Mapping Groundwater Dependent Ecosystems in Alberta's Oil Sands Region

Prepared for The Oil Sands Monitoring Program

March 28, 2024

Prepared by: Jacqueline Dennett[1], Cris Gray[1], Emily Herdman[2], Jen Hird[1], Cynthia McLain[1], David Roberts[1,2], Murdoch Taylor[1], Andrew Underwood[2], Michael Wendlandt[2]

[1] Alberta Biodiversity Monitoring Institute (ABMI)
[2] InnoTech Alberta

ABMI
ALBERTA BIODIVERSITY MONITORING INSTITUTE

InnoTech
ALBERTA

Table of Contents

List of Figures

List of Tables

Executive Summary

The relationship between groundwater and its receiving environment is of particular interest in the Alberta oil sands region (OSR) where industrial operations have the potential to affect both the quality and quantity of groundwater resources via e.g., landscape disturbance, groundwater withdrawals, and tailings pond seepage. Despite groundwater's importance to natural environments and species communities, monitoring of these interactions has been limited. Groundwater dependent ecosystems (GDEs) are ecosystems that are maintained by direct or indirect access to groundwater, and rely on the flow or chemical characteristics of groundwater for some or all of their water requirements (Rohde et al., 2017). Here, we present the results from the first year of a literature review and modeling effort to map aquatic GDEs within the OSR.

Our first year literature review included three components: (1) groundwater indicators of GDEs; (2) biological indicators of GDEs to support mapping, with a focus on aquatic environments; and (3) empirical methods for mapping GDEs.

The groundwater indicator review focused on 26 papers and some highly relevant grey literature, specific to the oil sands region and our study area (Bickerton et al., 2018; J. S. Birks et al., 2012). We summarized both direct groundwater indicators such as water levels and physicochemical properties of water (e.g., temperature, water quality, isotopic composition), and indirect indicators derived from topographic and hydrogeological mapping, numerical groundwater-surface water modeling and remote sensing. Many of these indicators are used as input to our GDE mapping workflow.

While the biological indicator literature review focused on aquatic GDEs, it also provided some preliminary knowledge of biological indicators for terrestrial and subterranean GDEs. The current literature on these topics is limited, with only 28 papers identified in our review, 7 of which are from Alberta. Despite coverage in the literature of some specific species, taxa, and other environmental features that could serve as useful GDE indicators in certain contexts, we conclude that GDE mapping is best informed by maps of wetland classes due to their known association with groundwater inputs.

The GDE mapping methods literature review included 22 papers and summarizes approaches used at a variety of scales (global to local) across the world, with an emphasis on methods appropriate for the boreal region, including studies from Finland. We summarize approaches using remote sensing (e.g., spectral vegetation indices and thermal imagery), integrated hydrological modeling, suitability mapping and machine learning. The selected method was machine learning, using the MLMapper tool (Martínez-Santos et al., 2021) because it is capable of leveraging multiple data sources of differing data types to achieve high predictive accuracy where data limitations exist, and is scalable to large spatial areas.

We identified and collated available geographic, geologic, hydrologic and landcover data for mapping GDEs. Over 50 datasets were identified, with over 40 datasets compiled. From the available data, we selected appropriate data to serve as training & validation data and explanatory variables in MLMapper model and identified data gaps. The key data gaps are access to the McKay River Integrated Surface Water-Groundwater Model, hydraulic head data, and higher resolution thermal data, among others.

Based on the results of the three literature reviews and data compilation, we undertook a machine-learning based modeling approach in the McKay and Steepbank River watersheds using a variety of topographic, hydrogeologic, and wetland/vegetation predictor data, with indicators from the literature review informing variable selection. Final variables included in the modeling were aquifer hosting sediment, bedrock, depth to water, elevation, flow accumulation, normalized difference vegetation index (NDVI), permeability, wetness index, slope, soil drainage, and wetland class. Model fit of the top-performing models, as assessed by internal cross-validation, was very high. Outputs from the top five models were averaged into a final ensemble model of GDE probability.

The GDE maps identify lower river reaches, riparian areas, and wetlands (e.g., fens) as GDEs, but do not capture lakes, likely due to the lack of training data in the modeling pipeline. Upland areas are mostly categorized as non-GDEs. We conclude with suggestions for next steps in model development and application, as well as for potential improved or additional datasets that could be integrated going forward.

1. Introduction and Background

Natural resource development in northwestern Alberta's oil sands region (OSR) continues to expand. Understanding the impacts of various related anthropogenic stressors on the region's landscapes, water resources and biota is crucial to effective land use planning and management. Since 2011, the federal and provincial governments have worked together on environmental monitoring in the OSR through the Oil Sands Monitoring (OSM) Program. In 2017, both governments renewed their commitment to working together with Indigenous communities and industry in the region. The OSM Program strives to improve and continue to add to current understanding of environmental conditions and potential oil sands-related effects, in the areas of air quality, terrestrial biology, wetlands, surface water and groundwater. The latter is a less visible, and therefore sometimes overlooked, but essential component of the hydrological cycle.

The Royal Society of Canada report on Environmental and Health Impacts of Canada's Oil Sands Industry (2010) noted that groundwater and surface water are often treated separately but are intimately linked and long-term environmental management should be based on an integrated approach. To implement these recommendations, the Joint Oil Sands Monitoring Plan (2011) for the Lower Athabasca River watershed included a groundwater component to improve understanding of groundwater-surface water interactions recognizing that this is essential knowledge for a program focused on aquatic ecosystem health impacts. Foundational work has been completed on assessing groundwater influence on selected river systems in the OSR (e.g., McKay) (Bickerton et al., 2018). Nevertheless, a 2022 "Condition of the Environment: Groundwater in the Oil Sands Region" report produced for the OSM Program by InnoTech Alberta reviewed important pathways (e.g., groundwater water recharge, flow, transport) by which stressors (e.g., landscape disturbance, groundwater withdrawals, spills, leeks, or seepage, etc.) can impact groundwater and groundwater discharge quality and/or quantity (J. S. Birks et al., 2022). While providing important insights on groundwater stressor-pathway-response in the OSR, this report highlights a continued knowledge gap that remains:

the occurrence and condition of groundwater dependent ecosystems (GDEs) within the area. In particular, key OSM program questions this work supports include:

- Do changes in groundwater have effects on the receiving environment?
- Do changes to groundwater impact harvesting and occupancy patterns, harvesting volumes, intergenerational transfer of knowledge, sharing of resources linked to the reinforcement of kinship bonds, people's relationship and obligations to the land?

With regard to the OSM's Groundwater Technical Advisory Committee, more specific key questions include:

- Where are the significant areas (e.g. groundwater dependent ecosystems) of groundwater connectivity (i.e. groundwater discharge/recharge) to surface waters such as streams, wetlands, springs and lakes?
- Has the quality and quantity of groundwater discharge to groundwater dependent ecosystems (GDEs), or other surface waters of interest, changed?
- What is the cause of any unexpected changes identified in preceding items?

GDEs are defined by Rohde et al. (2017) as "Ecosystems that are maintained by direct or indirect access to groundwater and rely on the flow or chemical characteristics of groundwater for some or all of their water requirements." Alternate definitions emphasizing slightly different qualities such as species composition are found elsewhere (e.g., Serov & Kuginis, 2017)), but here we rely on the definition from Rohde et al. (2017). GDEs themselves can be only partially, intermittently, or seasonally dependent on groundwater inputs. Within the context of the OSR, engineered or anthropogenic GDEs are likely to occur alongside natural GDEs, given the existence of wetlands resulting from reclamation practices.

GDEs are important features of the OSR landscape, fulfilling important ecological functions by supporting unique vegetation communities, maintaining local water quality and quantity, and acting as a mitigating factor in the face of climatic extremes (e.g., drought). They are of critical cultural and traditional significance to local Indigenous communities because groundwater-derived base flow supports navigation, and GDEs support vegetation and animal communities harvested by Indigenous communities (e.g., they provide ungulate watering holes, salt sources, waterbird habitat, base flow in fish habitat). Groundwater ecosystems themselves are far more complex than previously thought, showing high levels of trophic complexity and specialization often dominated by endemic microbial and other species (Saccò et al., 2024). GDE and general groundwater conservation efforts lag behind those for more visible surface water or terrestrial ecosystems, and where they exist, are in place because of the economic value of a given aquifer or other groundwater source (Rohde et al., 2017; Saccò et al., 2024).

GDEs by their nature are sensitive to changes in groundwater discharge, both in quantity and quality, and for this reason, act as an ecological assessment endpoint. While it is expected that GDEs may be impacted by oil sands development, based on the known impacts to groundwater, understanding of the stressor-pathway-response interactions that lead to changes in GDEs and associated monitoring have been limited by a lack of understanding regarding the extent and distribution of GDEs in the OSR. Improved identification of the location of GDEs in the OSR (i.e., mapping) will help support baseline assessments and develop appropriate long-term monitoring initiatives for cumulative impacts of local and regional oil sands activities.

The overall objective of the GDE Project is to map GDEs across the OSR and provide information on pathways in the conceptual model (impact of groundwater recharge and flow and transport of constituents of concern on terrestrial and aquatic ecosystem health) and identify opportunities to evaluate the response of biological communities to oil sands-related stressors (see Figure A.1 and Figure A.2 in Appendix A).

1.1 Objective for 2023-2024

The objective of this work for the 2023-2024 fiscal year is to *map aquatic GDEs across a pilot area in the OSR* so that they can be used to refine a long-term monitoring plan for groundwater and contribute to identification of cumulative effects in aquatic and terrestrial environments.

1.2 Phases & Deliverable Details for 2023-2024

The current work is being undertaken using a phased approach. Given the long-term objective of mapping GDEs across Alberta's OSR, this first phase consists of collaborative efforts between the OSM Groundwater Technical Advisory Committees (TACs), and the Alberta Biodiversity Monitoring Institute (ABMI) and InnoTech Alberta. Outcomes of the work will also be shared with the Terrestrial Biological Monitoring (TBM) Technical Advisory Committee in recognition of the important relationships to their work. It forms a scientific and practical foundation for future mapping phases, which are anticipated to incorporate wider OSM support of mutual monitoring plan integration and knowledge sharing.

Deliverables for this project for the 2023-2024 fiscal year include:

- A technical report (i.e., the current document), presenting the outcomes of the tasks listed below:
    - A review of the academic and grey literature on GDE mapping approaches and groundwater indicators that includes:
        - boreal GDE category definition;
        - use of methods or rules for identifying and mapping GDEs using examples from other jurisdictions; and
        - recommendations for validating approaches used for GDE mapping and identification of existing datasets that could be leveraged to support these approaches;
    - A review of the academic literature on biological (i.e., key species and community) indicators of aquatic GDEs in boreal systems (future work will include terrestrial and subterranean GDEs);
    - The identification, review, and collation of data sources currently available within the OSR to support GDE mapping using identified approaches, including the identification of data gaps and recommendations for filling these gaps;
    - Initial GDE mapping within a selected area of interest in the OSR, using the collated data and identified approaches;

- A digital, annotated geodatabase of the initial GDE map product, complete with metadata and methods documentation; and
- Summary presentation of the project's outcomes and results to the Groundwater TAC.

GDEs generally fall into the following broad categories:

- Aquatic (e.g., rivers, streams, lakes, wetlands, and springs);
- Terrestrial (e.g., riparian areas); and
- Subterranean (e.g., cave systems, aquifers).

The scope of the first year of the project's (2023/24) initial GDE mapping work focuses on aquatic GDEs in the OSR.

2. Definitions and GDE Categories

2.1 Introduction & Key Definitions

Sustained hydrological sources are imperative to ensuring healthy ecosystem function; conversely, during periods of hydrological scarcity, ecosystems can undergo drastic changes depending on their water source. Groundwater dependent ecosystems (GDEs) have a diverse range from aquatic, to terrestrial, to subterranean ecosystems. While their dependence on groundwater contributions can fluctuate throughout the year as annual precipitation and seasonal demand fluctuate, the presence of these ecosystems relies on a sustained source of groundwater for maintaining ecosystem function (e.g., by providing hydrological and nutrient inputs). GDE expressions are typically observed in both above-ground expressions (lakes, rivers, streams, springs, and seeps during base flows) as well as subsurface presence where phreatophytes (deep rooted plants) access water during periods of low hydrological availability (Klausmeyer et al., 2018) or where there are wet cave ecosystems. Although the definitions of GDEs evolve with the progression of the field and are defined differently within differing jurisdictions, the definition set out by Rohde et al. (2017):

> *"Ecosystems that are maintained by direct or indirect access to groundwater, and rely on the flow or chemical characteristics of groundwater for some or all of their water requirements"*

will be used for this project and encapsulates the generalized definition that GDEs may only be partially dependent on groundwater or may only demonstrate seasonal or intermittent dependence on groundwater. Serov and Kuginis (2017) provided a definition that emphasizes natural elements, and ecological aspects rather than the more generic reliance on water requirements:

> *"Natural ecosystems which have their species composition and natural ecological processes wholly or partially determined by groundwater"*.

Although natural ecosystem elements can be used to identify GDEs, this is significantly more challenging within cooler high-latitude environments (Autio et al., 2023), such as the boreal forest of Alberta. These environments typically have shortened growing seasons and less evaporative demands, as is the case of the wetland-dominated landscapes found in the OSR, where vegetation indicators are more difficult to apply. In contrast, identifying GDEs within southern Alberta, where higher hydrological demands occur, leverage the use of plant vigor for detection of surface and groundwater interactions (Van Der Kamp & Hayashi, 2009). Currently the definition of GDEs does not make distinctions between anthropogenic and naturally developed systems, however within the context of Alberta, engineered GDEs are likely an important component to consider as footprint is reclaimed.

Similar to quantifying whether an ecosystem is groundwater dependent, sensitivities of GDEs can fall into finer class segments, and determination of ecosystem sensitivities is correlated to the species and environmental conditions present within the GDE. The assessment of GDE sensitivities can be broken down first into climate classifications. According to the Thornthwaite climate regimes there are broadly five categories which are identified on the basis of monthly precipitation to evaporation ratios (P/E): (hyper humid “wet” (127), humid “forests” (127-64), subhumid “grasslands” (63-32), subarid (63-32), semi-arid “steppe” (31-16), arid “desert” (<16)). The impacts of climatic shifts are more pronounced within arid and semi-arid GDE environments, due to inherent water limitations present there. Disruptions to these particular ecosystems makes them highly sensitive to fluctuations of groundwater, with noticeable effects on ecosystem community composition such as changes in vegetation communities from aquatic to drought tolerant species (Beasley-Hall et al., 2023; Doody et al., 2017). As a result, the prevalence of GDEs on the landscape can easily be quantified in arid and semi-arid regions, as the water stress impacts plant and ecosystem functions, which is often first reflected in the degree of measurable vigor in vegetation.

In high latitude climates that are not hydrologically limited, the presence of GDEs can be more difficult to detect. Monitoring these ecosystems at a large scale serves as a critical pillar for the OSM program, helping to ensure that any changes to these sensitive systems are identified prior to significant adverse effects. However, the first step in developing an appropriate monitoring program is the identification of their location within the OSR (Strategic modeling plan TAC, 2019). GDEs’ inclusion into the OSM groundwater monitoring framework will serve to highlight possible pathways through which stressors may influence these systems, while understanding that the degree of reliance on groundwater has the potential for partitioning GDEs into subcategories, which may help determine monitoring priorities. Within the OSM Program’s *Technical Report Series* Bickerton et al.’s (2018) compilation of multiple techniques for assessing groundwater influences within the OSR of Alberta highlights the influences of groundwater on surface water expressions of tributaries to the Athabasca River and the need for future monitoring of surface water - groundwater interactions. Specifically, there are direct contributions of groundwater along various reaches of the MacKay River, and, as a whole, groundwater might contribute as much as 35% during under ice flow compared to 2-10% during low flows (Bickerton et al., 2018). Thus, consideration of both average and seasonal contributions of groundwater to GDEs should be considered.

2.2 GDE Categories in Alberta’s Boreal Systems

Groundwater dependent ecosystems fall into three categories: aquatic, terrestrial and subterranean.

2.2.1 Aquatic GDEs

Aquatic GDEs include all springs, rivers, streams, lakes, and wetlands with groundwater contributions. All these aquatic ecosystems occur in the Boreal region of Alberta.

Springs occur when groundwater overflows onto the land surface and can range in size from small seeps to pools. Springs with high mineral content can be associated with wet mineral licks (or "salt licks") which are utilized by ungulate species, can develop into muddy clearings ("wallows") when used by elk and moose, and may influence the spatial structure and movements of ungulate populations. In the OSR, discharge of saline groundwater occurs where Devonian carbonate bedrock intersects the land surface (e.g., along river valleys).

Rivers and streams can receive base flow from groundwater, which provides flow during low-flow and frozen conditions (e.g., supporting in-stream flow needs), constant-temperature water supply, and refugia for aquatic species such as fish and benthic invertebrates. Rivers and streams can have reaches that are "gaining" i.e. groundwater is contributing to the flow along these sections. During the winter, when many surface water bodies are frozen, springs and gaining sections of streams may remain unfrozen, providing access to liquid water for animals (e.g., ungulate watering holes, waterbird habitat). During frozen conditions, aufeis, or a layered mass of ice (also called icings), can also form from the freezing of successive flows of groundwater over previously formed layers of ice which can maintain unfrozen conditions beneath the insulating ice layer providing a perennial groundwater habitat (Huryn et al., 2020).

Lakes with subsurface inflow contributing to the lake water balance are GDEs. Topographic position, bathymetry, surficial and bedrock geology, and presence/absence of permafrost are some of the factors that influence the groundwater dependence of lakes in the OSR.

By definition, all wetlands classified as fens are GDEs. Fens are estimated to cover 21% of the recently mapped portion of the OSR (Alberta Biodiversity Monitoring Institute & Ducks Unlimited Canada, 2023) and are thus a critical component of GDE mapping within the OSR. Other classes of wetlands that may have groundwater input include shallow open water wetlands, marshes, and swamps.

The movement of water between groundwater and surface water provides a major pathway for chemical transfer between the subsurface and surface water. For example, groundwater can supply carbon, oxygen, and nutrients such as nitrogen that affect biological processes.

2.2.2 Terrestrial GDEs

Terrestrial GDEs include vegetated land such as uplands and riparian areas where groundwater provides water supply for plants but where surface water may not continually present. These habitats often occur along rivers and streams (Luke et al., 2007) and in floodplains, and in upland areas where there are phreatophytes: vegetation that depend for their water supply on groundwater that is within the reach of their root systems.

2.2.3 Subterranean GDEs

Subterranean GDEs include caves and aquifers. Caves are subterranean GDEs when the plant, animal, or microbial communities within depend on the presence of groundwater on a permanent or intermittent basis to meet all or some of their water needs. Underground caves and streams can form in karst landscapes, which in the OSR form from the dissolution of carbonate bedrock of Devonian age. At the surface, karst landscapes feature sinkholes. Aquifers occur in many different geological formations, and at varying depths in the OSR. These subterranean ecosystems are inhabited by microbial communities, and can also host stygofauna (i.e., aquatic animals such as arthropods and other invertebrates, as well as vertebrates including fishes and salamanders).

Within Alberta, surface karstification is more prominent within the Rocky Mountains, where both geological formations (limestones) are susceptible to dissolution from increased hydrological gradients that help facilitate subterranean ecosystems to form (D. Ford, 1987; D. C. Ford, 1997). In the far north-east of Alberta (e.g., in Wood Buffalo National Park) hundreds of sinkholes are common landscape feature formed from dissolution of the at or near-surface Middle Devonian Elk Point Group evaporites, as well as networks of underground cavernous systems and prominent escarpments making this area some of the most extensive karst landscapes in North America (Altosaar, 2013b; Parks Canada, 2022). Many of these "karstland" features were mapped and reported on in 2012 and 2013 by Suncor (Altosaar, 2013a). These systems contrast the OSR where karst formations are primarily caused via the dissolution of highly soluble evaporites (halite) which are remnants of the region being an ancient inland sea (S. J. Birks et al., 2022; Broughton, 2018),and are located deeper in the formation offering higher protection by the glacial till overburden from hydrological weathering (D. Ford, 1987; D. C. Ford, 1997).

3. Study Area

The oil sands region of Alberta covers 142,200 km^2 in northeastern Alberta, encompassing the Athabasca, Peace and Cold Lake oil sands regions. These regions fall almost entirely within the boreal ecoregion of Alberta. Terrestrial boreal habitats are peatland dominated mix of drier upland and lower wetland habitats. Forested uplands consist of boreal mixed woods, mostly spruce or aspen dominant stands with some pine stands in areas of sandy, well-drained soils. Wetland habitats in the oil sands region fall into a few hydrologically- and ecologically-defined classes, including fens, bogs, swamps, and shallow open-water wetlands. Fire is a major natural driver of change in the boreal and creates a mosaic of variable forest stand ages across the landscape. Forestry, energy development, and other industrial operations also create widespread and notable linear (e.g., roads, seismic lines, pipelines) and polygonal (e.g., cut blocks, mines, well pads) disturbance features throughout the region. Oil sands industrial operations are composed of two main processes of bitumen extraction: (1) surface mining in a relatively contained area north of Fort MacMurray, which creates the enigmatic mines and roads typically associated with oil sands operations; and (2) in situ mining for below-ground extraction, which creates a more widespread network of roads, wellpads, seismic lines, and pipelines across the entire region.

The first year of GDE work focuses on two tributaries of the Athabasca River in the Athabasca Oil Sands Area north of Fort McMurray. This study area spans 3,078 km^2 of the lower McKay and Steepbank River watersheds (Figure 1). Advice was solicited from the GoA and ECCC Technical Advisors for the project on Study Area selection. Of the multiple options considered for a study area within the Athabasca OSR, these particular watersheds represented areas where: 1) recent and relatively high resolution wetland inventory mapping had occurred (OSM Wetlands Inventory Project 2022/23); 2) multiple OSM-funded research and monitoring projects such as groundwater modeling, have been conducted (e.g., work by (Bickerton et al., 2018) along the McKay River); 3) isotope-based streamflow partitioning has revealed higher fractions of groundwater contributions in river tributary flows (i.e., on the east side of the Athabasca River (Gibson, Yi, et al., 2016)); and/or 4) a greater amount of relevant field or modeled data were available. These watersheds best leveraged the current knowledge and data available within the OSR that could support GDE mapping. The large majority (98.6%) of the selected study area falls within the area of the North Athabasca Oil Sands. The study area was selected to have upgradient areas without oil sands mining activities as well as areas with a variety of oil sands mining operations. This could allow for future evaluation of differences in GDE mapping in undisturbed and disturbed areas.

The study area is within the McMurray Lowlands and Regional Uplands hydrogeological regions of Alberta and the geological history and region maps can be explored in the Alberta Geological Survey's StoryMap (Alberta Geological Survey, 2021). These areas have been well studied from a geological and hydrogeological perspective given the prominence of oil sands extraction in the region. Lying between topographic high areas including Muskeg Mountain to the east and the Birch Mountains to the northwest, this region sits near the edge of the Western Canadian Sedimentary Basin where bedrock is closer to the surface than in most other regions of Alberta. Surficial sediment thickness varies greatly from <5m to >100 m. Local scale groundwater movement is driven by upland recharge areas and dynamic interactions of boreal wetlands with shallow groundwater. Cretaceous formations contain important nonsaline aquifers as well as the bitumen-bearing McMurray Formation. Devonian formations host saline groundwater that discharges in saline springs where the Athabasca and Clearwater rivers have eroded into these formations, and in some areas, ongoing dissolution of carbonate and evaporite bedrock continues to form karst landscapes appearing as circular ponds and wetlands that form above active sinkholes. Groundwater salinity is highly variable due to the complex geology and groundwater flow. Discontinuous permafrost exists in at least 87 small areas within the Study Area (Pawley & Utting, 2018).

In the last decade, groundwater-surface water interaction in the Study Area has been studied for rivers, open water wetlands and lakes (Gibson et al., 2019). In the McKay and Steepbank Rivers isotope-based streamflow partitioning and differential gauging (Bickerton et al., 2018; Gibson, Yi, et al., 2016) revealed 14-45% and 4-65% (McKay) and 29-69% (Steepbank) contribution from groundwater (notably higher on the east side of the Athabasca River). In the McKay River both studies estimate an approximately 3-fold increase in groundwater contribution in the winter compared to the fall. Between ~20-50% of shallow open water wetlands are groundwater reliant based on wetland water balance calculations using water isotope data collected by ABMI and analyzed by InnoTech Alberta between 2009 and 2019 (Gibson et al., 2022). In the study area northeast of Fort McMurray lake water quality was monitored for over 15 years under the Regional Aquatic Monitoring Program (RAMP, Joint Oil Sands Monitoring Program (JOSM) and OSM evaluating water balance, permafrost thaw and pH changes (Gibson et al., 2019) .

The study area map (Figure 1) shows the area of interest (focused on the McKay and Steepbank watersheds) and the larger analysis boundary area. Data was acquired for the analysis area and processed prior to clipping to the smaller area of interest. Analyzing the larger area of analysis allowed for the establishment of baseline environmental and geological conditions, offering background details that aid our understanding of the specific characteristics defining our area of interest. The expanded analysis area helps to identify potential external influences that might not be within the area of interest, but could have effects on it. In this report, maps reflect the larger area of analysis to provide the important context as identified above, evaluation and discussion of outcomes does, however, focus on the smaller area of interest.

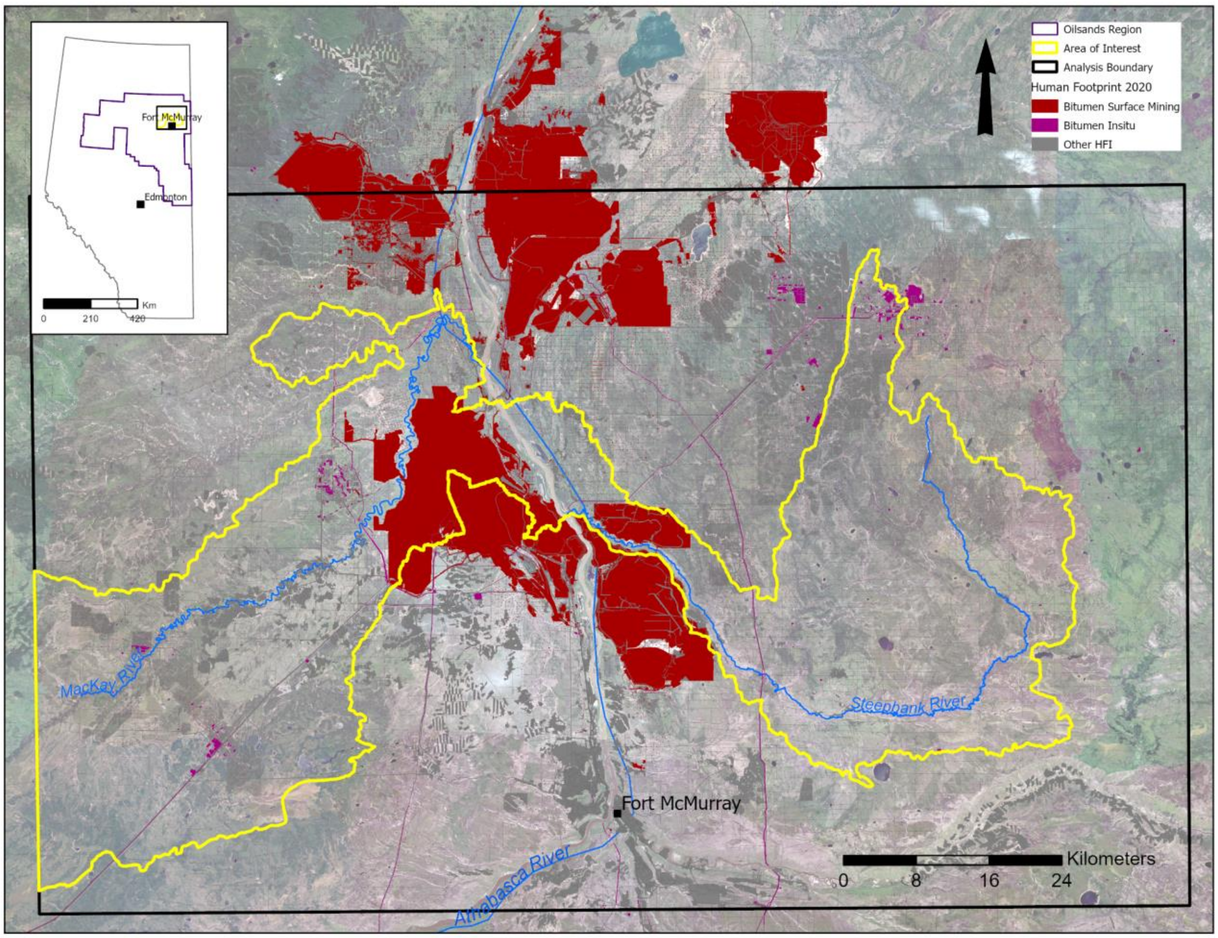

Figure 1. Map of the study area in the Alberta Oil Sands Monitoring program area, showing the area within the Northern Athabasca Oil Sands Region, the area of interest, the analysis boundary, and human footprint components.

4. Literature Review: Indicators of GDEs

4.1 Groundwater Indicators for GDEs

4.1.1 Overview

A comprehensive literature review was conducted by search for keywords “Groundwater dependent ecosystem*”, “mapping”, “machine learning”, “peatland*”, “groundwater surface water interaction”, “Athabasca oil sands region”, “karst”, “eskers”, “surficial geology”, “EM survey”, “spring*”, “seep*” within the peer-reviewed literature and leveraging key papers (Table 1).

Focus was given to surface and groundwater interaction, groundwater indicators, and mapping of GDEs within Boreal environments, with a total of 48 papers reviewed. These papers focused on the geographical regions of Alberta and outside of North America (World). A total of 4 review papers were used to guide current methods for mapping of GDE systems. The main focus areas of the literature broadly focused on five main topics (Isotope & Geochemical, GDE mapping, Modelling, Lake, Rivers, Wetlands), with papers able to fall within multiple topics (see Tables B.1 and C.2 in Appendices B and C, respectively).

Table 1. Summary of number of papers included in the literature review of groundwater indicators of GDEs, summarized by the geographic location of study or interest (top) and by topic focus area (bottom). Note that some manuscripts discussed more than one topic focus area and so may be counted in multiple groups. Full citations are provided in Appendix B.

