORUM is designed to address. First, despite the established theoretical linkage between WEF nexus stress and population displacement, models or frameworks incorporating migration as a thematic focus are lacking (Daher et al., 2021). Second, most migration models investigate single causal pathways rather than the multi-driver interactions that the empirical evidence consistently documents (Willekens et al., 2024; Niva et al., 2021). Third, the humanitarian response system continues to address climate-induced and conflict-induced displacement through fragmented, trigger-based classification schemes that obscure compounding displacement histories (See et al., 2026). Fourth, fewer than 20% of migration models provide calibrated uncertainty estimates alongside predictions, undermining both scientific credibility and operational utility for decision-makers (Willekens et al., 2024). Fifth, the tools that do exist are rarely designed with the human-systems integration principles necessary to make their outputs accessible and actionable for the communities most affected by displacement.

QUORUM addresses these five gaps through a systems engineering architecture described in Chapter 3, combining multi-source WEF data ingestion with lagged analytical methods, SHAP-based explainability for named indicator attribution, uncertainty quantification at every analytical stage, and a Design Module grounded in human-systems integration and CARE data sovereignty principles. The following chapter describes the methodology through which this architecture was developed, validated, and implemented.

Chapter 3
METHODOLOGY

3.1 Systems Engineering Framework

A system incorporating migration data, multiple World Bank metrics, spanning several years' time, with highly visible analytical reports, carries a necessary amount of complexity. Portions of the initial literature search and codebase debugging for the QUORUM framework were conducted with the assistance of Claude, a large language model developed by Anthropic (2025). All AI-generated outputs were reviewed, suggested articles verified against Google Scholar search availability, and codebases revised by the author.

The tools used here are derived from systems engineering principles, human-system design integration, and humanitarian engineering perspectives, and together they are well positioned to handle such complexity. The system architecture, functional requirements, alternatives analysis, and interface control documentation described here portray a composed and thoughtful system with measurable inputs, transformational functions, outputs, and a contextual and environmental scope, regardless of where and when it is used.

Ramírez-Márquez and Ponce-Ortega (2023) propose a layered, hierarchical classification of Process Systems Engineering (PSE) tools for WEF nexus challenges, ranging from resource flow analysis and integrated system modeling at its core, to decision support systems, resilience and risk assessment frameworks, and adaptive management methodologies at the outer layers. Their framework explicitly places "robust decision-making" strategies designed to remain resilient across a range of possible futures, and advocates for stochastic programming approaches to navigate uncertainty in a changing climate. QUORUM's V-Model implementation is structurally aligned with this vision with requirements decomposition proceeding from operational stakeholder needs through system-level performance criteria, culminating in validation against empirical migration data. The system's uncertainty quantification protocols, multi-level explainability architecture, and modular, independently verifiable subsystems align with what Ramírez-Márquez and Ponce-Ortega (2023) would classify as an integrated, adaptive, and participatory WEF support tool with a workflow that places migration risk at the analytical center. The tools used by Márquez and Ortega (2023) will likely be used as a framework in building out the programming required for the Design Module's behavior change communication goals. QUORUM's analytical pipeline also enables meta-algorithmic principles (Simske, 2013), in which multiple intentional analytical techniques are orchestrated to compensate for possible individual method limitations. The four-stage progression from descriptive characterization through correlation analysis, fixed-effects regression, and SHAP-based feature attribution represents a structured combination of methods whose outputs are interpreted collectively rather than in isolation. This approach aligns with Simske's argument that robust systems emerge from the principled combination of diverse algorithms rather than reliance on a single best method.

The complete QUORUM codebase, including the Information Module loaders, Analytics Module pipeline (lagged correlations, fixed-effects panel regression, Random Forest, and SHAP implementations), and Interface Control Document schemas described in this chapter, is openly available at [github.com/stefanhaugen/quorum]. This location will contain all future iterations of the program as well. The repository is licensed under [MIT/Apache-2.0/etc.] and includes the exact requirements.txt, scripts, and data ingestion procedures used to generate all results reported in Chapter 4. This release operationalizes the FAIR Accessibility and Reusability principles discussed in Section 2.5.2.

3.1.1 V-Model Implementation

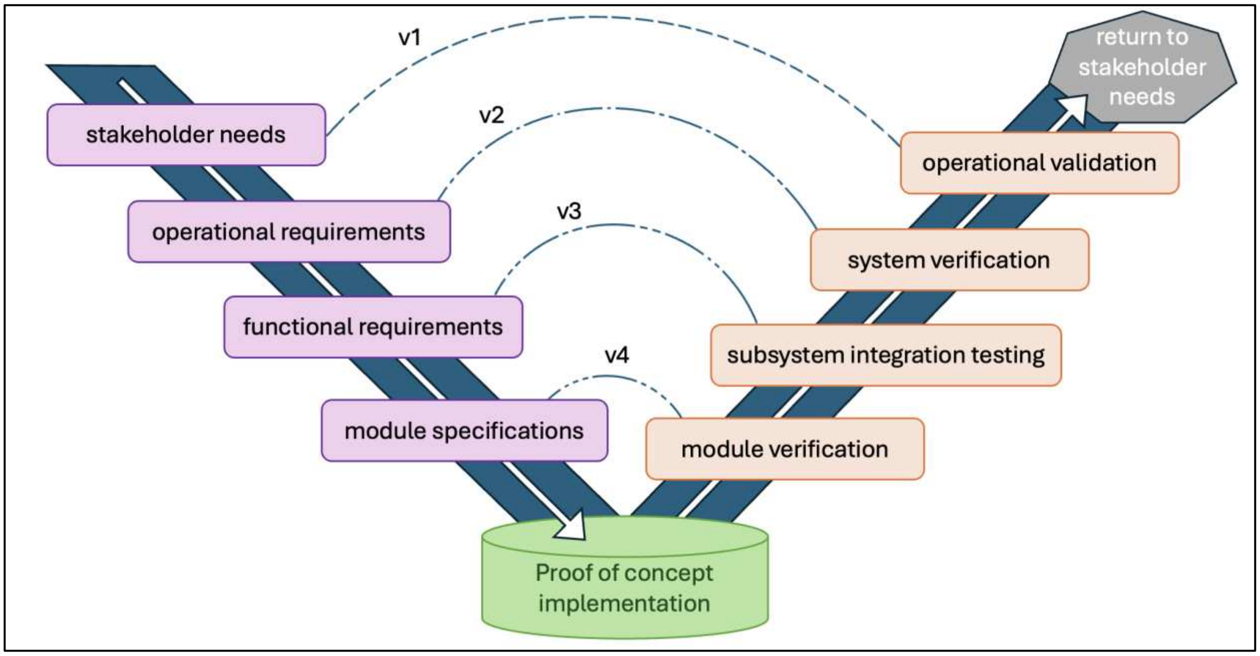

Figure 3.1. *QUORUM V-model implementation. The left descent represents system decomposition from stakeholder needs through functional requirements to module-level specs. The right-side ascent shows integration and verification, tracing analytical outputs back through ICD compliance checks to operational requirements. Arced connections (v#) are traceability verification activities explained below. The middle convergence of the V is the integration point, where build and proof of concept operations are executed.*

The canonical V-model in systems engineering applies very well to the development of QUORUM. The structure provides explicit relationships between requirements definition and system validation throughout the life cycle of QUORUM development. The reciprocal nature of this model is evident here, as system verification elements have counterparts in system decomposition, linked by their 'traceability arc' (v1 – v4). The proof-of-concept convergence point at the base of the V serves as the critical integration in which the system is executed end to end. For this thesis the implementation included the IM, AM, integrations, and system architecture applied to case-study test datasets.

Verification activities designed for system traceability are mapped to the V-model diagram as the following:

v1- Operational validation. Pipeline outputs are reviewed against stakeholder requirements.

v2- System verification. Functional requirements are traced to outputs directly.

v3- Subsystem integration. Cradle to grave pipeline execution with ICD compliance metrics.

v4- Module verification. ICD schemas are validated and acceptance criteria assessed.

3.1.2 System-Level Requirements Decomposition

Development of QUORUM begins with defining the system concept and assessing its feasibility before decomposing requirements into subsystem specifications. The system concept positions QUORUM as a decision-support tool that transforms heterogeneous open-source environmental and socioeconomic data into explainable migration risk indicators for communities, humanitarian organizations, and government agencies operating in climate-vulnerable regions. A concept feasibility assessment evaluated three dimensions. Technical feasibility was confirmed by the demonstrated availability of World Bank Development (WDI) indicators, Meta mobility data, SPEI drought indices, and WFP food price monitors covering the Central American Dry Corridor with sufficient temporal and geographic scope to support statistical modeling. Operational feasibility was assessed against the constraint that end users span a wide range of technical literacy, from rural community leaders to national-level analysts, requiring a modular architecture where analytical complexity is encapsulated within the Analytics Module while the Design Module translates outputs into stakeholder-appropriate formats. Economic feasibility was satisfied by the exclusive reliance on open-source data and open-source computational tools (Python, scikit-learn, SHAP), ensuring that the framework can be adopted by under-resourced organizations without proprietary licensing costs. This concept feasibility assessment is a standard practice in systems engineering (Kossiakoff et al., 2020), and its completion established the foundation for requirements decomposition.

Requirements decomposition proceeded from operational needs identified through the following stakeholder groups: (1) migration-candidate populations at personal and community levels, (2) humanitarian organizations requiring warning indicators, and (3) local government agencies responsible for resource allocation and disaster response. Operational requirements were decomposed into system performance requirements through structured analysis of gaps in existing systems identified in the literature (Daher et al., 2021; Shannak et al., 2018), and review of published stakeholder engagement findings from analogous humanitarian analytical systems (Nair et al., 2023). The resulting 22 functional requirements were allocated across four subsystem categories. The Information Module covers FR-1 through FR-8, the Analytics Module covers FR-9 through FR-13, the Design Module covers FR-14 through FR-18, and System Integration and Validation covers FR-19 through FR-22. Each requirement was assigned quantifiable performance criteria where feasible, with the understanding that criteria solidification requires direct stakeholder engagement before operational deployment.

3.1.3 Stakeholder Requirements Analysis

Requirements analysis revealed six critical performance dimensions that QUORUM must satisfy.

1. The system provides migration risk assessments with appropriate temporal resolution as defined by stakeholders. This resolution is monthly at most rapid and annually at least.
2. The system maintains full documentation and transparency regarding data sources, analytical methods and uncertainty propagation to earn stakeholder trust across user communities (active audit trail).
3. The design module interfaces must accommodate users across technical literacy levels, from rural community leaders with no formal data science training to national-level

analysts, ensuring that risk communication is contextually appropriate and actionable for each stakeholder group.

4. System output must achieve sufficient accuracy to distinguish between seasonal migration patterns and anomalous displacement events as indicated by climate or other measured forces. Performance criteria include a minimum feature relevance score greater than 0.3 and calibrated uncertainty estimates at the 90% confidence interval.
5. The system must comply with FAIR (Findable, Accessible, Interoperable, Reusable) and CARE (Collective Benefit, Authority to Control, Responsibility, Ethics) data governance principles, ensuring that all data sources are documented with full provenance metadata and that analytical outputs serve the interests of affected communities rather than external surveillance or enforcement purposes.
6. The system architecture must support modular independence, enabling each subsystem to be developed, tested, upgraded, or replaced without requiring structural modification to adjacent modules. Interface Control Documents must govern all inter-module data exchanges at 100% compliance.

3.1.4 Alternatives Analysis: Data Source Selection

A central decision in QUORUM's development was the selection of data sources for the primary analytical panel. Systems engineering practice calls for a structured alternatives analysis to evaluate competing design options against stakeholder requirements before committing to an architecture (Kossiakoff et al., 2020). For QUORUM, this decision shaped the temporal resolution, geographic coverage, and overall analytical power of the entire pipeline. Three candidate data source configurations were evaluated against the functional requirements established in Section 3.1.3, the ICD-01 data quality standards, and the overarching goal of producing a reproducible, open-source analytical framework.

Alternative A: World Bank WDI Annual Panel (Selected)

The World Bank World Development Indicators provide a comprehensive suite of environmental and socioeconomic indicators at annual resolution for all seven Central American Dry Corridor countries and are defined in Appendix 2. Coverage across the study period (2019 to 2022) exceeds 90% for the 14 selected WEF indicators, and the dataset is freely accessible through both bulk download and API without institutional gatekeeping. Annual WDI data can be directly aligned with Meta migration observations aggregated to yearly totals, producing a clean 14-observation panel (7 countries by analytical years). The primary limitation is temporal resolution, where annual aggregation obscures seasonal dynamics that are operationally important for early warning, and subsistence agricultural communities in the Dry Corridor respond to precipitation anomalies on timescales of months rather than years (Cai & Oppenheimer, 2013). Despite this constraint, the WDI panel satisfies FR-1 (environmental data acquisition), FR-2 (socioeconomic data acquisition), and FR-5 (data quality validation at less than 10% missingness) across all seven countries.

Alternative B: SPEI Drought Index and WFP Food Price Monthly Panel (Evaluated, Deferred)

Another environmental pipeline combined two data streams at monthly temporal resolution. The first was the SPEI Global Drought Monitor SPI-3 dataset (Vicente-Serrano et al, 2025), a NetCDF gridded product providing the Standardized Precipitation Index at a three-month accumulation window across a global 0.5-degree grid from 1901 to the present. Spatial averaging within country bounding boxes produced one SPI-3 value per country per month for all seven Dry Corridor nations, achieving complete geographic and temporal coverage across the study window. The SPI-3 index is the standard metric for agricultural drought stress (McKee et al., 1993), and its three-month window aligns with the growing season dynamics relevant to subsistence farming communities in the Dry Corridor. This data stream fully satisfied FR-1 and FR-5, with zero missing country-months across all seven countries.

The second monthly data stream was the WFP Global Market Monitor, which reports food commodity prices and year-over-year price changes for staple foods at monthly intervals. This dataset offered a direct operationalization of the food security dimension of the WEF nexus, with the year-over-year change metric capturing price pressure relative to baseline rather than short-term volatility. However, coverage analysis revealed substantial geographic gaps. Belize and Costa Rica were entirely absent from the WFP dataset, Guatemala had only three monthly observations, and coverage for the remaining four countries (Honduras, El Salvador, Panama, and Nicaragua) began only in October 2020. This produced a coverage pattern in which only 4 of 7 countries had usable food price data, and even those four had no observations for the first 21 months of the 2019 to 2022 study window. The year-over-year change variable itself carried approximately 28% missingness across available country-months. While this partial coverage could be preserved with data science NaN values rather than imputed, the asymmetric country coverage introduced a fundamental analytical constraint. Any model incorporating food prices as a feature would effectively reduce the panel to four countries, undermining the cross-country comparative logic central to the fixed-effects and leave-one-group-out cross-validation approaches. Additionally, the absence of data affected the region of interest (Dry Corridor) and therefore couldn’t be evaluated.

Alternative C: HDX Humanitarian API (HAPI) Feature Stream (Evaluated, Partially Integrated)

A third alternative replaced the WDI Excel-based feature stream with live API calls to the HDX Humanitarian API at hapi.humdata.org, sourcing IPC food security phase data and WFP commodity food price data programmatically. This approach offered the advantage of automated, repeatable data acquisition aligned with FAIR principles and eliminated manual download steps. However, geographic coverage for IPC food security classifications was similarly incomplete across the seven Central American countries, and the API introduced rate-limiting and pagination requirements that complicated automated batch processing. The HAPI integration was retained in the pipeline architecture as a supplementary data stream with paginated API client and JSON disk caching, but was not used as the primary analytical panel for the proof-of-concept results reported in Chapter 4. This was mainly a design operation issue, as connection to an API for iterative data testing was proven inefficient for rapid or offline test conditions.

Selection Rationale

Alternative A (WDI annual panel) was selected for the proof-of-concept implementation on the basis of these criteria, a complete geographic coverage across all seven countries ensured that the fixed-effects regression and LOGO cross-validation could exploit the full panel structure without asymmetric country dropout. Next, the breadth of WDI indicators (14 features spanning water, energy, food, emissions, sanitation, and forest cover) provided the multi-sector representation needed to test the WEF nexus hypothesis, whereas Alternative B offered only two environmental dimensions (drought severity and food prices). And finally, the accessibility of static datasets without internet connection requirements permitted rapid prototyping and development of the analytical module's programming and functionality. The relatively analog system described further validates the system's capability without requiring modern high-connectivity data streams. WDI dataset's long publication history and institutional backing by the World Bank provide a reproducibility guarantee that is important for a framework explicitly designed for adoption by under-resourced organizations.