	Geographic Location				
	Alberta	Canada	Outside Canada	Review	Total
Number of papers	22	2	20	4	48

	Topic Focus Area(s)*					
	Isotope and Geochemical	GDE Mapping	Modeling	Lakes	Rivers	Wetlands
Number of Papers	31	22	16	12	17	24

Foundational grey literature on GDEs from Alberta the OSR was included in the literature review as well including but not limited to an appendix developed by the Government of Alberta on GDEs for Groundwater Management Frameworks and Bickerton et al. (2018). Methods for measuring aquatic GDEs, and specifically surface water - groundwater interaction in the Athabasca oil sands region are described in Birks et al. (2012) including field and desktop methods.

4.1.2 Direct Groundwater Indicators

Birks et al. (2012) summarizes methods for monitoring direct indicators. These include, but are not limited to, physical indicators (e.g., groundwater levels, seepage measurements), locations of springs, aufeis (see also Ensom et al., 2020; Huryn et al., 2021),and other indicators discussed below.

The physicochemical properties of surface water can give insights into the presence of GDEs. More specifically, measurements of temperature and conductivity can aid in identifying GDE presence when assessing groundwater that has had mineral interactions. The temperature of groundwater primarily reflects the average seasonal temperature of the region (J. S. Birks et al., 2012; Hayashi & van der Kamp, 2023) such that during summer months in particular, there is a higher contrast between cool groundwater seeps and warmer surface waters (Bertrand et al., 2014; Pérez Hoyos et al., 2016). These temperature differentials can also be indicative of aquatic systems that are recharging or discharging groundwater, allowing for source and direction of surface and groundwater interactions to be inferred (Bertrand et al., 2014; Watts et al., 2023) . Temperature can be measured directly or via remote sensing. Solute loading within water increases its conductance, which are reflected in physicochemical differences with higher specific conductance, variation in pH and oxidation reduction potentials dependent on which mineral substrate water has come into contact with (A. Gue et al., 2018; A. E. Gue et al., 2015).

Additionally, direct indicators of GDEs can be expressed through hydrochemical facies that exhibit increased dissolved solutes in the form of dominant cation and anions, which are indicative of longer temporal scales of water and mineral interaction (S. J. Birks et al., 2022; Gibson et al., 2013; Manchuk et al., 2021; Wells & Price, 2015). Depending on the dominant cation and anions present, these ions can give indications about whether hydrological sources are from deep basin brines (primarily halite dissolutions), or shallow glacial tills (carbonate and silicates) (J. S. Birks et al., 2012; S. J. Birks et al., 2022; A. E. Gue et al., 2015). These deep aquifer systems access the surface through evaporite channels and karst systems which allow for the presence of highly saline groundwaters to migrate to the surface (Broughton, 2018; Hein & Cotterill, 2006; Walker et al., 2017; Wells & Price, 2015). Salinity can also be influenced through evaporation effects: minerals can precipitate, forming mineral deposits at the surface that can then be redissolved with precipitation or snowmelt. The latter process is seen in the prairie pothole regions of Alberta (Hayashi et al., 2016). Depending on hydrological contributions and processes, the characterization of GDEs can be limited when using salinity alone as an indicator of GDEs presence and should be used in conjunction with hydrological tracers.

The extensive use of stable isotopes and radioactive tracers can offer unique tracers to identify hydrological systems under the influence of groundwater. Both deuterium and oxygen-18 (^{2}H, ^{18}O) are stables isotopes which fractionate at predictable intervals and have been used to partition surface and groundwater contributions to many aquatic features (lakes, rivers, and wetlands) - as they are components of the water itself, they make ideal tracers to infer hydrological contributions and processes of aquatic systems (Gibson, Birks, et al., 2016; Gibson et al., 2019, 2020, 2022; Gibson, Yi, et al., 2016; Gibson & Peters, 2022). Isotope ratios of other solutes, including sulfate and strontium, among others, can also be helpful in tracing groundwater - surface water interactions.

The use of radioactive tracers is another tool for inferring contributions of groundwater to large aquatic systems. Radon (^{222}Rn) occurs from the decay of radium-226 parent material. As it is a noble gas, it does not participate in any biological or chemical processes with which it comes into contact, and has a relatively short half-life of ~72 hours. It has been used to quantify water budgets of lakes, most notably in two boreal lakes within the OSR, which found that groundwater contributions in the lakes varied from (0.5% to 14% annual flows) and that catchment size and benthic sediments are likely to play a controlling role in groundwater contributions (Schmidt et al., 2010). Due to the short half-life of Radon, it has the ability to be used to rapidly assess groundwater inputs, contrasting the longer temporal time frame for stable isotopes, which allows for complementary information on lake water budgets (Arnoux et al., 2017). A limitation of Radon is that it requires mineral parent materials and can be influenced by long transit times within peat, reducing some of its use in heavily organic substrates as a tracer for groundwater (Schmidt et al., 2010).

4.1.3 Indirect Groundwater Indicators

Geophysics and remote sensing approaches can be used for mapping the groundwater table (J. S. Birks et al., 2012). In the oil sands region, recent work by the Boreal Ecosystem Recovery & Assessment program demonstrated the use of orthophotography and photogrammetric point clouds for mapping groundwater level and depth to water (Rahman et al., 2017).

Within river and stream systems, gaining and losing reaches can be quantified during baseflow, giving indications of the potential for regions with both groundwater discharge (GDEs) and recharge areas. This can be accomplished via flow gauging, hydrographic analysis and isotope partitioning (Bickerton et al., 2018; J. S. Birks et al., 2012; Gibson, Yi, et al., 2016).

In addition to direct physicochemical indicators, physical landscape and subsurface features can also play an important role in assessing the presence of GDEs but are considered indirect indicators. Such features include the topography of the landscape, and surficial and bedrock composition. Topography plays a role in the assessment of GDEs with low lying areas having a higher chance for seeps and springs from groundwater to occur (Freeze & Cherry, 1979; Heagle et al., 2013) . Groundwater within these locations is more likely to collect and can indirectly be modeled through geospatial mapping techniques by assessing flow directions and flow accumulations, which highlight regions where water is more likely to collect. These low-lying areas are contrasted by higher elevation areas which are less likely to be GDEs and more likely to be perched hydrological features (Heagle et al., 2013).

Hydrogeological mapping of geological and hydrological properties of the subsurface provides conceptual/indirect understanding of GDEs (see Figure A.1 and Figure A.2 in Appendix A). Surficial and bedrock geological compositions influence the movement of groundwater in both discharge and recharging systems (S. J. Birks et al., 2019, 2022; Broughton, 2018; Freeze & Cherry, 1979; Hein & Cotterill, 2006). Differences in surficial materials primarily control the permeability of the soil structure (Hein & Cotterill, 2006). Features with coarse textures and larger pore spaces (e.g., sands and gravels) allowing for increased hydrological conductivity and a higher probability of being a GDE if positioned in low lying areas on the landscape. These contrast with fine-textured materials (silts, clays) which typically impede flows (S. J. Birks et al., 2019, 2022; Hein & Cotterill, 2006). Recent higher resolution mapping of quaternary surficial layers has highlighted the interconnectivity of both buried channels and surficial geological formations (Andriashek, 2001, 2007; Atkinson, 2022a, 2022b, 2022b; Atkinson et al., 2013; Atkinson & Pawley, 2022; Pawley & Atkinson, 2022; Utting, 2023), which likely play a controlling force on the presence of GDEs within the OSR through vertical groundwater discharge (Hein & Cotterill, 2006; Wells & Price, 2015). Bedrock structure can influence the hydrological connectivity with buried channels and fractured rock systems which allow for hydrological flow to migrate to the surface; these contrast with impervious layers which will act as hydrological barriers (S. J. Birks et al., 2022; Broughton, 2018). In areas where these channel and fracture features have been mapped there could be a higher probability of GDEs being present on the landscape.

These indirect indicators can be brought together in 3D numerical groundwater models and used to identify areas of groundwater interaction with the land surface (e.g., Aquanti HGS and McKay River Integrated Surface Water-Groundwater (J. S. Birks et al., 2012) models). These models can be very useful in testing scenarios of water use and climate change, and to inform field campaigns, but can be data hungry, costly and time consuming to develop. The most relevant types of numerical groundwater models for studying GDEs are integrated surface water - groundwater interaction models such as GS-FLOW, HydroGeoSphere (HGS), Parflow and MIKE SHE. A coarse-resolution regional-scale HGS model for the Athabasca River Basin was created by COSIA, and is used in this study. As well as a tributary-scale model using GS-FLOW for the McKay River watershed (originally created for CEMA, and updated for ECCC and OSM). Other types of groundwater models (e.g., MODFLOW, FEFLOW) are not as rigorous for estimating surface water contributions to aquatic ecosystems, but do exist in other areas of the OSR.

4.2 Biological Indicators for GDEs

4.2.1 Overview

A comprehensive literature review of aquatic GDEs was conducted using Web of Science and the keywords: groundwater; groundwater indicator; groundwater dependent ecosystem; indicator; bio*; biol*; Alberta; Canada; boreal; north*; fen; swamp; microb*; stygo*; macrophyte; vegetation; vascular plant; moss; bryo*; fauna; geophagy; mammal*. In total, 28 papers were used in review (Table 2). Selected papers were leveraged by reviewing both cited papers within a manuscript and papers that had cited that manuscript since its publication, i.e., the citation network function in Web of Science. Papers were evaluated based on their relevance in establishing empirical evidence to support indicator development for GDEs. 'Indicator' in this context was considered both from the perspective of biological metrics that could identify the location of GDEs and as biological metrics that could be used to monitor impacts of development and operation of industrial facilities (i.e., ecological endpoints). Review papers were considered valuable. Several papers that aimed to assess stressor impacts to GDEs simultaneously provided evidence in support of selected biological indicators of GDE presence. Relatively few assessed stressor impacts without providing such evidence. Papers from Canada were prioritized, followed by papers from areas with boreal, and then otherwise forested, landcover (e.g., Finland, Switzerland, France, United States). Papers from Australia were considered low priority due to fundamental differences in glacial history, ecology, and geology when compared to northern Alberta and due to their focus on arid environments.

Table 2. Summary of number of papers included in the literature review of biological indicators of aquatic GDEs, summarized by the geographic location of study or interest (top) and by topic focus area (bottom). "Other" is inclusive of Ionescu et al. 2022 and Driscoll et al. 2019, focused on biotic homogenization and natural range of variability, respectively. Note that some manuscripts discussed more than one focus area or indicator and so may be counted in multiple groups. Full citations are provided in Appendix C.

	Geographic Location				
	Alberta	Canada	Outside Canada	Review	Total
Number of papers	7	8	9	4	28
	Topic Focus Area				
	Vegetation	Macroinverts, Microbes, Stygofauna	Mammals	Other	Background
Number of Papers	11	12	2	2	2

The literature review made clear that there is limited research and therefore limited understanding of the role of groundwater in structuring and maintaining the biological components of aquatic GDEs in boreal ecosystems or limited understanding of what those components may be, particularly in the case of stygofauna and microbial communities. Broadly, the conceptualization of groundwater systems as ecosystems, with associated biota above and belowground, is an emerging theme in the literature. However, this work is more advanced in Australia than in other global locations (Hancock et al., 2005) and thus may not be specifically relevant to the boreal environments that were the primary focus of this review.

Biological indicators were rarely used to map aquatic GDEs (but see Graillot et al., 2014). Rather, biological indicators were more typically used to identify groundwater influence at local scales (Larocque et al., 2016; Munger et al., 2014)or as receptors of groundwater mediated effects, such as land use and contamination impacts (Ionescu et al., 2022; Lehosmaa et al., 2018). Changes in biological indicators can be monitored at a variety of ecological scales–individuals, populations, and communities–and metrics can include the condition, productivity, demographics, structure, or function of a range of indicators, from species to ecosystems (Eamus et al., 2006; Oiffer, n.d.). To address the complexity of ecological systems, indicator approaches such as sentinel species may be appropriate, particularly as larger scale change, such as to ecosystem structure and functions, may be slow to manifest (Rohde et al., 2017) or may be missed in the context of the natural range of variability of the larger GDE.

Grey literature was not targeted in the review and is not captured here but could be a valuable source of knowledge for future review. Notably, a draft report prepared by the Government of Alberta (Oiffer, n.d.) and made available for this project provides broad categories of biological indicators based on methods described in Eamus et al. (2006) for health of groundwater dependent vegetation.

4.3 Biological Indicators of Terrestrial and Subterranean GDEs

The 2023-24 biological indicator literature review was not comprehensive of subterranean and terrestrial GDEs, excluding manuscripts focused on ground- and surface-water interfaces that included subterranean indicators and associated discussion of subterranean groundwater resources. These indicators, specifically microbial and stygofaunal components, are discussed below in the context of aquatic GDEs. In the context of future work to review biological indicators of subterranean GDEs, the review paper of Mammola et al. (2020) poses expert-determined fundamental research questions related to subterranean GDEs and may make a good starting place. Conservation related questions posed by the authors, e.g., "What is the impact of above-ground disturbance on subterranean environments and their fauna" (Mammola et al., 2020)highlights the limited work that has been done and the outstanding gaps in our collective understanding of subterranean GDEs and their response to anthropogenic activity. Additional literature review in future years will allow us to provide a comprehensive review of biological components of subterranean GDEs and how they may respond to groundwater changes in the OSR.

Precursory scanning of the literature regarding biological indicators in terrestrial GDEs in Alberta and Canada resulted in papers focused on wooded riparian forests. Broadly, terrestrial GDEs are often tied to the presence of phreatophytes, a term that describes deep rooted trees and shrubs that can grow in dry environments by accessing sub-surface water and which have high transpiration rates. Research examples from Alberta focus on cottonwood-dominated riparian forests, where cottonwoods (*Populus deltoides* or *Populus trichocarpa*) are phreatophytes, in semi-arid southern Alberta. These forests are reliant on groundwater, especially during drought events (Tai et al., 2018; Zimmerman et al., 2023), with similar patterns observed in semi-arid to arid regions of the United States (Graup et al., 2022). It is possible that the related balsam poplar (*Populus balsamifera*), which occurs in the boreal, could be indicative of terrestrial GDEs, although this species is widespread in mesic to hygric forests. Cottonwoods, poplars and willows are considered keystone species for some Indigenous peoples in Alberta, and their distribution is known to have changed along riparian areas, even as recently as in the last 40 years.

Research specific to Alberta's boreal included mixed-wood forest stands on saline soils (trees are typically intolerant of saline soil) in proximity to salt pans and saline wetlands, which are likely GDEs (Lilles et al., 2010). Outside of Canada, the review of Chiloane et al. (2022) provides an excellent starting place. The authors list indicators related to vegetation, including phenology, advocate for the use of remote sensing in identifying terrestrial GDEs, and note that "So far, groundwater-vegetation interaction monitoring has been limited by the trade-off that exists between the costs, efficiency, and level of detail offered by the techniques employed" (Chiloane et al., 2022). Aside from vegetation, an assessment of microbial communities found distinct differences between groundwater and geological material samples, indicating that unique communities occupy subsurface terrestrial environments (Meyer et al., 2022). Meyer et al. (2022) observed a decrease in bacterial and archeal abundance and diversity with depth, but interpreting depth trends in eukaryotic microbial abundance was not possible due to low populations and sequence numbers in the deepest samples (approx. 50m).

A comprehensive literature review for biological indicators of terrestrial GDEs is scoped for completion in future years, concurrent with mapping of these ecosystems.

4.3.1 Vegetation Indicators of Aquatic GDEs

Vegetation indicators of GDEs are inclusive of vascular and non-vascular plants. Generally, vegetation is indicative of local conditions because individual plants are fixed in space and must respond to local conditions. The literature regarding vascular and non-vascular plant indicators for GDEs in Alberta and Canada is sparse. Springer et al. (2015) detected 25% of Alberta's native plant taxa in southern Alberta spring habitat. What may be most useful for future GDE mapping efforts in the boreal that use existing datasets and map products is the association of specific wetland types with groundwater. Among wetland classes in Alberta, fens are, by definition, influenced by groundwater, whereas bogs are disconnected from it (AESRD, 2015). Fens occur along a gradient from poor to extreme rich, which is understood to align with either or both the quantity or mineral composition of groundwater inputs (Vitt & Chee, 1990). The influence of groundwater in swamps remains poorly understood but presumably present (Elmes et al., 2021).

Existing documentation of species that are associated with water exchange or richer fen types can be leveraged when identifying GDEs within the boreal. In our study area, vascular and non-vascular species including bryophytes such as rusty peat moss (*Sphagnum fuscum*) and narrowleaf peatmoss (*Sphagnum angustifolium*), sedges such as few-seeded sedge (*Carex oligosperma*), and shrubs such as leatherleaf (*Chamadaphne calyculata*) are associated with bogs or very poor fens, while bryophytes such as small greasewart (*Aneura pinguis*), three-ranked thread moss (*Meesia triquetra*), Knieff's hook-moss (*Drepanocladus aduncus*), and Cosson's hook moss (*Scorpidium cossoni [syn: Limprichtia cossonii*), as well as sedges such as tufted clubrush (*Scirpus cespitosus*), and flowering plants such as sticky false asphodel (*Triantha glutinosa*) are indicative of alkaline, rich, wet fen types (Glaser et al., 2004; Vitt et al., 2022; Vitt & Chee, 1990). However, the use of vegetation indicators of groundwater influence should be tempered by several factors. First, fen communities that are out of sync with groundwater inputs can occur where surface runoff due to snowmelt is a substantial input (Cooper & Andrus, 1994). Second, the boreal is dominated by a stress-tolerant, generalist flora(Crisfield et al., 2019) ; the flora that occupies boreal peatlands shows high fidelity to peatlands, but single species or groups of species are rarely perfect indicators of specific peatland types and their underlying groundwater conditions. For example, Laroque et al. (2016) and Munger et al. (2014) investigated indicator species of groundwater exchange between a peatland and an aquifer in Quebec, with Laroque et al. (2016) stating that "the identified species (or combinations of species) do not have a 100% indicator value. There is a clear tendency for the species to be indicative of groundwater inflow, but they are not perfect indicators since an indicator species can be found in a peatland where there is no groundwater inflow, and vice versa." (Larocque et al., 2016).

A species summary table provided by (Jeglum, 1991), who aimed to classify wooded peatlands in Ontario using plant indicators, shows very few species being exclusively restricted to specific peatland types. In a study focused on vegetation patterning and landscape evolution in the Hudson Bay Lowlands, the authors note that “no species were solely restricted to bogs'' when assessing multiple peatland types (Glaser et al., 2004). Instead, it is often the presence, absence, and abundance of species within a community and species richness that together discriminate areas of greater or lesser nutrient status, i.e., groundwater input, and help to distinguish among peatland conditions. An investigation of plant assemblages and water flow from Sweden found higher species richness in areas of groundwater discharge, and that discharge effectively extended the distance at which riparian conditions persist from lotic systems (Kuglerová et al., 2016). In a review, Land & Peters (2023)note that the richness of aquatic vegetation tends to be greater in areas with groundwater inputs. Notably, aquatic vegetation has been proposed as an indicator of reclamation progress in the OSR, where aquatic indicator species have been shown to be responsive to salinity, water and sediment nutrient levels, and alkalinity (Rooney & Bayley, 2011).

Patterns among bryophytes, i.e., non-vascular plants, inclusive of mosses, hornworts, and liverworts, are consistent with those found in vascular plants, although bryophytes may display relatively greater indicator value for specific peatland types (Vitt et al., 2022). Lehosmaa et al. (2018) found that specialist aquatic bryophyte species declined significantly, but generalist aquatic species did not, in the presence of contaminants in boreal spring ecosystems in Finland. An interesting component of boreal flora is saline wetlands, which host a relatively unique vascular plant species assemblage and a lack of bryophytes that is rarely found elsewhere. Saline wetlands host species including flowering plants such as marsh samphire (*Salicornia europaea*), saline plantain (*Plantago eriopoda*), and willow (*Salix* sp.), which are rarely found in other wetland types, as well as grasses such as seaside arrow-grass (*Triglochin maritima*), sweetgrass (*Hierochloe hirta [syn: Anthoxanthum hirtum]*), foxtail barley (*Hordeum jubatum*), and Nuttall’s alkali grass (*Puccinellia nutalliana*), which are found in other saline or non-saline habitat types, including some uplands. As described in Wells & Price(2015)saline fens are exceptionally rare in boreal Alberta, typically found near rivers and, rarely, far from river systems. Species occupying these saline systems have few observations across the boreal and are therefore poor candidates for spatial modeling or generally for widespread monitoring initiatives, but their observation can provide a clear indicator of groundwater influence or otherwise unique environmental conditions in the boreal.

4.3.2 Invertebrates, Microbes, and Stygofauna as Indicators of Aquatic GDEs

Animal and microbial taxa are critical components of GDEs, contributing substantially to their functioning (Hancock et al., 2005). To date, very limited work to document or relate faunal and microbial components to groundwater attributes has been done in Canada. Examples from Canada have sought to understand pattern in macroinvertebrate taxa within areas of groundwater exchange in streams (i.e., the hyporheic zone) in Ontario (Fraser & Williams, 1998; Williams, 1993) and to characterize archaeal, bacterial, and eukaryotic community diversity and structure in aquifers and their connected surface water in Quebec (Groult et al., 2023). While these taxa are currently understudied in relation to GDE, advancements in techniques including genomic approaches, may provide effective monitoring opportunities.

It is expected that areas of upwelling groundwater should create niches for macroinvertebrates, but research in this area is limited (Land & Peters, 2023). Macroinvertebrate assemblages in the hyporheic zone in Ontario have not shown tight associations specifically with either groundwater or surface water, suggesting community (Fraser & Williams, 1998; Williams, 1993). In Finnish boreal springs, macroinvertebrates were shown to shift in taxonomic richness and community composition in response to contamination from nitrates and to contamination from nitrates and increased dissolved organic carbon from land drainage , respectively (Lehosmaa et al., 2018). In Japan, a recent study collected data of benthic invertebrates from literature on a global scale analyzing their taxonomic and biological habitats and presented biological indicators to evaluate the degree of dependency on groundwater springs (Sun et al., 2020).

Microbial communities in groundwater are unique relative to surface waters and are important for biodiversity, nutrient cycling, including carbon cycling, and contaminant mobility (Land and Peters, 2023). The work of Groult et al. (2023) in Quebec showed significant differences in the microbial communities of aquifers and surface waters, assessed using an amplicon sequencing approach, but the authors note that research in this area is in its infancy. Groundwater ecosystem diversity, based on similar taxa discussed in Groult et al. (2023), was assessed using eDNA sequencing in kettle hole wetlands in Germany, with the authors concluding that this approach was useful for cross-domain biodiversity assessment, but limited for single-taxa assessments (Ionescu et al., 2022). Febria et al. (2012) report that the bacterial community of the hyporheic zone of an intermittent stream in Ontario was responsive to various groundwater parameters, including water intermittency, temperature, and phosphate concentration. In a novel study relating macroinvertebrates, bryophytes, periphyton, and bacterial ecosystem components of Finnish boreal springs to land use intensity and groundwater contamination, Lehosmaa et al. (2023) found that bacterial communities shifted in response to groundwater contamination. Bacterial communities were assessed using DNA sequencing techniques (Lehosmaa et al., 2018). Recent work demonstrated that diverse microbial communities are widespread and surprisingly abundant in Albertan aquifers, particularly in older and deeper groundwaters (Ruff et al., 2023).

Stygofauna is a collective term describing animal species that are adapted to and live within groundwater. Generally, stygofauna are believed to be critical to groundwater ecosystems for their role in trophic structure, mediating microbial assemblages, and bioturbation (Hose et al., 2022). Observations of stygofauna inhabiting caves and groundwater in Alberta is limited to species descriptions from single locations near Rocky Mountain House and Castleguard Cave, a Cambrian limestone cave near Banff (Bousfield & Holsinger, 1981; Holsinger, 1980). However, stygofauna are better understood elsewhere, e.g., parts of Europe and Australia and have been strongly supported as indicators of GDEs that are known to decline with declining water quality (Hancock et al., 2005). Stygofauna may be relevant indicators of groundwater-surface water exchange. Research has identified that stygofauna may be both good indicators of GDE and useful in measuring changes in groundwater contribution and quality in GDE systems (Graillot et al., 2014; Hose et al., 2022). Stygofauna were used to identify groundwater upwelling in a Swiss study seeking to improve integrative mapping of surface and groundwater interactions. The authors used absolute and relative stygofauna richness, stygofauna abundance, and the ratio of stygobite to epigean species as indicators of groundwater presence, obtained from samples taken at 50 cm depths (Graillot et al., 2014). Interestingly, they found no significant correlation between hydraulic and groundwater fauna metrics, but found "surprisingly good congruence of results" between them (Graillot et al., 2014). The authors note that the indicator value of stygofaunal metrics may be more limited in regions affected by the most recent glaciations due to low presence of these species, although Holsinger (1980) posited that refugia for stygofauna may have been present in Alberta during glaciation. Finally, Graillot et al. (2014) note that sampling time and identification of stygofauna require high effort and expertise. A European-based global database providing stygofauna datasets from various research and other investigation, Stygofauna Mundi, is now in operation, but currently has no records from Canada and only a single record from the United States (Martinez et al., 2018). In Australia, eDNA has been proposed as a suitable approach for characterizing subterranean stygofauna (Saccò et al., 2022), in turn allowing for better understanding of subterranean ecosystems where they interact with aquatic systems. In a review, Hose et al. (2022) describe the utility of trait-based approaches to examining stygofaunal responses to change in groundwater quality and quantity. The authors describe that stygofaunal traits tend to be limited, with low variability, likely due to the immense selective pressure of their environment, which makes these taxa highly vulnerable to change (Hose et al., 2022). Finally, an assessment along a 40 km stretch of a large river in France concluded that stygobite fauna had the highest richness and abundance in upwelling zones, were tied to river features such as meanders and morainic hills, and showed little relationship with sediment size (Dole-Olivier et al., 2022).

Macroinvertebrates may be the most straightforward to sample, as several examples of sampling techniques targeted to groundwater-associated species/communities are described in the literature, although we note that some techniques require leaving sampling "pouches" out for most of a calendar year. Macroinvertebrate expertise is also available within the ABMI. There are some known macroinvertebrate data from the region (e.g., CABIN, OSM, ABMI, ECCC) but the utility of the existing data for GDE applications has yet to be investigated. Microbial and stygofaunal community sampling requires more complex sampling equipment that allows the observer to penetrate the substrates (for the former, similar to groundwater sampling methods). However, the literature does provide helpful guidance in understanding microbial operational taxonomic units and stygofaunal genera found elsewhere. Expertise in eDNA and DNA barcoding approaches is well established within InnoTech Alberta and the University of Calgary. And there is ongoing collaborative groundwater microbiome research in Alberta led by University of Calgary in collaboration with Environment and Protected Areas, the Alberta Geological Survey and ABMI through the Alberta Innovates Water Innovation Program (Ruff et al., 2023).

4.3.3 Ecological Endpoints of Aquatic GDEs

The literature review underscored the importance of understanding GDEs as ecosystems, which remains a novel concept in conservation. We currently lack enough understanding of boreal aquatic GDEs to be able to characterize or speculate on how they may respond to anthropogenic stressors, as we do not yet understand their components or interactions with other ecosystems or higher order taxa, e.g., upland areas and mammals. This is highlighted, for example, by Land and Peters (2023), who end their review with a call for research into the relationships between biodiversity and stream and groundwater ecosystems (Land and

Peters, 2023). These fundamental relationships are well understood, with decades of research and documentation, for many above-ground ecosystems.

It is difficult to speculate on the potential vulnerability and responses of biological components of aquatic GDEs to changes in groundwater quality and quantity with the current level of understanding of them in Canada. The Nature Conservancy, in a guide for practitioners seeking to determine groundwater thresholds for ecosystems, describes the importance of cause and effect chains when establishing groundwater thresholds (Rhode et al., 2020). These chains are described for species that directly rely on groundwater, e.g., aquatic fauna like fish or snails, and those that indirectly rely on it, e.g., birds or mammals that use riparian or wooded/shrubby peatland habitat (Rhode et al., 2020). Fish depend on groundwater because it provides stable temperature in the hot and cold extremes of temperate climates, maintains ice-free areas in winter, and provides flow (Power et al., 1999). As we gain insight into the extent of groundwater reliance in boreal ecosystems, we can begin to establish these chains, i.e., stressor-pathway-response relationships.

Licks are places where animals ingest mineral soil, a phenomenon known as geophagy. Licks can be wet or dry; in an exploration of animal use of licks in northern British Columbia, Ayotte et al.(2006) explains that "Wet licks are associated with apparent groundwater springs. Dry licks usually occur along streams or riverbeds, where unweathered deposits of soluble elements have concentrated above less impervious layers, and become exposed by erosion". In northern Alberta, Indigenous Knowledge understands relationships between mammals and licks, which are likely GDEs in many places, but little western science research has focused on this topic. Mammals in the boreal likely utilize wet licks, which are aquatic GDEs. One example from Alaska reported that "local observations suggest that… hare populations in areas with known licks appear to reach higher densities during the population high compared to areas where there is no known lick" (Worker et al., 2015). Based on this, the authors used a captive study that concluded that mineral soil appeared to allow snowshoe hares (*Lepus americanus*) to minimize body mass loss, and they consumed more food when mineral soil was made available to them (Worker et al., 2015). From a stressor-pathway-response perspective, groundwater drawdown in response to water withdrawals could reduce lick availability, which could in turn affect the forage use and functional responses of mammals like hare or moose(Ayotte et al., 2006; Worker et al., 2015). Indigenous insight into such relationships will be invaluable going forward.

An interesting example from Germany also highlights potential ecological endpoints resulting from change in groundwater. In brief, the authors speculated that groundwater-connected pothole wetlands in natural grassland and forest areas would be closer to the natural, pre-intensive agriculture conditions than those in agricultural fields. Instead, they found evidence of biotic homogenization across all ponds, presumably a product of intensive land use and landscape-level nutrient enrichment that was propagated across wetlands by groundwater connectivity among them (Ionescu et al., 2022).

As per the conceptual model (Figure 1), it is possible that stressors such as landscape disturbance, operational spills and leaks, and mine dewatering, may alter local groundwater ecosystems in the OSR, and that these effects could be realized at the landscape level due to groundwater connectivity, resulting in biotic homogenization as seen in Ionescu et al. (2022), i.e., negative effects to terrestrial and aquatic ecosystem health, but at this time we lack an understanding of homo- or heterogeneity in GDEs in the boreal. Further, broadly speaking, change in groundwater quantity may relate to or potentially advance peatland drying, which is known to increase vulnerability to wildfire, alter carbon dynamics, shift species composition, and alter the hydrological function of peatlands in their catchments such that stream flows are ultimately affected (Goodbrand et al., 2019; Miller et al., 2015)). While climate change is known to cause peatland drying in the boreal, anthropogenic factors (e.g., roads) associated with oil sands development are also culpable (Miller et al., 2015). The role of groundwater flow in promoting drying is an important area for further consideration, as this phenomenon may act as a pathway causing change in terrestrial and aquatic ecosystems under the stress of reduced groundwater quantity. Finally, it is likely that Indigenous Communities already hold knowledge on groundwater relationships in the boreal, forming the basis for their often-expressed concerns regarding landscape-level drying and a holistic understanding of water drawdown affecting multiple ecosystems.