The decision to defer Alternative B was not a rejection of its value. The SPEI SPI-3 drought index achieved perfect coverage and offers precisely the monthly temporal resolution that the WDI panel lacks. The food price signal, where available, provides a direct link to the household economic calculus described by Cai and Oppenheimer (2013). Both data streams have been integrated into the QUORUM pipeline architecture and are documented in ICD-01 as DS-03 (WFP food prices) and DS-04 (SPEI SPI-3). Their activation for a full monthly analytical panel, merging with Meta's monthly migration data at configurable lag offsets of 1, 2, 3, 6, 9, and 12 months, represents the most immediate expansion priority and is discussed further in Section 6.1. This expansion would increase the panel size from 14 country-year observations to potentially over 200 country-month observations, substantially improving statistical power and enabling detection of seasonal displacement dynamics that annual aggregation currently obscures.

This alternatives analysis illustrates the broader systems engineering principle that the modular architecture governed by Interface Control Documents allows the Information Module to accommodate multiple data source configurations without requiring structural changes to the Analytics or Design Modules. The ICD-01 specification defines how each data stream is ingested, validated, and harmonized, meaning that activating Alternative B or Alternative C in a future iteration requires only that the new data stream meet the schema and quality criteria already specified at the module boundary. This architectural flexibility was a deliberate design choice, informed by the recognition that open-source data ecosystems are dynamic and that coverage gaps in one period may close as organizations such as WFP and OCHA expand their monitoring footprints.

3.1.5 Verification and Validation Strategy

In order for QUORUM to perform as intended, foundational data collection is held with high importance. In this case, population data and migration data will be intricately analyzed for patterns and emergent behaviors, therefore sourcing a baseline of information from multiple legitimate sources.

The publicly available population and migration datasets used for baseline construction exhibit notable fragmentation. National population records from the United Nations, the U.S. Census Bureau, and the now statically archived CIA World Factbook (2020) differ in update cadence, if at all, methodological assumptions, and geographic granularity. These inconsistencies illustrate the heterogeneous data environment that necessitates QUORUM's dedicated Information Module and its harmonization protocols. Meta's Social Connectedness dataset (Chi et al., 2025) provides a contrasting source with substantially higher temporal resolution, estimating international migration volumes at monthly intervals through anonymized, aggregated Facebook usage patterns. Validation studies benchmarking Meta's estimates against high-accuracy national statistical bureaus in New Zealand, Germany, and Sweden reported alignment within approximately 10% error (Chi et al., 2025). This monthly resolution makes the Meta dataset a valuable validation benchmark for QUORUM.

A snapshot of migratory data is shown below, at a monthly resolution across the countries of the Central American Dry Corridor.

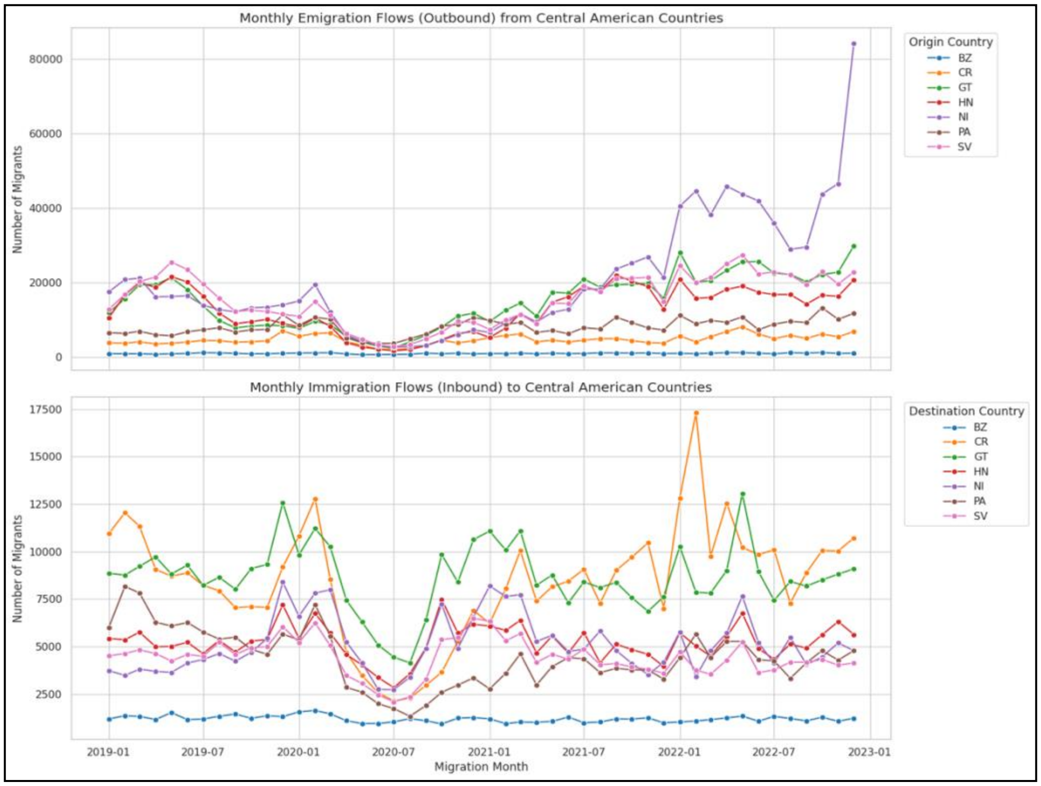

Figure 3.2. *Central American state-level outbound (top) and inbound (bottom) migratory volume from 2019 to 2023.*

With the verification baseline established through population data and Meta's monthly migration observations, the next methodological decision concerned the selection of environmental data sources for the primary analytical panel. The Analytics Module operates identically on any conformant input from the Information Module, so this decision affected only the IM configuration rather than the AM's analytical logic. A structured alternatives analysis was conducted to evaluate candidate configurations.

3.2 QUORUM System Architecture

Figure 3.2 illustrates the volume and directionality of migratory flows across the Central American corridor during the study period, providing geographic context for the system architecture described in this section. The concentration of northward flows from Honduras, Guatemala, and El Salvador reflects the regional displacement pressures that QUORUM is designed to analyze.

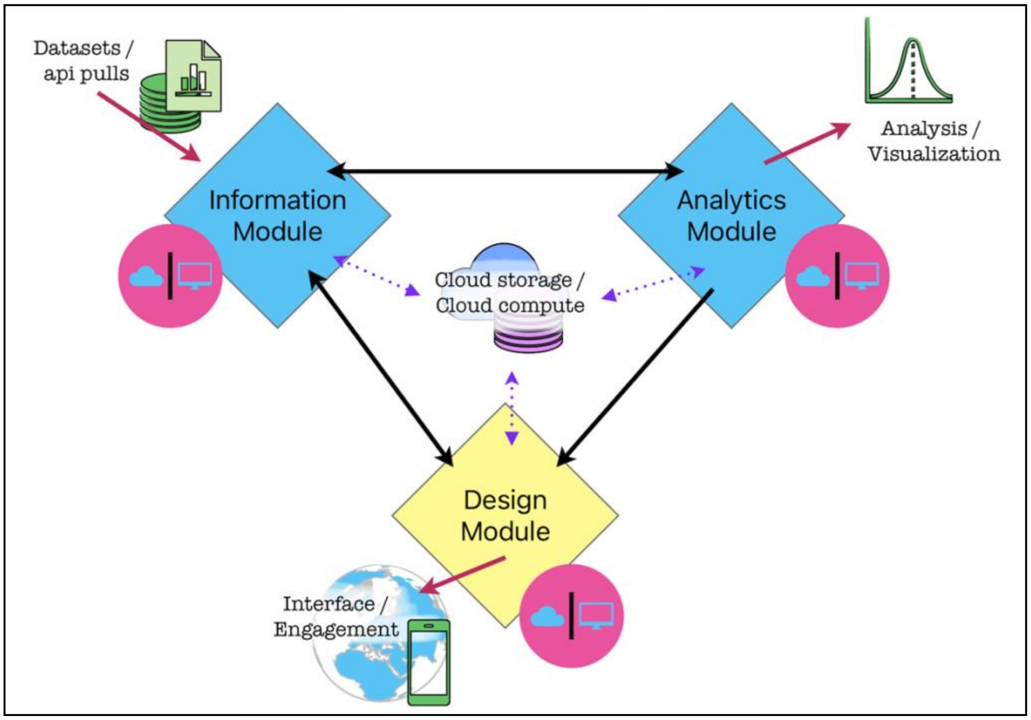

Figure 3.3. *Functional block diagram of QUORUM. Modules are mapped in diamond shapes, interactions by arrowed lines. Dotted lines indicate a cloud connected interaction. Circles indicate modality of performance at each module, either done remotely with cloud compute or locally prior to manual or remote communications.*

3.2.1 System Overview

QUORUM operates as three connected subsystems corresponding to the fundamental workflow of sourcing capable open-source data, cleaning, analysis, and delivery of useful risk assessment. This arrangement displayed visually in a simple functional block diagram (Figure 3.3).

Information Module (IM) - The IM acquires data from multiple source categories. These include environmental datasets from the World Bank World Development Indicators, climate re-analysis repositories, socioeconomic and food security indicators from UN OCHA, and migration reference data from Meta's Social Connectedness Index (Chi et al., 2025). These sources differ materially in update frequency, spatial resolution, temporal coverage, and data format, representing the heterogeneous data environment that necessitates a dedicated ingestion and harmonization module.

Functionally, the IM performs acquisition (FR-1 through FR-3), data governance (FR-4), quality assessment (FR-5), transformation and cleaning (FR-6), structured storage (FR-7), and documentation generation (FR-8). The latter functions, producing metadata, lineage tracking, and data dictionaries is impactful on the system's validation and verification requirements. The complete traceability of every data element must be seen from the data block back to its originating source, ensuring that risk attribution in the AM output can be independently audited if needed. The IM's output boundary is defined by ICD-02, which specifies the exact specifications and quality standards that the ABIB must satisfy before handoff to the AM.

Analytics Module (AM) - The AM receives the vetted information from the IM and applies a sequential analytical protocol to produce calibrated risk assessments with attributed drivers and quantified uncertainty. The AM is the primary analytical engine of the QUORUM system, and its boundaries are precisely defined. It accepts only conformant inputs and produces only ICD-03-conformant outputs. This interface discipline ensures that the AM's internal analytical choices like method selection, dimensionality reduction, or feature choices all remain encapsulated and do not create dependencies in either the IM or the DM.

The AM applies analyses in a consistent sequential order across all incoming information blocks. This procedural consistency provided by robust data science approaches allows for applying the same analytical stages to every dataset regardless of source. The module enables comparative assessment across data batches, geographic configurations, and indicator sets without requiring structural code modification. The analytical stages include:

1. general statistical characterization
2. linear correlation analysis
3. non-linear relationship detection

4. meta-algorithmic modeling
5. explainability attribution

Each stage is described in section 3.3 and in context in Chapter 4. The AM produces a standardized set of analytical output products, including correlation tables, regression summaries, feature importance rankings, and SHAP visualizations, all conforming to the ICD-03 schema for handoff to the Design Module.

Design Module (DM) - The DM is the QUORUM's external-facing constituent and the primary mechanism through which its Human-Systems Integration system requirements are satisfied. It receives the AM's outputs and transforms them into stakeholder-appropriate interface products, including dashboards, multilingual community communications, educational tools and scenario exploration tools. The DM is architecturally the most complex yet customized module in scope, as it serves stakeholders with substantially different technical literacy levels from rural community leaders with no data science background to technical analysts at national civil protection agencies without modifying the underlying analytical content produced by the AM.

The full deployment of QUORUM's Design Module incorporates the Behavior Change Communication strategy and the human-in-the-loop feedback mechanism that allow communities and field organizations to validate or challenge system outputs based on local knowledge. This feedback pathway creates a system-level validation and verification mechanism that provides a calibration from operators with direct knowledge of local conditions, complementing the statistical validation conducted within the AM. The DM is allocated functions FR-14 through FR-18, covering cultural context, stakeholder coordination, explainability and visualization, multilingual communication, and scenario exploration. System integration requirements (FR-19 through FR-22) govern operational governance, security, and system-wide validation.

User-interfacing and human-systems integration factors applied to analyzed or imported and processed information blocks occurs in the DM. In normal arrangements, this module interfaces highly with compute for development and maintenance of app-interface, website portal, and external collaborations for further research.

Modularity permits independent development, execution and assessment of each sub-system while maintaining clear interface specifications that apply globally. The architecture proposed displays SoS characteristics that allow for modules to function autonomously while contributing to the emergent capabilities of the fully integrated QUORUM system. The IM is the first intentional element in the system and defines the data ingestion pipeline for heterogeneous sources of information, UN OCHA datasets, climate re-analysis metrics, and Meta mobility data, with distinct update periods, data formats, universal data storage, and quality assurance protocols. ICDs specify the empirical transformations, temporal data matching, and spatial aggregation rules required to harmonize the data for understanding and further processing. The AM transforms standardized inputs into insights with quantified uncertainty.

3.2.2 Subsystem Functional Analysis

The decomposition from system-level requirements to subsystem specifications proceeds through a functional analysis that identifies the specific capabilities each module will provide to satisfy system objectives. By subsystem, requirements were generated based on the fundamental information transformations required, functions demanded to produce specific outputs, and input from stakeholders across modules. Appendix Table 1 includes each functional requirement, its definition, performance metric and is mapped to its respective location. Figure 3.4 maps each functional requirement to its assigned module, illustrating the distribution of system responsibilities across the IM, AM, and systemic network, prepared for DM integration.

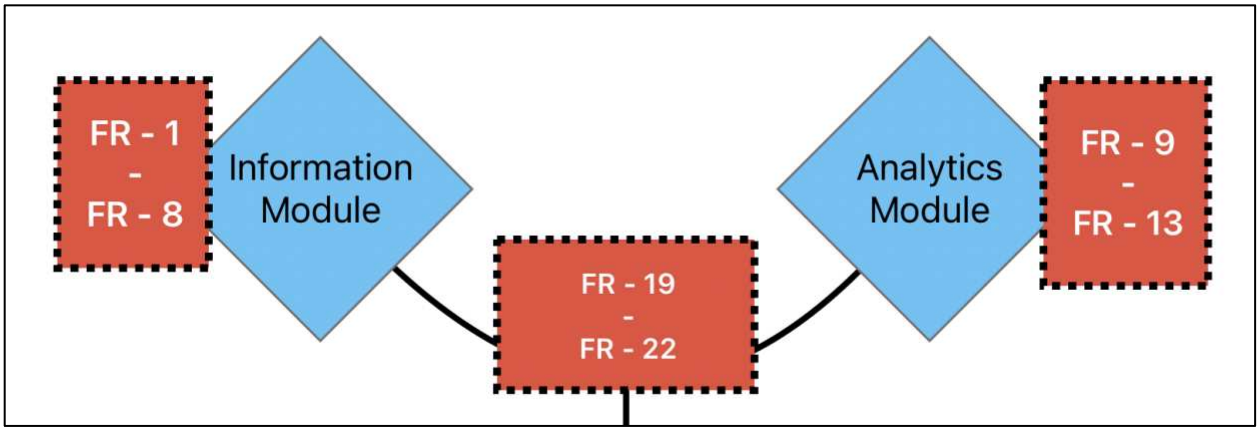

Figure 3.4. *Module functional requirement assignment. System integration and validation requirements are connected to each module through central threads as they are requirements at the data creation, transfer, and data visualization stages.*

3.2.3 Interface Control Document

Expected inputs and outputs are what establish consistent systems. Kossiakoff et al. (2020) define Interface Control Documents (ICDs) to specify the complete definition of each interface between configuration items, providing the basis on which individual component developers may design and test their products independently. These interfaces are woven within QUORUM as external data sources are ingested by the Information Module (ICD-01), the Information Module performs an Analytics Module handoff (ICD-02), and the Analytics Module to Design Module output is defined by (ICD-03). FR-22 (a System Integration and Validation requirement) requires ICD compliance at 100%.