5. Data Compilation

5.1 Data Considered

For this project, it was important to understand the broad set of data that could potentially be used to inform mapping of GDEs in the chosen study area. Various data that were considered for identifying the presence of GDEs are presented in Table 3. These datasets include data that was collected in the field (e.g., hydrologic data, isotope analysis, water quality), geologic mapping products, remote sensed data (e.g., DEM derived from lidar) and modeled data (e.g., Aquanty HGS output). Considering a broad set of data provided the best opportunity to develop an approach to mapping that would negotiate the variability in resolution, the relative value of the data in identifying GDE, and the availability of data.

Digital Elevation Models (DEMs) and flow-related data are important for providing the terrain layout which influences hydrological processes, and how water moves and accumulates in a landscape. Geological data including bedrock types and their permeability show where subsurface conditions could affect groundwater storage and flow; this impacts the availability of water to sustain GDEs. Groundwater data pertaining to water levels and chemistry help to describe the physical dynamics and quality of the groundwater, which both impact and indicate GDEs. Water use data gives a picture of human extraction patterns, which can alter the availability of groundwater for GDEs. Landcover data, such as ecosite classification, vegetation types, and historical forest fire record, provide context on biological diversity and physical indicators that may influence the likelihood of GDE presence. Isotope sampling and hydrometric monitoring from rivers and lakes provide data on the water cycle, which can support estimation of the relative contributions of groundwater and surface water. Finally, wetland mapping provides information on the types of wetlands present, which, by definition, can indicate the presence of a GDE. Together, these datasets may be used to support the mapping of GDEs but may also be useful for predicting risk and vulnerability of GDE to oil sands-related impacts.

We also requested thermal imagery from both ECCC studies from the McKay River and COSIA, however that data will not be available for broader use, including for GDE mapping, until 2024/25.

Table 3. Types of data that were considered with their purpose. A full list of all the datasets considered is found in Appendix D.

Category	Data Type	Purpose
Geography	DEM, Flow Accumulation/Direction	Terrain analysis for hydrologic studies
	Slope, TWI	Surface inclination and wetness evaluation
Geology	Quaternary Units, Bedrock, Permeability	Geological composition and water transmission analysis
Groundwater	Hydraulic Head, Chemistry	Groundwater pressure and quality assessment
Groundwater & Surface Water	Water Use Data	Monitoring of water usage
Landcover	Ecosite Classification, Fire Polygons	Ecosystem classification and post-fire land assessment
River & Lake Surveys	Water Quality, Isotope Sampling, Hydrometric Monitoring	Water source study and flow measurement
Wetlands	Wetland Monitoring	Wetland water quality monitoring
	Inventory of OSM Area	Wetland mapping and study

5.2 Data Acquired

This dataset (Table 4) is made up of various types of acquired geographical, geological, groundwater, landcover, and wetland information pertinent to the boreal region's aquatic GDEs. This multi-variable dataset, acquired from multiple organizations, ranges from point and line data to more complex rasters and polygons. Geography parameters were available for the analysis area, and are also available for the broader oil sands region. Geology data is mostly available throughout Alberta, however, more detailed information was available for some parts of the analysis area, and availability at the scale of the oil sands region is patchy. There is a concentration of groundwater data available in the oil sands region, when compared to other areas of boreal Alberta, due to energy sector monitoring and targeted work by the AGS and others. Landcover data like soils and ecological classification is in general more detailed in

Southern Alberta, when compared to that available in the oil sands region. There is relatively good coverage of river and lake survey data in the oil sands region.

Table 4. Summary of the data compiled to support the GDE mapping approach.

Category	Data Name	Description	Data source	Date of collection/ publishing	Type of data
Geography	Annual Unit Runoff	Measures the annual amount of runoff in a unit area.	GOC	2013	Point, Line
	Digital Elevation Model (DEM) - Advanced Land Observing Satellite (ALOS)	3D representation of a terrain's surface created from satellite data.	JAXA	2015	Raster
	Flow Accumulation - ALOS Derived	Indicates the accumulation of water flow across a surface.	JAXA	2015	Raster
	Flow Direction - ALOS Derived	Shows the direction of water flow derived from elevation data.	JAXA	2015	Raster
	HUC 8, 10	Hydrologic Unit Codes that identify hydrological features.	GOA	2024	Polygon
	Slope - ALOS Derived	Measures the steepness or incline of a surface.	JAXA	2015	Raster
	Topographic Wetness Index (TWI)- ALOS Derived	Predicts the accumulation of water in a geographic area.	JAXA	2015	Raster
Geology	Paleogeography, Evaporite Karstification	Studies historical geology related to salt cavern potential.	AER/AGS	2020	Point, Report
	Bedrock (Map 600)	Maps the distribution of bedrock.	AGS	2013	Polygon
	Modeled Surfaces and Unit Picks of Quaternary Units	Represents the geological composition of Quaternary units in NAOS.	AGS	2023	Raster, Point

Category	Data Name	Description	Data source	Date of collection/ publishing	Type of data
	Aquifer Hosting Sediments	Sediments above bedrock known or inferred to contain aquifers (sand, gravel, or water supply wells)	AGS	2023	Polygon
	Permeability - derived from geological materials	Indicates the capacity of rock materials to transmit water.	AGS	2013; 2022	Polygon
	Surficial geological maps (bedrock)	Maps the surface geology.	AGS	2013	Polygon
	Surficial geological maps (bedrock) - updated - Maps 618-621	Maps the surface geology in higher detail.	AGS	2022	Polygon
Groundwater	Integrated Surface Water-Groundwater Model for the Athabasca River Basin	Simulates water table depths and groundwater flows. (Aquanty/HGS)	OSM, Aquanty	2022	Raster, Point
	Groundwater Protection Data	Provides estimated elevation for the base of the deepest formation that is likely to contain non-saline groundwater.	AER	2016	Point
	Total Dissolved Solids Distribution	Indicates the concentration of dissolved substances in groundwater.	AGS	2021	Raster
	Distribution of Hydraulic Head in the Peace River / Viking / Bow Island Hydrostrati-graphic Unit	Measures the pressure exerted by groundwater at various locations.	AGS	2021	Raster

Category	Data Name	Description	Data source	Date of collection/ publishing	Type of data
	Map 596 (Distribution of Total Dissolved Solids in the Grand Rapids Hydrostrati-graphic Unit)	Indicates the concentration of dissolved substances in groundwater.	AGS	2020	Raster
	Map 597 (Distribution of Hydraulic Head in the Grand Rapids Hydrostrati-graphic Unit)	Measures the pressure exerted by groundwater at various locations.	AGS	2020	Raster
	Map 612 (Distribution of Total Dissolved Solids in the McMurray Hydrostrati-graphic Unit)	Indicates the concentration of dissolved substances in groundwater.	AGS	2021	Raster
	Map 613 (Distribution of Hydraulic Head in the McMurray Hydrostrati-graphic Unit)	Measures the pressure exerted by groundwater at various locations.	AGS	2021	Raster
	Operators/EIAs GW Chemistry	Chemical analysis of groundwater by operators or EIAs.	Various	2021	Point
	Operators/EIAs Water Levels	Water level measurements taken by operators or EIAs.	Various	2021	Point
	Spring Compilation (AGS)	Compiles locations and data of springs.	AGS	2014	Point
	Spring Compilation (InnoTech)	Compiles locations and data of springs.	InnoTech	2022	Point
	Thalwegs	Depicts the path of the deepest part of a stream or valley.	AGS	2018	Line

Category	Data Name	Description	Data source	Date of collection/ publishing	Type of data
Groundwater and Surface Water	2022 Water Use Data	Data on the usage of water resources.	AER	2022	Polygon
	Water Quality	Groundwater and surface water chemistry.	OSM	2016-2022	Point
Landcover	Eco_AB_10TM - 4 scales of ecosite	Provides scales of ecosite classification.	Agricul-ture Canada	2021	Polygon
	Ecosystem Based Management	Includes various layers related to land management and classification.	ABMI	2022	Polygon
	Forest Fire Polygons	Maps the areas affected by forest fires.	AFP	2022	Polygon
	Soil Landscapes of Canada	Maps the distribution and types of soil landscapes in Canada.	Agricul-ture Canada	2011	Polygon
River and Lake Surveys	Isotope Sampling	Collects data on isotopes for water sources studies.	ISO-ABMI; InnoTech	2009-2018	Point
	RAMP Hydrometric Monitoring Locations	Locations where water level and flow are monitored.	RAMP	2017	Point
	RAMP Water Quality Monitoring Locations	Surveys to measure and track long-term water quality.	RAMP	2017	Point
River Surveys	Electromagnetic Terrain Conductivity Mapping	Delineates zones of groundwater pore fluid in river bottom sediments with elevated salinity.	Advisian, InnoTech, OSM	2014, 2015	Raster, Point
	Water quality/LTRN	Surveys to measure and track long-term water quality.	AEP	2023	Point

Category	Data Name	Description	Data source	Date of collection/ publishing	Type of data
	WSC/RAMP Stream Gauging	Measures the volume of water flowing through rivers.	WSC	2023	Point
Wetland Surveys	Wetland Monitoring Surface Water Quality	Monitors the quality of surface water in wetlands.	OSM	2022	Point, Table
	OSM Wetland Inventory Pilot Area	Inventory area for pilot studies on wetlands using OSM data.	ABMI/ DUC	2022	Raster

5.3 Data Gaps

There were multiple data sources that would have been useful in the GDE approach that was applied in the area of interest, or which could have supported other approaches. Some key gaps include the McKay River Integrated Surface Water-Groundwater Model, hydraulic head maps, and higher resolution thermal data; all of which are continuing to be pursued (Table 5). InnoTech Alberta will continue to work to gain access to the output from the McKay River model, improved hydraulic head maps for quaternary aquifers may be available from AGS in 2024/25 for a portion of the oil sands region, and thermal imagery access opportunities from additional sources will be further explored. InnoTech Alberta is discussing opportunities to access data collected by the Fort McKay Métis Nation to support future GDE mapping initiatives; focus of discussions includes clarity on data protection and management, data use limitations, and requirements to support community interests in sharing back outcomes.

Table 5. Ongoing gaps in data availability that could potentially contribute to enhanced GDE mapping.

Data Type	Description
Biological data	Biomass, species, diversity, population, and productivity
Climate data	Information on climate patterns
McKay River Integrated Surface Water -Groundwater Model	GIS model outputs, including groundwater levels and surface water - groundwater interactions, are needed from EarthFX
Higher resolution thermal data	Airborne or satellite derived thermal data at a higher resolution than what is available through GEE

Hydraulic head maps	Hydraulic head data may be available from AGS, particularly for the quaternary units, at a later date
LTRN Data is outside of our AOI	Only two points near the AOI, data would need to be acquired.
Surficial and near-surface karst mapping	Not much info available in the OSR for surficial or subterranean karst feature mapping
Springs	Additional locations of springs
Temporal resolution data	Lack of time series data

6. Literature Review: Methods for Mapping GDEs

Global approaches to GDE mapping have been limited to date, however, with the rise in population and economic growth, strain on water resources may drive an increased need for national or multinational approaches. The mapping of GDEs has been identified as important to support both ecosystem protection and human health. Saccò et al. (2024) found that ~75 percent of the global land surface has an interaction with groundwater resources when high mountain and desert terrain are excluded. Additionally, the authors draw attention to the transboundary nature of groundwater and the impact of declines on surrounding biodiversity and human water needs (Saccò et al., 2024). As demands on groundwater resources via human abstraction intensifies, the need for reliable GDE mapping will increase to enable use of these systems as critical proxies of aquifer health. Link et al., (2023) globally grid-mapped key GDE potentials, with indicators related to GDEs based on type (streams, wetland, vegetation) and further refined using 16 GDE indexes. These were coupled with stressors to groundwater and used to highlight regions where GDEs are predicted to be at a higher risk of impact. Although globally mapped northern hemisphere regions show lower risk scores in remote areas, regions with higher human footprint reflect higher GDE impact risk, showing the need to better map GDEs in less disturbed areas that have potential of being impacted by groundwater withdrawals as they are developed (Link et al., 2023). Global-scale mapping is primarily limited to low resolution approaches to date, for instance these authors used a spatial resolution of 0.5°. For local and regional-scale monitoring, higher resolution outputs are required.

Building foundational knowledge around the occurrence and locations of different types of GDEs within the OSR that can support groundwater monitoring efforts and stressor-response pathways, requires accurate and spatially comprehensive regional-level mapping. The methods used to map GDEs vary according to the category of GDE being mapped (i.e., terrestrial, aquatic, or subterranean). Existing efforts have largely focused on terrestrial and aquatic GDEs and can rely on indicators visible or measurable at the surface (e.g., water chemistry, phreatophytic vegetation). Subterranean GDEs, however, require an alternate approach that focuses on less easily observable subsurface factors including geology, lithology, and structure. As is the case for previous sections, the following focuses on data and approaches used for mapping aquatic GDE systems.

Aquatic GDE identification can come from a variety of indicators. For instance, the review of surface water expressions (e.g., springs, seeps, wetlands, riverine systems) for signs of links with groundwater discharge, or data on surface water temperatures and chemistry, are used to infer GDE occurrence. Using these data to understand the hydrogeological foundation of an area is a common approach to GDE mapping (Martínez-Santos et al., 2021; Saccò et al., 2024), and can also be complemented by the use of biological indicators, such as through comparisons with ecological responses (e.g., wildlife or vegetation communities linked with a reliance on water table levels (Doody et al., 2017; Link et al., 2023). Often such data are collected directly on the ground and are only available for a limited number of discrete locations across an area due to the high costs of acquisition. For this reason, the use of remote sensing datasets and geospatial technology have become important tools for broader-scale, spatially explicit GDE mapping as they can offer comprehensive, larger-scale, repeating views of the Earth's surface at a fraction of the cost of ground-based data collection.

6.1 Remote Sensing-Based GDE Mapping

Within the context of mapping GDEs, progress in their delineation and detection have drastically improved with advances in remote sensing technology and computational power, and range from simple spatial analysis methods combined with expert opinions (Doody et al., 2017), to a greater reliance on computing power using machine learning approaches to map GDEs (Fildes et al., 2023; Martínez-Santos et al., 2021; Rohde et al., 2021; Rosa et al., 2023). The following sections first describe remotely-sensed datasets that are most commonly used for mapping GDEs, and then summarize three types of mapping approaches in which these datasets are typically used.

6.1.1 Remote Sensing Data

There are two main types of remotely-sensed data that are found to be particularly useful for mapping GDEs: spectral vegetation indices, and thermal imagery. The following sections introduce these two types of datasets, describe their use in various jurisdictions for mapping GDEs.

Spectral Vegetation Indices

Much of the published mapping efforts that leverage remote sensing data have focused on arid and semi-arid regions, where differences in vegetation vigor (e.g., greenness) are notable between more hydrologically-stable, greener aquatic or terrestrial GDEs, such as wetlands or riparian areas, and drought-prone non-GDE systems. Under these conditions, remote sensing approaches are able to rely on spectral vegetation indices (SVIs) that often combine visible red and near infrared wavelengths, which respond to levels of vegetation health or vigor. Popular SVIs used for this purpose include the Normalized Difference Vegetation Index (NDVI), and the Normalized Difference Wetness Index (NDWI). GDEs remain more vegetatively green throughout the year because of their more consistent supply of water, even during dry seasons, and these indices leverage this phenomenon. Additional SVIs that have also been used for the detection of both aquatic and terrestrial GDEs are listed in Table 6. While these indices may leverage different spectral bands or different combinations of bands, these are nevertheless more often utilized for detecting GDEs within regions where the differentials between the source of water (surface water vs. groundwater) are distinct in nature (i.e., arid, semi-arid regions). In these environments, higher contrasts can be found between ecosystems receiving consistent hydrological sources (e.g., GDEs) and those which have high seasonal variation in water availability (Fildes et al., 2023; Martínez-Santos et al., 2021; Rohde et al., 2021). While SVIs like NDVI or NDWI are particularly useful in arid and semi-arid regions, additional data sources are also used in combination with these to support GDE detection and mapping. Such ancillary datasets with information on: geology, lithology, piezometric surfaces, elevation, slope, aquifer permeability, soils, and flow accumulation potential (e.g., Martinez-Santos et al., 2021).

Table 6. List of spectral vegetation indices used in the remote sensing-based mapping of GDEs.

Spectral Index	Recent Publications Using the Index
Normalized Difference Vegetation Index (NDVI)	Rohde et al. (2021); Martinez-Santos et al. (2021); LaRocque & Leblon (2022)
Normalized Difference Vegetation Index (NDVI) - Coefficient of Variation ($NDCVI_{NDVI}$)	Fildes et al. (2023)
Normalized Difference Vegetation Fractional Cover Photosynthetic Vegetation - Coefficient of Variation ($NDCVI_{PV\ FC}$)	Fildes et al. (2023)
Difference Vegetation Index (DVI)	LaRocque & Leblon (2022)
Green Difference Vegetation Index (GDVI)	LaRocque & Leblon (2022)
Green Ratio Vegetation Index (GRVI)	LaRocque & Leblon (2022)
Normalized Green (NG)	LaRocque & Leblon (2022)
Normalized Near Infrared (NNIR)	LaRocque & Leblon (2022)
Normalized Red (NR)	LaRocque & Leblon (2022)
Red Simple-Ratio Vegetation Index (RVI)	LaRocque & Leblon (2022)
Enhanced Vegetation index (EVI)	Rohde et al. (2021)
Normalized Difference Aquatic Vegetation Index (NDAVI)	LaRocque & Leblon (2022)
Normalized Difference Moisture Index (NDMI)	Rohde et al. (2021)

Normalized Difference Water Index (NDWI)	Rohde et al. (2021)
Normalized Burn Ratio (NBR)	Rohde et al. (2021)
Normalized Difference Evapotranspiration (NDET)	Fildes et al. (2023)
Near Infrared reflectance of Vegetation (NiRv)	Rohde et al. (2021)
Soil Adjusted Vegetation Index (SAVI)	Rohde et al. (2021)
Water Adjusted Vegetation Index (WAVI)	LaRocque & Leblon (2022)
Tasseled Cap (TCAP) derivatives: Brightness, Greenness, Wetness, Angle	Rohde et al. (2021)

Thermal Imagery

Groundwater-surface water interactions often involve distinct water temperature differences between the two water sources. While groundwater retains a steady temperature throughout the year as it is insulated from surface seasonal and daily atmospheric temperature changes, the temperatures of surface water vary with the seasons and local weather conditions (e.g., amount of incoming solar radiation, levels of precipitation, etc.). Thus, depending on the time of year, groundwater can be much cooler than surface water or vice versa. Variations in water temperature can be used as an indicator of groundwater presence, and have even been leveraged for empirically quantifying groundwater-surface water interactions and fluxes (Anibas et al., 2011; Kløve et al., 2011). Examining spatial variations in water temperatures lends itself well to remote sensing approaches - levels of thermal infrared energy is commonly captured using a variety of space-based, airborne, drone-based, or handheld sensors and cameras.

Ala-aho et al. (2015)used airborne thermal images captured from a helicopter over an unconfined aquifer study area in central Finland, to manually identify temperature gradients along a lake shoreline. From these they identified locations of groundwater inflow into the lake, which were used to validate a fully integrated hydrological model of water fluxes in the area. Isokangas et al. (2017)similarly use helicopter-based thermal imagery and temperature thresholds to delineate locations of groundwater seepage in a Finnish peatland. They compared this with the outputs of an isotope mass balance mapping approach and showed success in mapping groundwater contributions to peatland pore water in the area. Both Autio et al. (2023) and Watts et al. (2023) leverage newer drone technology to capture local and highly detailed thermal image mosaics of wetlands in northern Finland and Massachusetts, U.S.A, respectively. The former show that temperature anomalies identified through manual and threshold-based delineation from thermal imagery can align well with the stable water isotope-based identification of groundwater seeps, and with a physically-based flow model of the area (Autio et al., 2023). The latter used thermal imagery for more than model validation, however, mapping groundwater seeps with drone-based thermal imagery both pre- and post-restoration to show the successful removal of barriers to surface expression at the site (Watts et al., 2023).

Using thermal imagery to identify groundwater-surface water interactions and therefore the GDEs that accompany them shows great potential in the literature. However, as is the case for other approaches, it does not come without its limitations and challenges. For instance, the time of year and time of day at which the thermal imagery is captured are both important for ensuring maximum temperature differences between different water sources to enhance attribution ability. That is, times of year and day when surface water is at its coldest or hottest in comparison to more stable groundwater (e.g., mid-winter or late summer; evening, early morning) should be selected so as to maximize groundwater detection. Thermal imagery also requires significant amounts of calibration and processing post-acquisition if actual water or surface temperatures are being extracted and, even with careful post-processing, variabilities across images can come from weather conditions, camera angles, within-waterbody temperature stratification, differing vegetation thermal properties, etc. (Autio et al., 2023; Isokangas et al., 2017; Watts et al., 2023) .

Remote Sensing Mapping Approaches

Approaches leveraging remote sensing for the mapping of GDEs are divided into three main types of methods: integrated hydrological modeling, suitability /risk mapping, and machine learning approaches. Each of these is described in more detail in the sections below, with approaches used in the boreal identified.

Integrated Hydrological Modelling Approaches

The combination of fully-integrated hydrological models, isotopes, and thermal mapping has shown promise for determining GDEs over a relatively small catchment (approximately 100 km^2) in peatland environments, but lacks the spatial capacities of larger regional studies due to cost (Ala-aho et al., 2015; Autio et al., 2023; Eskelinen et al., 2015). Despite these advances, mapping GDEs within peatland regions is still hindered by the low variability in vegetation vigor, and therefore spectral vegetation properties as seen in arid or semi-arid areas, due to the relative hydrological stability in more humid peatlands. Furthermore, despite the availability of open access satellite-based thermal imagery for larger-scale mapping, the lower resolution of these data (i.e., ≥ 60 m pixel sizes) makes the detection of smaller GDEs more difficult and the delineation of larger GDEs less exact. Most studies which have utilized thermal imagery as a mapping approach in peatlands are typically small in nature and leverage high resolution drone thermal infrared imagery to detect groundwater seeps on the landscape (Isokangas et al., 2017; Watts et al., 2023). This approach would not currently be feasible for larger-scale applications (e.g., across the OSR).

Suitability Mapping Approaches

Suitability maps have been used to identify terrestrial and aquatic GDEs in both arid and boreal settings (Doody et al., 2017; Eskelinen et al., 2015; Kuginis et al., 2016), and typically rely on three major remote sensing data groups: vegetation community mapping and associated data (e.g. NDVI, NDWI), groundwater levels, and layers pertaining to aquifer types and soil parameters (Eskelinen et al., 2015; Doody et al., 2017). These approaches rely on a workflow that uses user-defined scores and decision rules to normalize variables. Users then assign weight relative to each variable, which calculates the likelihood that any given location within an area of interest is a. These methods are susceptible to bias from experts in assigning weights, and do have limitations with regards to only having confirmed (positive) points, which can influence computational power in regions with sparse data availability (Doody et al., 2017; Fildes et al., 2023; Malczewski, 2004).

Within the context of a national approach for mapping GDEs in Australia's arid environments, Doody et al., (2017) follow a workflow incorporating 7 methodologies:

- a literature review involved assessing 200 reports from Eco-Hydrological Zones (EHZ) and expert consultation;
- collation of continental spatial data with the division into 57 eco-hydrogeological zones fundamentally based on climate, geology, and groundwater flow systems;
- the development of rules on groundwater dependency, based on both literature and expert consultation to determine the criteria required for GDE systems;
- collation of vegetation and hydrological spatial data (e.g., locations of wetland, river, springs), in conjunction with remote sensing derived products (MODIS satellite imagery at a 250 m resolution, to make inflow dependent ecosystem products);
- an assessment of where spatial layers intersect with locations classified as potential GDEs (e.g., surficial geology and vegetation features);
- the normalization and weighting of developed rules, based on expert opinions; and
- calculation of GDE potential across the area of interest, based on expert-derived rules multiplied by the weightings, and divided by the sum of total weightings.

The validation of GDEs within this Australian study required the utilization of "known GDE" locations garnered from previous literature, with the assumption that they were highly accurate in nature, and also depended heavily on expert opinions (Doody et al. 2017). Some of the limitations of using this method were the broadness of the study, resulting in lower spatial detail in the detection of GDEs, and challenges with knowledge gaps, where mapping layers did not overlap within all regions. Expert opinion was heavily relied upon in both the validation approaches and in initial data sourcing (Doody et al., 2017). The heavy reliance on vegetation data to corroborate GDE presence, in conjunction with large dependence on expert opinion, makes it more difficult to apply this methodology within the boreal region.

The use of GDE likelihood mapping within boreal peatland in Finland by Eskelinen et al., (2015) shows some promise for use within the boreal, despite the small scale of the study. Mapping a relatively small catchment area of approximately 7 km^2, Eskelinen et al. (2015), used a methodology that relied more heavily on hydrological inputs and models (Darcy's Law for flow, and Hydrogeosphere (HGS) for model validation), and leveraged the following inputs:

- Slope, utilizing Darcy's Law, whereby hydrological flux is dependent on discharge area and hydrological gradient (Freeze & Cherry, 1979)), with verification from a previous study in the region verifying that groundwater level within an esker peatland system followed the topographic plane (Rossi et al., 2012).
- Natural spring systems were used as a metric for areas which might have discontinuous geological structures allowing for groundwater seeps to be assigned a higher potential of being a GDE; this was done in a stepwise function (100 m) to a total of 1 km distance. Inversely, the probability of discharge was reduced at 500 m intervals until 3 km from the boundary of the recharge area.

- Peat thickness, as a function of interpolations between measurements with likelihood determined at slope, divided by the inverse peat thickness layer.

Validation of the models were conducted using two methods: (1) field assessments of known GDE points with existing base flow measurements and 15-cm LiDAR digital elevation model delineations over two years; and (2) the use of HydroGeoSphere groundwater modeling software based on Ala-Aho et al. (2015), which describes a fully integrated groundwater model (Eskelinen et al., 2015). Of the two models generated, the model using basic inputs (i.e., slope and springs) had good predictive abilities and resolution, while including the third input (i.e., peat thickness) increased GDE detection resolution in areas with data, but lowered resolutions in regions lacking accurate data and increased computational demands (Eskelinen et al., 2015). The use of likelihood maps for mapping of GDEs with the boreal has potential. However, the reliance on detailed spatial data, and hydrological modelling inputs (detection (Darcy's law) and validation (HGS)) might limit the accuracy of GDE predictions in areas that lack substantial data coverage and high-quality data layers.

Machine Learning Approaches

There have been marked improvements in the use of machine learning approaches combining multiple data sources to detect GDEs, which enable the processing of large amounts of remote sensing and other geospatial data (Rampheri et al., 2023). Recent studies focused on leveraging the use of machine learning in the mapping of GDEs in both arid and boreal environments. Within these studies, model variability ranged >20 with the most common machine learning algorithms being random forest, support vector machine, artificial neural network, naive Bayes classification, and maximum entropy modeling. This section will review three papers in depth. Of the three studies reviewed, most implemented tools and packages for common geospatial rendering software including QGIS - MLMapper 2.0 (Martínez-Santos et al., 2021), Google Earth Engine - Shallow Groundwater Estimation Tool (SAGE) (Rohde et al. 2021), and Maxent software (https://biodiversityinformatics.amnh.org/open_source/maxent/, Gerlach et al., 2022. The following paragraphs describe some of these methods in more detail.

Within the context of arid and semi-arid environments, Martínez -Santos et al. (2021) developed and employed the ML Mapper tool - a multi-layered supervised classification approach that leveraged several data layers to map GDEs within a 6100 km^2 groundwater aquifer system in Spain, similar to the size of our Study Area. Their approach utilized the following explanatory inputs in the MLMapper software: a digital elevation model (DEM), a piezometric surface, topographic wetness index, slope, NDVI, flow accumulation, geology, aquifer permeability, and soil type. These were each broken into 4-5 subclassifications typically ranging from very low to very high potential for GDEs, with quantifiable metrics such as slope, groundwater table, NDVI classed into numerical bins. The approach then trains, tunes, and cross-validates a multitude of machine learning algorithms simultaneously. These include: support vector machines (SVM), linear vector machines (LVM), logistic regression (LRG), decision tree classifier (CRT), random forest classifier (RFC), K-nearest neighbor classification (KNN), linear discriminant analysis (LDA), gaussian naïve Bayes classification (NBA), multilayer perceptron neural network (MLP), ada-boost classifier (ABC), quadratic discriminant analysis (QDA), gradient boosting classification (GBC), gaussian process (GPC), ridge (RID), stochastic gradient descent linear classifier (SGD), perceptron (PRC), nearest centroid classifier (NCC), multinomial naive Bayes (MNB), complement naive Bayes (CNB), and extra-trees classifier (EXT). This allows for both low-suitability models and collinearity between variables to be excluded, and for the effective leveraging of multiple spatial input layers in one pass.

Model tuning and validation includes automated parameter tuning using 10-fold cross-validation, and the removal of counterproductive and redundant variables. The authors used a 50/50 training vs. test data split in this study (Martínez -Santos et al., 2021), with 150 total reference data points split into 75 known points over the 6 major wetlands, and the remaining 75 points spread over non-GDE points across the aquifer. The MLMapper plugin allows the user to extrapolate the results to produce a predictive map. Martinez-Santos et al. (2021) found that tree-based classifiers (e.g., RFC, EXT), in addition to LRG, SVM, and KNN all performed well in their ability to map GDEs in their study area. The highest model confidences were produced by the tree-based classifiers and the use of only 4 explanatory variables: DEM, lithology, permeability, and water table elevation. This method, MLMappper, has the potential to be useful for mapping GDEs in the boreal as there is less dependence on vegetation derived indices, and more of a focus on hydrological and hydrogeological factors (wetlands, lakes, stream, water table, lithology, permeability). This makes it an ideal choice when combined with its high computational capability to leverage multiple geospatial input layers at once.

Google Earth Engine (GEE) and machine learning have been used to map terrestrial and aquatic GDEs in California, alongside risks associated with groundwater level changes. Rohde et al. (2021) leveraged random forest modeling to assess risk for GDEs in groundwater level changes. The methodology used a number of inputs, divided into two types of variables: dynamic or categorical. Dynamic variables included: groundwater elevations from shallow wells over multiple years (roughly 55.6K sample points total), and GDE maps based on vegetation indices derived from Landsat data. The indices used were: NDVI, NDMI, NDWI, Normalized Burn Ratio NBR, NiRv, SAVI, EVI, and TCAP indices (see Table 6). GDE mapping also employed a downscaled climate surface (NASA's Daymet; https://daymet.ornl.gov/) for predictions of climate. Both the Landsat and climate data were fed into a temporal segmentation, which distributes the data into time series segments feeding into the model. Categorical variables used by Rohde et al., (2021) included: watershed boundaries at a HUC 8 level, hydrological region, ecoregion, and vegetation typing. Both categorical and dynamic variables are fed into a random forest machine learning model to detect risks of groundwater level changes on local GDEs (2021).

The results of the study leveraged large data sources to map GDEs to demonstrate how machine learning can inform risks to GDE once they are successfully mapped. Although this approach is ultimately assessing GDE risk to groundwater fluctuations and not directly mapping them, a similar method has been incorporated into a study mapping GDEs across the globe with a random forest approach. The latter is set to be published in the spring of 2024 (M. M. Rohdes, personal communication, January 23, 2024). This new approach has not been used to map boreal Alberta but has mapped approximately 61,600 km^2 of southern Alberta prairie pothole region (Nature Conservancy & Desert Research Institute, 2023), which exemplifies the power of GEE and machine learning for scaling up GDE mapping efforts. This approach could be suitable for the boreal region in the future, depending on data availability.

Mapping of GDEs across landscapes, like all predictive modeling, requires validation, preferably using independent data or methods. These data can range from known GDEs points from previously studied areas ranging from known seeps such as springs, wetlands, reaches of river, and lakes system to informative point source data like stable isotopes and geochemistry (Doody et al., 2017; Eskelinen et al., 2015; Fildes et al., 2023; Gerlach et al., 2022; Isokangas et al., 2017; Klausmeyer et al., 2018; Lidberg et al., 2020; Martínez-Santos et al., 2021; Rohde et al., 2021). These known points within a machine learning approach serve as an important source of training data for the models and highlight the need for both positive (known GDEs) and negative (non-GDE) points for both training and model verification (Lidberg et al., 2019; Martínez-Santos et al., 2021; Rohde et al., 2021; Fildes et al., 2023). There is also the potential for Indigenous communities to support identification of known GDEs such as seeps, salt licks, and other prominent groundwater features to assist in mapping efforts to be used as additional training or validation data, with limitations on data use and protection that reflect community requirements.

6.2 Approaches in Boreal Environments

Much of the effort using remote sensing to map GDEs has focused in arid and semi-arid environments, like Australia or drier portions of the U.S.A., where differences in vegetation vigor are noticeable between GDE and non-GDE environments because the former maintain some level of green vegetation given their more stable access to groundwater. SVIs are particularly useful for mapping these systems. Unlike arid or semi-arid regions, however, more humid boreal regions at higher latitudes must rely on alternative approaches that often leverage local biotic or abiotic context. With high-latitude climates like the boreal forest, reliance on methods that are dominated by NDVI and dependencies on phreatophytes (deep rooting plants) are far less efficient, posing a challenge as water availability (surface and groundwater) is sufficiently high so as not to cause high amounts of water stress outside GDE systems, thereby making negligible contrast in vegetation vigor between GDE and non-GDE systems. The mapping and detection of GDEs in high-latitude boreal areas has begun to grow as concerns around water security and anthropogenic impacts on these systems increase (Kløve et al., 2011). Some of this work has centered around the use of thermal imaging from helicopters and drones, in conjunction with stable isotopes (^{2}H, ^{18}O), as tools for accurately identifying where GDEs are present (Isokangas et al., 2017). While these approaches have shown good performance in Finnish examples (Ala-Aho et al., 2015; Autio et al., 2023), they are limited in their spatial extent as they rely not only on ground-based isotope sampling, but their helicopter- and drone-based approaches cannot be scaled easily to a large area. As an alternative approach, machine learning has played an important role in mapping GDEs in other jurisdictions.

Lidberg et al. (2019) implemented machine learning to map the wet areas of a boreal forest landscape in Sweden (an area of 450,295 km^2). An important challenge of mapping these ecosystems in this area is that the region is dominated by peatlands and dense forests, where visible differences between GDEs and non-GDE areas are less distinct. Model inputs included average soil moisture regimes according to the Swedish national forest inventory, using the following categories: dry, mesic, mesic-moist, moist, and wet. As peatlands such as fens typically have high water table heights close to the surface, soil moisture information can be useful for predicting areas which would have higher potential to be GDEs. The model also used inputs from a national digital elevation model (DEM) at 2 m resolution, local topography, and flow grids based on deterministic-8 (D8) for hydrological conditioning. Streams, lakes and rivers were then rasterized to create a source layer for surface water. Elevation above stream, depth to water table, and topographic wetness were then also calculated. Additional inputs included: quaternary deposits parsed into five categories (till soils, peat soils, coarse sediments, fine sediments, rock outcroppings); open wetlands, which were used to aid in peat delineations; and climate variables that capture runoff seasonality. Out of the four common machine learning algorithms used (artificial neural networks, random forest, support vector machines, and naive bayes classification) the random forest and artificial neural networks were able to account for 84% of the wet and dry areas correctly with a relatively high Kappa coefficient, which accounts for random chance in its measure of accuracy (0.65; Lidberg et al., 2019). Although this method does not directly map GDEs themselves, it would identify larger areas that can be further delineated into GDE specific regions within the boreal (i.e., areas with high potential to contain GDEs), and demonstrates an approach that could be effective for supporting GDE mapping in Alberta's boreal region.

In the higher latitude regions of Alaska, Gerlach et al.,(2022) used machine learning to locate areas where groundwater discharged into salmon-bearing streams. This approach compiled existing data and literature, identified well logs (>800 points), which covered 40% of the study area, and compiled high resolution lidar (1x1 m resampled to 3x3 m) for the entire study area (1655 km^2). The authors then subset those areas where geological data existed, to determine groundwater discharge and locations. The use of field-based observations also highlighted that groundwater features could be identified via a combination of topographic variables (narrow gullies, abrupt starts of deep incised stream channels along topographic contours intervals on hill slopes). All data was processed using ESRI ArcPro. DEM-derived inputs included: Terrain Ruggedness Index (TRI), Topographic Wetness Index (TWI), Flow-Weighted Slope (FWS), and flow lines. The machine learning algorithm used only topographic data to predict groundwater likelihood, using maximum entropy modeling to determine this likelihood. The Maxent modeling tool (http://biodiversityinformatics.amnh.org/open_source/maxent) was leveraged for this, and relies on presence-only data (point, and layer form) to infer maximum likelihood by minimizing the relative entropy between the predicted density and input points, based on the probability densities of data inputs. Data was split into a 70/30 division with 51 locations (n=36 used as training data, and n=15 for test data). Verification was initially done via ground-truthed points, with the overall accuracy of the models having a high ability to predict groundwater discharge likelihoods (AUC scores = 0.95 training data, and 0.91 test data; Gerlach et al., 2022). The method was able to predict the location of seeps and groundwater discharge into streams and rivers by leveraging geological and topographic inputs. Although this study's region has greater topographic relief than is found in Alberta's boreal region, this method could be useful in determining GDEs along flowing stream and river systems. If applied within an Alberta context, however, the approach is likely to be limited in its usefulness, given the lower topographic relief found in these areas.

A recent use of machine learning within the Fort McKay watershed of northeastern Alberta illustrates the only such example within the OSR. It was a two part study conducted by Larocque and Leblon (2022), and comprised the use of remote sensing for mapping landcover and wetlands (some of which are GDEs), followed by water level mapping within the same study area . The landcover mapping approach by LaRocque and Leblon (2022) targeted a study area of approximately 5,600 km^2, and leveraged three types of remote sensing data over multiple season (spring, summer, and fall): Landsat optical (https://glovis.usgs.gov), SAR (Copernicus Sentinel-1), and LiDAR imagery (Alberta Geospatial Centre 2009-2013). Landsat-8 imagery was used to calculate a range of 11 spectral vegetation indices (see Table 6), while LiDAR data at 15 m resolution was used to extract a digital terrain model (DTM) as well as additional slope (%), and slope curvature input layers. These inputs were combined in a random forest non-parametric supervised classification, applied using R statistical software and the Random Forest code package (Breiman, 2001, 2003). Data was split into 810 training areas, spread over 21 classes (e.g., forest types, fen and bog types, burned areas, etc.). The method produced an overall accuracy of 94% and a Kappa coefficient of 93.29%. The two variables with greatest model importance based on mean accuracy were: the DTM and slope, which, when removed, decreased classification accuracies by 93.29% and 41.86%, respectively (LaRocque and Lablon, 2022).

The second phase of this work focused specifically on water level mapping using Sentinel-1 SAR data to derive high and low open water levels in mapped wetlands areas. Although this study showed high classification accuracies with machine learning, the categorical map output does not quantify how likely a mapped wetland system in the resulting map is a GDE. Several of the wetland classes (e.g., types of fens) are defined within the Alberta Wetland Classification System (AWCS) as relying on groundwater inputs; the output itself is not specifically mapping GDEs. Nevertheless, the study does highlight the importance of topographical variables in predicting GDEs, with the DTM and slope inputs accounting for the highest accuracy in predictions, while spectral measures of vegetation vigor were less useful. As the boreal is not a water-limited environment like arid or semi-arid areas, the use of vegetation indices such as NDVI is less important, while topographical, geological and hydrogeological inputs are more important. The high number of classes delineated in this study also might introduce overfitting, as can happen when distinguishing a high number of classes with machine learning algorithms. These categories also do not directly equate the definite presence or absence of a GDE.

6.3 Recommendations for Mapping GDEs in the OSR

Of the three types of remote sensing-based GDE mapping efforts reviewed here, machine learning is likely to be the most practical approach to mapping GDEs within the OSR. It is capable of leveraging multiple data sources of differing data types to achieve high predictive accuracy where data limitations exist and is scalable to large spatial areas. The integrated hydrological modeling approach, while successful, is not a feasible approach on its own for larger-scale applications across the OSR. It relies on ground-based water samples that are costly and time-consuming to collect, as well as detailed thermal imagery from airplane, helicopter, or drone platforms, which is neither easily scalable nor easy to repeat for future updating. While thermal imagery is acquired regularly at large scales through various satellite-based sensors, the spatial resolution of these (e.g., 100 m for Landsat-8 or -9) is too coarse to capture anything but the largest of GDEs. The risk mapping approach, also known as suitability modeling, can be used for mapping of GDEs, however can have limitations with leveraging multiple data sources, association of weighting of explanatory variables, and is typically adapted to a presence only mapping (Malczewski, 2004). These methods are also heavily dependent on expert opinion and can be biased (Doody et al., 2017; Fildes et al. 2023).

Of the machine learning approaches described here, the multilayer supervised classification approach used by the MLMapper tool (Martínez-Santos et al., 2021) shows the most promise for mapping GDEs within the OSR. The tool is able to incorporate a wide range of geospatial data layers simultaneously, including a mixture of numeric and categorical variables, and could include many of the same physical geological inputs in the OSR as those used within the paper. It would also enable the addition of other high-resolution layers from recent Alberta Geological Society products, which are available within the OSR and could offer improved spatial resolution in a GDE inventory. Furthermore, the MLMapper approach implements twenty machine learning model algorithms simultaneously and can create an ensemble product that leverages those offering the highest accuracy (Martínez-Santos et al., 2021), to successfully map GDEs in a multitude of environments, including the boreal. While other platforms such as GEE, Maxent, or R statistical software are regularly used to implement multiple machine learning algorithms, is it unknown whether any exist that provide the same degree of ease-of-use for applying multiple machine learning algorithms for modeling the presence/absence of a phenomenon, as well as integrating a range of input layers within one program.

Although suitability or risk-based mapping approaches could also be implemented they were not preferred due to their inherent limitations (demand for expert opinion, bias in assigning weighting of variables, standardization of criterion maps, and limited computational power for regions that do not possess data coverage), leading to stochastic mapping probabilities, and lack the ability to leverage all available data sources (Malczewski 2004). Machine learning was selected as the method that would allow for the maximum amount of data to be leveraged, limiting or avoiding many of the limitations present in suitability or risk-based mapping. Specifically, the use of multiple models to derive a binary and probability GDE map of the OSR study area, specifically the MLMapper platform as it can leverage up to 20 algorithms to calculate the best fitted models (Martínez-Santos et al., 2021).

7. Methods for GDE Mapping in the Study Area

7.1 Model: MLMapper 2.0

The MLMapper 2.0 method and tool developed and described by Martínez-Santos et al. (2021) was used to map GDEs within the selected OSR study area (Figure 1). Updated Python code was provided by the authors and run within a Python development environment on a local desktop machine. The approach leverages point-source reference data in the form of known presence and absence locations of the phenomenon of interest (i.e., GDEs in this case), alongside explanatory variable values associated with each of these same point sources. From this, the tool applies up to 20 machine learning algorithms to perform supervised multilayer pattern recognition and produce predictive models of GDE occurrence (Martínez-Santos et al., 2021). The best performing of these, based on a user-defined threshold, are then combined to produce an ensemble model output map of relative GDE likelihood. Ensembles or averaged models have been shown to typically perform better than single method approaches (Dormann et al., 2018). The overall workflow of the MLMapper tool and the steps that were followed here for mapping GDEs in the study area, are outlined in the diagram given in Figure 2. The machine learning modeling approaches tested in our study area included the following 15 methods: linear support vector machines (LSVM), logistic regression (LRG), a decision tree classifier (CRT), a random forest classifier (RFC), linear discriminant analysis (LDA), a K-nearest neighbor classifier (KNN), a gradient boosting classifier (GBC), an Ada-boost classifier (ABC), an extra-trees classifier (EXT), a passive aggressive classifier (PAC), quadratic discriminant analysis (QDA), multilayer perception neural networks (MLP), a ridge classifier (RID), a stochastic gradient descent linear classifier (SGD), and perceptron (PRC). Analyses and output map products were both produced using a 50 m pixel resolution. This provided an effective scale that balanced the desire for detailed mapping and the resolutions of the various input datasets (see the following sections for more detail on these).

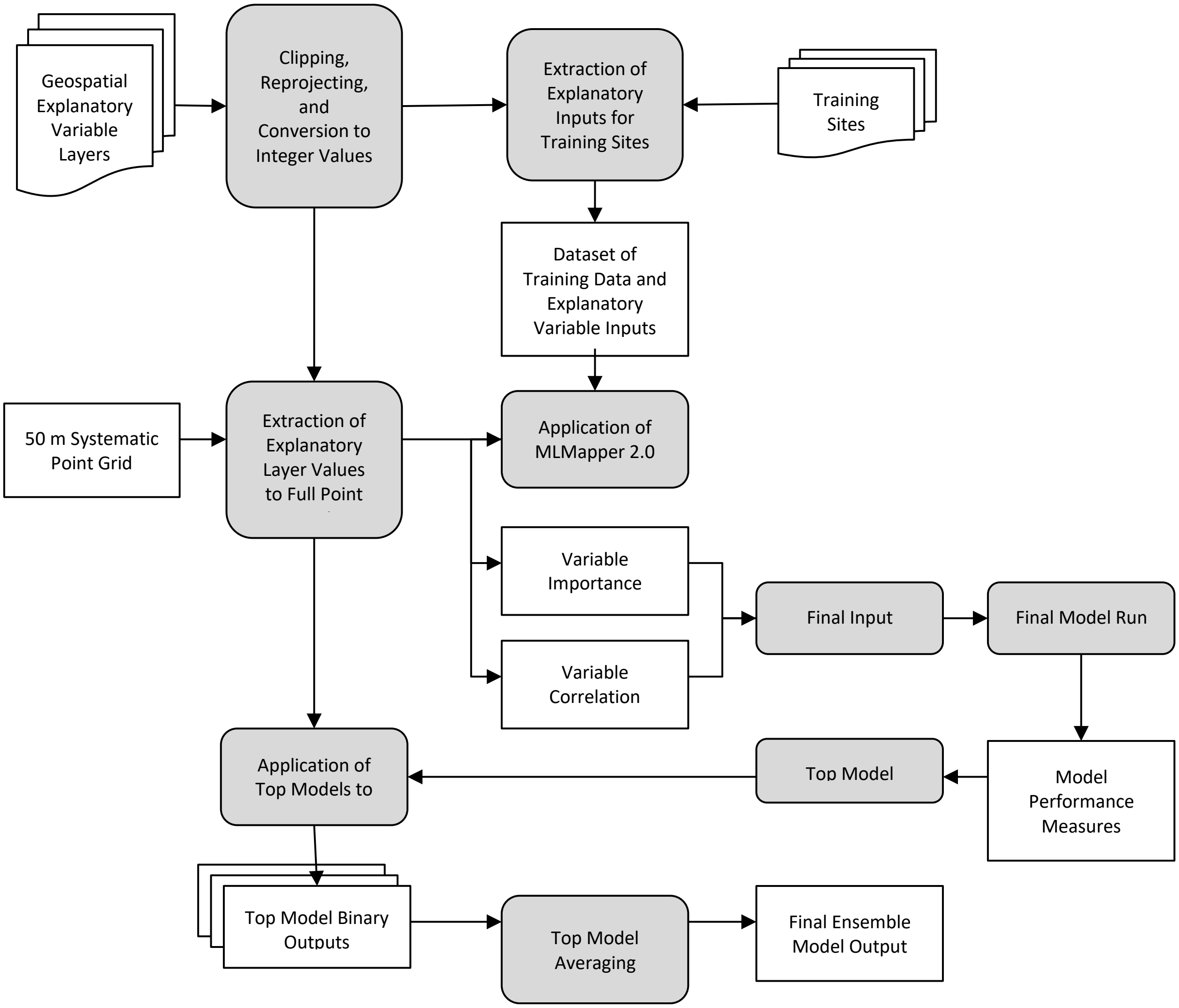

Figure 2. Flowchart showing the main components of the workflow used to map GDEs across the study area, leveraging the MLMapper 2.0 tool from MartÍnez-Santos et al. (2021).

7.2 Explanatory Variables

The 13 explanatory input variables used in this preliminary GDE mapping work each fall into one of three categories: topographic, hydrogeological, and wetland. Data were reprojected to a common coordinate reference system (i.e., NAD 1983 CSRS UTM Zone 12N (EPSG 2956)), so as to ensure geolocational alignment. Martínez-Santos et al. (2021) reclassified their input variable datasets into integer category values before using them in the MLMapper tool. The use of reclassified inputs versus original, scaled inputs was tested here. For some model runs, variable values were binned into one of several output categories and assigned a relative integer value (see TablesTOPO VAR CLASSES, HYDROGEOL VAR CLASSES, and WETLAND VAR CLASSES). Reclassification was based on subject matter expertise and knowledge (e.g., John Gibson, personal communication). For other model runs wherein inputs were not reclassified, the values were simply scaled before being used in the MLMapper tool, so as to remove negative and decimal values (TablesTOPO VAR CLASSES, HYDROGEOL VAR CLASSES, and 8).

7.2.1 Topographic

Information on the regional and local terrain is key to identifying likely locations for GDEs since terrain is a strong influencer of water movement and depth to groundwater, which is a critical component of groundwater presence near, or expression at, the surface. Five topographic input variables were used in this work: elevation, flow accumulation, slope, terrain ruggedness index (TRI), and the SAGA wetness index (SWI), which is comparable to a more generic topographic wetness index but specific to the SAGA software. Maps of these inputs are provided in Figure 3. A satellite-based digital surface/elevation model from the Japan Aerospace Exploration Agency's Advanced Land Observing Satellite (ALOS) (Tadono et al., 2014) provided the source for elevation, slope, TRI, and SWI inputs. These data are publicly available in a 30 m resolution for the globe, and were accessed using Google's Earth Engine platform (Gorelick et al., 2017). SWI and TRI datasets derived from ALOS already existed for the study area within the ABMI's geospatial data archive, and were used here, while slope was derived for this project from the same source. Flow accumulation data were compiled from a separate source, however, and originated from the MERIT Hydro global hydrography datasets (https://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT_Hydro/). Table 7 describes how these topographic inputs were classified and scaled for inclusion in the MLMapper tool.

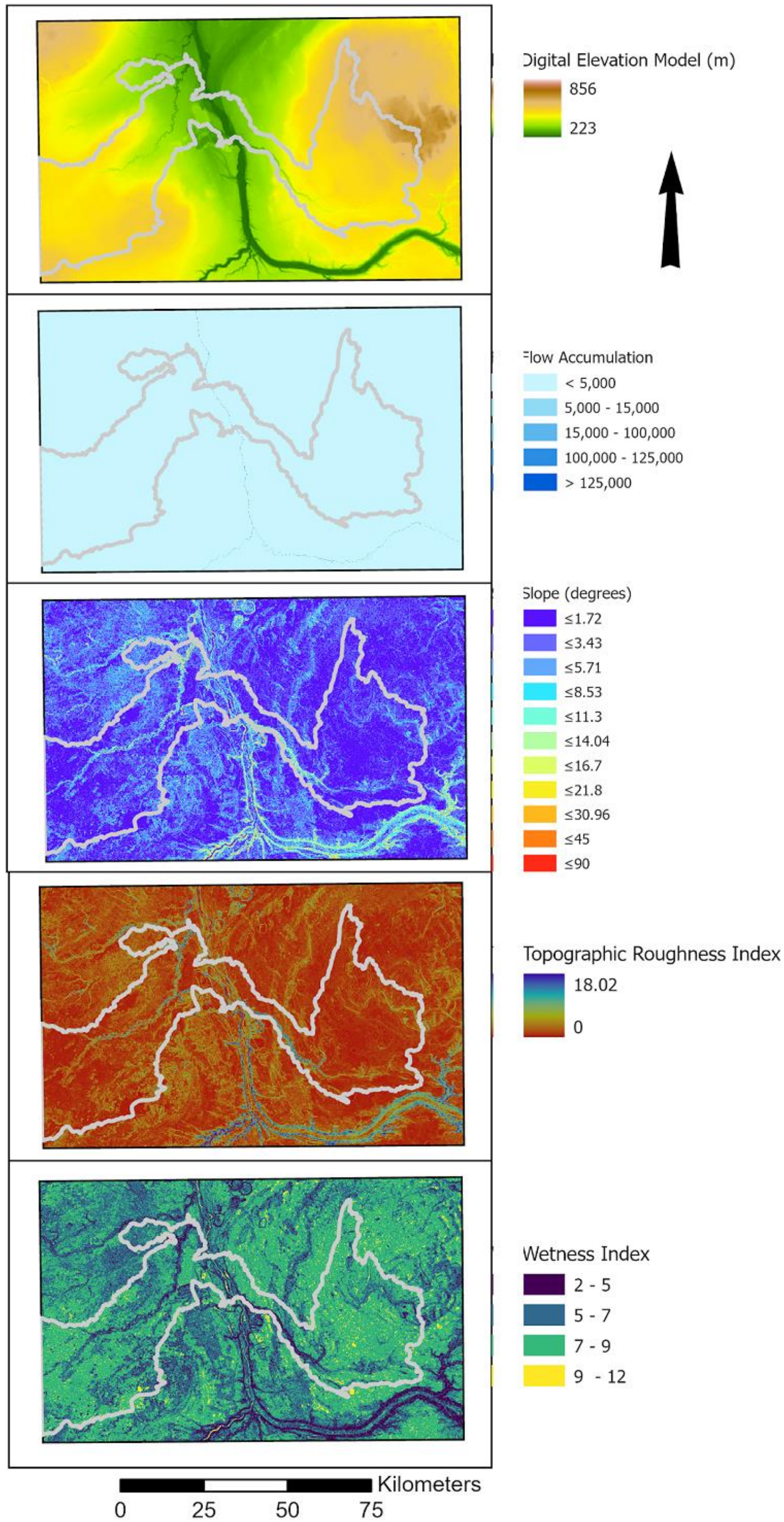

Figure 3. Maps of the five topographic explanatory input variables used in mapping GDEs over the study area.

Table 7. Table describing the scaling and reclassification of topographic input explanatory variable values into integer classes, for inclusion in the MLMapper tool.

Input Variable	Integer Reclassification	Spatial Resolution	Data Source
Elevation Scaling: rounded to nearest integer	1: 222 - 350 m	30 m	Satellite-based ALOS DEM (https://developers.google.com/earth-engine/datasets/catalog/JAXA_ALOS_AW3D30_V3_2)
	2: 350 - 450 m		
	3: 450 - 550 m		
	4: 550 - 650 m		
	5: 650 - 750 m		
	6: 750 - 860 m		
SAGA Wetness Index (unitless) Scaling: multiplied by 100	1: 1 - 5	10 m	Calculated from ALOS DEM (https://developers.google.com/earth-engine/datasets/catalog/JAXA_ALOS_AW3D30_V3_2)
	2: 5 - 7		
	3: 7 - 9		
	4: 9 - 12		
Terrain Ruggedness Index (unitless) Scaling: multiplied by 10	1: 0	10 m	Calculated from ALOS DEM (https://developers.google.com/earth-engine/datasets/catalog/JAXA_ALOS_AW3D30_V3_2)
	2: >0 - 0.5		
	3: 0.5 - 1		
	4: 1 - 2		
	5: 2 - 5		
	6: 5 - 18		
Slope (degrees) Scaling: multiplied by 100	1: 0 - 5 degrees	30 m	Calculated from ALOS DEM (https://developers.google.com/earth-engine/datasets/catalog/JAXA_ALOS_AW3D30_V3_2)
	2: 5 - 10 degrees		
	3: 10 - 15 degrees		
	4: 15 - 20 degrees		
	5: 20 - 90 degrees		
Flow accumulation (cell units)	1: 0-5000	50 m	MERIT Hydro (https://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT_Hydro/)
	2. 5000-15000		
	3. 15,000-10,0000		
	4. 100,000-125,000		
	5. 125,000-150,000		

7.2.2 Hydrogeologic

Explanatory variables under a hydrogeological focus included: permeability derived from surficial geology, soil drainage, aquifer hosting sediment, depth to water, and bedrock. Surficial geology plays a crucial role in the movement of groundwater, where highly permeable geological features are likely to have an increased probability of groundwater upwelling and are thus more likely to coincide with the presence of GDEs. A surficial geology dataset from the Alberta Geological Survey (AGS) maps 601, 618-621 was converted into a permeability scale through classification into five permeability classes, based on expert opinion from AGS. In these classes a 0 (zero) signified low permeability, while five was considered to represent high permeability (Figure 4; Table 8).

Soil drainage influences the hydrological recharge of GDEs and the likelihood that an area is prone to water table fluctuations. These data were obtained from Soil Landscapes of Canada V3.2 which includes a max depth of 1-2 m below the surface. Soil types were classified into five categories normalized along a range from 1 = well drained to 5 = very poorly drained (Figure 4; Table 8).

Aquifer hosting sediment indicates regions that are likely to have large aquifer systems and thereby be more permeable, and more likely to contain GDEs. These data were obtained from AGS map 632 and classified into 4 categories, wherein: 0 = no values, 1 = known plains upland, 2 = potential plains upland, 3 = inferred buried valley, and 4 = known buried valleys (Figure 4; Table 8).

Depth to water was derived from the Aquanty HGS model of the Athabasca River Basin at a 500 m resolution. Areas closer to high water tables are more likely to be GDEs as there is both a greater potential for a groundwater seep or spring to occur, and for the water table to be in closer proximity to the rooting zone of local vegetation. The data were scaled by 100 and then normalized along a 1-5 scale with a range of 1 indicating near surface (e.g., 0 to 0.1 m), and 5 indicating greater depths (e.g., 2 to 5 m) to groundwater (Figure 4; Table 8).

Bedrock formation data was pulled from the AGS 600 map and categorized into 12 unique classes. Bedrock formations influence regional groundwater flow and have potential to infer channeling of deep basin groundwater based on the unique hydrogeochemical and hydrogeological features of each formation. These were therefore included to assess to what extent GDEs presence is derived from bedrock formation (Figure 4; Table 8).

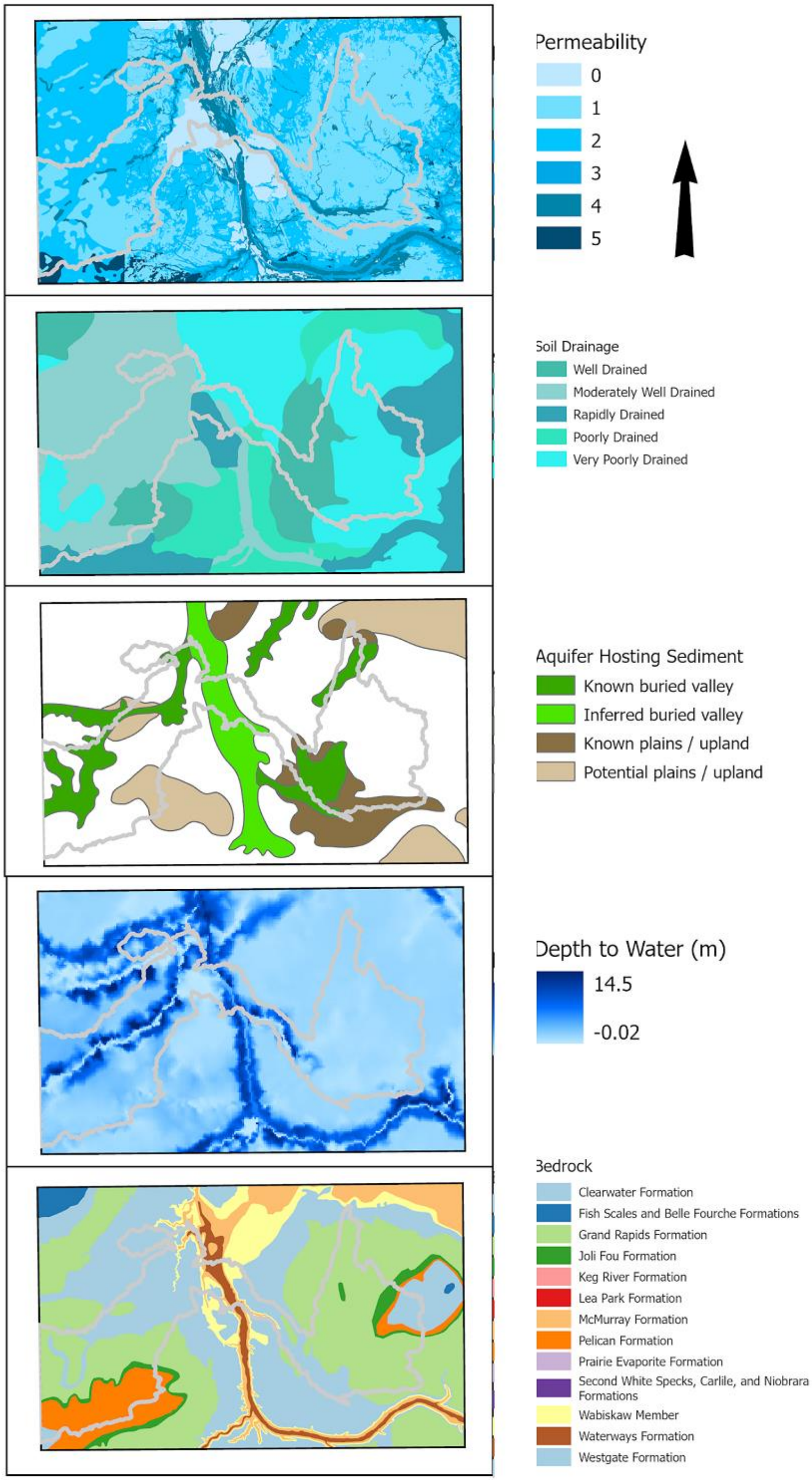

Figure 4. Maps of the five hydrogeologic explanatory input variables used in mapping GDEs over the study area.