ICD-01 External Sources to Information Module

Interface: Data acquisition boundary of the system. This interaction governs all inbound data streams before any transformation.

Stream ID	Source	Pipeline Variable	Temporal Res.	Spatial Scope	Transformation Rule
DS-01	World Bank WDI (World Bank, 2025)	wdi_raw	Annual (2020-2024)	7 CA countries ISO3	Pivot long to wide; retain FEATURE_CODES

					keys; rename columns per FEATURE_CODES values; deduplicate on iso3 by year
DS-02	Meta Movement Dataset (Chi et al., 2025)	mig_raw	Monthly (2019-2022)	7 CA countries ISO2 to ISO3	Filter country_from within CA_ISO2; map ISO2 to ISO3; aggregate num_migrants by iso3 x year x month; compute net_outbound
DS-03	WFP Global Market Monitor	df (food loader) YoYChangeMonth	Monthly (Oct 2020 onward)	4 of 7 countries (BLZ, CRI absent)	Skip HXL row 1; map CountryName to iso3 via WFP_TO_ISO3; mean YoY across staple foods per iso3 x month; use NaN for absent countries
DS-04	SPEI Global Drought Monitor (v2)	spi_df spi3 column	Monthly (1901 to present)	7 CA countries 0.5 degree grid	Load NetCDF; filter COUNTRY_BOXES by latitude/longitude index; spatial mean per country per month; mask fill values less than -9990

Compliance check: FR-22 requires ICD compliance at 100%. Each data stream above is validated by the print statements in Steps 2a-2d of the pipeline. Observed row counts (monthly_mig: 336 rows; wdi_wide: 35 rows × 17 cols) confirm schema adherence. data_note column in the data loader provides per-row provenance flags (Coverage Flag, CF) satisfying (FAIR/CARE compliance).

ICD-02 - Information Module to Analytics Module

Interface: One of the most important interactions, delivering organized and clean information to analyses. This is interface control generates the primary analytical dataset from IM to AM.

Product ID	Variable Name	Schema	N Rows	Contents and Rules

ABIB-01 (primary)	primary_panel = lag_panels[1]	iso3, year, outbound_total, outbound_to_us, + 14 WDI features	14	Country-year panel at lag=1 year. Primary analytical dataset for all 3B-3E analyses. Saved as quorum_panel_lag1.csv. Features drawn from available_feats (>=10 non-null observations). WDI year shifted +1 before merge so WDI[t-1] aligns with migration[t].
ABIB-02 (lag variants)	lag_panels[0] lag_panels[2]	Same as ABIB-01	217	Lag-0 and lag-2 panels used exclusively in 3B correlation analysis. Not used as RF training data. Allow testing contemporaneous and 2-year-lead relationships against the 3-12 month theoretical prior.
ABIB-03 (v2 monthly)	monthly_env (v2 loader)	iso3, year, month, spi3, food_price_yoy, data_note	336	Monthly environmental panel produced by quorum_v2_monthly_loader.py. Merged with monthly_mig on [iso3, year, month] at configurable LAG_MONTHS offsets [1,2,3,6,9,12]. Replaces annual lag_panels in v2 pipeline.

Handoff verification: The IM outputs ABIB-01 as a CSV confirmation (quorum_panel_lag1.csv) before any AM computation begins. This decoupling ensures the AM can be re-run independently without re-executing the IM, satisfying the modular independence requirements, as well as minimizing compute demands. The FEATURE_COLS list, computed reactively by its inputs at the ICD-02 boundary, constitutes the Allocated Feature Set (AFS) transferred to the AM.

ICD-03 – Analytics Module to Design Module

Interface: Design Module outputs as designed. All AM outputs are written to OUTPUT_DIR before DM import.

Product ID	Filename Pattern	Format	Producing Variable	DM Presentation
DM-01	~prefix…_3B_lagged_correlations.csv	CSV	lag_corr_df	Tabular display in DM dashboard; drives

				significance filtering (p<0.05); source for LCVR entries in VCRM. Dual-prefix (tot_, us_) supports comparative analysis.
DM-02	~prefix…_3B_lagged_correlation_heatmap.png	PNG 150dpi	pivot_corr pivot_p	Primary visualization for RQ2 evidence. Asterisks mark p<0.05 cells. Embedded as Figure in thesis Chapter 4.
DM-03	~prefix…_3C_fixed_effects_results.csv	CSV	fe_df	FEVR source; documents beta, SE, t, p per feature with Significant flag. Used to populate VCRM FR-11 row.
DM-04	~prefix…_3C_fixed_effects_coefplot.png	PNG 150dpi	fe_df	Coefficient plot with 95% CI error bars; red = p<0.05. Embedded as Figure 4.3.
DM-05	~prefix…_3D_shap_importance_pca.png	PNG 150dpi	mean_shap	SHAP bar chart in original WDI feature space (back-projected). Primary RQ3 visualization. Feature ranking answers stakeholder requirement for named indicator attribution.
DM-06	~prefix…_3E_pca_variance.png	PNG 150dpi	pca_obj	Scree plot of explained variance ratio per principal component. Documents dimensionality reduction decision for AM audit trail.

3.3 Statistical and Machine Learning Methods

QUORUM's AM specifically requires layered analytical methods to result in outputs that drive risk mitigation strategies. The four-stage pipeline proceeds from descriptive characterization, lagged cross-correlation, fixed-effects panel regression, and Random Forest modeling with SHAP-based feature attribution. Each method was selected to address the composition of the datasets in use, and to address specific research questions while accommodating the constraints imposed by the limited panel size.

3.3.1 Lagged Cross-Correlation Analysis

Pearson correlation coefficients were computed between each WDI feature and log-transformed outbound migration. This lag assessment was done across three intervals, meant to simulate the real-world delay in mobilization. A zero-year lag, one-year and two-year lag were chosen based on reported outmigration delays of 6 months to one year established in literature, as well as by the confines of the data resolution (Cai & Oppenheimer, 2013; Niva et al., 2021). Statistical testing was conducted at $\alpha = 0.05$ with no correction for multiple comparisons, consistent with the structural intent of the proof of concept.

3.3.2 Fixed-Effects Panel Regression

A fixed-effects panel specification was chosen over random effects based on the premise that unobserved country-level characteristics are correlated with WDI metrics. In the current configuration of QUORUM and considering these inputs, a random-effects estimator could produce biased coefficients, and is likely an appropriate option for further expansion of this analytical approach. A Hausman Specification Test can be used to evaluate whether random-effects or fixed-effects panel approaches should be used, where a significant result ($p < 0.05$) of this test indicating that fixed effects is preferred, and an insignificant result pointing towards random effects applications (Baltagi, 2014). Using the fixed-effects approach, each model regressed log-transformed total outbound migration on a single standardized WDI feature lagged one year, with the full set of country dummy variables included. The one-feature-at-a-time technique was adopted due to the panel dimensions (n=14, with six degrees of freedom used by country fixed effects) are insufficient to support multivariate regression without risk of overfitting. Log transformation was applied to the dependent variable to address right-skewed migration count distributions (Nicaragua, Honduras) and improve understanding of these related factors. The one-feature-at-a-time specification follows the following model for each regression:

$$\log(\text{outbound\_migration})_{it} = \alpha + \beta_k \cdot \text{WDI\_feature}_k(i, t\text{-}1) + \delta_i + \varepsilon_{it}$$

where $\log(\text{outbound\_migration})_{it}$ is the logarithmic migration per country (i), per year (t), and is the dependent variable. α defines the baseline log migration for the reference country (intercept), δ_i represents the country fixed effects (six dummy variables for seven total countries), β_k is the coefficient for single WDI features at a time, and the subscript k changes for each of the 14 separate regressions. Residual error per country (i), per year (t) is consequently captured by the term ε_{it}. Using this schema, in which a feature is run in isolation, each individual regression only requires one predictor coefficient plus the six country dummies. The subscript k is the most important variable in the equation above as β_k instructs indexing by feature, fitting a different model for each development indicator. Fitting one model with all k features does not apply well for this evaluation as all degrees of freedom are consumed and negative correlations can be the only conceivable fit. By performing this regression one-at-a-time, the limited degrees of freedom are used to generate one slope plus six country dummies plus the intercept, allowing (n-1 = 6) degrees of freedom for the error term. This is a constrained circumstance, explained further in Chapter 4 surrounding *Forest Cover* and it's within-country indications. Importantly, fixed-effects is unable to control for confounding among the WDI features but serves as a useful evaluation for the framework intended for scaled-up datasets. In this case, if *Forest Cover* and *Freshwater per Capita* are correlated within countries over time, the *Forest Cover* regression cannot distinguish its own effect from the portion attributable to the associated undulating freshwater availability. This is a certainly a fixed-effects limitation, and it is one reason the SHAP analysis is deployed as a valuable complement.

3.3.3 Random Forest Modeling

A Random Forest (RF) regressor (Breiman, 2001) was trained on the 14-observation lagged panel using all 14 standardized WDI features as predictors and log outbound migration as the dependent variable. The regressor built 500 decision trees with maximum depth constrained to four levels and a minimum leaf size of two observations. These hyperparameters were chosen to best balance model overfitting from effective model performance, a strategy in-line with regularization recommendations for low-sample ecological and panel data (Probst, Wright & Boulesteix, 2019). The python package scikit-learn was used for this model fitting (Pedregosa et al., 2011).

In order to cross-validate this approach, suggested by Willekens et al, leave-one-group-out (LOGO) was used. This tactic holds out all observations from one country in each fold of the Random Forest, and tests whether the pattern observed can be derived from any six countries in Central America, to the seventh.

3.3.4 SHAP Feature Assignments

Shapley Additive Explanations were computed from the trained Random Forest to produce a meaningful feature importance ranking, bringing interpretability to the often esoteric machine learning workflow (Lundberg & Lee, 2017). SHAP values quantify the marginal contribution of each feature to the model's prediction for each observation, averaged across all possible WDI orderings. The TreeExplainer tool (Lundberg et al, 2020) was utilized to compute precise SHAP values for the generated tree-based models. Mean absolute SHAP attribution values were elected for in this case over other approaches (Gini importance), as these values account for feature interactions and provide stable rankings on correlated indicators (Lundberg & Lee, 2017; Molnar, 2020). This property is important given the known collinearity among WDI indicators within the considered WEF nexus.

These approaches (fixed-effects, RF and SHAP) are used separately in order to answer different questions. The fixed-effects one-at-a-time regressions ask "does a single feature have a detectable within-country alignment with outmigration after removing all time-invariant country characteristics?" Whereas the SHAP analysis following RF modeling is able to answer "when all features are considered together, which indicators contribute most to the model's predictions?" As *Forest Cover* shows up strongly in both approaches, and *Carbon Intensity* ranks highest only in SHAP, the nature of the signal in each feature can be distinguished, indicating meaningful insights for QUORUM's framework with these proof-of-concept data.

3.3.5. Computational Environment

All analyses were implemented in Python 3.11 using the following libraries and packages: pandas, numpy, and matplotlib for data manipulation, scikit-learn (Pedregosa et al., 2011) for RF modeling and LOGO cross-validation, statsmodels for fixed-effects regression, and the shap library (Lundberg & Lee, 2017) for SHAP value attributions.

Chapter 4
RESULTS

4.1 Systems Engineering Results (Framework Validation)

The systems engineering implementation produced several deliverables that support the viability of the V-Model approach for humanitarian analytical systems. Twenty-two functional requirements were decomposed across four subsystem categories. The Information Module covers FR-1 through FR-8, the Analytics Module covers FR-9 through FR-13, the Design Module covers FR-14 through FR-18, and System Integration and Validation covers FR-19 through FR-22. Each requirement was assigned quantifiable performance criteria where feasible, with stakeholder engagement identified as a prerequisite before criteria solidification and deployment.

Three Interface Control Documents were developed to govern distinct data exchange at module boundaries, as specified in Section 3.2.3. ICD-01 governs data ingestion from external sources, ICD-02 governs the Information-to-Analytics Module handoff, and ICD-03 governs Analytics Module outputs delivered to the Design Module as defined by thesis scope.

The modular architecture was validated through a proof-of-concept execution in which the full pipeline was run end-to-end from raw data ingestion through AM visualizations. The Information Module successfully harmonized four data sources with different temporal resolutions (monthly and annual), spatial scopes (country-level and grid-based), and data formats (CSV, Excel, and NetCDF) into a unified analytical panel. The Analytics Module processed this panel through all analytical stages in sequence without requiring structural code modification between stages, supporting the modular independence requirement. ICD compliance was verified through row-count confirmation, schema validation, achieving the 100% compliance target specified by FR-22.

The V-Model's (Figure 3.1) verification-validation arced line pairings were assessed by tracing each analytical output back to its originating functional requirement. The lagged correlation analysis satisfies FR-9 (feature engineering) and FR-10 (WEF stressor detection). The fixed-effects regression and Random Forest modeling satisfy FR-11 (ML workflow execution). The SHAP analysis satisfies FR-12 (uncertainty quantification) and FR-16 (explainability). This traceability supports the claim that the left side of the V (decomposition) and the right side (integration and verification) remain connected through standardized functions and documents.

4.2 Analytics Module Proof-of-Concept: Overview and Descriptive Results

A note on interpretation is necessary prior to presenting results. The merged analytical panel for this proof-of-concept comprises 14 country-year observations across seven nations and two analytical years. This sample size is an acknowledged constraint of the current implementation, arising from the temporal availability overlap between the WDI annual data and the Meta monthly migration data. This approach was chosen for its manageability in the associated alternatives analysis. Discussion in Chapter 5 addresses QUORUM's case-study analyses in context.

Results shown here represent the QUORUM analytics pipeline as it ingests WDI features for the seven nations of Central America with international migration flow data derived from Meta's dataset (Chi et al., 2025). The analysis proceeds in stages beginning with a descriptive characterization of migration patterns across the study region, followed by a lagged cross-correlation analysis testing the temporal relationship between WDI indicators and outbound migration, a fixed-effects panel regression isolating within-country predictor effects and finally a Random Forest model with Shapley Additive Explanations to identify the relative importance of each environmental and socioeconomic feature. SHAP is a methodology adapted to machine learning which calculates the marginal contributions of each WDI feature and maps the contribution of that feature to the resulting outmigration metrics. Together, these analyses are designed to address thesis research questions 1 through 3, with specific emphasis on RQ3: 'which publicly available indicators illuminate migratory patterns in the Central American Dry Corridor?'.

The study period spans 2019 to 2022, covering 48 months of monthly migration data for all seven countries. The analytical panel merges migration observations with WDI features lagged by one year, consistent with the empirical literature documenting a 3 month to one year delay in response time between environmental stressors and household migration decisions in subsistence agricultural communities (Cai & Oppenheimer, 2013; Niva et al., 2021). The resulting merged panel comprises 14 country-year observations across seven nations and two analytical years (2020-2022). Implications of this sample size for inferential power are discussed in Section 4.5.

4.2.1 Descriptive Analysis of Migration Patterns

QUORUM firstly produces an overview of out-migration in the region. Across the 2019-2022 study period, a total of 2,667,810 outbound migration events were recorded across the seven nations of Central America (Figure 4.1). Guatemala exhibited the highest cumulative outbound volume (n = 721,912), followed by Honduras (n = 616,252), Nicaragua (n = 495,709), and El Salvador (n = 371,552). These four countries collectively accounted for approximately 83% of total regional outbound migration, aligning with their geographic positioning within the

Dry Corridor ecoregion and their documented vulnerability to drought-related agricultural stress (Nakamura et al., 2024; Huber et al., 2023). The contextual assessment here serves as a qualitative validation that the analytics pipeline produces outputs that match the regional circumstances of migration flux.

The United States was the dominant destination across all countries of origin. Outbound flows to the U.S. comprised the majority of outbound volume for Guatemala (76.4%), Honduras (63.2%), and El Salvador (64.2%), a pattern aligned with the gravitational migratory pull of the U.S. / Mexico border corridor noted by Huber et al. (2023) and the Climate Refugees initiative. Intra-regional migration was also revealed, as Nicaragua recorded 125,479 migrants arriving in Costa Rica, representing a second major migratory corridor within the study region.

Nicaragua exhibited a notable behavioral anomaly beginning in mid-2021, with outbound migration surging substantially relative to prior trends. This pattern aligns with the compounded political and economic instability documented in the country during this period and is suggestive of the non-linear "extreme stress" response pattern identified in the climate-migration literature, wherein systemic shocks amplify baseline migration pressure in a non-proportional manner (Cattaneo et al., 2019; Richardson et al., 2023). The following figure is templatized within the AM, and is generated consistently each time.

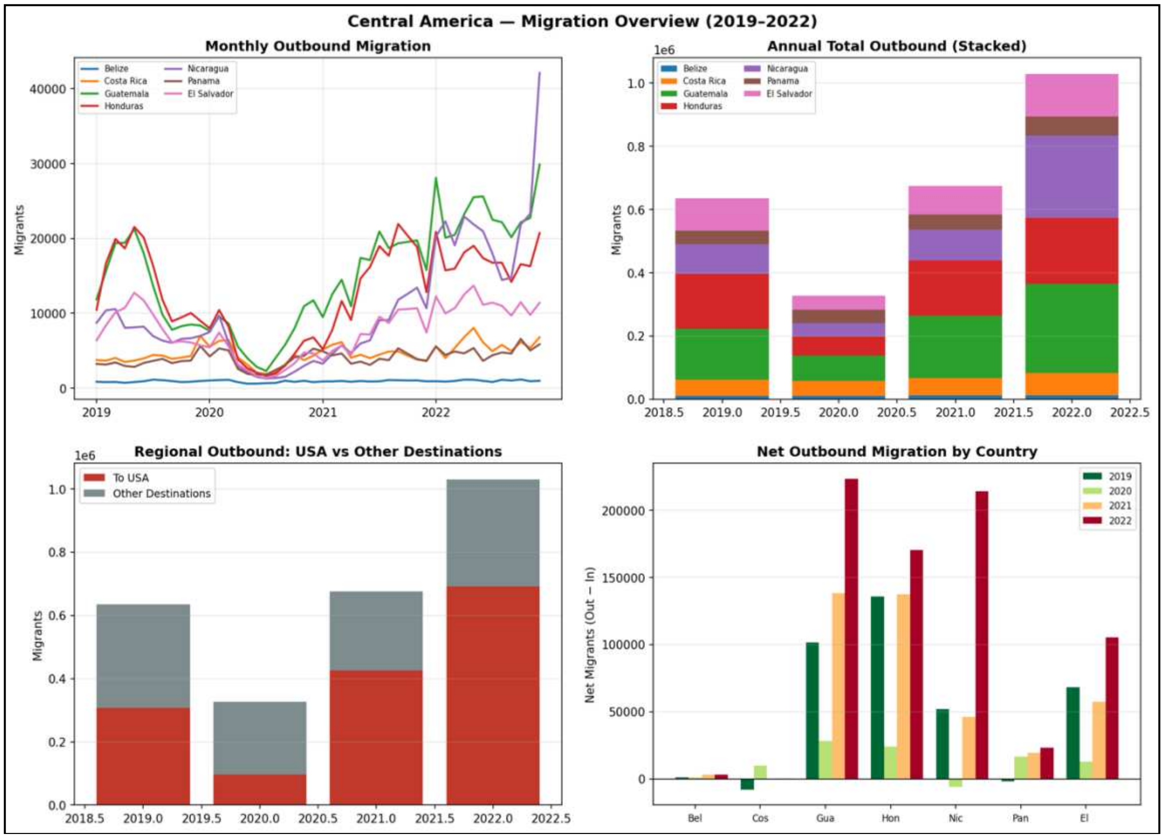

Figure 4.1. *Central American migration overview, 2019–2022. Panels display (clockwise from top left): monthly outbound migration by country; annual outbound stacked by country; net outbound migration (outbound minus inbound) by country and year; and annual outbound disaggregated by destination (U.S. vs. all other). Data source: Chi et al. (2025).*

4.3 Lagged Cross-Correlation Analysis

Comparative analyses within QUORUM reveal the time-based factors of the relationship between WDI environmental and socioeconomic indicators and total outbound migration by computing Pearson and Spearman correlation coefficients (Table 4.1). This was done at varying lag intervals, zero years (contemporaneous), one year, and two years. Both WDI indicators and migration totals were aggregated annually. Fifteen WDI features were evaluated across all three lag conditions, yielding 45 correlation tests. Statistical significance was assessed at $\alpha = .05$.

Table 4.1. *Statistically Significant Lagged Pearson Correlations: WDI Features and Total Outbound Migration. Pearson r coefficients are reported for WDI features that reached statistical significance ($p < .05$) at one or more lag intervals. Fifteen WDI features were evaluated at lag 0 (contemporaneous), lag 1 (one year prior), and lag 2 (two years prior), yielding 45 total tests across the country-year panel. Sample size varies by feature ($N = 7$ to 21) depending on WDI data coverage. WDI lead time indicates the number of years by which the indicator precedes the migration measurement. All p-values are two-tailed. Features that did not reach significance at any lag interval are excluded from this table.*

lagged correlations								
Feature	Lag (years)	Pearson_r	Pearson_p	Spearman_r	Spearman_p	N	Sig_Pearson	Sig_Spearman
Forest Cover (%)	0	-0.406	0.0676	-0.335	0.1376	21		
Energy Intensity	0	0.133	0.6225	-0.193	0.4744	16		
Rural Electricity Access (%)	0	-0.215	0.362	-0.068	0.7774	20		
Electricity Access (%)	0	-0.183	0.4264	-0.107	0.6453	21		
Renewable Energy (%)	0	0.567	0.0274	0.457	0.0867	15	1	
PM2.5 Exposure	0	-0.29	0.5288	0.143	0.7599	7		
GHG per Capita	0	-0.245	0.2836	-0.225	0.3275	21		
Carbon Intensity of GDP	0	0.411	0.0644	0.456	0.0378	21		1
Water Stress (%)	0	0.043	0.8528	0.405	0.0688	21		
Freshwater per Capita	0	-0.544	0.0109	-0.673	0.0008	21	1	1
Net National Savings (% GNI)	0	-0.066	0.8224	-0.051	0.8637	14		
Natural Resource Rents (% GDP)	0	0.272	0.3476	0.609	0.0209	14		1
Basic Water Access (%)	0	-0.242	0.3325	-0.139	0.5814	18		
Rural Sanitation (%)	0	-0.344	0.1625	-0.131	0.6042	18		
Basic Sanitation (%)	0	-0.47	0.0489	-0.044	0.8612	18	1	
Forest Cover (%)	1	-0.52	0.0567	-0.433	0.122	14		
Energy Intensity	1	0.381	0.1795	0.103	0.7253	14		
Rural Electricity Access (%)	1	-0.321	0.2852	-0.261	0.3884	13		
Electricity Access (%)	1	-0.351	0.2186	-0.295	0.3056	14		
Renewable Energy (%)	1	0.745	0.0022	0.613	0.0197	14	1	1
PM2.5 Exposure	1	0.038	0.9349	0.214	0.6445	7		
GHG per Capita	1	-0.336	0.2409	-0.31	0.2809	14		
Carbon Intensity of GDP	1	0.561	0.0371	0.64	0.0138	14	1	1
Water Stress (%)	1	0.048	0.8697	0.376	0.1856	14		
Freshwater per Capita	1	-0.661	0.01	-0.688	0.0065	14	1	1
Net National Savings (% GNI)	1	-0.04	0.8931	-0.037	0.8991	14		
Natural Resource Rents (% GDP)	1	0.461	0.0967	0.71	0.0045	14		1
Basic Water Access (%)	1	-0.358	0.2539	-0.329	0.2969	12		
Rural Sanitation (%)	1	-0.455	0.1371	-0.224	0.4845	12		
Basic Sanitation (%)	1	-0.615	0.0334	-0.266	0.4038	12	1	
Forest Cover (%)	2	-0.665	0.1031	-0.357	0.4316	7		
Energy Intensity	2	0.453	0.3078	0.143	0.7599	7		
Rural Electricity Access (%)	2	-0.464	0.294	-0.179	0.7017	7		
Electricity Access (%)	2	-0.541	0.2095	-0.429	0.3374	7		
Renewable Energy (%)	2	0.843	0.0172	0.75	0.0522	7	1	
PM2.5 Exposure	2	-0.031	0.947	0.143	0.7599	7		
GHG per Capita	2	-0.353	0.4372	-0.393	0.3833	7		
Carbon Intensity of GDP	2	0.518	0.2336	0.643	0.1194	7		
Water Stress (%)	2	0.031	0.9475	0.429	0.3374	7		
Freshwater per Capita	2	-0.65	0.1137	-0.643	0.1194	7		
Net National Savings (% GNI)	2	0.055	0.9065	-0.25	0.5887	7		
Natural Resource Rents (% GDP)	2	0.925	0.0028	0.929	0.0025	7	1	1
Basic Water Access (%)	2	-0.407	0.4229	-0.6	0.208	6		
Rural Sanitation (%)	2	-0.507	0.3041	-0.257	0.6228	6		
Basic Sanitation (%)	2	-0.667	0.1482	-0.486	0.3287	6		

Within the exploratory panel (n = 7 to 21 country-year observations depending on lag and WDI coverage), nine of 45 total tested correlations reached the $p < 0.05$ threshold. Given the small sample sizes involved, these results should be interpreted as suggestive patterns rather than stable estimates. The strongest and most consistent signals involved *Renewable Energy* (% of total final energy consumption), *Natural Resource Rents* (% of GDP), *Freshwater per Capita*, and *Basic Sanitation* access. Results are presented in Appendix Table 1 and visualized in Figure 4.2 which are again templatized within the AM.

QUORUM found the strongest single association to be between *Natural Resource Rents* (% of GDP) lagged two years and total outbound migration ($r = .925$, $p = .003$, $n = 7$). This indicates that countries with higher dependence on extractive natural resource revenues two years prior exhibited substantially greater outmigration. This pattern may be consistent with the resource curse hypothesis and the WEF nexus literature documenting how resource-dependent economies exhibit reduced adaptive capacity to environmental shocks (Daher et al., 2021; Niva et al., 2021), though the small sample size ($n = 7$) warrants caution in interpretation.

Renewable Energy (% of total final energy consumption) demonstrated the most temporally consistent positive association across all three lag periods ($r = .567$ at lag 0, $r = .745$ at lag 1, $r = .843$ at lag 2). This counterintuitive positive feedback loop, in which higher renewable energy share correlates with higher outmigration could likely reflect a confounding country-level effect. This is seen commonly nations with greater renewable capacity (Costa Rica and Panama) also attract higher-volume internal and transit migration, inflating outbound counts. This interpretation is explored further in the fixed-effects analysis (Section 4.4), which controls for baseline country-level differences.

Freshwater per capita and *Basic Sanitation* access both exhibited statistically significant negative associations with outbound migration at a one-year lag ($r = -.661$, $p = .010$; $r = -.615$, $p = .033$, respectively). These patterns appear broadly aligned with the theoretical position that deteriorating WASH (water, sanitation, and hygiene) access contributes to migration pressures, though the panel dimensions limit the strength of inference at this level of data coverage (Daher et al., 2021; Huber et al., 2023).

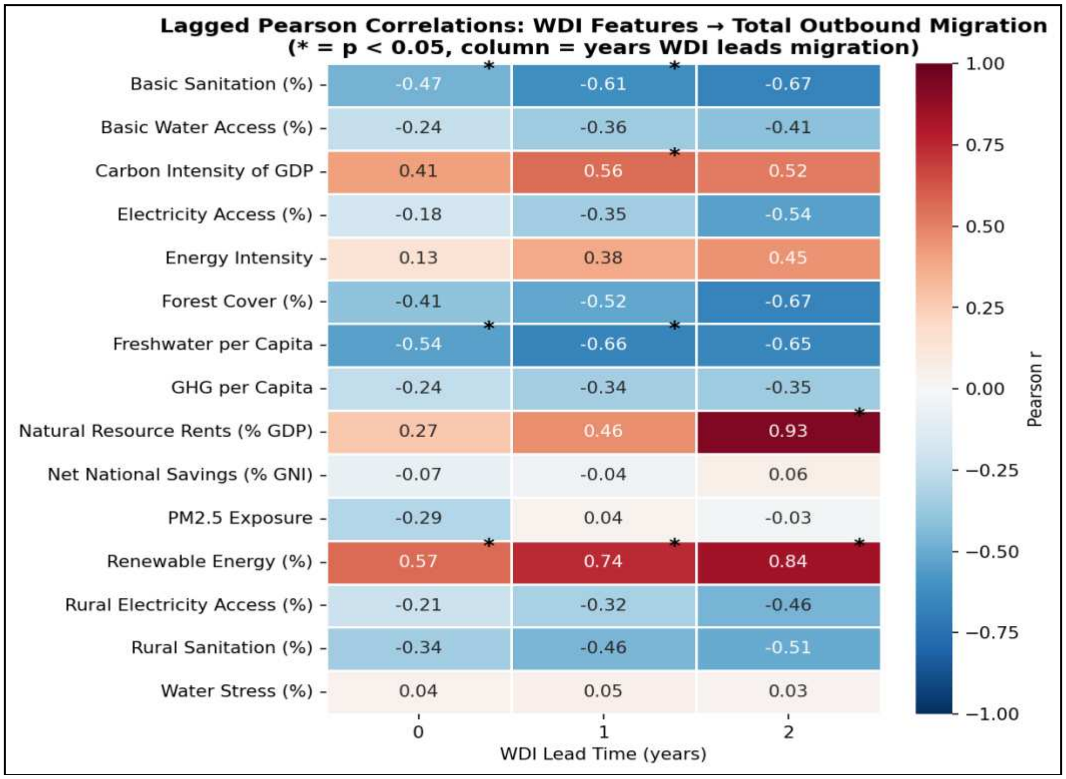

Figure 4.2. *Lagged Pearson correlation heatmap: WDI features versus total outbound migration at lag 0, 1, and 2 years. Asterisks () denote significance at p < .05. Color scale ranges from dark red (r = −1.0) to dark blue (r = +1.0).*

4.4 Fixed-Effects Panel Regression

To isolate intra-country effects of WDI features on migration while controlling for time-independent country-level characteristics (geographic distance to the U.S. border, baseline economic structures of the region), a series of ordinary least squares regressions (OLS) were estimated with country fixed effects modeled as binary indicator variables. Log transformations were then applied to address the right-skewed distribution of migration counts to improve interpretability of the regression coefficients.

Forest Cover (% of land area) was the only feature to achieve statistical significance within the fixed-effects framework ($\beta = -11.53$, $SE = 4.57$, $t = -2.53$, $p = .029$). The negative standardized coefficient indicates that within a given country and controlling for all time-independent country characteristics, a one-standard-deviation decline in *Forest Cover* is associated with an 11.5-unit increase in log outbound migration. This finding aligns with the literature on deforestation-driven livelihood displacement, wherein forest degradation disrupts subsistence agricultural systems and erodes the natural capital base on which rural communities depend (Nakamura et al., 2024; National Academy of Sciences, 2024).

The limited number of significant predictors under the fixed-effects specification relative to the raw correlations reported in Table 4.2 should be expected given the study's panel dimensions. With degrees of freedom constrained by country fixed effects, only predictors with large within-country effect sizes are eligible to achieve significance here. This result does not indicate that other development indicators are disproportionately unimportant but shows the within-country signal is more likely produced by the short observation window. The SHAP-based feature importance analysis (Section 4.5) provides a back projection ranking not subject to this constraint.