Table 8. Table describing the scaling and reclassification of hydrogeological input explanatory variable values into integer classes, for inclusion in the MLMapper tool.

Input Variable	Integer Reclassification	Spatial Resolution	Data Source
Surficial Geology	0: Anthropogenically excavated materials, infilled or made materials	1:100,000 to 1:1,000,000	Alberta Geological Survey, Maps 618-621, and 601
	1: Bedrock (general or thin/absent), Organic deposits (general or peat)		
	2: Fluted moraine (general or clayey-silt diamicton), Glaciolacustrine deposits (general or silt-clay), Lacustrine deposits (general, or sand, silt, and clay, minor deposits), Littoral and nearshore sediments, Moraine (clayey-silt diamicton or silty-sand diamicton), Stagnant ice moraine (general, clayey-silt diamicton, or silty-sand diamicton)		
	3: Colluvial deposits (general or diamicton), Fluted moraine (pebbly diamicton, or silty-sand diamicton), Ice-contact sediments (stratified sand and silt), Ice-thrust moraine (general, stratified sand and silt, or syngenetic diamict and displaced sediment and/or bedrock), Moraine (general, pebbly diamicton, sandy-silt diamicton, or massive to stratified silty sand, pebbly sand and minor gravel)		
	4: Fluted moraine (predominantly sand) Fluvial deposits (general, or stratified sand, gravel silt, clay and organic sediments) Glaciofluvial deposits (general, or sand with minor gravel) Ice-contact sediments (sand) Moraine (predominantly sand, or sand and gravel)		
	5: Eolian deposits (general, or sand)		
Soil Drainage	1: Very poorly drained	1:1,000,000	Soil Landscapes of Canada v3.2, Agriculture and Agri-Food Canada
	2: Poorly drained		
	3: Moderately well drained		
	4: Well drained		
	5: Rapidly drained		

Aquifer Hosting Sediment	0: *no value*	1:3,000,000	AER/AGS Map 632, Alberta Geological Survey
	1: Known plains upland		
	2: Potential plains upland		
	3: Inferred buried valley		
	4: Known buried valley		
Depth to Water Scaling: multiplied by 100	1: 0 - 0.1 m	500 m	Aquanty
	2: 0.1 - 0.5 m		
	3: 0.5 - 1.0 m		
	4: 1 - 2 m		
	5: 2 - 5 m		
Bedrock	1: Pelican Formation	1:1,000,000	Alberta Geological Survey, Map 600
	2: Joli Fou Formation		
	3: Grand Rapids Formation		
	4: Clearwater Formation		
	5: Wabiskaw Member		
	6: McMurray Formation		
	7: Waterways Formation		
	8: Westgate Formation		
	9: Lea Park Formation		
	10: Fish Scales and Belle Fourche Formations		
	11: Second White Specks, Carlile, and Niobrara Formations		
	12: Keg River Formation		

7.2.3 Wetlands

Two wetland inventories covering the study area and developed for the OSM program (Alberta Biodiversity Monitoring Institute & Ducks Unlimited Canada, 2023), were used here to map GDEs. The first is a map of wetland classes according to the AWCS, while the second is more detailed and maps the area down to AWCS wetland form (Figure 5). These represent the most recent wetland mapping efforts in our study area (e.g., reflecting 2020-2022 conditions), and possess a 10 m resolution with a 0.04 ha minimum mapping unit. Other wetland inventories available for the study area, such as the Alberta Merged Wetland Inventory (Alberta Environment and Parks, Government of Alberta, 2022), are either quite dated (e.g., 1999-2009), less thematically detailed (i.e., not to the form level), and/or provided at a coarser resolution (e.g., 30 m). As the wetland class and form inputs were already provided as integer rasters, wherein each integer value represents a class or form, no scaling or further reclassification was needed for input into the MLMapper tool.

In addition to wetland, the last explanatory input variable comprised a 2023 growing season NDVI composite covering the study area (Figure 5). This was derived from Sentinel-2 optical imagery acquired during May through September of 2019 to 2023 that was cloud-masked and composited using a per-pixel median statistical filter to produce a “best available pixel” type NDVI image of the 2019 to 2023 growing season greenness. Processing was done using the Google Earth Engine platform. Table 9 describes how the NDVI input raster was scaled and classified for use in the MLMapper tool.

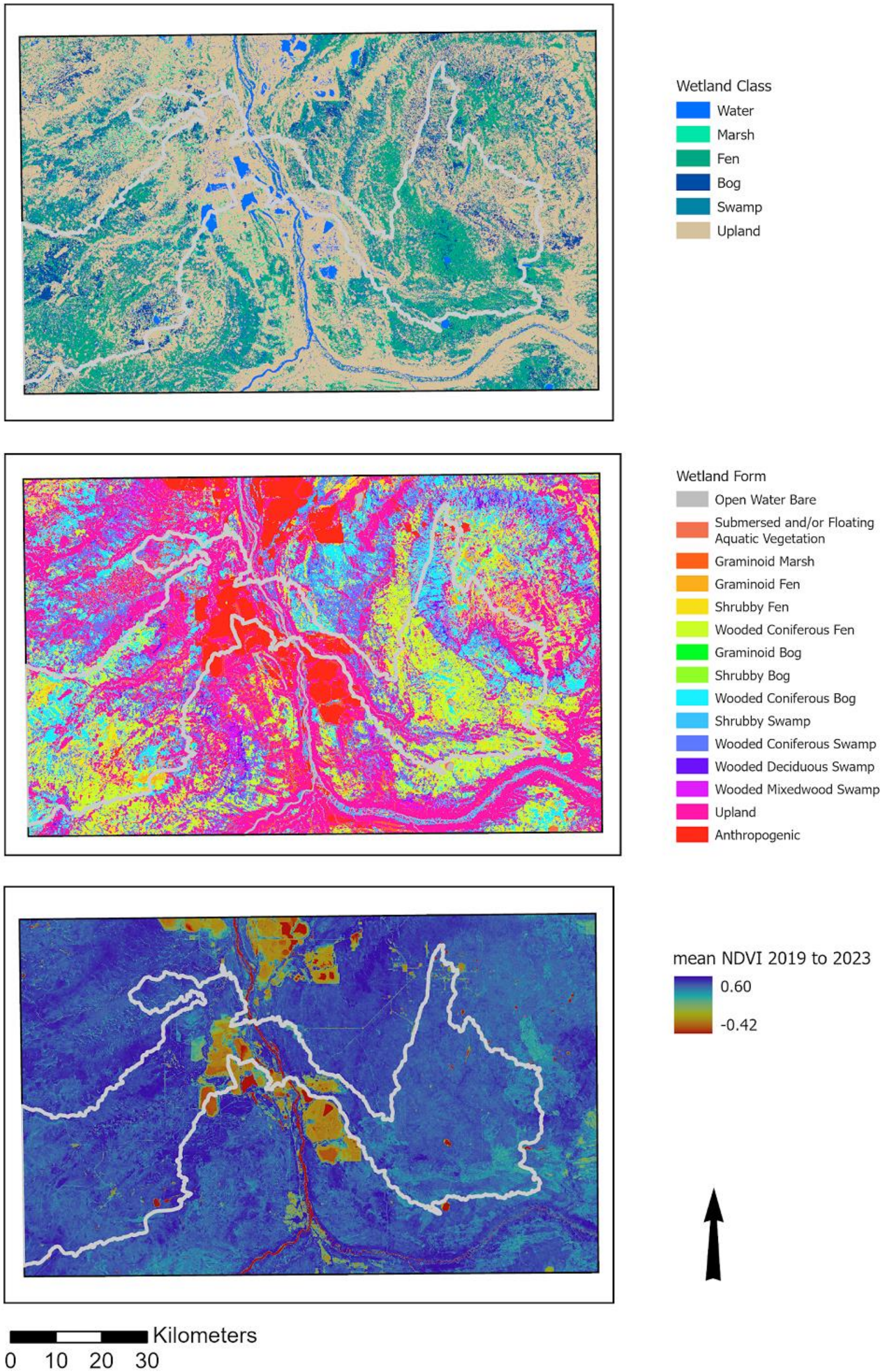

Figure 5. Maps of the two wetland inventory and one NDVI explanatory input variable used in mapping GDEs over the study area.

Table 9. Table describing the scaling and reclassification of NDVI input explanatory variable values into integer classes, for inclusion in the MLMapper tool.

Input Variable	Integer Reclassification	Spatial Resolution	Data Source
Normalized Difference Vegetation Index (unitless) Scaling: added 1, then multiplied by 1000	1: -1.0 - 0.0	10 m	Calculated from 2023 ESA Sentinel-2 summertime imagery (https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR_HARMONIZED)
	2: 0.0 - 0.5		
	3: 0.5 - 1.0		

7.3 Training and Test Data

The data used for model training and cross-validation comprised 227 known GDE presence locations, and 227 known absence locations. These data were compiled from a combination of ABMI open water wetland water isotope analysis samples from the collaborative isoABMI project with InnoTech Alberta, Alberta Geological Survey spring and fen locations, McKay River differential gauging locations (Bickerton et al., 2018), local high-elevation points, and a recent AWCS wetland class and form inventory(Alberta Biodiversity Monitoring Institute & Ducks Unlimited Canada, 2023). The ABMI open water wetlands are confidential site locations and not provided in the corresponding data files. The wetlands were categorized using isotope mass balance approaches relative to their groundwater input where thresholds/categories for GDE presence and absence were the isotopic ratio of groundwater > surface water, and groundwater ≤ surface water, respectively based on water yield and run off calculations. Greg Bickerton (ECCC) provided the locations of differential gauging measurements along the McKay River and indicated that the reach between stations 2-4 had no evidence of significant groundwater discharge (i.e., surface water inputs can explain the gain in flow, whereas the reach between stations 5-9 had significant groundwater input). Points categorized as GDE absence locations were randomly placed between stations 2-4, and GDE presence locations were randomly placed between stations 5-9. The Alberta Biodiversity Monitoring Institute and Ducks Unlimited Canada (2023 OSM wetland inventory considered fen locations as GDE presence, and bog and upland locations as GDE absence points. Figure 6 shows the locations of training point data, while the numbers of samples extracted for use from these various sources are provided in Table 10.

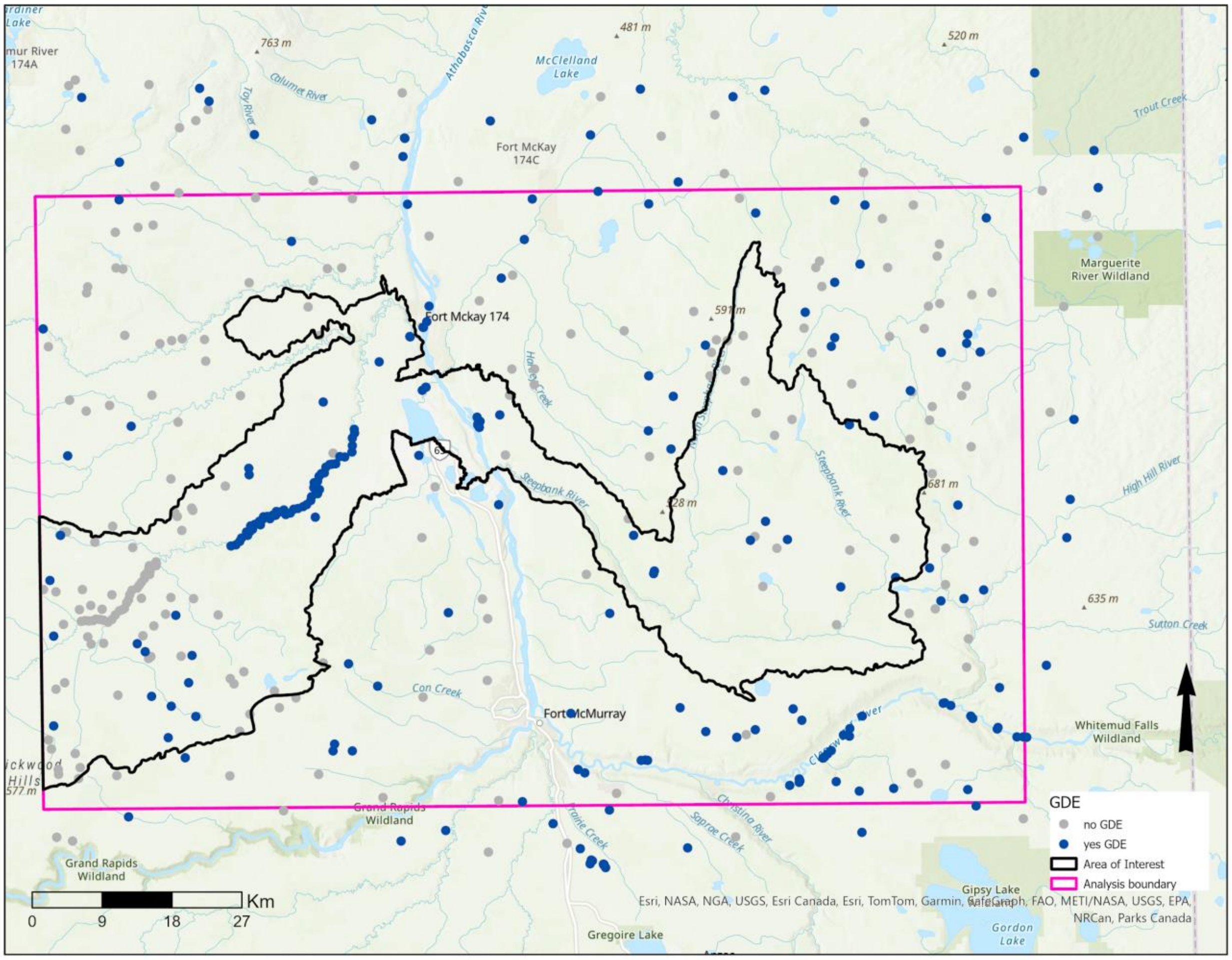

Figure 6. Figure showing the distribution of GDE presence (blue, "yes") and absence (grey, "no") training and test points across the study area and greater area of analysis.

Table 10. Table listing the sources of GDE presence and absence training and test data used to model GDEs in the study area.

Reference Data Source	GDE Presence Points	GDE Absence Points
Shallow Open Water Wetlands (isoABMI)	17	21
OSM Wetland Inventory (ABMI & DUC 2023)	Fens - 98	Bogs - 130
	-	Uplands - 41
Alberta Geological Survey Springs & Fens	Springs & Fens - 62	-
MacKay River differential gauging (ECCC)	Drive stations - 50	Between drive stations - 30
ALOS DEM	-	Elevation > 700 m - 5
Total	227	227

Reclassified data from the 13 explanatory variable input layers was extracted for each of the 454 GDE presence/absence training locations, and placed into a comma-delimited text file, for input into the MLMapper tool. The tool was run with 100 iterations, and varying combinations of input variables were tested on the basis of variable importance and correlation metrics output by the tool after each model run. Model runs also included the use of reclassified input variables or scaled, original input variables, as well. Model performance metrics were used to select the number of iterations and which explanatory input variables to include for the final model run. After optimization, fine-tuning, and iteration and input variable selection were complete, a final model run was performed and the resulting outputs from the top-performing machine learning models were selected using several common model performance metrics, most notably the area under the curve of the receiver operating characteristic (AUC), which balances model sensitivity and specificity, as well as optimized test scores.

Once selected, each of the top models was then mapped over the study area using a systematic grid of points placed 50 m apart to produce a binary map of GDE predicted presence and absence, which was then converted into a 50-m resolution raster map product. These binary outputs (the 1s and 0s, reflecting presence and absence, respectively) from the top-performing models were then averaged to produce a final ensemble map of predicted relative GDE likelihood or probability across the study area.

As a final step, the final binary and ensemble GDE map outputs were overlaid with information on known, non-vegetated human footprint features so as to identify where likely GDEs overlap with these types of disturbances. These features were extracted from the ABMI's Human Footprint Inventory 2021 (Alberta Biodiversity Monitoring Institute & Alberta Human Footprint Monitoring Program, 2023), and included features from the following sublayers: reservoirs; borrow pits, sumps, dugouts, and lagoons; roads; railways; canals; mines; and, industrial features.

8. Results and Discussion

8.1 Model Tuning and Performance

The results of the pairwise correlation analysis performed on all 13 of the potential explanatory input variables is found in Figure 6. The results of the matrix are displayed using two color schemes along a normalized scale with green (R = 1) showing strong positive correlation and purple (R = -1) showing strong negative correlation. The discrimination thresholds used for removing correlated variables were adopted from MartÍnez-Santos et al. (2021), which are commonly accepted at a range between 0.7 to -0.7, although results as high as 0.84 have been known to be acceptable (Dormann et al., 2013). Initial results showed that wetland class and wetland form had a positive correlation higher than 0.75, while TRI and SWI showed a strong negative correlation (< -0.75; Figure 6). Outputs from the initial model run were used to assess which variables had higher importance to the models overall, and subsequently remove one in each pair of highly correlated variables. Both wetland form and TRI were removed from the model as they contributed less to all models when compared to wetland class and surface wetness index, and were highly correlated with them, respectively (Appendix E; R1E). Once these two had been removed, subsequent model runs showed the input variable correlations remained between 0.7 to -0.7 for all pairwise comparisons (Figure 6).

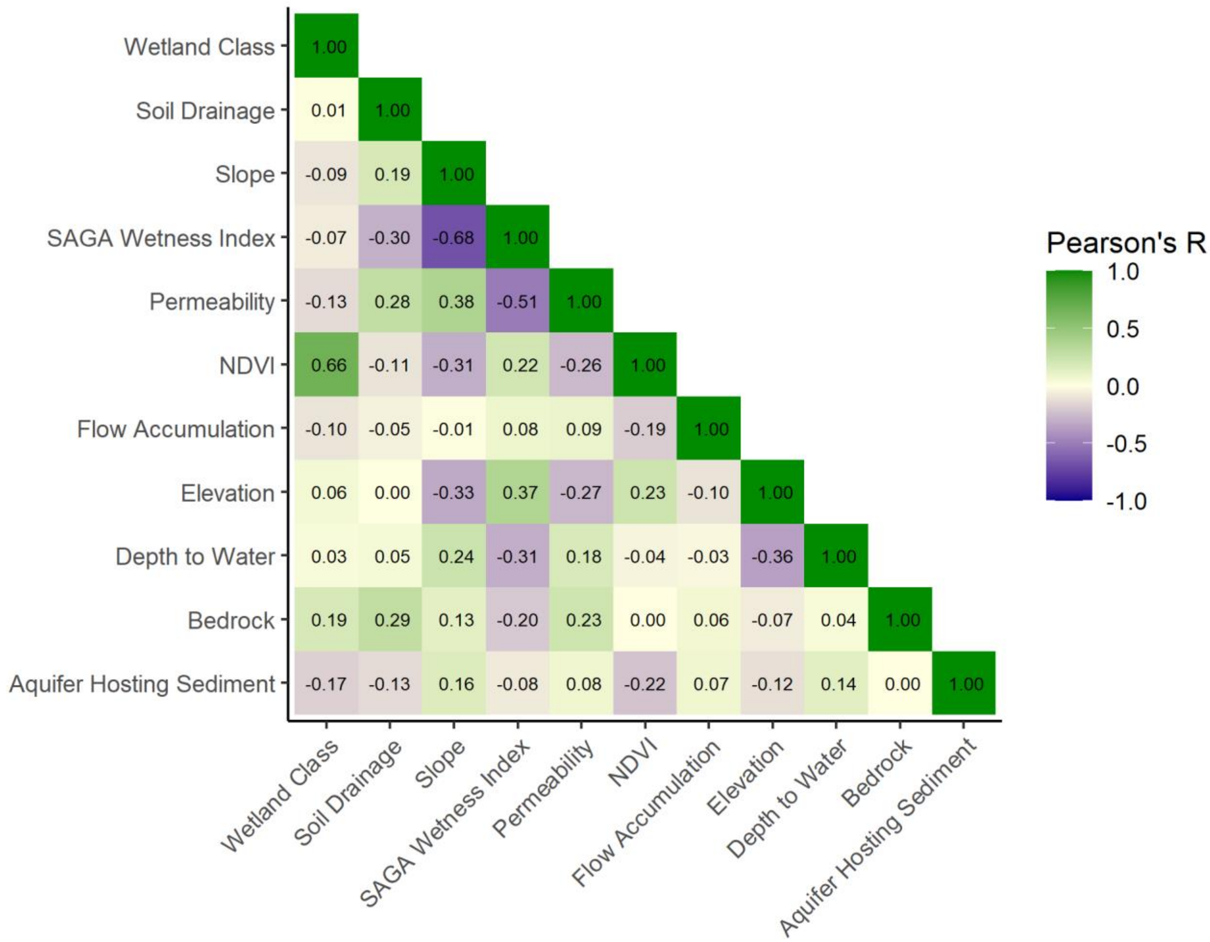

Figure 6: Results of pairwise correlation analysis of the final selection of 11 explanatory variables used in mapping GDEs in the study area. Numbers in the plot show the Pearson correlation coefficient. These included: aquifer hosting sediment, bedrock, depth to water, elevation, flow accumulation, normalized difference vegetation index (NDVI), permeability, SAGA wetness index, slope, soil drainage, and wetland class.

Table 11 lists the tuning parameters used for each of the separate machine learning algorithms applied to the GDE dataset, as well as the optimum number of input variables found to produce the best results during cross-validation.

Table 11. *MLMapper algorithm tuning parameters and optimum number of input variables based on outputs from cross-validation.*

Algorithm	Optimized Parameter Values	Optimum Number of Input Variables
Random forest classifier	max_depth = 4 max_features = 0.6 min_samples_leaf = 5 n_estimators = 141 random_state = 0	9
Gradient boosting classifier	max_depth = 9 max_features = 0.3 min_samples_leaf = 29 n_estimators = 70 random_state = 0	11
Ada-boost classifier	algorithm = 'SAMME' learning_rate = 0.232065 n_estimators = 250 randomt_state = 0	4
Decision tree classifier	max_depth = 6 max_features = 0.8 min_samples_leaf = 18 min_samples_split = 0.2 randome_state = 0	7
Extra-trees classifier	max_depth = 6 max_features = log2 min_samples_leaf = 2 n_estimators = 590 random_state = 0	3

Ranked feature importance from the top five algorithms indicated that wetland classification, followed by the digital elevation model (DEM) were the most important features. There is a steep drop off in importance to the third variable which was surface wetness index, and then a gradual decline in importance with regards to permeability and NDVI. Depth to water table, bedrock and soil drainage were among the lowest ranked explanatory variables. Both aquifer and flow scored the lowest rank among the 11 explanatory variables (Figure 7). Individually each algorithm ranked the explanatory variables slightly differently, however both wetland class and DEM remained in the top positions across all five consistently (Figures E.1, E.2, E.3, E.4, and E.5 in Appendix E).

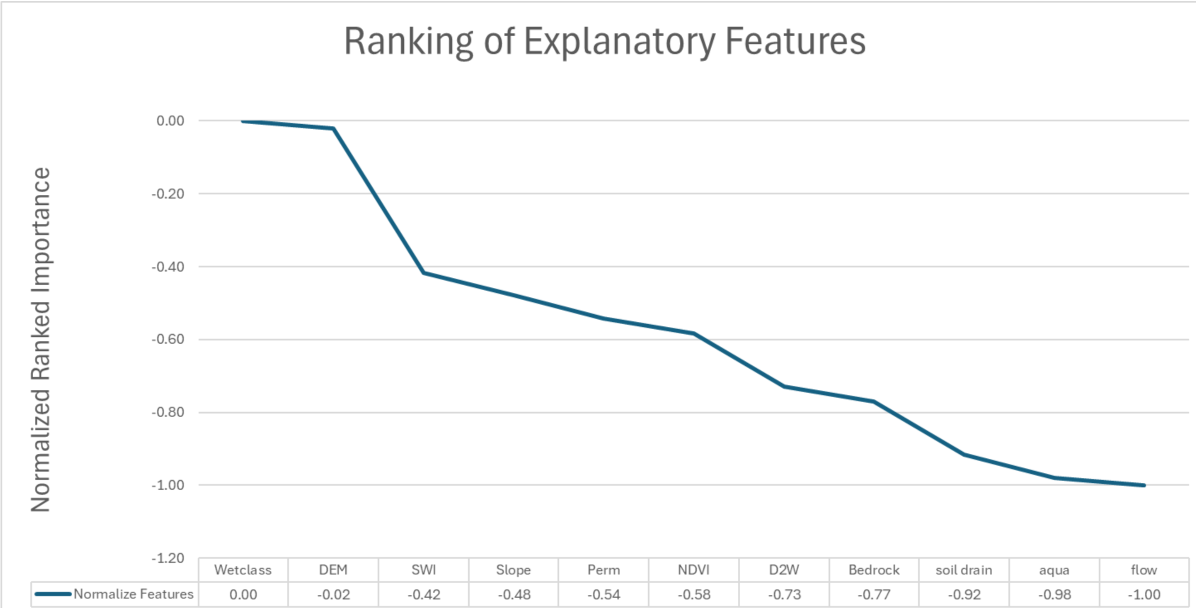

Figure 7. Normalized explanatory feature importance based on rank from the top five algorithms.

The model tuning and correlations show that some remote sensing products are correlated (Wetland Form and Wetland Class; SAGA wetness Index and Terrain roughness index) and highlight that some explanatory variable can have higher contributions to the model and that there should be moderation between input variables to balance tuning (Fildes et al., 2023). Consideration of derivative variables (e.g., Normalized Difference Coefficients of Variation Index $NDCVI_{NDVI}$ from NDVI) may identify additional products which prove more sensitive for detection of GDES within the boreal (Fildes et al., 2023). Access to high resolution thermal imagery could improve detection of GDEs within aquatic and terrestrial systems (Birks et al., 2012; Ala-aho et al., 2015; Autio et al., 2023; Watts et al., 2023). Although each model had a unique feature reduction, the limitation for all models improved predictions capability, and while some models were able to perform well on test scores (0.92 to 0.89) each had unique strengths in their ability to predict aquatic and terrestrial systems.

The standardized performance metrics output by MLMapper showed a group of five5 algorithms with that yielded test score results greater than 0.89 which included: random forest classifier (RFC), gradient boosting classification (GBC), AdaBoost classifier (ABC), decision tree classifier (CRT), extra-trees classifier (EXT). There is a steep drop off from below test scores of 0.89, with the remaining models only able to predict from 0.76 to 0.55. The remaining models were logistic regression (LRG), linear discriminant analysis (LDA), ridge classifier (RID), K-neighbor classification (KNN), stochastic gradient descent linear classifier (SGD), perceptron (PRC), multilayer perceptron neural network (MLP), quadratic discriminant analysis (QDA), Passive Aggressive Classifier (PAC), and Linear Support Vector Classifier (LSVC) which struggled to predict GDEs with a test score of only 0.55. The discrimination threshold was therefore set at 0.89 with algorithms scoring below being discarded from the ensemble map (Table 12).

Table 12. Performance metrics of supervised algorithms (Train = optimized training score; Test = optimized test score; Prec. F = precision false; Prec. T = precision true; Rec. F = recall false; Rec. T = recall true; F1. Sc. F = f-1 score false; F1. Sc. T = f-1 score true; AUC = area under curve; TN = true negatives; TP = true positives; FP = false positives; FN = false negatives).

Algorithm	Train	Test	Prec. F	Prec. T	Rec. F	Rec. T	F1. Sc. F	F1. Sc. T	AUC	TN	TP	FP	Fn
RFC	0.91	0.92	0.92	0.92	0.90	0.93	0.91	0.93	0.97	57	69	6	5
GBC	0.95	0.92	0.91	0.93	0.92	0.92	0.91	0.93	0.95	58	68	5	6
ABC	0.89	0.91	0.89	0.93	0.92	0.91	0.91	0.92	0.94	58	67	5	7
CRT	0.90	0.91	0.89	0.92	0.90	0.91	0.90	0.91	0.92	57	67	6	7
EXT	0.91	0.89	0.86	0.92	0.90	0.88	0.88	0.90	0.94	57	65	6	9
LRG	0.69	0.76	0.70	0.83	0.83	0.70	0.76	0.76	0.80	52	52	11	22
LDA	0.66	0.74	0.68	0.80	0.79	0.69	0.74	0.74	0.82	50	51	13	23
RID	0.68	0.72	0.67	0.78	0.78	0.68	0.72	0.72	-	49	50	14	24
KNN	1.00	0.69	0.63	0.80	0.83	0.58	0.71	0.67	0.77	52	43	11	31
SGD	0.63	0.68	0.66	0.70	0.63	0.72	0.65	0.71	-	40	53	23	21
PRC	0.63	0.64	0.59	0.73	0.76	0.54	0.66	0.62	-	48	40	15	34
MLP	0.54	0.61	0.58	0.64	0.57	0.65	0.58	0.64	0.64	36	48	27	26
QDA	0.68	0.61	0.54	0.88	0.95	0.31	0.69	0.46	0.81	60	23	3	51
PAC	0.62	0.57	0.52	0.94	0.98	0.22	0.68	0.35	-	62	16	1	58
LSVC	0.50	0.55	0.67	0.55	0.06	0.97	0.12	0.70	-	4	72	59	2

The test scores of the best-performing models were all within 0.92 to 0.89, with the top algorithms (random forest and gradient boosting classifiers) scoring 0.92. The next highest was Ada boost and Decision tree classifiers at 0.91, and lastly Extra trees classifier scoring 0.89 on test data (Table 12). The recursive feature elimination outputs for the top five algorithms indicated that random forest was able to optimize test results using only 9 of the variables (Figure E.2A in Appendix E), while gradient boosting leveraged all 11 to achieve the same result of 0.92 (Figure E.2B in Appendix E), Ada-boost was able to predict an accuracy of 0.91 using only 4 variable (Figure E.2C in Appendix E), while decision tree needed to leverage 7 to obtain the same result (Figure E.2D in Appendix E). The Extra-Trees classifier was optimized at 3 variables but achieved the lowest accuracy at 0.89 compared to the other top five (Table 12; Figure E.2E in Appendix E).