Table 4.2. *Fixed-Effects Panel Regression Results: Standardized WDI Coefficients Predicting Log Outbound Migration (Selected Features). Each row represents a separate ordinary least squares regression of log-transformed total outbound migration on a single standardized WDI feature lagged one year, with country fixed effects included. Standardized β coefficients reflect within-country effect sizes after time-invariant country-level characteristics. N = 14 country-year observations across seven nations and two analytical years (2020-2022). WDI features were standardized to μ = 0, σ = 1 prior to estimation. Statistical significance is evaluated at α = .05.*

Fixed effects data panel

Feature	coef	se	p	t	R^2	n	Significant
Freshwater per Capita	-12.5686	7.2919	0.1355	-1.724	0.965	14	FALSE
Forest Cover (%)	-11.5326	4.0349	0.0289	-2.858	0.978	14	TRUE
Rural Sanitation (%)	2.8148	2.5861	0.3182	1.088	0.957	14	FALSE
Basic Sanitation (%)	2.2332	2.1373	0.3364	1.045	0.956	14	FALSE
Basic Water Access (%)	2.1629	1.7044	0.2514	1.269	0.959	14	FALSE
Energy Intensity	-2.1437	1.1905	0.1218	-1.801	0.966	14	FALSE
Water Stress (%)	1.7264	6.6004	0.8024	0.262	0.949	14	FALSE
Renewable Energy (%)	-1.143	0.6912	0.1493	-1.654	0.964	14	FALSE
GHG per Capita	1.1341	0.5976	0.1065	1.898	0.968	14	FALSE
Net National Savings (% GNI)	-0.3584	0.279	0.2463	-1.284	0.959	14	FALSE
Natural Resource Rents (% GDP)	0.2033	0.1025	0.0946	1.983	0.969	14	FALSE
Carbon Intensity of GDP	0.1955	1.2574	0.8816	0.155	0.948	14	FALSE
Electricity Access (%)	0.1845	0.7903	0.8231	0.233	0.948	14	FALSE
Rural Electricity Access (%)	0.108	0.4433	0.8156	0.244	0.949	14	FALSE

Country fixed effects are implemented here as a full set of indicator variables arranged in binary. This is done with one indicator per country, and one omitted as the reference category using the Least Squares Dummy Variable (LSDV) estimator. This is the standard realization of a fixed-effects panel regression and is mathematically equivalent to applying the within-transformation prior to OLS (Wooldridge, 2010). The dummy variables absorb all time-invariant country-level characteristics for the observation window scope. In this case, those characteristics include proximity to the migration destination, long-term economic trends, or governmental structures.

Reported R^2 values in Table 4.2 are relatively high (≈ 0.95 to 0.98) primarily because the country fixed effects absorb substantial between-country variance in baseline migration levels. A specification with country dummies alone and no development indicator features yields a R^2 of comparable magnitude. The R^2 column therefore should not be interpreted as evidence that individual WDI features explain migration. The coefficient, standard error, and p-value columns together (in the context of dataset quality and scale) are the informative quantities to be driving insights from this table. Figure 4.3 plots the standardized β coefficients with their 95% confidence intervals.

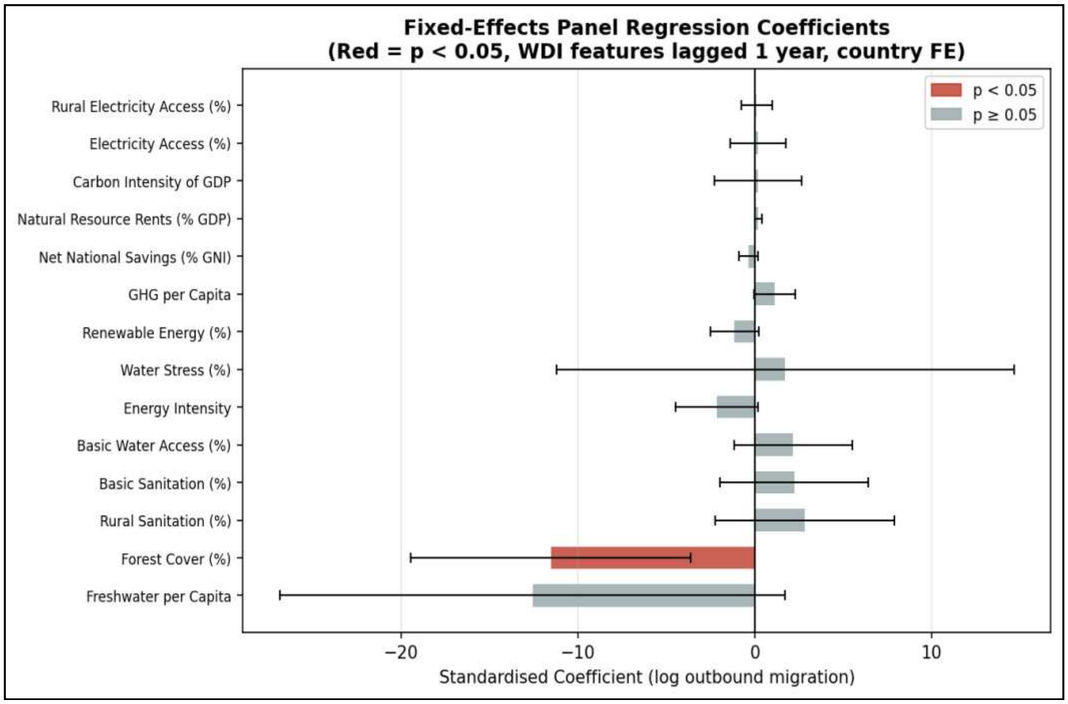

Figure 4.3. *Fixed-effects panel regression coefficient plot. Horizontal bars represent standardized β coefficients; error bars represent ±1.96 standard errors (95% confidence intervals). Red indicates p < .05. WDI features lagged one year; country fixed effects included.*

A specification note here is warranted for explanation and further use. With the proof of concept dataset including only two analytical years per country, a country-fixed-effects regression is algebraically equivalent to a first-differences regression on the seven country-level year-over-year changes. The reported N = 14 reflects total country-year observations, while the effective information content identifying the slope is the set of seven within-country differences. This equivalence is a known property of two-period panels (Wooldridge, 2010, ch. 10) and motivates the cautious interpretive stance taken throughout this use of methodology.

As the fourteen separate feature regressions were estimated, the possibility that one would appear statistically significant by chance alone must be acknowledged. At an α of 0.05, roughly one in twenty independent tests returns a "significant" result under the null. The *Forest Cover* coefficient ($p = 0.029$) is the only one of fourteen to cross this threshold and falls within the range of chance findings for this number of tests. It is therefore reported as a directionally relevant signal consistent with the deforestation-migration literature, not as a concrete predictor.

Formal multiple-testing correction should be deployed (Bonferroni or False Discovery Rate) with future work for an expanded panel, where adjusted thresholds would leave sufficient statistical power to meaningfully apply them.

4.5 Explainability Workflow Demonstration: Random Forest and SHAP

To address QUORUM's research question 3 directly: which data indicators illuminate migratory patterns in the Central American Dry Corridor? The Random Forest regressor exhibited in-sample model fit at (R^2 = .880), suggesting that the feature set captures some variance in migration outcomes within the constrained training data. With this data limited test environment, this model performance is susceptible to memorization of the results, or ‘over-fitting’.

Leave-one-group-out cross-validation was used to best assess out-of-sample over-fitting. LOGO is a cross-validation strategy that serves as an appropriate tool for these panel data with a grouped structure. It tests whether the model can predict migration for a held-out country based on patterns learned from the remaining six (Willekens et al., 2024). The mean LOGO cross-validated R^2 was undefined and effectively negative across folds. This result suggests that the model does not generalize to unseen countries within the current data dimensions and is again an expected consequence of the small panel size (N = 14 across 7 countries, 2 years).

Despite the generalization constraint, Shapley Additive Explanations were computed from the trained model to extract a theoretically interpretable feature importance ranking. SHAP values measure the marginal contribution of each WDI feature to the model's prediction for each observation, averaged across all possible feature configurations. Mean absolute SHAP values are reported below in Table 4.3 and Figure 4.4.

Table 4.3. *SHAP Feature Importance: Mean Absolute SHAP Values Predicting Log Outbound Migration. SHAP values computed from trained Random Forest (500 trees, max depth = 4). Features standardized prior to training. Mean |SHAP| is the average absolute Shapley value across all 14 observations.*

Rank	**Feature**	**Mean │SHAP│**	**Directional Migration Application**
1	*Carbon Intensity of GDP (kg CO_2e per PPP $)*	0.254	Higher carbon intensity → more migration
2	*Freshwater per Capita (m^3)*	0.152	Lower freshwater availability → more migration
3	*Energy Intensity (MJ / $2021 PPP GDP)*	0.076	Higher energy inefficiency → more migration
4	*Renewable Energy (% of total)*	0.067	Positive; likely country-level confound
5	*Natural Resource Rents (% GDP)*	0.066	Higher resource dependence → more migration
6	*Forest Cover (% of land)*	0.058	Lower forest cover → more migration
7	*Water Stress (% of resources)*	0.047	Higher stress → more migration
8	*Basic Sanitation (%)*	0.043	Lower sanitation access → more migration
9	*GHG per Capita (t CO_2e)*	0.038	Higher emissions → more migration
10	*Rural Electricity Access (%)*	0.031	Lower rural access → more migration

Within this exploratory analysis, '*Carbon Intensity of GDP*' emerged as the highest-ranked feature (mean SHAP = 0.254), followed by '*Freshwater per Capita*' (0.152) and '*Energy Intensity'* (0.076). These rankings are produced by the SHAP workflow as a demonstration of how the pipeline attributes feature influences. The pattern is broadly aligned with the WEF nexus stress framework advanced by Daher et al. (2021) and Niva et al. (2021), which posits that migration is driven not by single environmental shocks but by the compounded interaction of resource stress across the WEF system.

Figure 4.4 presents these mean absolute SHAP values as a horizontal bar chart, with features ordered by descending importance. The visualization makes the dominance of *Carbon Intensity of GDP* visually apparent, where its mean SHAP value is nearly twice that of the second-ranked *Freshwater per Capita* and roughly three times the third-ranked *Energy Intensity*. The steep drop-off after the top three features suggests that the Random Forest concentrated its predictive signal within a small subset of the WDI panel, with the remaining features each contributing modestly to the model's predictions. The directionality of each feature's effect,

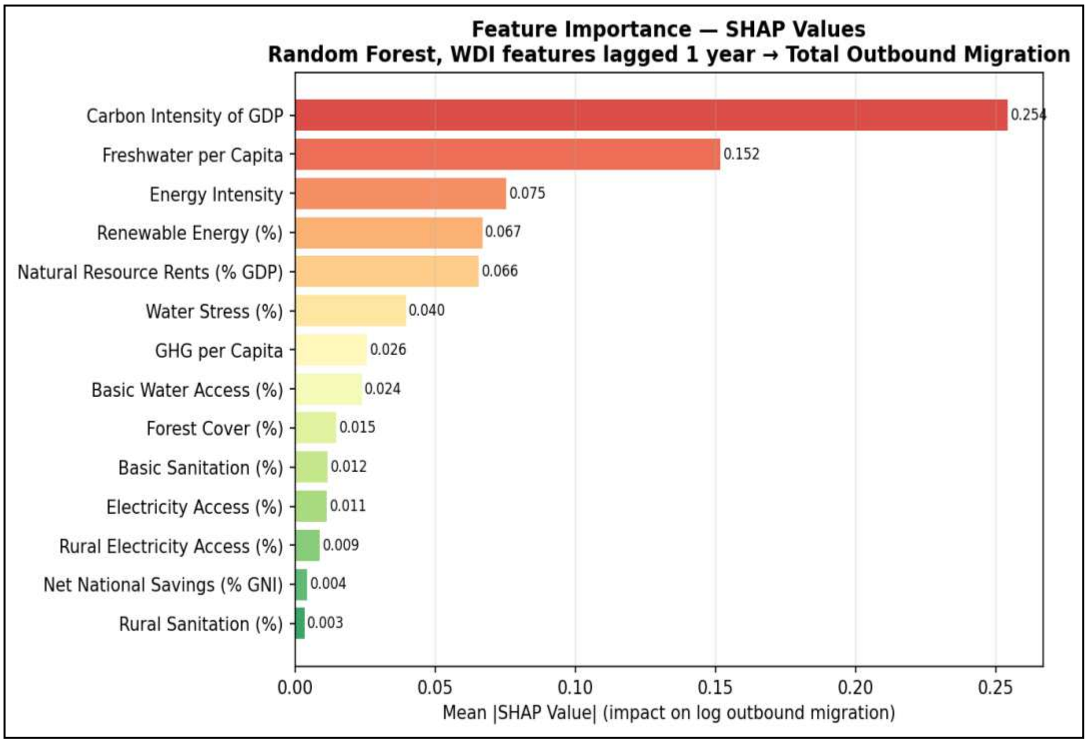

Figure 4.4. *SHAP feature importance (mean absolute SHAP values). Features sorted by descending importance. Larger values indicate greater average marginal contribution to log outbound migration predictions. Random Forest model with WDI features lagged one year.*

summarized in the rightmost column of Table 4.3, indicates that high carbon intensity, high energy intensity, low freshwater availability, and poor sanitation access are all associated with increased migration predictions, directionally aligned with theoretical expectations. *Forest Cover* displays a negative directional effect, aligning with the fixed-effects finding in Section 4.4 and the deforestation and displacement linkage identified by Nakamura et al. (2024).

Figure 4.5 extends importance ranking by displaying partial dependence plots for the four highest-ranked features, illustrating how predicted log outbound migration changes as each feature varies with the remainder held at their average values. The dashed trend line summarizes the overall directional relationship between feature value and outmigration. *Renewable Energy* displays a positive relationship that, as discussed further in Section 5.1, likely reflects a country-level confound rather than a causal mechanism, and bolsters the notion of human-in-the-loop requirement.

Two precautions apply to the partial dependence visualizations. First, the country coloring reveals that the apparent trend lines are heavily anchored by between-country differences rather than within-country temporal variation, a pattern consistent with the limitations discussed in Section 4.4. Another is the partial dependence plots assume feature independence while computing marginal effects, an assumption that is violated to varying degrees by the known interconnected nature of WDI indicators within the WEF nexus. The plots are therefore presented as descriptive visual support for the SHAP rankings rather than causal evidence.

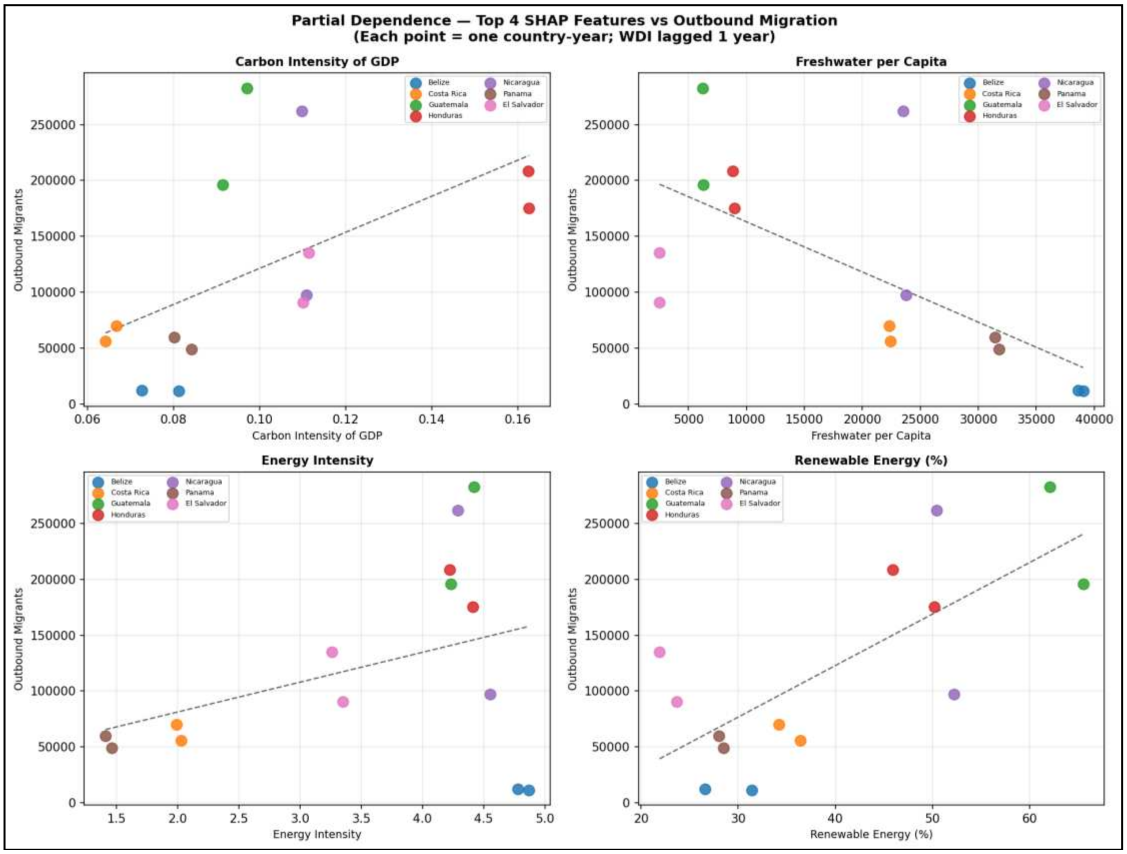

Figure 4.5. *Partial dependence plots for the four highest-importance SHAP features. Each point represents one country–year observation, colored by country. The dashed line is an OLS trend. WDI features lagged one year, dependent variable is total outbound migration.*

4.6 Summary of Findings and Research Question Responses

Systems engineering principles converged within the V-model, modular design, interface control documentation to allow the data science pipeline to perform as intended. It reliably ingested messy open-source inputs, harmonized them, applied lags, and ran the analytics in a traceable, modular workflow. The data are stored, processed, packaged insularly, passed between modules, analyzed, and subsequently presented. The panel architecture within QUORUM's Analytics Module permits new generation of statistical assessment and visualization in one operation. Entry of data requires basic manual data wrangling before upload to the Information Module.

Together, these results suggest that publicly available environmental and socioeconomic indicators, when processed through the QUORUM pipeline, produce outputs broadly aligned with existing field findings. The patterns observed are encouraging as a proof-of-concept, and the pipeline's modular design means that these results can be strengthened substantially as temporal coverage and panel size increase. This thesis presents QUORUM applied to Central America. The modular architecture, governed by the ICD specifications, is designed to accommodate any geographic region with available open-source data. Adaptation to other locations will not require structural changes to any module. The analytical frameworks used to create statistically correlative features to migration provide preliminary support for the defined research question: open-source WDI datasets appear capable of surfacing migration risk indicators with explanatory relevance, though the small panel size limits the strength of this inference. Considering the data constraints of this exploratory analysis, the SHAP output provides a preliminary response to research question 3: carbon intensity, freshwater availability, and energy intensity emerge as the highest-influence indicators in the pipeline's feature attributions within this proof-of-concept panel. These results suggest that the SHAP workflow functions as designed and produces theoretically coherent rankings, rather than establishing these variables as definitive predictors of migration. Deforestation within a country is discussed in relation to the broader literature and QUORUM system design implications in Chapter 5.

The following responses to each research question synthesize the evidence presented in Chapters 3 and 4:

RQ1: How can systems engineering principles be applied to design a transparent and adaptive risk assessment architecture for migration? The V-Model decomposition, modular three-subsystem design, 22 functional requirements with quantifiable performance criteria, and three Interface Control Documents collectively support the position that systems engineering principles provide a viable and valuable framework for structuring humanitarian analytical systems. The proof-of-concept pipeline executed end-to-end without structural code modification between stages, and all analytical outputs are traceable to originating functional requirements. This architectural contribution is the primary result of the thesis.

RQ2: Can widely accessible open-source datasets provide migration correlation with risk indicators with statistical power? Partially supported under constrained panel conditions. Nine of 45 tested correlations reached statistical significance ($p < .05$), and fixed-effects regression identified *Forest Cover* as a statistically significant within-country predictor ($p = 0.029$). However, the small panel size ($N = 14$ country-year observations) limits statistical power, the leave-one-group-out cross-validation produced negative R-squared values indicating that the model does not generalize to unseen countries at this data scale, and degrees of freedom are severely constrained once country fixed effects are included. The evidence suggests that open-source WDI indicators carry information relevant to migration patterns, but statistical power is insufficient for generalizable inference at the current panel dimensions. Expanding to monthly resolution (Section 6.1) would substantially improve this assessment.

RQ3: Which data indicators illuminated migratory patterns in the Central American Dry Corridor for this case example? Within this exploratory analysis, SHAP feature attribution ranked *carbon intensity of GDP*, *freshwater per capita*, and *energy intensity* as the three highest-influence indicators. Fixed-effects regression identified *Forest Cover* as the sole statistically significant within-country predictor. These rankings should be interpreted as preliminary evidence of which WEF dimensions the pipeline identifies as most relevant under current data constraints, not as definitive causal attributions.

Chapter 5
DISCUSSION

5.1 Interpretation of Results in the Context of the Literature

The analytical results presented in Chapter 4 are broadly aligned with several theoretical positions advanced in the climate-migration literature, while also revealing nuances that merit careful interpretation. The results support successful operation of the QUORUM pipeline, as each analytical stage executed as specified, outputs were traceable to originating requirements, and ICD compliance was confirmed at each module boundary. The limited statistical power arising from the 14-observation panel is acknowledged throughout with the results presented as evidence that the architecture functions as designed, not as definitive empirical conclusions.

The SHAP feature rankings presented in Section 4.5 offer preliminary qualitative support for the multi-driver interaction hypothesis. No single WEF sector dominated the top positions, a spread across each sector is consistent with the argument that compounding resource stress across water, energy, and food systems drives displacement in this context (Daher et al., 2021; Niva et al., 2021).

The fixed-effects result identifying *Forest Cover* as the sole statistically significant within-country predictor warrants careful interpretation. While this finding aligns with the deforestation-livelihood displacement literature (Nakamura et al., 2024; National Academy of Sciences, 2024), it should not be read as evidence that other indicators are unimportant in comparison. As previously stated, the panel dimensions impose severe constraints on degrees of freedom once country fixed effects absorb between-country variance. Rigorous regression requires higher degrees of freedom, and under these conditions, it is likely that only predictors with large within-country effect sizes can reach statistical significance. The SHAP analysis, which is not subject to the regression degrees-of-freedom variable, reveals a broader and more granular feature importance landscape that includes *Carbon Intensity*, *Freshwater Availability*, and *Energy Intensity* alongside *Forest Cover* (Figure 4.4).

The four leading indicators identified by the SHAP analysis correspond to specific World Bank World Development Indicators with standardized definitions listed for consistent documentation in Appendix Table 2. *Carbon intensity of GDP* measures carbon dioxide equivalent emissions per unit of economic output in kilograms of CO2e per 2021 dollar of GDP, capturing how carbon reliant an economy's production structure is. Higher values indicate that a country produces more greenhouse gas emissions for each dollar of economic output, reflecting reliance on fossil fuel combustion and carbon-intensive manufacturing (World Bank, 2024). *Renewable internal freshwater resources per capita* measures the total volume of internal river flows and groundwater from rainfall available per person per year in cubic meters, representing the renewable freshwater endowment that sustains agriculture, sanitation, and drinking water systems. Declining values indicate growing water stress as either supply contracts or population increases. *Energy intensity level of primary energy* is the ratio of total primary energy supply to GDP measured at purchasing power parity, expressed in megajoules per 2021 purchasing power parity. This indicates how much energy is consumed to produce one unit of economic output, with higher values reflecting less efficient energy use. The World Bank notes that this indicator serves as an imperfect proxy for energy efficiency, as it can be influenced by factors such as climate and economic structure beyond pure efficiency (World Bank, 2024). *Forest area (%)* measures land under natural or planted stands of trees of at least five meters in distance (whether botanically productive or not) and excludes tree stands in agricultural production systems such as fruit plantations and larger agricultural forestry systems. Changes in forest cover reflect deforestation pressures that directly affect subsistence agricultural systems, watershed integrity, and the natural capital base on which rural communities in the Central America depend.

The counterintuitive positive association between renewable energy share and outbound migration across all three lag periods illustrates the importance of framing one's context. Controlling for country-level confounders even in this proof-of-concept result is necessary. Countries with higher renewable energy capacity in the study region (notably Costa Rica and Panama) also serve as transit and destination hubs that attract and subsequently re-export higher volumes of migrants, inflating their outbound counts. Most persuasively, this observed artifact disappears under the fixed-effects specification, further suggesting that it reflects between-country structural differences rather than a causal relationship between renewable energy and displacement.

These findings carry implications that extend beyond Central America. See et al. (2026) document that humanitarian practitioners in the Philippines, a country experiencing both frequent climate disasters and protracted conflict displacement, recognize the overlapping and compounding nature of displacement drivers but lack the analytical tools to act on that recognition. Their concept of "complex displacement" resonates with QUORUM's empirical results. The structural blind spots that See et al. identify in the humanitarian response system, including trigger-based classification that erases prior displacement history, reliance on operational presence that leaves emerging displacement corridors underserved, and logistical biases that prioritize centralized evacuation centers over dispersed populations, are precisely the types of informational gaps that QUORUM is architected to fill. The "triple nexus" framework (Joireman & Haddad, 2023; OECD, 2024), calls for integrated approaches that bridge short-term humanitarian response with long-term development and peacebuilding objectives. QUORUM's multi-sector analytical pipeline, with its capacity to trace compounding resource stress across water, energy, and food systems over lagged time intervals, offers a data-driven operationalization of this triple nexus vision. Where the HDP Nexus has remained largely a policy aspiration, QUORUM illustrates how systems engineering methodology can translate that aspiration into a traceable, testable analytical architecture that has the potential to operate effectively with iterative design into the future.

Placing this system in the context of existing migration modeling frameworks highlights both what distinguishes QUORUM, and what it can offer. Agent-based models such as those reviewed by Cai and Oppenheimer (2013) simulate individual household decision-making under heterogeneous conditions and are well suited to capturing behavioral dynamics, however they require substantial calibration data and are rarely designed for operational deployment in data-scarce areas. The statistical panel models created by Hoffmann et al. (2021) offer transparency and interpretability but treat migration as the dependent variable in a single equation framework. This approach also ignores the multi-sector nexus integration that characterizes QUORUM's data design. Machine learning approaches reviewed by Best et al. (2022) and Tarraga et al. (2020) achieve strong predictive performance on high-quality data but are susceptible to interpretability and reproducibility problems without a defined workflow. QUORUM occupies a unique position where it is not optimized for any single capability but is instead designed to combine multi-source WEF data ingestion, interpretable lagged analysis, and SHAP-explainability within a modular architecture. This combination is rare in the literature space and is prepared for adoption by under-resourced humanitarian organizations that require analytical power and operational documentation and transparency.

QUORUM most clearly improves on existing modeling efforts in its architectural governance. Models are delivered as research codebase and papers, without formal documentation of data contracts, requirements traceability or validation criteria. The ICD's presented here and functional requirements are design choices that permit new data sources and analytical methodology to evolve over time. Additionally, upgrades can occur modularly without overall system interruption. This design is relevant for organizations operating in contexts where data availability, staff capacity, and computational resources fluctuate, often the case of humanitarian organizations.

5.2 Systems Engineering Contributions and the Value of the V-Model Approach

This systems engineering implementation shows that the V-Model provides a viable and valuable framework for designing humanitarian analytical systems. The structured decomposition from stakeholder needs through functional requirements to subsystem specifications, combined with the verification-validation pairing produces a level of traceability and accountability that is largely absent from existing climate-migration modeling efforts. Each analytical output is traced back to a specific functional requirement, and each functional requirement is shown to trace forward to a validation. The nomenclature presented here will be adapted to future locales, however stay structurally identical for the required maintenance of modularity. The traceability shown here provides the basis for stakeholder trust, a critical requirement for systems operating in low-trust environments prepared for areas where institutional credibility in humanitarian efforts can be scarce.

The Interface Control Documents proved particularly valuable as a design mechanism. By formally specifying the data schema, transformation rules, and acceptance criteria at each module boundary, the ICDs enabled each module to be developed, tested, and validated independently before integration. This modular independence is essential for a system intended to evolve over time, as it allows individual modules to be upgraded or replaced without requiring redesign of the entire pipeline. The successful end-to-end execution of the proof-of-concept pipeline, from raw heterogeneous data ingestion through final SHAP visualization, provides initial validation for this architectural approach.

The systems engineering architecture contributes capabilities beyond what a standard data science workflow would produce. A practitioner could in principle, load the WDI and Meta datasets into a Python notebook, run a Random Forest model, compute SHAP values, and produce similar statistical results. The distinction of QUORUM lies in what the systems engineering framework provides around and beyond that analytical core. First, the formal requirements traceability ensures that every analytical output can be traced to a specific stakeholder need and validated against a specific acceptance criterion, a property that ad hoc analyses lack. Second, the Interface Control Documents create enforceable contracts at module boundaries, meaning that the Information Module can be reconfigured for new data sources (the SPEI monthly panel, the HAPI API stream, or future subnational data products) without modifying or re-validating the Analytics Module. The modular architecture positions the Design Module as a dedicated system component rather than an afterthought, ensuring that stakeholder communication, cultural context, and feedback mechanisms are governed by the same requirements discipline as the statistical modeling. These properties, traceability, interface discipline, documented rationale, and parity between technical and human-facing components, are precisely what distinguishes a systems-engineered analytical framework from an equivalent data science pipeline.

QUORUM's architecture also addresses a critique raised by Correa-Cano et al. (2022) regarding the difficulty of translating complex multi-sector model outputs into accessible stakeholder communication. Where their WEF toolkit acknowledged this as a limitation, QUORUM treats it as a core design requirement through the Design Module's allocation of functional requirements for cultural context (FR-14), stakeholder coordination (FR-15), explainability (FR-16), and inclusive communication (FR-17).

5.3 Data Sovereignty Principles

QUORUM operationalizes data sovereignty principles through deliberate architectural decisions embedded across all three modules. At the Information Module level, the system enforces FAIR compliance by maintaining complete metadata lineage for every ingested data element, documenting source provenance, transformation history, and coverage flags through the data_note column in the harmonized panel. Every dataset used in the proof-of-concept implementation is drawn from publicly accessible repositories, ensuring that the analytical pipeline can be independently audited and reproduced without institutional gatekeeping.

CARE principles are addressed through the Design Module's architecture, which positions affected communities not merely as subjects of analysis but as intended co-interpreters of system outputs. The multilingual communication capability, cultural context, and human-in-the-loop feedback mechanism are designed so that communities retain interpretive authority over the risk assessments generated about their regions. The system is explicitly designed so that local knowledge can validate, challenge, or contextualize analytical outputs, preventing the scenario in which externally generated data products override community-level understanding. A local understanding of saliency in design choices for the adaptable DM workflow and presentation of information will help to satisfy these functional requirements and extend cognitive impact (Lee,Wickens, Liu & Boyle, 2017).

The architectural choice to rely on open-source data rather than proprietary mobility datasets for the primary analytical pipeline also reflects a sovereignty commitment. While Meta's migration data serves as a validation benchmark, the core QUORUM workflow suggests that meaningful risk indicators can be derived from World Bank, UN, and WFP datasets that any government, NGO, or community organization can access independently. This design principle ensures that the communities most affected by climate-related displacement are not dependent on the continued goodwill or data-sharing policies of multinational technology corporations for information critical to their own resilience planning. During the design and execution of this thesis, the CIA World Factbook was removed (an archived 2020 version is available on GitHub), NOAA has been threatened to be dissolved, and USAID has faced severe budget cuts from the US Congress. These highly used and important database housings are under threat of extinction, and highlight what open-source and available datasets we use now but may be at risk moving into the future.

Taken together, these properties reflect QUORUM's systemic contribution: a replicable transparent analytical architecture that can be implemented, validated, and expanded.

5.4 Limitations

Several limitations qualify the conclusions drawn from this work and define the boundaries of appropriate inference and effective system outputs. The most consequential is the small panel size. With 14 country-year observations across 7 countries and 2 analytical years, the study's statistical power is constrained, particularly for the fixed-effects regression where country placeholders absorb substantial variance. The negative cross-validated R-squared from the leave-one-group-out analysis confirms that the Random Forest model does not generalize to unseen countries within the current data dimensions. This result should be interpreted as a sample-size constraint rather than a failure of the feature relationships, consistent with Hoffmann et al.'s (2021) observation that quantitative climate-migration models routinely face generalization challenges when panel dimensions are short.

The reliance on annual WDI data for the primary analytical panel limits temporal resolution and may mask seasonal dynamics that are operationally important for higher resolution accuracy. Agricultural communities in the Dry Corridor respond to seasonal precipitation anomalies on timescales of months, not years, and the annual aggregation required to align WDI data with Meta's migration observations likely attenuates signals that would be stronger at finer temporal resolution. The monthly environmental pipeline (incorporating SPEI drought indices and WFP food prices) represents a step toward addressing this limitation, but full integration with monthly migration data remains a future objective.

Finally, the Design Module, while architecturally specified and allocated functional requirements, has not yet been implemented or tested with end users. The human-systems integration and behavior change communication capabilities that distinguish QUORUM from purely technical modeling efforts remain at the design stage. Validating these capabilities through participatory engagement with affected communities represents a critical step forward. Implementation of the Design Module will require Institutional Review Board (IRB) approval and field validation with human subjects, placing it outside the scope of this thesis but positioning it as a priority for future work.