8.2 Individual Model Results: Binary GDE Maps

Random forest (RFC) was able to predict aquatic flowing systems and identified the McKay River as a GDE location for approximately 22 km upstream from the mouth of the Athabasca River. It was similarly able to identify GDE presence in the High Hill River at the confluence of the Clearwater River mouth approximately 10 kms into the headwaters. When assessing lake features, RFC labeled the centre of some lakes (e.g., two small unnamed lakes east of Fort McMurray (56.769341, -110.908782; 56.896575, -110.896889) as Non-GDE, while the perimeter of these same lakes is labeled as GDE (Figure 8: RFC). The gradient boosting classifier (GBC) was able to predict similar outputs to those seen in random forest (RFC) with a slightly wider buffer for confirmed GDEs along flowing systems, and slight differences in classifications of terrestrial GDEs (Figure 8: GBC). However, the AdaBoosting classifier (ABC), struggled to pick up flowing systems and underpredicted the McKay River (approximately 10 km from the confluence of the Athabasca River), and High Hill River (approximately 2 km from confluence with the Clearwater River). Nevertheless, it seemed to be able to predict both terrestrial and lake features similar to the random forest and gradient boosting classifiers (Figure 8: ABC).

The Decision Tree classifier (CRT) was similar to AdaBoosting classifier (ABC) in its abilities to identify GDEs in flowing systems, and terrestrial and lakes delineations. The greatest noticeable difference is that it predicted GDE locations slightly larger and predicted somewhat more terrestrial GDEs (Figure 8). In contrast, The Extra Tree classifier (ETC) was able to predict GDEs in flowing systems with the highest accuracy and captured approximately 38 km of the McKay River and performed similarly in capturing GDEs along the High Hill River from the confluence of the Athabasca and Clearwater Rivers respectively. It delineated both terrestrial and lake feature GDEs similar to all preceding models (Figure 8).

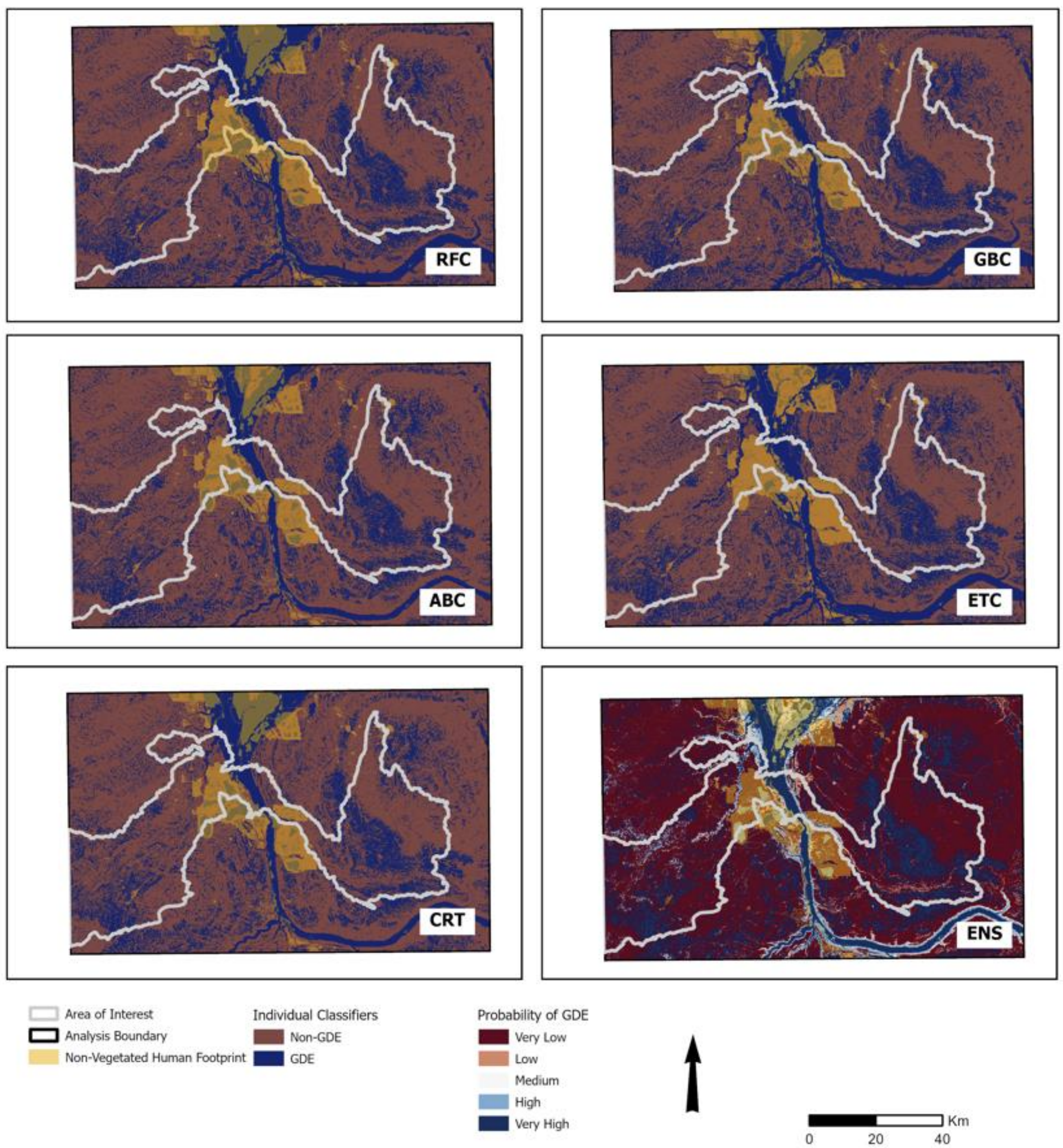

Figure 8. Binary predictive GDE outputs maps from the five best algorithms: random forest classifier (RFC), gradient boosting classification (GBC), AdaBoost classifier (ABC), decision tree classifier (CRT), extra-trees classifier (EXT). The ensemble map (ENS) averages all five algorithms into a predictive GDE occurrence (Very Low to Very High) probability.

The results from the models indicate a bias in training data. Model performance in delineating aquatic flowing systems and terrestrial (open water wetlands) GDEs varied considerably, but all binary models tended to label the centre of the lake as non-GDEs while labeling the shorelines as GDEs. This is likely a result of the training sets including only wetland and river features (i.e., no lake training data), and highlights the need to include datasets from boreal lakes into the training model, such as the Regional Aquatic Monitoring Program (RAMP) lakes datasets, to improve GDE delineation in non-flowing systems(Gibson et al., 2019).

8.3 Ensemble GDE Probability Map

The ensemble map (ENS) averages the outputs of all five binary models and is then able to produce a predictive map of GDE the likelihood of a GDE based on five classes (very low to very high), broken down into probability steps of 20% (e.g., 1 - 20% = very low; 80% - 100% = very high). The ENS map highlights flowing systems well and shows lower GDE probability with both distance from the river and while moving into the headwaters (Figure 8). The ENS map also highlights small regions of terrestrial GDEs that are not consistent between all models by showing lower GDE probabilities in these areas (e.g., regions west of McKay River headwaters are identified as high GDE probability, while along the Steepbank River area some sections show lower GDE probability). Edge effects are seen at the two small unnamed lakes east of Fort McMurray, where the interior of the lake showed a lower GDE probability and the shoreline showed a higher GDE probability (Figure 8).

In terms of human footprint, the top 5 models all predict several of the mining areas as being GDEs, resulting in their having a high probability of being GDEs in the resulting ensemble map (Figure 9). This is most likely a result of the input DEM data, which captures the lower elevations in features such as mining pits and tailings ponds and thus suggests to the models higher GDE likelihood. This highlights the strong role the DEM input layer plays in predicting GDE location and occurrence.

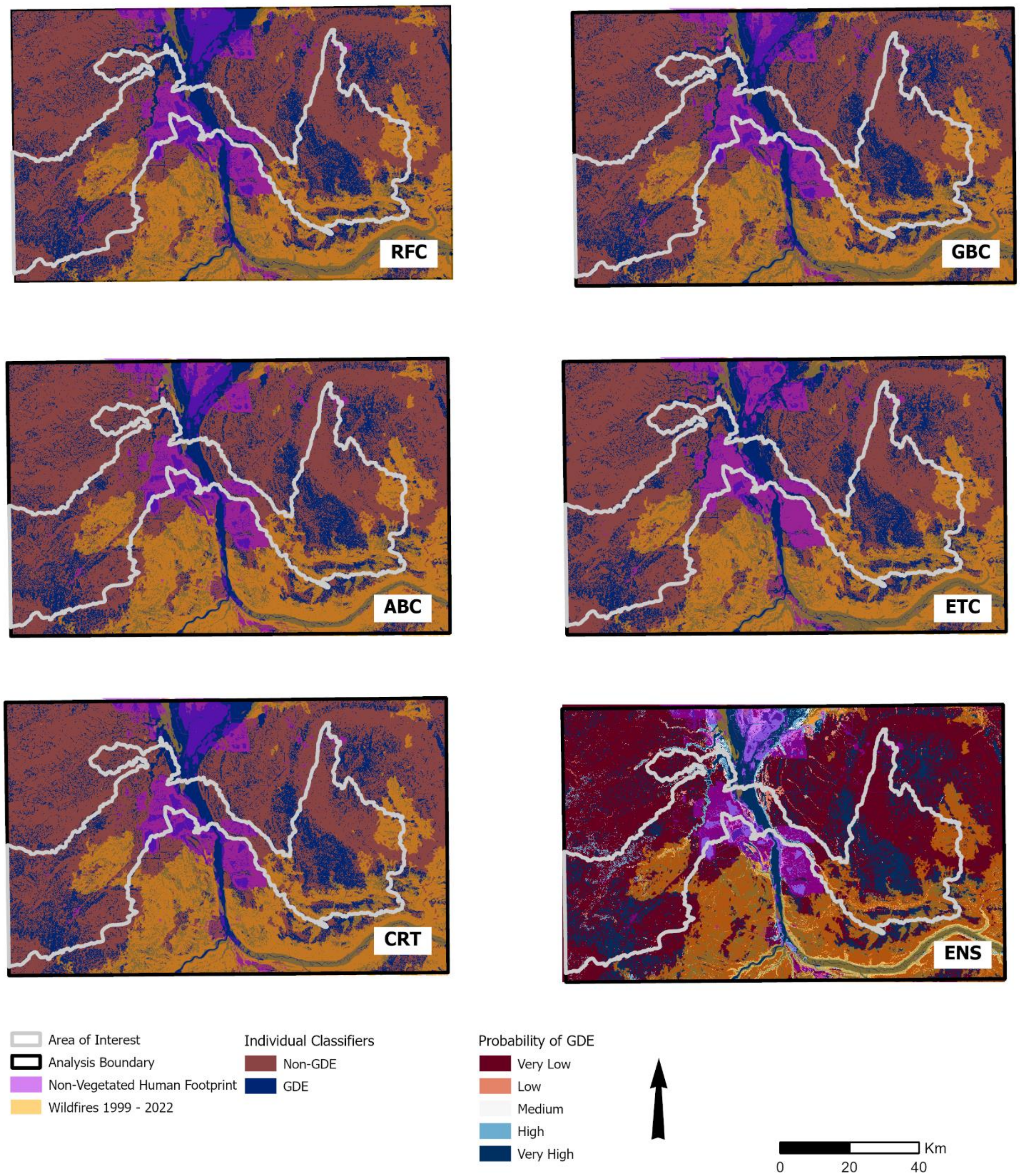

Figure 9. Binary predictive GDE outputs maps from the five best algorithms: random forest classifier (RFC), gradient boosting classification (GBC), AdaBoost classifier (ABC), decision tree classifier (CRT), extra-trees classifier (EXT). The ensemble map (ENS) averages all five algorithms into a predictive GDE occurrence (Very Low to Very High) probability. Both human footprint (mining areas; purple) and previous fire records (1999-2020; orange) overlay the area of analysis.

Overlaying fire data onto the final map seems to show that many of the previously burned areas are predicted to be non-GDEs or sparsely populated GDE systems on the ENS map (Figures 9 and 10). This could be further explored. Thompson et al., (2019) noted that peatlands can have a fragmenting effect during wet years, protecting the landscape from fires, with the inverse being true during times of drought.

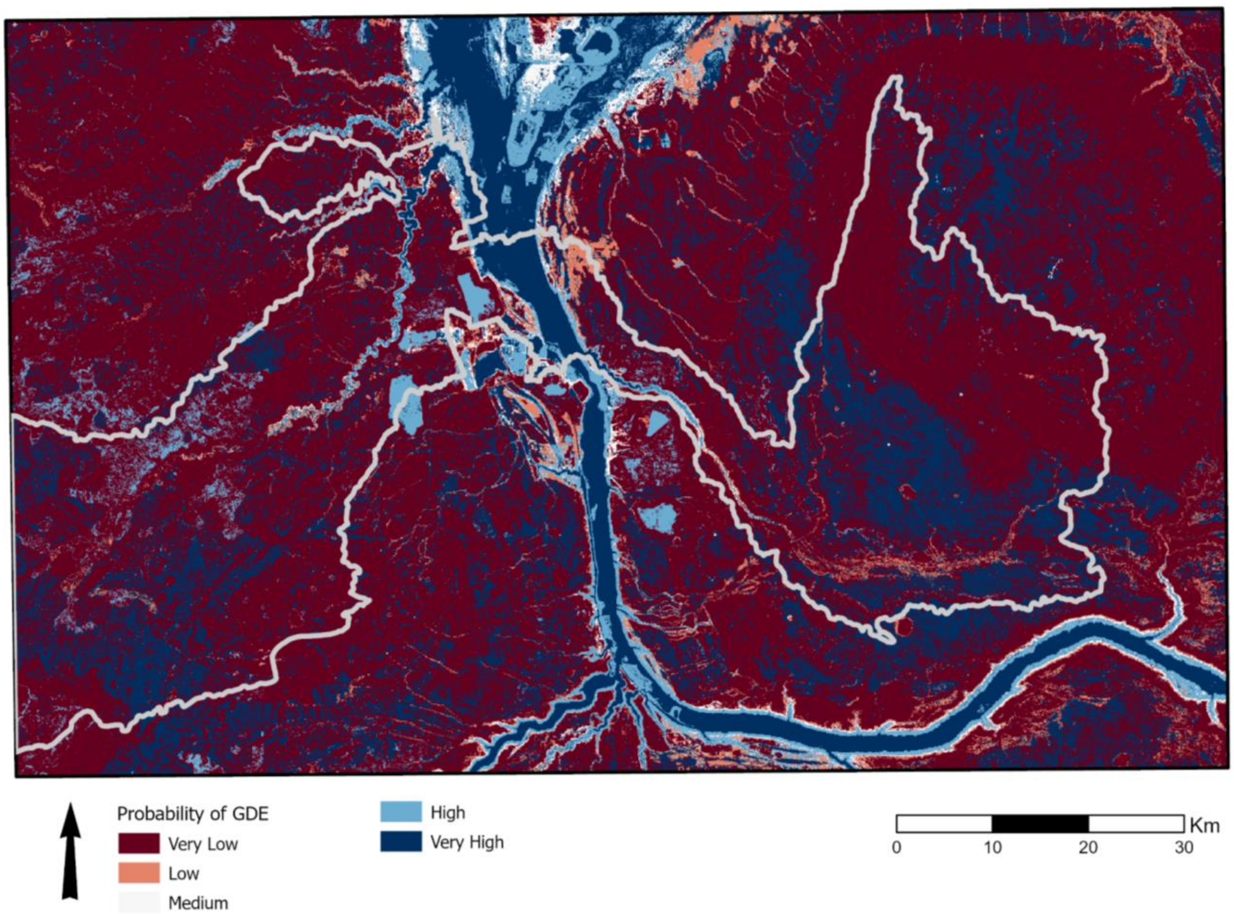

Figure 9. Map of GDE probability across the study area (white line) and area of analysis (full extent) produced by the ENS model.

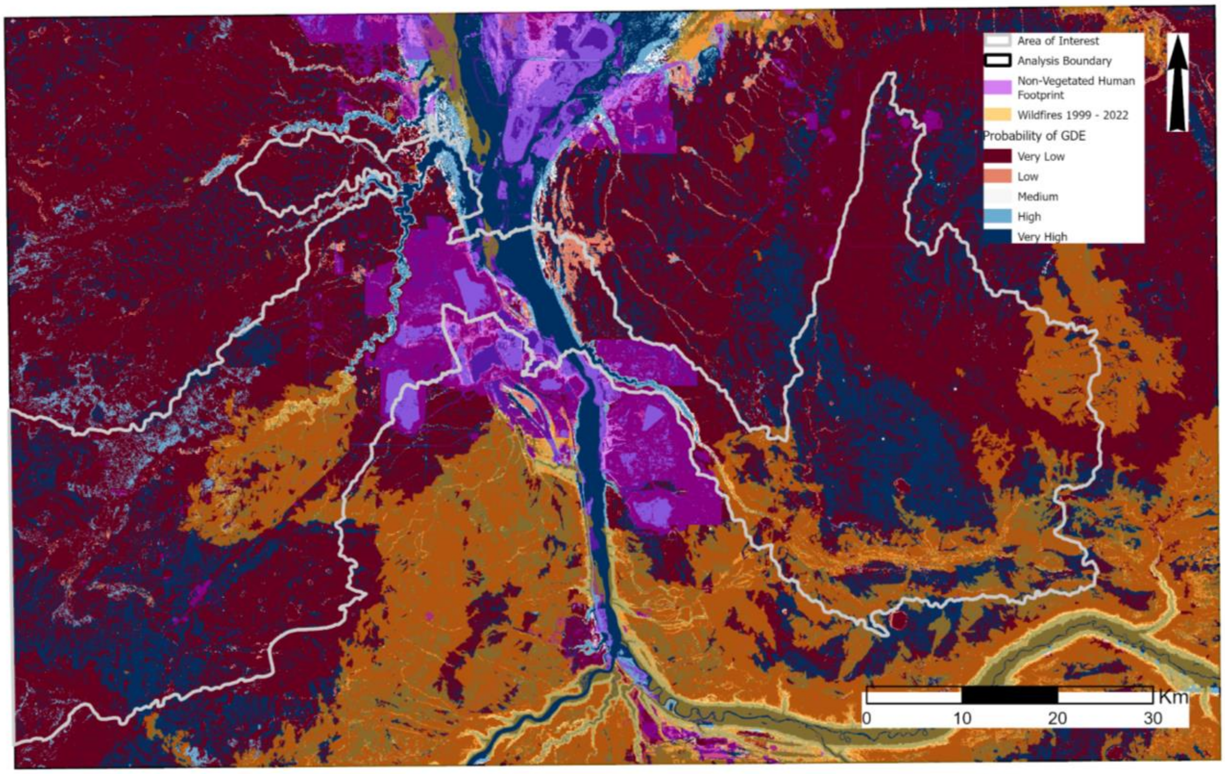

Figure 10. Map of GDE probability across the study area (white line) and area of analysis (full extent) produced by the ENS model. Both human footprint (mining areas; purple) and previous fire records (1999-2020; orange) overlay the area of analysis.

Figures 11 and 12 show the ENS map over a portion of the study area at a more detailed scale, with or without human footprint and fire features, respectively. These close-ups focus on the McKay River, where a buffer area around the river can be more clearly seen, and the channel of the river showing high GDE probability. This probability decreases with distance from the channel, although there are sections where shift is very less gradual and very sudden. In the downstream reaches of the McKay River, the width of the high probability GDE areas undulates slightly in certain regions, however overall they appear to be slightly widening (Figure 11). Fire and human footprint do not seem to play a role in GDE predicted likelihood, but rather, reflect regions where fewer GDEs are present (Figure 12).

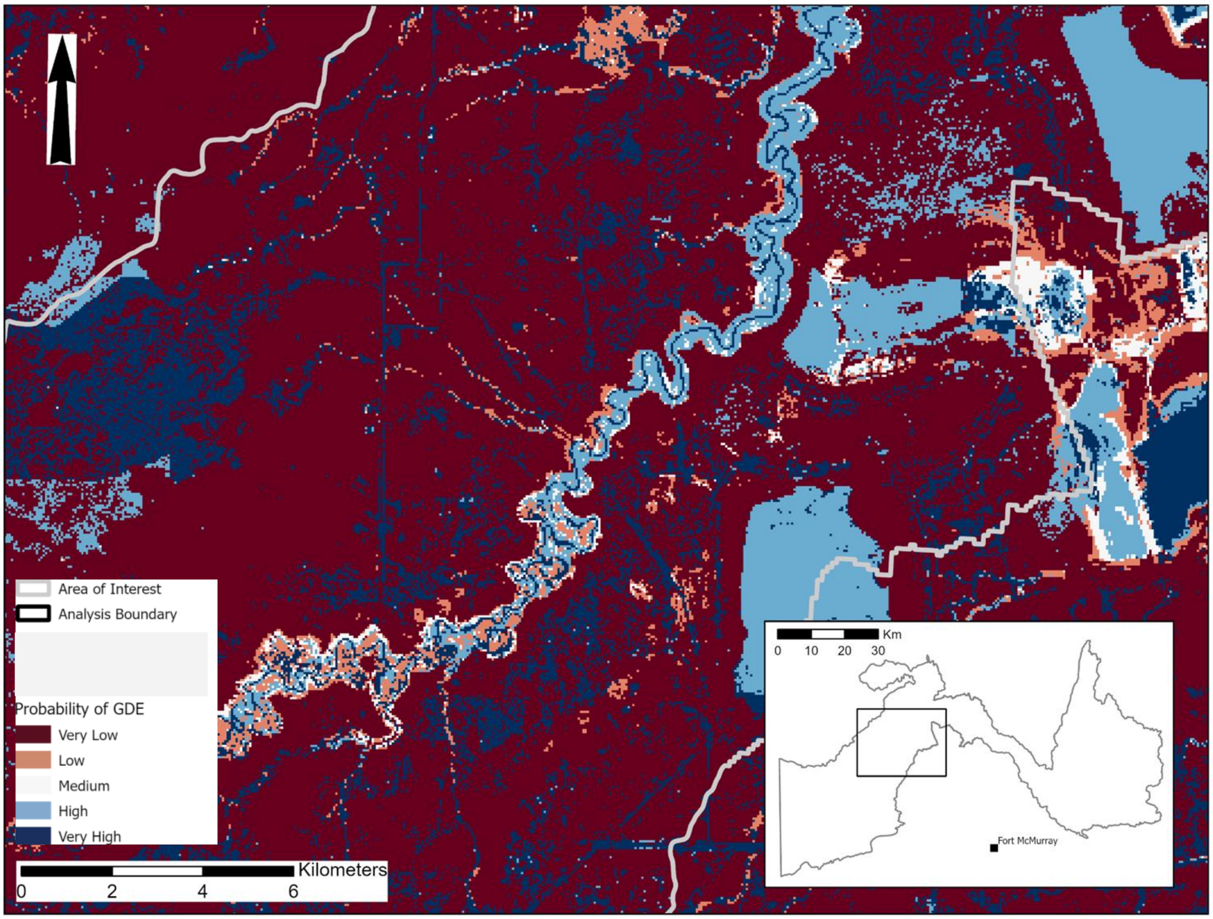

Figure 11. A focused look at the Fort McKay River within the study region, based on the ensemble of the top five mapping algorithms.

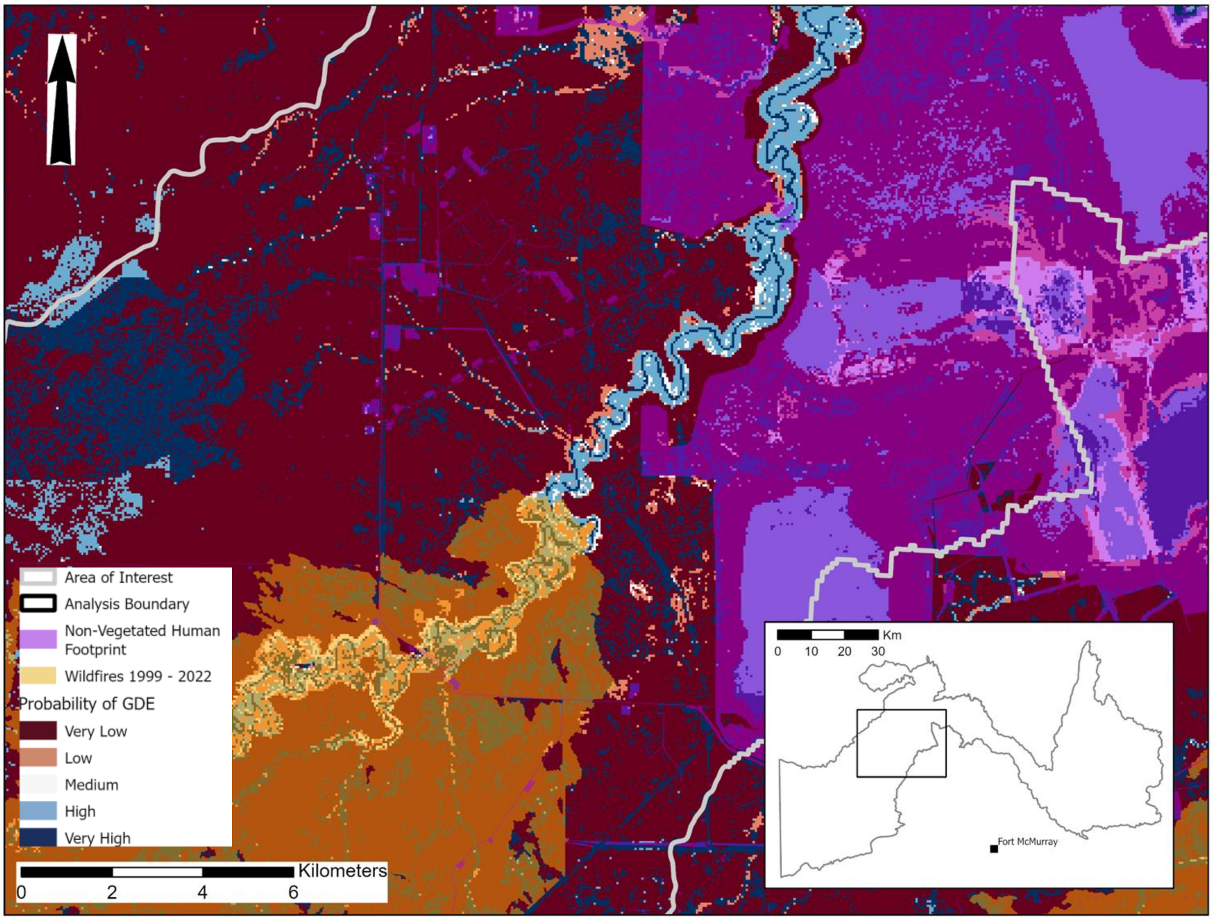

Figure 12. A focused look at the Fort McKay River within the study region, with both human footprint (purple) and previous fires 1999-2020 (orange), based on the ensemble of the top five mapping algorithms.

8.4 Next Steps

8.4.1 Alternative Modeling Approaches

While ensemble models generally score higher in validation assessments (Dormann 2018), this gain in overall model fit must be considered against potential drawbacks of model averaging in general. Averaging outputs of multiple models has the effect of converging different values in individual models that may be more or less accurate for certain locations, classes, etc. So, whereas the ensemble model is generally more accurate, individual models may be more accurate for certain applications. To assess these specific applications requires a more comprehensive and data intensive validation process that could form part of the modeling work going forward. Further, to get at more accurate models for specific classes or region, instead of averaging all model outputs together into a single ensemble, separate modeling efforts could be undertaken for specific environmental features, habitats, or environments (e.g., separate modeling efforts for riverine vs. wetland vs. upland environments), then combined. While this could result in more accurate models, it also compounds the effort and computational intensity of the modeling process and generally requires larger datasets, specific to each modeling effort, which could be a major barrier to this approach within this project.

8.4.2 Model Generalization and Validation

Models built for this project were trained and tested only within our relatively small study area, so their capacity for generalization to other regions (i.e., predictions in new spatial locations) throughout the OSR has not been rigorously tested. All model projections beyond the training data locations assume consistency in how explanatory variables (i.e., our 11 predictors) relate to response variables (i.e. existence of GDEs or not) between the model training and validation locations–an assumption known as stationarity. In reality, environments seldom behave in this way and stationarity is not maintained over larger areas, so using locally trained models to project to larger areas or to locations farther away in space should be done with extreme caution and should always, if possible, be validated with independent data from the projection area. There are alternative statistical cross-validation approaches that can produce better accuracy estimates for model projection to new areas, and these could be explored in the future if additional data are not available to test model projections to new areas.

In addition to testing model capacity for projection, enhanced model validations could be implemented to provide more specific accuracy statistics. For example, GDE predictions could be validated against mapped wetland classes to provide a more comprehensive understanding of where (i.e., for which wetland classes) the model was performing better or worse. Such validations would help inform modeling next steps (including the potential composite modeling approach described above) by identifying the strengths and weaknesses of the individual models. Such a validation could also provide ecological understanding by comparing which predictor variables are more important in the best model for certain wetland classes (or other mapped features). Again, such validations would require adequate data resources and additional validation data could be required for underrepresented classes in the existing model training data.

This first GDE project phase focused exclusively on the boreal region, with the study area entirely with the Athabasca OSR. Because water resources tend to be readily available for plants in this region, productivity tends to be relatively consistent across the area. For this reason, NDVI is a less powerful indicator of plant stress (i.e., discriminator of wetland vs. upland habitats) than it would be in, for example, more arid environments where gradients of productivity are more closely tied to water availability. For this reason, in arid environments, phreatophytes (species able to access deep water resources) would be stronger indicators of GDEs than they are in the boreal where the local source of their water is more uncertain. An additional confounder of leveraging NDVI data in the boreal is the natural fire regime, which can complicate the interpretation and modeling integration of NDVI values if not paired with knowledge of recent burns or regenerating stand ages.