To state this boundary clearly this thesis presents a proof-of-concept implementation of the Information Module and Analytics Module, validated through end-to-end pipeline execution on the WDI annual panel with Meta migration data for seven Central American countries. This distinction between demonstrated capability and specified architecture is intentional and consistent with V-Model practice, where left-side decomposition (specifying what the system must do) can proceed ahead of right-side integration (confirming that it does so) for subsystems not yet under active development.

5.5 Decision-Maker Context and Policy Implications

The analytical outputs produced by QUORUM are designed to serve distinct decision-maker tiers, adaptable to different scales of authority and with different information requirements.

At the community level, local leaders and municipal authorities represent the first tier of decision-makers. QUORUM outputs such as the SHAP-attributed risk indicators inform community-level adaptation planning, including decisions about crop diversification, water harvesting infrastructure, or the timing of seasonal labor migration. The Design Module's plain-language risk communication is specifically architected with functional requirements to serve this group (Section 1.7.1).

Humanitarian organization analysts at agencies such as UNHCR, IOM, and WFP constitute the second tier. These actors allocate resources across multiple countries and communities requiring comparative risk assessments to set priorities in local operations. QUORUM's cross-country panel structure and exportable data formats directly support this comparative monitoring workflow. For this tier, the identification of *Forest Cover* as the strongest within-country predictor suggests that reforestation programs, payments for ecosystem services, and agricultural land-use regulation represent policy-relevant intervention points that could be monitored through the pipeline's temporal tracking capabilities.

National government civil protection agencies form the third tier. These decision-makers set macro-level policy, allocate disaster preparedness budgets, and coordinate international development assistance. The SHAP results identifying *Carbon Intensity of GDP* and *Energy Intensity* as high-ranking features point toward energy transition policy, industrial diversification strategies, and infrastructure investment as structural interventions that address the economic dimensions of displacement pressure. The scenario exploration workflow defined in Section 1.7.3 is designed to support this tier's need for what-if analysis under varying policy landscapes.

Across all tiers, QUORUM outputs should be understood as inputs to existing decision frameworks rather than standalone policy directives. The exploratory nature of this current proof-of-concept, combined with the acknowledged panel size limitations, means that these indicators identify areas warranting further investigation and targeted data collection rather than prescribing specific policy responses. As the system's temporal coverage and geographic resolution expand (Section 6.1), the precision and actionability of these policy connections will strengthen correspondingly. The expansion of QUORUM is effectively limitless, as local sensor networks or IoT device integration can readily provide a useful granularity that tightens any outputs into insights for specific regions or industries depending on stakeholder scope.

5.6 Ethical Considerations

The development of migration risk assessment systems raises ethical questions that extend beyond data governance. Models capable of identifying communities likely to experience displacement could, if misapplied, be used for predatory or deterrence purposes rather than humanitarian support. QUORUM's design mitigates this risk by aggregating risk indicators at the country level rather than community-level tracking, obscuring views into targeting. The FAIR and CARE governance framework requires that analytical outputs serve the interests of affected communities, and the human-in-the-loop feedback mechanism ensures that community voices are embedded in the validation process. However, the potential for dual use of migration prediction systems must be acknowledged as an ongoing ethical concern that requires vigilance as the system progresses and involves deployment.

Chapter 6
FUTURE DIRECTIONS

6.1 Expanding Temporal and Spatial Resolution

The most immediate priority for QUORUM's analytical output is the expansion of temporal resolution from annual to monthly analytical panels. The monthly environmental pipeline, which integrates SPEI drought indices and WFP food price data at monthly intervals, provides the foundation for this transition. Merging these monthly environmental indicators with Meta's monthly migration observations at configurable lag offsets (1, 2, 3, 6, 9, and 12 month windows) would substantially increase the panel size, improve statistical power, and enable detection of seasonal dynamics that annual aggregation currently obscures. This expansion is critical for achieving mathematical resolution to drive the most influential insights as possible, a central objective of QUORUM's use case.

Spatial resolution improvements represent a complementary priority. The current country-level analysis treats each nation as a homogeneous unit, masking any more granular variation in both environmental stress and migration response. Future iterations should incorporate subnational administrative units, particularly for Guatemala and Honduras where the Dry Corridor's geographic footprint does not cover the entire national territory. A well coordinated IoT sensor network placed across the CADC would serve as a useful data collection input vector for the Information Module in QUORUM applications without public access to outputs (aforementioned security concern). Current packaged climate data (such as the SPEI data already integrated in the QUORUM IM pipeline) provide the spatial granularity needed for this disaggregation, and the ICD-01 architecture is designed to accommodate additional API data streams without modification to downstream modules, ensuring that this appropriate expansion will have enough data for real-world application.

6.2 Enhanced Analytical Methods

The current analytical pipeline uses a Random Forest regressor with SHAP-based feature attribution. While this approach is well-suited for the proof-of-concept phase due to its interpretability and robustness to small sample sizes, future iterations should explore additional modeling approaches. Gradient boosting methods (such as XGBoost or LightGBM) may improve predictive performance on expanded panels. Bayesian hierarchical models would provide a natural framework for partially pooling information across countries while respecting country-specific heterogeneity, an approach particularly well-suited to the small-N, multi-group structure of Central American migration data.

The meta-algorithmic frameworks described by Simske (2013) offer a principled approach to combining multiple analytical methods within QUORUM's modular architecture. Rather than selecting a single best model, meta-algorithmic patterns enable the parallel deployment of multiple models whose outputs are combined through structured decisioning. This approach aligns with QUORUM's existing emphasis on methods and could improve robustness by leveraging the complementary strengths of different analytical approaches. Specifically, a meta-algorithmic sensitivity analysis could be applied to identify which analytical methods perform best under different data conditions, enabling adaptive model selection as new data sources are integrated. This avenue has promise to mimic high performing systems at a fraction of the cost.

The integration of agent-based modeling capabilities would complement the current statistical approach by enabling simulation of household-level migration decisions under different climate scenarios. This extension would support the scenario exploration functionality allocated to the Design Module and provide a mechanism for testing policy interventions in a computational test-bed before real-world deployment.

6.3 Design Module Implementation and Community Validation

The Design Module represents the most substantial remaining development effort. While the module's functional requirements and interface specifications (ICD-03) have been defined in Section 3.2, implementation and testing with actual stakeholders will not occur for the case study. Because the Design Module requires direct engagement with human subjects across vulnerable populations, its implementation will require IRB approval and structured field validation protocols before deployment. Priority development tasks include the construction of a web-based dashboard that presents SHAP-attributed risk indicators with interactive drill-down education capability, multilingual output generation (minimally Spanish and indigenous languages relevant to the Dry Corridor), and the human-in-the-loop feedback interface that allows community operators to annotate, validate, or challenge system outputs.

Community validation will proceed through structured participatory workshops with stakeholder groups like migration-candidate communities in the Dry Corridor, humanitarian organizations operating in the region (such as WFP and UNHCR field offices), and local government civil protection agencies. These workshops test the usability and comprehensibility of system outputs as well as the cultural appropriateness of visualization choices, the perceived trustworthiness of the risk communication, and the actionability of the information provided. Findings from these workshops are structurally fed back into the requirements decomposition process, iterating the V-Model to refine the system based on real stakeholder input in its very next iteration.

The operational workflows for each stakeholder group are defined in the Concept of Operations (Section 1.7), which specifies the risk inquiry sequence for community leaders, the comparative monitoring sequence for humanitarian analysts, and the scenario exploration sequence for government civil protection agencies. Translating these operational specifications into implemented, tested interface components represents a primary deliverable for future application. Detailed task flow diagrams, cognitive walkthrough protocols, and usability test scripts should be developed in conjunction with representative users from each stakeholder group before interface design proceeds.

6.4 Generative AI Integration

Current data ingestion and wrangling within the Information Module requires manual preprocessing with data science programming tools. Advances in generative artificial intelligence (genAI) present opportunities to automate portions of this workflow, particularly for tasks such as format detection, schema mapping, missing value imputation, and natural language documentation generation. A future iteration of QUORUM could incorporate a genAI-assisted data steward that ingests raw files, identifies their structure, proposes transformation rules aligned with ICD-01 specifications, and generates human-readable provenance documentation, all subject to human review and approval before data enters the analytical pipeline.

GenAI capabilities could also enhance the Design Module by enabling natural language query interfaces that allow non-technical stakeholders to interrogate the system outputs in plain or tailored language. Rather than navigating a dashboard, a community leader could ask a question such as "What are the biggest risks for my region this season?" and receive a contextually appropriate, SHAP-attributed response. This capability would meaningfully lower the barrier to system adoption in communities where information literacies vary considerably.

6.5 Geographic Expansion and Transferability

While this thesis focuses on the Central American Dry Corridor as a bounded empirical context, QUORUM's modular architecture is designed to be location-agnostic. The ICD-01 data ingestion protocol accommodates any geographic region for which open-source environmental and socioeconomic indicators are available. Priority regions for expansion include the Sahel corridor of West Africa, the Horn of Africa, and South Asian river delta regions, all of which exhibit the compounded WEF nexus stressors and climate-migration dynamics that QUORUM is designed to assess.

Geographic expansion will require recalibration of the analytical pipeline for region-specific dynamics, including different lag structures between environmental stress and migration response, different dominant migration corridors, and different socioeconomic mediating factors. The modular architecture supports this recalibration by isolating region-specific parameters in the Information Module and ICD-01 configuration while preserving the analytical and design modules' structural integrity. This transferability represents one of the primary contributions of the systems engineering approach by separating regional data and context from analytical methods and interface design. QUORUM provides a reproducible template for climate-migration risk assessment that can be adapted to diverse contexts without fundamental redesign.

REFERENCES

Anthropic. (2025). *Claude* (Opus 3 version) [Large language model]. https://claude.ai

Baltagi, B. H. (2014). Hausman-Test - an overview | ScienceDirect Topics. Www.sciencedirect.com. https://www.sciencedirect.com/topics/social-sciences/hausman-test

Besinga, L. E., & Theophilus. (2025). Technological Innovations for Climate Adaptation and Peacebuilding: A Holistic Approach to Resource Conflict and Environmental Challenges. Open Journal of Applied Sciences, 15(01), 285–304. https://doi.org/10.4236/ojapps.2025.151019

Best, K. (2019). A Data-Driven Analysis of Environmental Migration in Coastal Bangladesh. Vanderbilt University. https://ir.vanderbilt.edu/available/etd-07172019-112002/

Best, K., Gilligan, J. M., Baroud, H., Carrico, A. R., Donato, K. M., & Mallick, B. (2022). Applying machine learning to social datasets: A study of migration in southwestern Bangladesh using random forests. Regional Environmental Change, 22(2). https://doi.org/10.1007/s10113-022-01915-1

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.

Brown, A., & De Regil Sánchez, I. L. (2021, December 22). *Structure and outcomes of a virtual international humanitarian engineering collaborative program with workshop.* ResearchGate. https://www.researchgate.net/publication/357299066

Cai, R., Feng, S., Oppenheimer, M., & Pytlikova, M. (2016). Climate variability and international migration: The importance of the agricultural linkage. Journal of Environmental Economics and Management, 79, 135–151. https://doi.org/10.1016/j.jeem.2016.06.005

Cai, R., & Oppenheimer, M. (2013). Agent-based modelling of climate-migration linkages: A review and a proposal. Institute for the Study of Labor (IZA), Discussion Paper No. 7592.

Carpenter, J., & Robbett, A. (2022). Game Theory and Behavior. MIT Press.

Cattaneo, C., et al. (2019). Human migration in the era of climate change. Review of Environmental Economics and Policy, 13(2). https://doi.org/10.1093/reep/rez008

Chi, G., Abel, G., Johnston, D., Giraudy, E., & Bailey, M. (2025). Measuring global migration flows using online data. Proceedings of the National Academy of Sciences, 122(18). https://doi.org/10.1073/pnas.2409418122

Conrad, S., Geisenhoff, J., Brueck, T., Volna, M., Brink, P., Hall, M., Cook, S., & Kenway, S. (2011). *Decision support system for sustainable energy management.* Water Research Foundation.

Couldry, N., & Mejias, U. A. (2019). The costs of connection: How data is colonizing human life and appropriating it for capitalism. Stanford University Press.

Curtis, A., Mills, J. W., & Leitner, M. (2007). Katrina and vulnerability: The geography of stress. Journal of Health Care for the Poor and Underserved, 18(2), 315–330. https://doi.org/10.1353/hpu.2007.0029

Duijndam, S., et al. (2025). Global determinants of coastal migration under climate change. Nature Communications, 16(1), 6866. https://doi.org/10.1038/s41467-025-59199-y

Gaikwad, S. S., Iyer, S., Lunga, D., & Bondi, E. (2021). Data-driven Humanitarian Mapping: Harnessing Human-Machine Intelligence for High-Stake Public Policy and Resilience Planning. https://doi.org/10.1145/3447548.3469461

García, J. M. R., & Valencia, J. E. R. (2024). The Ripple Effects of Climate Change on Migration Patterns. IntechOpen. https://doi.org/10.5772/intechopen.1004698

Haugen, S. (2026). *QUORUM: Quantitative Understanding of Regional Outmigration Using Multivariate data* (v1-v2) [Python program]. GitHub. https://github.com/stefanhaugen/quorum

Hoffmann, R., Šedová, B., & Vinke, K. (2021). Improving the evidence base: A methodological review of the quantitative climate migration literature. Global Environmental Change, 71, 102367. https://doi.org/10.1016/j.gloenvcha.2021.102367

Huber, J., et al. (2023). Climate-related migration and the climate-security-migration nexus in the Central American Dry Corridor. Climatic Change, 176(6). https://doi.org/10.1007/s10584-023-03549-6

IASC. (2010). IASC Framework on Durable Solutions for Internally Displaced Persons. Inter-Agency Standing Committee.

IDMC. (2023). Global Report on Internal Displacement. Internal Displacement Monitoring Centre.

IDMC. (2024). Global Report on Internal Displacement 2024. Internal Displacement Monitoring Centre.

IDMC & NRC. (2015). Understanding the Root Causes of Displacement: Towards a Comprehensive Approach to Prevention and Solutions. Internal Displacement Monitoring Centre.

IPCC. (2023). Climate Change 2023: Synthesis Report. Intergovernmental Panel on Climate Change.

Johnson, A. F., Lamontagne, N. D., Glickson, D., & Luhachack, L. (2024). Climate Change and Human Migration: An Earth Systems Science Perspective. National Academies Press.

Johnson, R. A., Miller, E. E., & Conrad, S. (2022). Technology adoption and acceptance of urban air mobility systems: Identifying public perceptions and integration factors. *The International Journal of Aerospace Psychology, 32*(4), 240–253. https://doi.org/10.1080/24721840.2022.2100394

Joireman, S. F., & Haddad, F. (2023). The Humanitarian-Development-Peace Nexus in practice: Building climate and conflict sensitivity into humanitarian projects. Current Opinion in Environmental Sustainability, 62, 101272. https://doi.org/10.1016/j.cosust.2023.101272

Kossiakoff, A., Sweet, W. N., Seymour, S. J., & Biemer, S. M. (2020). Systems Engineering Principles and Practice (3rd ed.). Wiley.

Lee, J. D., Wickens, C. D., Yili Liu, & Linda Ng Boyle. (2017). Designing for people : an introduction to human factors engineering (3rd ed.). Createspace.

Lenton, T. M., et al. (2023). The Global Tipping Points Report 2023. University of Exeter.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, 2(1), 56-67.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30.

Lynch, R., et al. (2023). 'The tears don't give you funding': Data neocolonialism in development in the Global South. Third World Quarterly, 44(5), 911. https://doi.org/10.1080/01436597.2023.2166482

McAuliffe, M., & Triandafyllidou, A. (2023). World Migration Report 2022. International Organization for Migration.

McKee, T. B., Doesken, N. J., & Kleist, J. (1993). The relationship of drought frequency and duration to time scales. In Proceedings of the 8th Conference on Applied Climatology (Vol. 17, No. 22, pp. 179–183). American Meteorological Society.