8.4.3 Other Data & Knowledge

There are additional data not immediately available for this first modeling effort that could improve model accuracies (overall or in specific environments) and potentially increase model generalizability or transferability. These could include other data such as:

- Higher resolution depth to water, which could improve both aquatic and terrestrial GDE detection;
- Lake datasets (i.e., RAMP lakes) to train the model to better predict lake features;
- High resolution thermal imagery, which would better capture groundwater inputs of aquatic reaches such as lakes, rivers, and open water wetlands;
- Full lidar coverage of the OSR (in progress via ABMI), which supports the development of very high resolution bare earth model DEMs capable of detecting smaller GDEs;
- Updated wetland inventories for the OSR, created using high-resolution lidar; and
- Additional consideration of burned and human footprint areas and any confounding effects they may have on outcomes.

Finally, local Indigenous communities have developed deep knowledge of their traditional lands, including the location and importance of many GDEs (e.g., mineral licks used by local mammal populations). Integrating western science approaches with other ways of knowing, including Indigenous Knowledge, would strengthen this work. Further, there are valuable data being collected via Indigenous Community Based Monitoring programs within the OSR (e.g., isotopes, water geochemistry) that could be applicable to mapping GDEs. Opportunities for collaboration between western science approaches and Indigenous communities are likely to result in beneficial understanding for both local communities and the Oil Sands Monitoring program.

9. Conclusions and Recommendations

We developed the first aquatic GDE map (Figure 9) for a study area in the OSR to fulfill a key knowledge gap in the OSM program related to identifying the locations of GDEs. This work supports long-term planning for groundwater monitoring and could be used to assist in identifying where baseline, change and effects-based monitoring could be considered for GDE receptors. This milestone marks the successful completion of Year 1 of a multi-year project with the long-term objective of mapping GDEs across the OSR. While the focus of Year 1 was on mapping aquatic GDEs, future years aim to focus on developing methods for mapping other GDE categories (e.g., terrestrial, subterranean) and scaling-up the application of these approaches across the larger OSR.

All three phases of work were completed: 1) developing an approach for mapping GDEs, 2) evaluating data availability, and 3) preliminary mapping of one GDE category in a pilot study area. The first phase, developing the approach, involved defining the categories of GDEs in Alberta's boreal (aquatic, terrestrial, subterranean), conducting a modest literature review of both groundwater and biological indicators (with a focus on aquatic GDEs), and reviewing and selecting geospatial approaches for mapping GDEs with careful attention to methods that may be suitable in boreal environments (because much more work has been done in arid regions globally). The literature review for biological indicators of aquatic GDEs was underscored by the consistent messaging, across approaches, locations, and taxa that our collective understanding of GDEs is lacking due to limited research and the often cryptic nature of these ecosystems. Suitable biological indicators for the specific identification and mapping of GDEs in boreal Alberta cannot be firmly backed by empirical evidence at this time, nor are the potential relationships between stressors, pathways, and ecological endpoint responses clear. The use of wetland types, e.g., bogs and various fen types, in mapping GDEs is supported by our understanding of the dependence of these wetlands on groundwater (AESRD, 2015). Considering macroinvertebrate, microbial, and stygofauna indicators, the paucity of information from Canada presents a limitation. That GDEs house unique assemblages of these taxa is clear from the literature, however defining the components and distribution of assemblages in boreal Alberta would require further work.

We selected a machine-learning based geospatial approach for mapping aquatic GDEs using MLMapper developed by Martinez-Santos (2019) because remote sensing-based GDE mapping approaches are the most cost effective and accessible for large-scale application, machine learning approaches allow the leveraging of multiple data sources with different data types to support higher predictive accuracy even where data may be limited and they avoid challenges of bias associated with dependency on expert opinion. The specific machine learning approach chosen offers the ability to incorporate a wide range of geospatial data layers, applies a broad set of machine learning model algorithms simultaneously, and offers the advantage of an ensemble product that combines the highest accuracy individual models. Finally, the flexibility of this approach means that additional datasets can easily be incorporated into future applications, to test opportunities to improve model accuracy.

The second phase, evaluating data availability, included identifying and collating available data for mapping GDEs, selecting appropriate data to serve as training & validation data and explanatory variables in MLMapper model, and identifying data gaps. Over 50 datasets were identified, with over 40 datasets compiled. The key data gaps are access to the McKay River Integrated Surface Water-Groundwater Model, hydraulic head data, and higher resolution thermal data. Recommended next steps to filling these data gaps include ongoing communication with the producers of the McKay River Model, working with AGS to access updated hydraulic head data in 2024/25, and evaluating additional options for higher resolution thermal data from satellite or aerial collection. In addition, we will continue to work with the Fort McKay Métis Nation to enable considerate access to their data to increase validation datasets and consider the use of the RAMP lake datasets to improve prediction for boreal lakes.

The third phase was completed by mapping aquatic GDEs in the McKay and Steepbank River watersheds using the methods identified in phase 1 and the data collated in phase 2. The MLMapper model was trained and tested with binary GDE presence/absence data including from wetlands, springs, and reaches of the McKay River with differential gauging measurements. The final models included 11 explanatory variables, the most important being wetland class, DEM, surface wetness index, slope, permeability (derived from surficial geology) and NDVI. We selected the top five individual classifier models (GDE presence/absence), each of which had slightly different results (particularly for riverine GDEs), and created an ensemble map of GDE probability (with five classes from low to high). GDEs primarily occur along the lower reaches of rivers, riparian areas and fens. Maps are presented both with and without non-vegetated human footprint to visualize the juxtaposition of oil sands footprint with GDE probability. Future modeling work aims to explore individual model accuracies for certain GDE classes or sub-classes, evaluate the generalizability of the model by selecting another Study Area with independent validation data, and test model performance with different subsets of explanatory variables and spatial resolution of explanatory variable datasets (because the Year 1 Study Area likely had the highest quality data within the OSR).

This project has highlighted the ability to leverage existing data to effectively map GDEs in a portion of the OSR that is fairly well studied. This proof of concept supports the expansion of GDE mapping to the broader OSR, while recognizing that addressing some data gaps and ensuring a broader set of test data is necessary to enable successful expansion. The mapping of GDEs completed to date and the opportunities identified for future application of these techniques will support the Groundwater Technical Advisory Committee in answering key questions as they develop a monitoring approach that will evaluate the impact of oil sands-related stressors on GDEs as a key environmental endpoint.

10. References

AESRD. (2015). *Alberta Wetland Classification System*. Water Policy Branch, Policy and Planning Division, Alberta Environment and Sustainable Resource Development, Government of Alberta.

Ala-aho, P., Rossi, P. M., Isokangas, E., & Kløve, B. (2015). Fully integrated surface–subsurface flow modelling of groundwater–lake interaction in an esker aquifer: Model verification with stable isotopes and airborne thermal imaging. *Journal of Hydrology*, *522*, 391–406. https://doi.org/10.1016/j.jhydrol.2014.12.054

Alberta Biodiversity Monitoring Institute & Alberta Human Footprint Monitoring Program. (2023). *ABMI Human Footprint Inventory (HFI) for Alberta 2021 (version 1.0)* (1.0) [Geodatabase].

Alberta Biodiversity Monitoring Institute & Ducks Unlimited Canada. (2023). *2022/23 Wetland Inventory Program Annual Report* (Prepared for the Oil Sands Monitoring Program, p. 59).

Alberta Environment and Parks, Government of Alberta. (2022). *Alberta Merged Wetland Inventory*. https://www.arcgis.com/home/item.html?id=6933a7b62a11440bbc2d69a0d6d047c0

Alberta Geological Survey. (2021). *Hydrogeological Regions of Alberta Story Map Collection: Interactive App and Map 009 (Website)*. Alberta Energy Regulator / Alberta Geological Survey. https://ags.aer.ca/data-maps-models/interactive-maps-and-apps/iam-009

Altosaar, A. K. (2013a). *Stratigraphy, sedimentology, and karst features associated with Devonian carbonates and evaporites found within Wood Buffalo National Park, northeast Alberta* (Summer Field Report WB-2013-13333; p. 115). Suncor Energy Inc.

Altosaar, A. K. (2013b, June 17). *The Geologic Wonders of Wood Buffalo National Park, Alberta*. GeoConvention 2013: Integration, Calgary, AB, Canada.

Andriashek, L. D. (2001). *Quaternary Stratigraphy of the Buried Birch and Willow Bedrock Channels, NE Alberta* (ESR 2000-15; Earth Sciences Report, p. 61). Alberta Energy and Utilities Board / Alberta Geologic Survey. https://static.ags.aer.ca/files/document/ESR/ESR_2000_15.pdf

Andriashek, L. D. (2007). *Buried Channels and Glacial-Drift Aquifers in the Fort McMurray Region, Northeast Alberta* (ESR 2007-01). Alberta Energy and Utilities Board / Alberta Geologic Survey. https://ags.aer.ca/publication/esr-2007-01

Anibas, C., Buis, K., Verhoeven, R., Meire, P., & Batelaan, O. (2011). A simple thermal mapping method for seasonal spatial patterns of groundwater–surface water interaction. *Journal of Hydrology*, *397*(1), 93–104. https://doi.org/10.1016/j.jhydrol.2010.11.036

Arnoux, M., Gibert-Brunet, E., Barbecot, F., Guillon, S., Gibson, J., & Noret, A. (2017). Interactions between groundwater and seasonally ice-covered lakes: Using water stable isotopes and radon-222 multilayer mass balance models. *Hydrological Processes*, *31*(14), 2566–2581. https://doi.org/10.1002/hyp.11206

Atkinson, N. (2022a). *Surficial Geology of the Telephone Lake Area (NTS 74E/SE)* [Map]. Alberta Geologic Survey. https://ags.aer.ca/publication/map-621

Atkinson, N. (2022b). *Surficial Geology of the Whitemud Falls Area (NTS 74D/NE)* [Map]. Alberta Geologic Survey. https://ags.aer.ca/publication/map-620

Atkinson, N., Andriashek, L. D., & Slattery, S. R. (2013). Morphological analysis and evolution of buried tunnel valleys in northeast Alberta, Canada. *Quaternary Science Reviews*, *65*, 53–72. https://doi.org/10.1016/j.quascirev.2012.11.031

Atkinson, N., & Pawley, S. M. (2022). *Surficial Geology of the Fort McMurray Area (NTS 74D/NW)* [Map]. Alberta Geologic Survey. https://ags.aer.ca/publication/map-619

Autio, A., Ala-Aho, P., Rossi, P. M., Ronkanen, A.-K., Aurela, M., Lohila, A., Korpelainen, P., Kumpula, T., Klöve, B., & Marttila, H. (2023). Groundwater exfiltration pattern determination in the sub-arctic catchment using thermal imaging, stable water isotopes and fully-integrated groundwater-surface water modelling. *Journal of Hydrology*, *626*, 130342. https://doi.org/10.1016/j.jhydrol.2023.130342

Ayotte, J. B., Parker, K. L., Arocena, J. M., & Gillingham, M. P. (2006). Chemical composition of lick soils: Functions of soil ingestion by four ungulate species. *Journal of Mammalogy*, *87*(5), 878–888. https://doi.org/10.1644/06-MAMM-A-055R1.1

Beasley-Hall, P. G., Murphy, N. P., King, R. A., White, N. E., Hedges, B. A., Cooper, S. J. B., Austin, A. D., & Guzik, M. T. (2023). Time capsules of biodiversity: Future research directions for groundwater-dependent ecosystems of the Great Artesian Basin. *Frontiers in Environmental Science*, *10*, 1021987. https://doi.org/10.3389/fenvs.2022.1021987

Bertrand, G., Siergieiev, D., Ala-Aho, P., & Rossi, P. M. (2014). Environmental tracers and indicators bringing together groundwater, surface water and groundwater-dependent ecosystems: Importance of scale in choosing relevant tools. *Environmental Earth Sciences*, *72*(3), 813–827. https://doi.org/10.1007/s12665-013-3005-8

Bickerton, G., Roy, J. W., Frank, R. A., Spoelstra, J., Langston, G., Grapentine, L., & Hewitt, L. M. (2018). *Assessments of groundwater influence on selected river systems in the oil sands region of Alberta*. Alberta Environment and Parks.

Birks, J. S., Gibson, J. J., Fennell, J. W., & Munoz, F. C. (2022). *Condition of the Environment Report: Groundwater in the Oil Sands Region (DRAFT)* (p. 129). Innotech Alberta, Edmonton, AB, Canada.

Birks, J. S., Jones, J., & Gibson, J. J. (2012). *Surface water-groundwater interactions in the Lower Athabasca Region* (pp. 5–83). Cumulative Environmental Management Association Working Group/Task Group: Groundwater Working Group.

Birks, S. J., Fennell, J. W., Gibson, J. J., Yi, Y., Moncur, M. C., & Brewster, M. (2019). Using regional datasets of isotope geochemistry to resolve complex groundwater flow and formation connectivity in northeastern Alberta, Canada. *Applied Geochemistry*, *101*, 140–159. https://doi.org/10.1016/j.apgeochem.2018.12.013

Birks, S. J., Manchuk, J., Yi, Y., McClain, C. N., Moncur, M. C., Gibson, J. J., Deutsch, C. V., Taylor, E. B., & Bayegnak, G. (2022). Groundwater monitoring near oil sands development: Insights from regional water quality datasets in the Alberta Oil Sands Region (AOSR). *Journal of Hydrology: Regional Studies*, *41*, 101079. https://doi.org/10.1016/j.ejrh.2022.101079

Bousfield, E. L., & Holsinger, J. R. (1981). A second new subterranean amphipod crustacean of the genus Stygobromus (Crangonyctidae) from Alberta, Canada. *Canadian Journal of Zoology*, *59*(9), 1827–1830. https://doi.org/10.1139/z81-250

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L. (2003). *Manual-Setting Up, Using and Understanding Random Forests V4.0*. https://www.stat.berkeley.edu/~breiman/Using_random_forests_v4.0.pdf

Broughton, P. L. (2018). Ghost-rock karstification of Devonian limestone flooring the Athabasca Oil Sands in western Canada. *Geomorphology*, *318*, 303–319. https://doi.org/10.1016/j.geomorph.2018.07.001

Chiloane, C., Dube, T., & Shoko, C. (2022). Impacts of groundwater and climate variability on terrestrial groundwater dependent ecosystems: A review of geospatial assessment approaches and challenges and possible future research directions. *Geocarto International*, *37*(23), 6755–6779. https://doi.org/10.1080/10106049.2021.1948108

Cooper, D. J., & Andrus, R. E. (1994). Patterns of vegetation and water chemistry in peatlands of the west-central Wind River Range, Wyoming, U.S.A. *Canadian Journal of Botany*, *72*(11), 1586–1597. https://doi.org/10.1139/b94-196

Crisfield, V. E., Haughland, D. L., & Pyle, L. A. (2019). Reproductive ecology of impoverished pinweed (*Lechea intermedia* var. *Depauperata*), a fire-associated narrow endemic from the boreal forest. *Botany*, *97*(10), 559–570. https://doi.org/10.1139/cjb-2019-0097

Dole-Olivier, M.-J., Creuzé des Châtelliers, M., Galassi, D. M. P., Lafont, M., Mermillod-Blondin, F., Paran, F., Graillot, D., Gaur, S., & Marmonier, P. (2022). Drivers of functional diversity in the hyporheic zone of a large river. *Science of The Total Environment*, *843*, 156985. https://doi.org/10.1016/j.scitotenv.2022.156985

Doody, T. M., Barron, O. V., Dowsley, K., Emelyanova, I., Fawcett, J., Overton, I. C., Pritchard, J. L., Van Dijk, A. I. J. M., & Warren, G. (2017). Continental mapping of groundwater dependent ecosystems: A methodological framework to integrate diverse data and expert opinion. *Journal of Hydrology: Regional Studies*, *10*, 61–81. https://doi.org/10.1016/j.ejrh.2017.01.003

Dormann, C. F., Calabrese, J. M., Guillera-Arroita, G., Matechou, E., Bahn, V., Bartoń, K., Beale, C. M., Ciuti, S., Elith, J., Gerstner, K., Guelat, J., Keil, P., Lahoz-Monfort, J. J., Pollock, L. J., Reineking, B., Roberts, D. R., Schröder, B., Thuiller, W., Warton, D. I., … Hartig, F. (2018). Model averaging in ecology: A review of Bayesian, information-theoretic, and tactical approaches for predictive inference. *Ecological Monographs*, *88*(4), 485–504. https://doi.org/10.1002/ecm.1309

Eamus, D., Froend, R., Loomes, R., Hose, G., & Murray, B. (2006). A functional methodology for determining the groundwater regime needed to maintain the health of groundwater-dependent vegetation. *Australian Journal of Botany*, *54*(2), 97–114. https://doi.org/10.1071/BT05031

Elmes, M. C., Davidson, S. J., & Price, J. S. (2021). Ecohydrological interactions in a boreal fen–swamp complex, Alberta, Canada. *Ecohydrology*, *14*(7), e2335. https://doi.org/10.1002/eco.2335

Ensom, T., Makarieva, O., Morse, P., Kane, D., Alekseev, V., & Marsh, P. (2020). The distribution and dynamics of aufeis in permafrost regions. *Permafrost and Periglacial Processes*, *31*(3), 383–395. https://doi.org/10.1002/ppp.2051

Eskelinen, R., Ala-aho, P., Rossi, P. M., & Kløve, B. (2015). A GIS-based method for predicting groundwater discharge areas in esker aquifers in the Boreal region. *Environmental Earth Sciences*, *74*(5), 4109–4118. https://doi.org/10.1007/s12665-015-4491-7

Febria, C. M., Beddoes, P., Fulthorpe, R. R., & Williams, D. D. (2012). Bacterial community dynamics in the hyporheic zone of an intermittent stream. *The ISME Journal*, *6*(5), 1078–1088. https://doi.org/10.1038/ismej.2011.173

Fildes, S. G., Doody, T. M., Bruce, D., Clark, I. F., & Batelaan, O. (2023). Mapping groundwater dependent ecosystem potential in a semi-arid environment using a remote sensing-based multiple-lines-of-evidence approach. *International Journal of Digital Earth*, *16*(1), 375–406. https://doi.org/10.1080/17538947.2023.2176557

Ford, D. (1987). Effects of glaciations and permafrost upon the development of karst in Canada. *Earth Surface Processes and Landforms*, *12*(5), 507–521. https://doi.org/10.1002/esp.3290120508

Ford, D. C. (1997). Principal features of evaporite karst in Canada. *Carbonates and Evaporites*, *12*(1), 15–23. https://doi.org/10.1007/BF03175798

Fraser, B. G., & Williams, D. D. (1998). Seasonal Boundary Dynamics of a Groundwater/ Surface-Water Ecotone. *Ecology*, *79*(6), 2019–2031. https://doi.org/10.1890/0012-9658(1998)079[2019:SBDOAG]2.0.CO;2

Freeze, R., & Cherry, J. (1979). *Groundwater*. Prentice-Hall.

Gerlach, M. E., Rains, K. C., Guerrón-Orejuela, E. J., Kleindl, W. J., Downs, J., Landry, S. M., & Rains, M. C. (2022). Using Remote Sensing and Machine Learning to Locate Groundwater

Discharge to Salmon-Bearing Streams. *Remote Sensing*, *14*(1), Article 1. https://doi.org/10.3390/rs14010063

Gibson, J. J., Birks, S. J., Yi, Y., Moncur, M. C., & McEachern, P. M. (2016). Stable isotope mass balance of fifty lakes in central Alberta: Assessing the role of water balance parameters in determining trophic status and lake level. *Journal of Hydrology: Regional Studies*, *6*, 13–25. https://doi.org/10.1016/j.ejrh.2016.01.034

Gibson, J. J., Eby, P., Birks, S. J., Twitchell, C., Gray, C., & Kariyeva, J. (2022). Isotope-based water balance assessment of open water wetlands across Alberta: Regional trends with emphasis on the oil sands region. *Journal of Hydrology: Regional Studies*, *40*, 101036. https://doi.org/10.1016/j.ejrh.2022.101036

Gibson, J. J., Fennell, J., Birks, S. J., Yi, Y., Moncur, M. C., Hansen, B., & Jasechko, S. (2013). Evidence of discharging saline formation water to the Athabasca River in the oil sands mining region, northern Alberta. *Canadian Journal of Earth Sciences*, *50*(12), 1244–1257. https://doi.org/10.1139/cjes-2013-0027

Gibson, J. J., & Peters, D. L. (2022). Water and environmental management in oil sands regions. *Journal of Hydrology: Regional Studies*, *44*, 101274. https://doi.org/10.1016/j.ejrh.2022.101274

Gibson, J. J., Yi, Y., & Birks, S. J. (2016). Isotope-based partitioning of streamflow in the oil sands region, northern Alberta: Towards a monitoring strategy for assessing flow sources and water quality controls. *Journal of Hydrology: Regional Studies*, *5*, 131–148. https://doi.org/10.1016/j.ejrh.2015.12.062

Gibson, J. J., Yi, Y., & Birks, S. J. (2019). Isotopic tracing of hydrologic drivers including permafrost thaw status for lakes across Northeastern Alberta, Canada: A 16-year, 50-lake assessment. *Journal of Hydrology: Regional Studies*, *26*, 100643. https://doi.org/10.1016/j.ejrh.2019.100643

Gibson, J. J., Yi, Y., & Birks, S. J. (2020). Watershed, climate, and stable isotope data (oxygen-18 and deuterium) for 50 boreal lakes in the oil sands region, northeastern Alberta, Canada, 2002–2017. *Data in Brief*, *29*, 105308. https://doi.org/10.1016/j.dib.2020.105308

Glaser, P. H., Siegel, D. I., Reeve, A. S., Janssens, J. A., & Janecky, D. R. (2004). Tectonic drivers for vegetation patterning and landscape evolution in the Albany River region of the Hudson Bay Lowlands. *Journal of Ecology*, *92*(6), 1054–1070. https://doi.org/10.1111/j.0022-0477.2004.00930.x

Goodbrand, A., Westbrook, C. J., & van der Kamp, G. (2019). Hydrological functions of a peatland in a Boreal Plains catchment. *Hydrological Processes*, *33*(4), 562–574. https://doi.org/10.1002/hyp.13343

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Graillot, D., Paran, F., Bornette, G., Marmonier, P., Piscart, C., & Cadilhac, L. (2014). Coupling groundwater modeling and biological indicators for identifying river/aquifer exchanges. *SpringerPlus*, *3*, 68. https://doi.org/10.1186/2193-1801-3-68

Graup, L. J., Tague, C. L., Harpold, A. A., & Krogh, S. A. (2022). Subsurface Lateral Flows Buffer Riparian Water Stress Against Snow Drought. *Journal of Geophysical Research: Biogeosciences*, *127*(12), e2022JG006980. https://doi.org/10.1029/2022JG006980

Groult, B., St-Jean, V., & Lazar, C. S. (2023). Linking Groundwater to Surface Discharge Ecosystems: Archaeal, Bacterial, and Eukaryotic Community Diversity and Structure in Quebec (Canada). *Microorganisms*, *11*(7), 1674. https://doi.org/10.3390/microorganisms11071674

Gue, A. E., Mayer, B., & Grasby, S. E. (2015). Origin and geochemistry of saline spring waters in the Athabasca oil sands region, Alberta, Canada. *Applied Geochemistry*, *61*, 132–145. https://doi.org/10.1016/j.apgeochem.2015.05.015

Gue, A., Grasby, S. E., & Mayer, B. (2018). Influence of saline groundwater discharge on river water chemistry in the Athabasca oil sands region – A chloride stable isotope and mass balance approach. *Applied Geochemistry*, *89*, 75–85. https://doi.org/10.1016/j.apgeochem.2017.10.004

Hancock, P. J., Boulton, A. J., & Humphreys, W. F. (2005). Aquifers and hyporheic zones: Towards an ecological understanding of groundwater. *Hydrogeology Journal*, *13*(1), 98–111. https://doi.org/10.1007/s10040-004-0421-6

Hayashi, M., & van der Kamp, G. (2023). The role of Canadian research in advancing groundwater hydrology: Historical sketches from the past 75 years. *Canadian Water Resources Journal / Revue Canadienne Des Ressources Hydriques*, *48*(4), 363–378. https://doi.org/10.1080/07011784.2023.2177197

Hayashi, M., van der Kamp, G., & Rosenberry, D. O. (2016). Hydrology of Prairie Wetlands: Understanding the Integrated Surface-Water and Groundwater Processes. *Wetlands*, *36*(2), 237–254. https://doi.org/10.1007/s13157-016-0797-9

Heagle, D., Hayashi, M., & Kamp, G. van der. (2013). Surface–subsurface salinity distribution and exchange in a closed-basin prairie wetland. *Journal of Hydrology*, *478*, 1–14. https://doi.org/10.1016/j.jhydrol.2012.05.054

Hein, F. J., & Cotterill, D. K. (2006). The Athabasca Oil Sands—A Regional Geological Perspective, Fort McMurray Area, Alberta, Canada. *Natural Resources Research*, *15*(2), 85–102. https://doi.org/10.1007/s11053-006-9015-4

Holsinger, J. R. (1980). Stygobromus canadensis, a new subterranean amphipod crustacean (Crangonyctidae) from Canada, with remarks on Wisconsin refugia. *Canadian Journal of Zoology*, *58*(2), 290–297. https://doi.org/10.1139/z80-034

Hose, G. C., Chariton, A. A., Daam, M. A., Di Lorenzo, T., Galassi, D. M. P., Halse, S. A., Reboleira, A. S. P. S., Robertson, A. L., Schmidt, S. I., & Korbel, K. L. (2022). Invertebrate traits, diversity and the vulnerability of groundwater ecosystems. *Functional Ecology*, *36*(9), 2200–2214. https://doi.org/10.1111/1365-2435.14125

Huryn, A. D., Gooseff, M. N., Hendrickson, P. J., Briggs, M. A., Tape, K. D., & Terry, N. C. (2021). Aufeis fields as novel groundwater-dependent ecosystems in the arctic cryosphere. *Limnology and Oceanography*, 66, 607–624. https://doi.org/doi: 10.1002/lno.11626

Ionescu, D., Bizic, M., Karnatak, R., Musseau, C. L., Onandia, G., Kasada, M., Berger, S. A., Nejstgaard, J. C., Ryo, M., Lischeid, G., Gessner, M. O., Wollrab, S., & Grossart, H.-P. (2022). From microbes to mammals: Pond biodiversity homogenization across different land-use types in an agricultural landscape. *Ecological Monographs*, *92*(3), e1523. https://doi.org/10.1002/ecm.1523

Isokangas, E., Rossi, P. M., Ronkanen, A., Marttila, H., Rozanski, K., & Kløve, B. (2017). Quantifying spatial groundwater dependence in peatlands through a distributed isotope mass balance approach. *Water Resources Research*, *53*(3), 2524–2541. https://doi.org/10.1002/2016WR019661

Jeglum, J. K. (1991). Definition of trophic classes in wooded peatlands by means of vegetation types and plant indicators. *Annales Botanici Fennici*, *28*(3), 175–192.

Klausmeyer, K., Howard, J., Keeler-Wolf, T., Davis-Fadtke, K., Hull, R., & Lyons, A. (2018). *Mapping Indicators of Groundwater Dependent Ecosystems in California: Methods Report*. The Nature Conservancy. https://www.groundwaterresourcehub.org/where-we-work/california/mapping-indicators-gdes/

Kløve, B., Ala-aho, P., Bertrand, G., Boukalova, Z., Ertürk, A., Goldscheider, N., Ilmonen, J., Karakaya, N., Kupfersberger, H., Kvœrner, J., Lundberg, A., Mileusnić, M., Moszczynska, A., Muotka, T., Preda, E., Rossi, P., Siergieiev, D., Šimek, J., Wachniew, P., … Widerlund, A. (2011). Groundwater dependent ecosystems. Part I: Hydroecological status and trends. *Environmental Science & Policy*, *14*(7), 770–781. https://doi.org/10.1016/j.envsci.2011.04.002

Kuginis, L., Dabovic, J., Byrne, G., Raine, A., & Hemakumara, H. (2016). *Methods for the identification of high probability groundwater dependent vegetation ecosystems*. Department of Primary Industries: Water, Government of New South Wales. https://nla.gov.au/nla.obj-3144390969/

Kuglerová, L., Dynesius, M., Laudon, H., & Jansson, R. (2016). Relationships Between Plant Assemblages and Water Flow Across a Boreal Forest Landscape: A Comparison of Liverworts, Mosses, and Vascular Plants. *Ecosystems*, *19*(1), 170–184. https://doi.org/10.1007/s10021-015-9927-0

Land, E., & Peters, C. N. (2023). Groundwater impacts on stream biodiversity and communities: A review. *Journal of Freshwater Ecology*, *38*(1), 2260801. https://doi.org/10.1080/02705060.2023.2260801

Larocque, M., Ferlatte, M., Pellerin, S., Cloutier, V., Munger, J. L., Paniconi, C., & Quillet, A. (2016). Chemical and botanical indicators of groundwater inflow to *Sphagnum*-dominated peatlands. *Ecological Indicators*, *64*, 142–151. https://doi.org/10.1016/j.ecolind.2015.12.012

Larocque, M., & Leblon, B. (2022). *Remote Sensing of Ground-Water Dependent Ecosystems in the Oil Sands Area (Final Report)* (pp. 1–28). University of New Brunswick.