Mejia, V. D. B., et al. (2022). Efectos del Cambio Climático en Centroamérica. Revista Iberoamericana de Bioeconomía y Cambio Climático, 8(16). https://doi.org/10.5377/ribcc.v8i16.15227

Molnar, C. (2020). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. https://christophm.github.io/interpretable-ml-book/

Monteiro de Carvalho, C. (2026). Participatory Mapping as a Tool for Social Justice and Mobilization in Brazilian Informal Settlements. Media and Communication, 14. https://doi.org/10.17645/mac.11175

Nair, R., Madsen, B. E., & Kjærum, A. (2023). An Explainable Forecasting System for Humanitarian Needs Assessment. Proceedings of the AAAI Conference on Artificial Intelligence, 37(13), 15569. https://doi.org/10.1609/aaai.v37i13.26846

Nakamura, J., et al. (2024). Recent trends in agriculturally relevant climate in Central America. International Journal of Climatology, 44(8), 2701. https://doi.org/10.1002/joc.8476

National Academy of Sciences. (2024). Climate Change and Human Migration. National Academies Press. https://doi.org/10.17226/27930

Nguya, G., & Siddiqui, N. (2020). Triple nexus implementation and implications for durable solutions for internal displacement: On paper and in practice. Refugee Survey Quarterly, 39, 466–480.

Niva, V., et al. (2021). Global migration is driven by the complex interplay between environmental and social factors. Environmental Research Letters, 16(11), 114019. https://doi.org/10.1088/1748-9326/ac2e86

OECD. (2024). The Humanitarian-Development-Peace Nexus and Forced Displacement: Progress, Insights and Recommendations for Operational Practice. OECD Publishing. https://www.oecd.org/en/publications/the-humanitarian-development-peace-nexus-and-forced-displacement_3e493170-en.html

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O. & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825-2830.

Peters, K., Gray-Meral, A., Opitz-Stapleton, S., & Measures, H. (2021). Climate Change, Conflict and Displacement: Implications for Protection Agencies. ODI.

Probst, P., Wright, M. N., & Boulesteix, A.-L. (2019). Hyperparameters and tuning strategies for random forest. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(3), e1301.

Ramírez-Márquez, J. E., & Ponce-Ortega, J. M. (2023). Process systems engineering tools for the sustainable management of the water–energy–food nexus. Current Opinion in Chemical Engineering, 41, 100952.

Reimann, L., et al. (2023). Exploring spatial feedbacks between adaptation policies and internal migration patterns due to sea-level rise. Nature Communications, 14(1), 2630. https://doi.org/10.1038/s41467-023-38278-y

Richardson, K., et al. (2023). Earth beyond six of nine planetary boundaries. Science Advances, 9(37). https://doi.org/10.1126/sciadv.adh2458

See, S., Opdyke, A., & Banki, S. (2026). Navigating the overlap of climate-induced and conflict-induced displacement: Perspectives from humanitarian practitioners in the Philippines. npj Climate Action, 5, 8. https://doi.org/10.1038/s44168-025-00326-2

Shannak, S., Mabrey, D., & Vittorio, M. (2018). Moving from theory to practice in the water–energy–food nexus: An evaluation of existing models and frameworks. Water-Energy Nexus, 1(1), 17–25. https://doi.org/10.1016/j.wen.2018.04.001

Simpson, N. P., et al. (2021). A framework for complex climate change risk assessment. One Earth, 4, 489–501. https://doi.org/10.1016/j.oneear.2021.03.005

Simske, S. J. (2013). *Meta-algorithmics: Patterns for robust, low-cost, high-quality systems.* Wiley-IEEE Press.

Tarraga, J. M., Piles, M., & Camps-Valls, G. (2020). Learning drivers of climate-induced human migrations with Gaussian processes. arXiv. https://doi.org/10.48550/arxiv.2011.08901

Tower, A. (2021). Central American Climate Migration is a Human Security Crisis. The Center for Climate and Security. https://www.climate-refugees.org/advocacy/central-america

Vicente-Serrano, S. M., Gracia, F. R., Beguería, S., & Pajuelo, D. P. (2025). An Optimal and Flexible Approach for Drought Quantification Based on Standardised Indices. International Journal of Climatology, 45(10). https://doi.org/10.1002/joc.8906

Weerasinghe, S. (2021). Bridging the Divide in Approaches to Conflict and Disaster Displacement: Norms, Institutions and Coordination in Afghanistan, Colombia, Niger, the Philippines and Somalia. UNHCR.

Willekens, K., et al. (2024). Migration Modeling Research Landscape: Current Methods and Critical Knowledge Gaps.

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data (2nd ed.). MIT Press.

World Bank. (2024). World Development Indicators. https://databank.worldbank.org/source/world-development-indicators

Yar, M. A., Hamdan, M., Anshari, M., Fitriyani, N. L., & Syafrudin, M. (2024). Governing with Intelligence: The Impact of Artificial Intelligence on Policy Development. Information, 15(9), 556. https://doi.org/10.3390/info15090556

APPENDIX

Appendix Table 1. *QUORUM's Functional Requirements. Functional requirements mapped by module, function and description. Performance criteria are as quantifiable as is reasonable and require stakeholder engagement before solidification and deployment.*

Code	Module	FR Name	Description	Performance Criteria
FR-1	IM	Acquire environmental data	Procure and access open-source environmental datasets from OCHA, UN, World Bank, and other databases	Data validated by benchmarking
FR-2	IM	Acquire socioeconomic data	Obtain food pricing, economic indicators, population dynamics, and infrastructure data	Open-source, reliable output schedule, widely accepted data collection procedures.
FR-3	IM	Acquire migration data	Access Meta/Facebook mobility data or other migration datasets for validation	Temporal resolution <=1 month
FR-4	IM	Manage data access	Establish and maintain data sharing agreements, APIs, and access protocols	FAIR/CARE compliance
FR-5	IM	Validate data quality	Assess completeness, accuracy, and reliability of incoming datasets	Missing data <10%
FR-6	IM	Clean and transform data	Remove noise, handle missing values, standardize formats	Data loss during cleaning <5%
FR-7	IM	Organize data storage	Structure data in optimized formats for analysis and retrieval	Query response for tidy data time <2 sec
FR-8	IM	Generate documentation	Create metadata, lineage tracking, and data dictionaries	Documentation completion 100% across ICDs

FR-9	AM	Feature engineering	Extract features, calculate time-lagged correlations, engineer predictive variables	Feature relevance score > 0.3. Mutable by stakeholder agreement.
FR-10	AM	Detect WEF stressors	Identify water, energy, food system stress indicators and threshold violations	Detection sensitivity >80% and all data metrics published.
FR-11	AM	Execute ML workflows	Train, validate, and deploy ML models for migration risk prediction	R-squared >0.7 or RMSE acceptable
FR-12	AM	Quantify uncertainty	Calculate prediction confidence intervals and model uncertainty	Calibrated estimates (90% CI)
FR-13	AM	Analytical output	Identifies correlative indicators and panel outputs complete	Full AM feature set covered, defined formally in ICD
FR-14	DM	Cultural Awareness	Ensure all system outputs are presented with culturally appropriate framing, local language options, and context-relevant visualization for target region communities	Output in minimum 3 languages of region; cultural review prior to deployment (human-in-loop)
FR-15	DM	Coordinate Stakeholder Communication	Enable structured data export in formats compatible with humanitarian coordination platforms; maintain ICD-traceable audit trail from displayed indicators to source data	Export validated against ICD-03 schema; audit trail completeness 100%
FR-16	DM	Explainability and Visualization	Translate all SHAP-attributed analytical outputs into stakeholder-appropriate visualizations with plain-language directional interpretations for each ranked feature; support scenario exploration workflows for	All top-ranked SHAP features presented with plain-language directional interpretation; 90% confidence intervals displayed for scenario outputs; community-level workflow. Linked educational tools present and jargon

			government users with confidence intervals displayed alongside point estimates	vocabulary definitions outlined
FR-17	DM	Inclusive Visualization	Deliver system outputs in a minimum of two visualization modalities with audience appropriate framing; support human-in-the-loop annotation at any point in a workflow so communities can flag outputs inconsistent with local knowledge	Visual outputs validated by cultural review prior to deployment; annotation events logged at 100% and routed to validation pipeline
FR-18	DM	Scenario Exploration	Enable what-if analysis by allowing users to adjust input parameters and observe model response; prevent deterministic misinterpretation by displaying confidence intervals alongside all point estimates	Scenario interface prevents fixed-point forecasts; all scenario outputs carry 90% CI; interface requires no technical vocabulary for community-level task flows
FR-19	SI	End-to-End Pipeline Execution	Execute the full pipeline from raw data ingestion through AM visualization outputs without structural code modification between stages; modular independence confirmed by independent re-run of AM from saved ABIB	Zero structural modifications required between pipeline stages; AM re-runnable from saved ICD-02 output without re-executing IM
FR-20	SI	Requirements Traceability	Maintain documented mapping from each analytical output to its originating functional requirement and stakeholder need; traceability matrix	100% of analytical outputs traceable to at least one FR; traceability matrix complete and version-controlled

			current at each delivery milestone	
FR-21	SI	Operational Security and Access Control	Restrict all migration data public outputs to aggregate country-level data at minimum resolution; prohibit individual-level tracking, re-identification, or any output that could support surveillance or enforcement use cases	No individual-level data exposed in any output; aggregate-only constraint confirmed by data audit at each pipeline execution
FR-22	SI	ICD Compliance Verification	Verify that all inter-module data exchanges conform to the schemas specified in ICD-01, ICD-02, and ICD-03 at each pipeline execution; log compliance flags with provenance metadata	100% ICD schema compliance confirmed by row-count and column-type validation at each module boundary; compliance logged in data_note provenance column

Appendix Table 2. *World Bank Development Indicators. Categorical alignment, technical code depiction, WDI indicator name, and a working definition as defined by the World Bank (WDI, 2024).*

Category	WDI Code	Indicator Name	Definition
Water & Sanitation	ER.H2O.INTR.PC	*Renewable internal freshwater resources per capita (m^3)*	Mean annual volume of internally generated freshwater (rivers and precipitation-fed groundwater) normalized by total population.
Water & Sanitation	ER.H2O.FWST.ZS	*Level of water stress: freshwater withdrawal as a proportion of available freshwater resources (%)*	Ratio of total freshwater abstracted by all major economic sectors (agriculture, forestry and fishing, manufacturing, electricity, services) to total renewable freshwater resources.
Water & Sanitation	SH.H2O.BASW.ZS	*People using at least basic drinking water services (% of population)*	Proportion of the population whose primary source of drinking water is an improved facility (piped supply, boreholes or tubewells, protected dug wells, protected springs, or packaged/delivered water), on the condition that collection time does

			not exceed 30 minutes for a round trip.
Water & Sanitation	SH.STA.BASS.ZS	*People using at least basic sanitation services (% of population)*	Percentage of the population using improved sanitation facilities not shared with other households. Improved facilities include piped sewer connections, septic systems, ventilated improved pit latrines, or composting toilets.
Water & Sanitation	SH.STA.BASS.RU.ZS	*People using at least basic sanitation services, rural (% of rural population)*	Same definition as SH.STA.BASS.ZS, restricted to only rural inhabitants as defined by national statistical authorities.
Energy	EG.EGY.PRIM.PP.KD	*Energy intensity level of primary energy (MJ / $2021 PPP GDP)*	Megajoules of primary energy supply per dollar of GDP at 2021 purchasing power parity (PPP). Lower values here indicate more efficient economies.
Energy	EG.FEC.RNEW.ZS	*Renewable energy consumption (% of total final energy consumption)*	Percentage of total final energy consumption supplied by renewable sources including solar, wind, hydroelectric, geothermal, modern bioenergy, and traditional biomass.
Energy	EG.ELC.ACCS.ZS	*Access to electricity (% of population)*	Share of the total population with access to electricity, as derived from representative household surveys, censuses, and administrative records compiled in the Global Electrification Database.
Energy	EG.ELC.ACCS.RU.ZS	*Access to electricity, rural (% of rural population)*	Percentage of rural inhabitants with access to electricity.
Emissions & Land	AG.LND.FRST.ZS	*Forest Cover (% of land area)*	Land spanning more than 0.5 hectares with trees of at least 5 m height in situ and canopy cover exceeding 10 percent, expressed as a percentage of total land area. Excludes agroforestry systems, fruit plantations, and trees in urban parks and gardens.

Emissions & Land	EN.GHG.ALL.PC.CE.AR5	*Total greenhouse gas emissions excluding LU/LUC/F per capita (t CO_2e)*	National greenhouse gas emissions in metric tons of CO_2-equivalent per person, omitting land-use (LU), land-use change (LUC), and forestry (F) sources.
Emissions & Land	EN.GHG.CO2.RT.GDP.PP.KD	*Carbon intensity of GDP (kg CO_2e per 2021 PPP \$)*	Kilograms of CO_2-equivalent emitted for each dollar of GDP at 2021 purchasing power parity. This indicates the carbon dependency of an economic structure.
Economic Structure	NY.ADJ.NNAT.GN.ZS	*Adjusted savings: net national savings (% of GNI)*	Gross national savings minus depreciation of fixed capital, expressed as a share of Gross National Income.
Economic Structure	NY.GDP.TOTL.RT.ZS	*Total natural resources rents (% of GDP)*	Combined oil, gas, coal, mineral, and forest rents as a percentage of GDP, where each rent is the difference between commodity price and average extraction cost.
Air Quality	EN.ATM.PM25.MC.M3	*PM2.5 air pollution, mean annual exposure (µg/m³)*	Population-weighted average concentration of fine particulate matter (<2.5 µm aerodynamic diameter) that the average resident breathes annually.

LIST OF ABBREVIATIONS

Abbreviation	Definition / Meaning
IPCC	Intergovernmental Panel on Climate Change
IDMC	Internal Displacement Monitoring Centre
WEF	Water–Energy–Food (nexus)

NRC	Norwegian Refugee Council
HDP	Humanitarian–Development–Peace (Nexus)
OECD	Organization for Economic Co-operation and Development
CLIMDEX	Climate Extremes Indices
WDI	(World Bank) World Development Indicators
SPEI	Standardized Precipitation Evapotranspiration Index
WFP	World Food Programme
SHAP	Shapley Additive Explanations
IASC	Inter-Agency Standing Committee
QUORUM	Quantitative Understanding of Regional Outmigration Using Multivariate (data)
RQ1 / RQ2 / RQ3	Research Question 1 / 2 / 3
CADC	Central American Dry Corridor
SE	Systems Engineering
V-Model	Verification & Validation Model
FR	Functional Requirement
UNHCR	United Nations High Commissioner for Refugees
IOM	International Organization for Migration
ICD	Interface Control Document
FAIR	Findable, Accessible, Interoperable, Reusable
CARE	Collective benefit, Authority to control, Responsibility, Ethics
HSI	Human-Systems Integration
BCC	Behavior Change Communication
AI	Artificial Intelligence
WEFE	Water–Energy–Food–Environment
SDG	Sustainable Development Goal
IFRC	International Federation of Red Cross and Red Crescent Societies
CIA	Central Intelligence Agency
API	Application Programming Interface
ISO3	ISO 3166-1 alpha-3 (three-letter country code)
USAID	United States Agency for International Development
PSE	Process Systems Engineering
IM	Information Module
AM	Analytics Module

DM	Design Module
SI	System Integration (and Validation)
SPI-3	Standardized Precipitation Index, 3-month accumulation
HDX	Humanitarian Data Exchange
HAPI	Humanitarian API
IPC	Integrated Food Security Phase Classification
JSON	JavaScript Object Notation
NetCDF	Network Common Data Form
UN	United Nations
OCHA	(UN) Office for the Coordination of Humanitarian Affairs
ABIB	Analytical Block of Information from IM to AM
AFS	Allocated Feature Set
CSV	Comma-Separated Values
LOGO	Leave-One-Group-Out (cross-validation)
RF	Random Forest
ML	Machine Learning
WASH	Water, Sanitation, and Hygiene
GDP	Gross Domestic Product
PPP	Purchasing Power Parity
GHG	Greenhouse Gas
CO_2 / CO_2e	Carbon dioxide / Carbon dioxide-equivalent
MJ	Megajoule
OLS	Ordinary Least Squares
LSDV	Least Squares Dummy Variable
CI	Confidence Interval
NGO	Non-Governmental Organization
RMSE	Root Mean Square Error
GenAI	Generative Artificial Intelligence