Lehosmaa, K., Jyväsjärvi, J., Ilmonen, J., Rossi, P. M., Paasivirta, L., & Muotka, T. (2018). Groundwater contamination and land drainage induce divergent responses in boreal spring ecosystems. *Science of The Total Environment*, *639*, 100–109. https://doi.org/10.1016/j.scitotenv.2018.05.126

Lidberg, W., Nilsson, M., & Ågren, A. (2020). Using machine learning to generate high-resolution wet area maps for planning forest management: A study in a boreal forest landscape. *Ambio*, *49*(2), 475–486. https://doi.org/10.1007/s13280-019-01196-9

Lilles, E. B., Purdy, B. G., Chang, S. X., & Macdonald, S. E. (2010). Soil and groundwater characteristics of saline sites supporting boreal mixedwood forests in northern Alberta. *Canadian Journal of Soil Science*, *90*(1), 1–14. https://doi.org/10.4141/CJSS08040

Link, A., El-Hokayem, L., Usman, M., Conrad, C., Reinecke, R., Berger, M., Wada, Y., Coroama, V., & Finkbeiner, M. (2023). Groundwater-dependent ecosystems at risk – global hotspot analysis and implications. *Environmental Research Letters*, *18*(9), 094026. https://doi.org/10.1088/1748-9326/acea97

Luke, S. H., Luckai, N. J., Burke, J. M., & Prepas, E. E. (2007). Riparian areas in the Canadian boreal forest and linkages with water quality in streams. *Environmental Reviews*, *15*(NA), 79–97. https://doi.org/10.1139/A07-001

Malczewski, J. (2004). GIS-based land-use suitability analysis: A critical overview. *Progress in Planning*, *62*(1), 3–65. https://doi.org/10.1016/j.progress.2003.09.002

Mammola, S., Amorim, I. R., Bichuette, M. E., Borges, P. A. V., Cheeptham, N., Cooper, S. J. B., Culver, D. C., Deharveng, L., Eme, D., Ferreira, R. L., Fišer, C., Fišer, Ž., Fong, D. W., Griebler, C., Jeffery, W. R., Jugovic, J., Kowalko, J. E., Lilley, T. M., Malard, F., … Cardoso, P. (2020). Fundamental research questions in subterranean biology. *Biological Reviews*, *95*(6), 1855–1872. https://doi.org/10.1111/brv.12642

Manchuk, J. G., Birks, J. S., McClain, C. N., Bayegnak, G., Gibson, J. J., & Deutsch, C. V. (2021). Estimating Stable Measured Values and Detecting Anomalies in Groundwater Geochemistry Time Series Data Across the Athabasca Oil Sands Area, Canada. *Natural Resources Research*, *30*(2), 1755–1779. https://doi.org/10.1007/s11053-020-09801-5

Martínez-Santos, P., Díaz-Alcaide, S., De La Hera-Portillo, A., & Gómez-Escalonilla, V. (2021). Mapping groundwater-dependent ecosystems by means of multi-layer supervised classification. *Journal of Hydrology*, *603*, 126873. https://doi.org/10.1016/j.jhydrol.2021.126873

Meyer, J., Zakhary, S., Larocque, M., & Lazar, C. S. (2022). From Surface to Subsurface: Diversity, Composition, and Abundance of Sessile and Endolithic Bacterial, Archaeal, and Eukaryotic Communities in Sand, Clay and Rock Substrates in the Laurentians (Quebec, Canada). *Microorganisms*, *10*(1), 129. https://doi.org/10.3390/microorganisms10010129

Miller, C. A., Benscoter, B. W., & Turetsky, M. R. (2015). The effect of long-term drying associated with experimental drainage and road construction on vegetation composition and productivity in boreal fens. *Wetlands Ecology and Management*, *23*(5), 845–854. https://doi.org/10.1007/s11273-015-9423-5

Munger, J., Pellerin, S., Larocque, M., & Ferlatte, M. (2014). Espèces végétales indicatrices des échanges d'eau entre tourbière et aquifère. *Le Naturaliste canadien*, *138*(1), 4–12. https://doi.org/10.7202/1021038ar

Nature Conservancy, & Desert Research Institute. (2023, June). *GlobalGDE*. Global Groundwater Dependent Ecosystems Map, Version 1.2.0. https://kklausmeyer.users.earthengine.app/view/global-gde

Oiffer, A. (n.d.). *Appendix 3: Groundwater Dependent Ecosystems* [Draft Report]. Government of Alberta.

Parks Canada. (2022). *Wood Buffalow National Park—Nature (Website)*. Government of Canada. https://www.pc.gc.ca/pn-np/nt/woodbuffalo/nature/natcul1

Pawley, S. M., & Atkinson, N. (2022). *Surficial Geology of the Fort McKay Area (NTS 74E/SW)* [Map]. Alberta Geologic Survey. https://ags.aer.ca/publication/map-618

Pawley, S. M., & Utting, D. J. (2018). *Permafrost classification model for Northern Alberta* [Gridded data, GeoTIFF format]. Alberta Energy Regulator, AER/AGS Digital Data 2018-2008. https://ags.aer.ca/publication/dig-2018-0008

Pérez Hoyos, I. C., Krakauer, N. Y., Khanbilvardi, R., & Armstrong, R. A. (2016). A Review of Advances in the Identification and Characterization of Groundwater Dependent Ecosystems Using Geospatial Technologies. *Geosciences*, *6*(2), Article 2. https://doi.org/10.3390/geosciences6020017

Power, G., Brown, R. S., & Imhof, J. G. (1999). Groundwater and fish—Insights from northern North America. *Hydrological Processes*, *13*(3), 401–422. https://doi.org/10.1002/(SICI)1099-1085(19990228)13:3<401::AID-HYP746>3.0.CO;2-A

Rahman, M. M., McDermid, G. J., Strack, M., & Lovitt, J. (2017). A New Method to Map Groundwater Table in Peatlands Using Unmanned Aerial Vehicles. *Remote Sensing*, *9*(10), 1057. https://doi.org/10.3390/rs9101057

Rampheri, M. B., Dube, T., Dondofema, F., & Dalu, T. (2023). Progress in the remote sensing of groundwater-dependent ecosystems in semi-arid environments. *Physics and Chemistry of the Earth, Parts A/B/C*, *130*, 103359. https://doi.org/10.1016/j.pce.2023.103359

Rhode, M. M., Saito, L., & Smith, R. (2020). *Groundwater Thresholds for Ecosystems: A Guide for Practitioners* (p. 37). Global Groundwater Group (G3), The Nature Conservancy. https://www.scienceforconservation.org/products/groundwater-thresholds-for-ecosystems/

Rohde, M. M., Biswas, T., Housman, I. W., Campbell, L. S., Klausmeyer, K. R., & Howard, J. K. (2021). A Machine Learning Approach to Predict Groundwater Levels in California Reveals Ecosystems at Risk. *Frontiers in Earth Science*, *9*. https://doi.org/10.3389/feart.2021.784499

Rohde, M. M., Froend, R., & Howard, J. (2017). A Global Synthesis of Managing Groundwater Dependent Ecosystems Under Sustainable Groundwater Policy. *Groundwater*, *55*(3), 293–301. https://doi.org/10.1111/gwat.12511

Rooney, R. C., & Bayley, S. E. (2011). Setting reclamation targets and evaluating progress: Submersed aquatic vegetation in natural and post-oil sands mining wetlands in Alberta, Canada. *Ecological Engineering*, *37*(4), 569–579. https://doi.org/10.1016/j.ecoleng.2010.11.032

Rosa, E., Larocque, M., Hatch, C. E., & Springer, A. E. (2023). Editorial: “Novel approaches for understanding groundwater dependent ecosystems in a changing environment.” *Frontiers in Earth Science*, *11*, 1165061. https://doi.org/10.3389/feart.2023.1165061

Rossi, P. M., Ala-aho, P., Ronkanen, A.-K., & Kløve, B. (2012). Groundwater–surface water interaction between an esker aquifer and a drained fen. *Journal of Hydrology*, *432–433*, 52–60. https://doi.org/10.1016/j.jhydrol.2012.02.026

Ruff, S. E., Humez, P., De Angelis, I. H., Diao, M., Nightingale, M., Cho, S., Connors, L., Kuloyo, O. O., Seltzer, A., Bowman, S., Wankel, S. D., McClain, C. N., Mayer, B., & Strous, M. (2023). Hydrogen and dark oxygen drive microbial productivity in diverse groundwater ecosystems. *Nature Communications*, *14*(1), 3194. https://doi.org/10.1038/s41467-023-38523-4

Saccò, M., Guzik, M. T., van der Heyde, M., Nevill, P., Cooper, S. J. B., Austin, A. D., Coates, P. J., Allentoft, M. E., & White, N. E. (2022). eDNA in subterranean ecosystems: Applications, technical aspects, and future prospects. *Science of The Total Environment*, *820*, 153223. https://doi.org/10.1016/j.scitotenv.2022.153223

Saccò, M., Mammola, S., Altermatt, F., Alther, R., Bolpagni, R., Brancelj, A., Brankovits, D., Fišer, C., Gerovasileiou, V., Griebler, C., Guareschi, S., Hose, G. C., Korbel, K., Lictevout, E., Malard, F., Martínez, A., Niemiller, M. L., Robertson, A., Tanalgo, K. C., … Reinecke, R. (2024). Groundwater is a hidden global keystone ecosystem. *Global Change Biology*, *30*(1), e17066. https://doi.org/10.1111/gcb.17066

Schmidt, A., Gibson, J. J., Santos, I. R., Schubert, M., Tattrie, K., & Weiss, H. (2010). The contribution of groundwater discharge to the overall water budget of two typical Boreal lakes in Alberta/Canada estimated from a radon mass balance. *Hydrology and Earth System Sciences*, *14*(1), 79–89. https://doi.org/10.5194/hess-14-79-2010

Serov, P. A., & Kuginis, L. (2017). A groundwater ecosystem classification—The next steps. *International Journal of Water*, *11*(4), 328–362. https://doi.org/10.1504/IJW.2017.088043

Springer, A. E., Stevens, L. E., Ledbetter, J. D., Schaller, E. M., Gill, K. M., & Rood, S. B. (2015). Ecohydrology and stewardship of Alberta springs ecosystems. *Ecohydrology*, *8*(5), 896–910. https://doi.org/10.1002/eco.1596

Sun, Y., Takemon, Y., & Yamashiki, Y. (2020). Freshwater spring indicator taxa of benthic invertebrates. *Ecohydrology & Hydrobiology*, *20*(4), 622–631. https://doi.org/10.1016/j.ecohyd.2019.02.003

Tadono, T., Ishida, H., Oda, F., Naito, S., Minakawa, K., & Iwamoto, H. (2014). Precise Global DEM Generation by ALOS PRISM. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *II–4*, 71–76. https://doi.org/10.5194/isprsannals-II-4-71-2014

Tai, X., Mackay, D. S., Sperry, J. S., Brooks, P., Anderegg, W. R. L., Flanagan, L. B., Rood, S. B., & Hopkinson, C. (2018). Distributed Plant Hydraulic and Hydrological Modeling to Understand the Susceptibility of Riparian Woodland Trees to Drought-Induced Mortality. *Water Resources Research*, *54*(7), 4901–4915. https://doi.org/10.1029/2018WR022801

Utting, D. J. (2023). *Modelled Surfaces of Quaternary Units in the North Athabasca Oil Sands (NAOS) Region (gridded data, ASCII format)*. Alberta Energy Regulator / Alberta Geological Survey. https://ags.aer.ca/publication/dig-2023-0017

Van Der Kamp, G., & Hayashi, M. (2009). Groundwater-wetland ecosystem interaction in the semiarid glaciated plains of North America. *Hydrogeology Journal*, *17*(1), 203–214. https://doi.org/10.1007/s10040-008-0367-1

Vitt, D. H., & Chee, W.-L. (1990). The relationships of vegetation to surface water chemistry and peat chemistry in fens of Alberta, Canada. *Vegetatio*, *89*(2), 87–106. https://doi.org/10.1007/BF00032163

Vitt, D. H., House, M., & Glaeser, L. (2022). The response of vegetation to chemical and hydrological gradients at a patterned rich fen in northern Alberta, Canada. *Journal of Hydrology: Regional Studies*, *40*, 101038. https://doi.org/10.1016/j.ejrh.2022.101038

Walker, J., Almási, I., Stoakes, F., Potma, K., O'Keefe (Nee Cranshaw), J., & Schneider, C. (2017). Hypogenic karst beneath the Athabasca Oil Sands: Implications for oil sands mining operations. *Bulletin of Canadian Petroleum Geology*, *65*(1), 115–146. https://doi.org/10.2113/gscpgbull.65.1.115

Watts, C. L., Hatch, C. E., & Wicks, R. (2023). Mapping groundwater discharge seeps by thermal UAS imaging on a wetland restoration site. *Frontiers in Environmental Science*, *10*. https://doi.org/10.3389/fenvs.2022.946565

Wells, C. M., & Price, J. S. (2015). A hydrologic assessment of a saline-spring fen in the Athabasca oil sands region, Alberta, Canada – a potential analogue for oil sands reclamation. *Hydrological Processes*, *29*(20), 4533–4548. https://doi.org/10.1002/hyp.10518

Williams, D. D. (1993). Nutrient and flow vector dynamics at the hyporheic/groundwater interface and their effects on the interstitial fauna. *Hydrobiologia*, *251*(1), 185–198. https://doi.org/10.1007/BF00007178

Worker, S. B., Kielland, K., & Barboza, P. S. (2015). Effects of geophagy on food intake, body mass, and nutrient dynamics of snowshoe hares (Lepus americanus). *Canadian Journal of Zoology*, *93*(4), 323–329. https://doi.org/10.1139/cjz-2014-0237

Zimmerman, O. R., Pearce, D. W., Woodman, S. G., Rood, S. B., & Flanagan, L. B. (2023). Increasing contribution of alluvial groundwater to riparian cottonwood forest water use through warm and dry summers. *Agricultural and Forest Meteorology*, *329*, 109292. https://doi.org/10.1016/j.agrformet.2022.109292

Appendix A. OSM Groundwater Conceptual Models

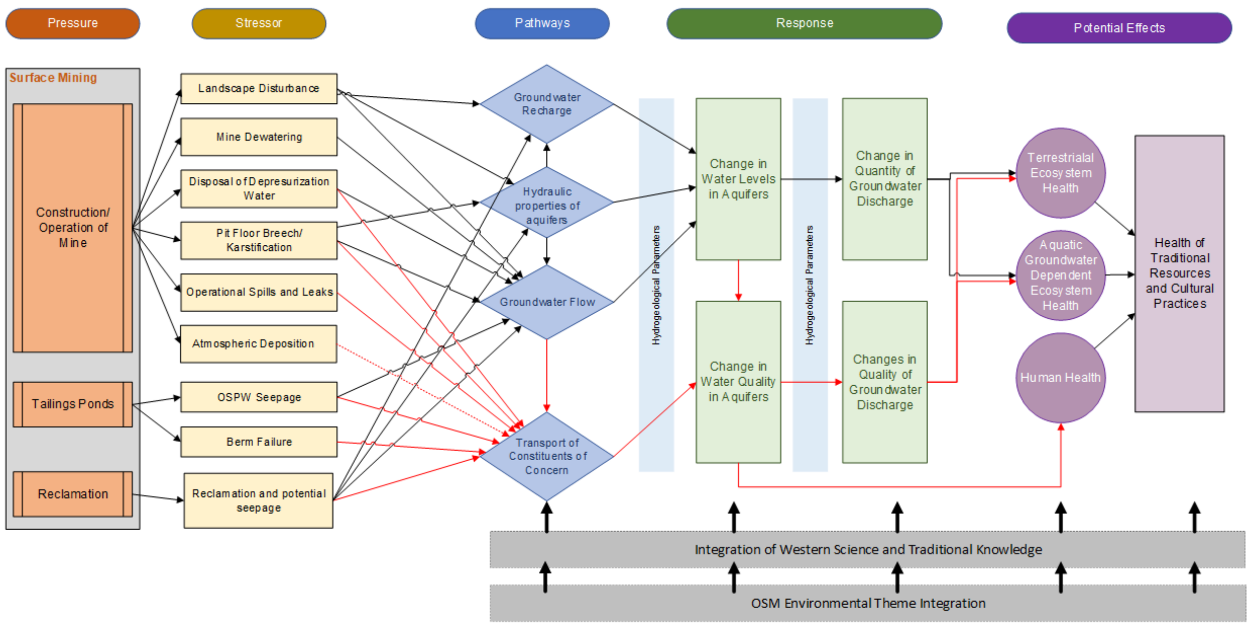

Figure A.1. Flowchart illustrating the Oils Sands Monitoring groundwater conceptual model.

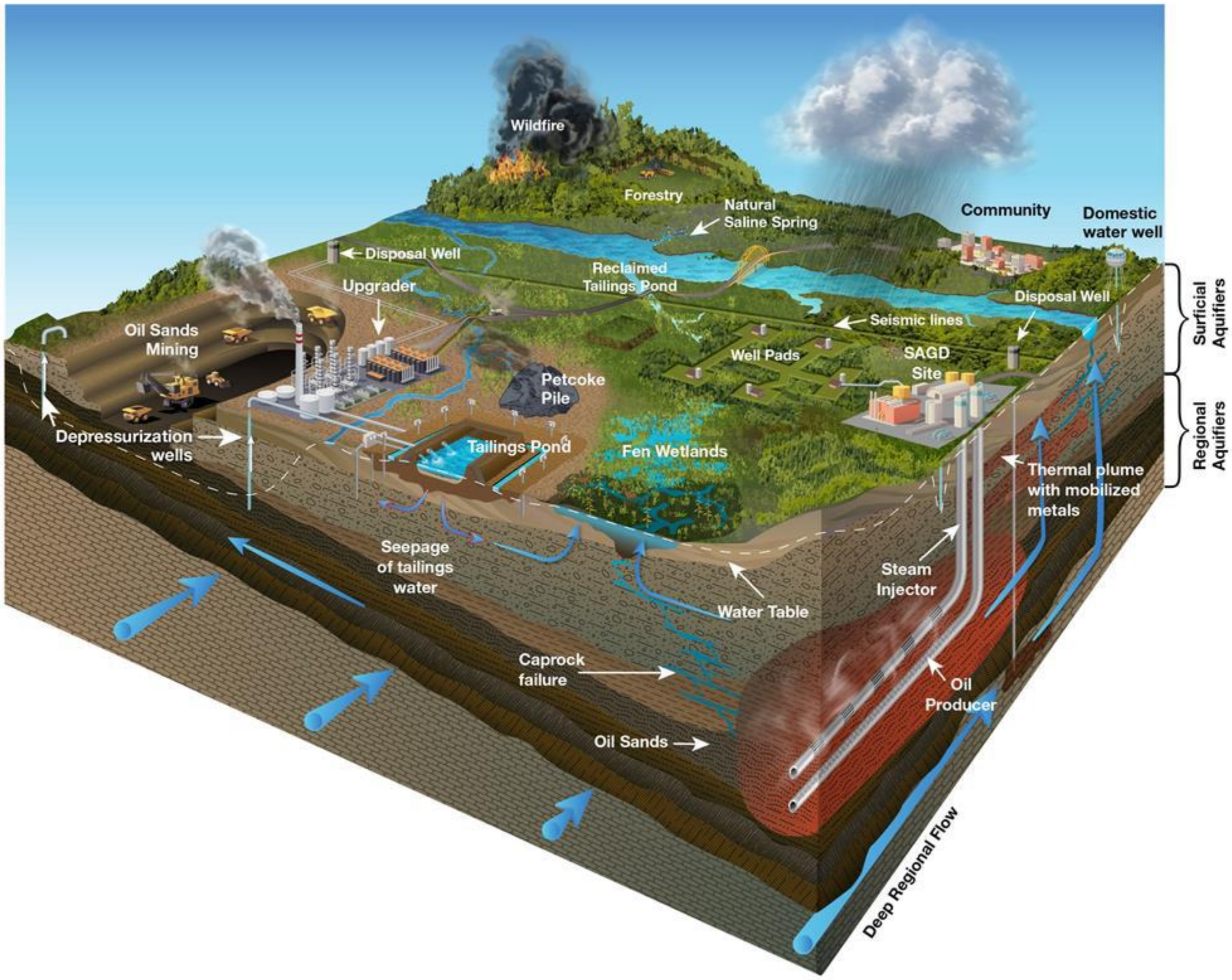

Figure A.2. Schematic diagram illustrating the Oils Sands Monitoring groundwater conceptual model.

Appendix B. Groundwater Indicator Literature Review

Table B.1. List of papers included in the groundwater literature review.

Full Citation	Indicator	Location
Manuscripts in Aquatic GDE Indicators and Mapping Literature Review		
Ala-aho, P., Rossi, P. M., Isokangas, E., & Kløve, B. (2015). Fully integrated surface–subsurface flow modelling of groundwater–lake interaction in an esker aquifer: Model verification with stable isotopes and airborne thermal imaging. Journal of Hydrology (Amsterdam), 522, 391–406. https://doi.org/10.1016/j.jhydrol.2014.12.054	Groundwater, Boreal, Esker, Peatlands, Isotope, Thermal	Finland
Arnoux, M., Gibert-Brunet, E., Barbecot, F., Guillon, S., Gibson, J., & Noret, A. (2017). Interactions between groundwater and seasonally ice-covered lakes: Using water stable isotopes and radon-222 multilayer mass balance models. Hydrological Processes, 31(14), 2566–2581. https://doi.org/10.1002/hyp.11206	Groundwater, Boreal, Lakes, Isotope, Radon	Canada
Autio, A., Ala-Aho, P., Rossi, P. M., Ronkanen, A.-K., Aurela, M., Lohila, A., Korpelainen, P., Kumpula, T., Kløve, B., & Marttila, H. (2023). Groundwater exfiltration pattern determination in the sub-arctic catchment using thermal imaging, stable water isotopes and fully-integrated groundwater-surface water modelling. Journal of Hydrology (Amsterdam), 626, 130342-. https://doi.org/10.1016/j.jhydrol.2023.130342	Groundwater, Boreal, Isotope, Thermal, Integrated Hydrological Model	Finland
Bertrand, G., Siergieiev, D., Ala-Aho, P., & Rossi, P. M. (2014). Environmental tracers and indicators bringing together groundwater, surface water and groundwater-dependent ecosystems: importance of scale in choosing relevant tools. Environmental Earth Sciences, 72(3), 813–827. https://doi.org/10.1007/s12665-013-3005-8	GDE, SW-GW interaction, Tracer section	World

Full Citation	Indicator	Location
Beasley-Hall, P. G., Murphy, N. P., King, R. A., White, N. E., Hedges, B. A., Cooper, S. J. B., Austin, A. D., & Guzik, M. T. (2023). Time capsules of biodiversity: Future research directions for groundwater-dependent ecosystems of the Great Artesian Basin. Frontiers in Environmental Science, 10. https://doi.org/10.3389/fenvs.2022.1021987	GDE Mapping Arid Environment	Australia
Bickerton, G., Roy, J.W., Frank, R.A., Spoelstra, J., Langston, G., Grapentine, L. & L.M. Hewitt. 2018. Assessments of Groundwater Influence on Selected River Systems in the Oil Sands Region of Alberta. Oil Sands Monitoring Program Technical Report Series No. 1.5. 32 p.	GW indicators, Rivers, Flow, Geochemistry	Alberta Oil Sands Region
Birks, J., Jones, J. P., & Gibson, J. J. (2012). (rep.). SURFACE WATER - GROUNDWATER INTERACTIONS IN THE LOWER ATHABASCA REGION (pp. 5–83). Fort McMurray, Alberta: Cumulative Environmental Management Association Working Group/Task Group: Groundwater Working Group.	Methods (Flow, Geochemical, Thermal, Water Budget), Rivers, Wetlands, Lakes	Alberta Oil Sands Region
Birks, S. J., Fennell, J. W., Gibson, J. J., Yi, Y., Moncur, M. C., & Brewster, M. (2019). Using regional datasets of isotope geochemistry to resolve complex groundwater flow and formation connectivity in northeastern Alberta, Canada. Applied Geochemistry, 101, 140–159. https://doi.org/10.1016/j.apgeochem.2018.12.013	Groundwater, Isotope, Geochemistry	Alberta Oil Sands Region
Birks, S. J., Manchuk, J., Yi, Y., McClain, C. N., Moncur, M. C., Gibson, J. J., Deutsch, C. V., Taylor, E. B., & Bayegnak, G. (2022). Groundwater monitoring near oil sands development: Insights from regional water quality datasets in the Alberta Oil Sands Region (AOSR). Journal of Hydrology. Regional Studies, 41, 101079-. https://doi.org/10.1016/j.ejrh.2022.101079	SW-GW Interaction, Groundwater, Isotope, Geochemistry	Alberta Oil Sands Region
Broughton, P. L. (2018). Ghost-rock karstification of Devonian limestone flooring the Athabasca Oil Sands in western Canada. Geomorphology (Amsterdam, Netherlands), 318, 303–319. https://doi.org/10.1016/j.geomorph.2018.07.001	Karst formations	Alberta Oil Sands Region

Full Citation	Indicator	Location
Kuginis, L., Dabovic, J., Byrne, G., Raine, A., & Hemakumara, H. (2016). Methods for the identification of high probability groundwater dependent vegetation ecosystems. Department of Primary Industries: Water, Government of New South Wales. https://nla.gov.au/nla.obj-3144390969/	GDE Mapping Arid Environment	Australia
Doody, T. M., Barron, O. V., Dowsley, K., Emelyanova, I., Fawcett, J., Overton, I. C., Pritchard, J. L., Van Dijk, A. I. J. M., & Warren, G. (2017). Continental mapping of groundwater dependent ecosystems: A methodological framework to integrate diverse data and expert opinion. Journal of Hydrology. Regional Studies, 10(C), 61–81. https://doi.org/10.1016/j.ejrh.2017.01.003	GDE Mapping Arid Environment	Australia
Eskelinen, R., Ala-aho, P., Rossi, P. M., & Kløve, B. (2015). GIS-based method for predicting groundwater discharge areas in esker aquifers in the Boreal region. Environmental Earth Sciences, 74(5), 4109–4118. https://doi.org/10.1007/s12665-015-4491-7	GDE Mapping, Boreal, Isotope, Darcy's Law, Hydrological Model,	Finland
Fildes, S. G., Doody, T. M., Bruce, D., Clark, I. F., & Batelaan, O. (2023). Mapping groundwater dependent ecosystem potential in a semi-arid environment using a remote sensing-based multiple-lines-of-evidence approach. International Journal of Digital Earth, 16(1), 375–407. https://doi.org/10.1080/17538947.2023.2176557	GDE Mapping, NDVI, NDCVI, Semi-Arid Environment, Machine Learning	Australia
Gerlach, M. E., Rains, K. C., Guerrón-Orejuela, E. J., Kleindl, W. J., Downs, J., Landry, S. M., & Rains, M. C. (2022). Using Remote Sensing and Machine Learning to Locate Groundwater Discharge to Salmon-Bearing Streams. Remote Sensing (Basel, Switzerland), 14(1), 63-. https://doi.org/10.3390/rs14010063	GDE mapping, Streams, Machine Learning	USA - Alaska
Gibson, J. J., Fennell, J., Birks, S. J., Yi, Y., Moncur, M. C., Hansen, B., & Jasechko, S. (2013). Evidence of discharging saline formation water to the Athabasca River in the oil sands mining region, northern Alberta. Canadian Journal of Earth Sciences, 50(12), 1244–1257. https://doi.org/10.1139/cjes-2013-0027	SW-GW interaction, Flow, Geochemistry, Rivers	Alberta Oil Sands Region

Full Citation	Indicator	Location
Gibson, J. J., Birks, S. J., Yi, Y., Moncur, M. C., & McEachern, P. M. (2016a). Stable isotope mass balance of fifty lakes in central Alberta: Assessing the role of water balance parameters in determining trophic status and lake level. Journal of Hydrology. Regional Studies, 6(C), 13–25. https://doi.org/10.1016/j.ejrh.2016.01.034	SW-GW interaction, isotopes, mass balance, Lakes	Alberta
Gibson, J. J., Yi, Y., & Birks, S. J. (2016b). Isotope-based partitioning of streamflow in the oil sands region, northern Alberta: Towards a monitoring strategy for assessing flow sources and water quality controls. Journal of Hydrology. Regional Studies, 5, 131–148. https://doi.org/10.1016/j.ejrh.2015.12.062	SW-GW interaction, isotopes, mass balance, Rivers	Alberta Oil Sands Region
Gibson, J. J., Birks, S. J., & Moncur, M. C. (2019). Stable isotope data (oxygen-18 and deuterium) from surveys of lakes, wetlands, rivers, and input waters across the South Athabasca Oil Sands region, Alberta, 2007–2009. Data in Brief, 22, 781–786. https://doi.org/10.1016/j.dib.2018.12.074	SW-GW interaction, Isotopes, Wetlands, Lakes, Rivers, data	Alberta Oil Sands Region
Gibson, J. J., Yi, Y., & Birks, S. J. (2020). Watershed, climate, and stable isotope data (oxygen-18 and deuterium) for 50 boreal lakes in the oil sands region, northeastern Alberta, Canada, 2002–2017. Data in Brief, 29, 105308-. https://doi.org/10.1016/j.dib.2020.105308	SW-GW interaction, Isotopes, Lakes, data	Alberta Oil Sands Region
Gibson, J. J., & Peters, D. L. (2022a). Water and environmental management in oil sands regions. Journal of Hydrology. Regional Studies, 44, 101274-. https://doi.org/10.1016/j.ejrh.2022.101274	Geochemistry, isotopes, review	Alberta Oil Sands Region
Gibson, J. J., Eby, P., Birks, S. J., Twitchell, C., Gray, C., & Kariyeva, J. (2022b). Isotope-based water balance assessment of open water wetlands across Alberta: Regional trends with emphasis on the oil sands region. Journal of Hydrology. Regional Studies, 40, 101036-. https://doi.org/10.1016/j.ejrh.2022.101036	Wetlands, isotopes, water balance	Alberta Oil Sands Region
Gue, A. E., Mayer, B., & Grasby, S. E. (2015). Origin and geochemistry of saline spring waters in the Athabasca oil sands region, Alberta, Canada. Applied Geochemistry, 61, 132–145. https://doi.org/10.1016/j.apgeochem.2015.05.015	SW-GW interaction, Geochemistry, Isotopes, Springs	Alberta Oil Sands Region

Full Citation	Indicator	Location
Gue, A., Grasby, S. E., & Mayer, B. (2017). Influence of saline groundwater discharge on river water chemistry in the Athabasca oil sands region – A chloride stable isotope and mass balance approach.	SW-GW interaction, Geochemistry, Isotopes, Rivers	Alberta Oil Sands Region
Hayashi, M., van der Kamp, G., & Rosenberry, D. O. (2016). Hydrology of Prairie Wetlands: Understanding the Integrated Surface-Water and Groundwater Processes. Wetlands (Wilmington, N.C.), 36(Suppl 2), 237–254. https://doi.org/10.1007/s13157-016-0797-9	SW-GW interaction, Wetlands, Geochemistry	Alberta
Hayashi, M., & van der Kamp, G. (2023). The role of Canadian research in advancing groundwater hydrology: historical sketches from the past 75 years. Canadian Water Resources Journal, 48(4), 363–378. https://doi.org/10.1080/07011784.2023.2177197	SW-GW interaction, isotopes, Geochemistry, Flow, Modelling, Lakes, Rivers, Wetlands	Canada
Heagle, D., Hayashi, M., & Kamp, G. van der. (2013). Surface–subsurface salinity distribution and exchange in a closed-basin prairie wetland. Journal of Hydrology (Amsterdam), 478, 1–14. https://doi.org/10.1016/j.jhydrol.2012.05.054	SW-GW Interaction, Geochemistry, Modelling	Alberta
Hein, F. J., & Cotterill, D. K. (2006). The athabasca oil sands -: A regional geological perspective, Fort McMurray Area, Alberta, Canada. Natural Resources Research (New York, N.Y.), 15(2), 85–102. https://doi.org/10.1007/s11053-006-9015-4	Geology, mapping, Stratigraphy	Alberta Oil Sands Region
Isokangas, E., Rossi, P. M., Ronkanen, A., Marttila, H., Rozanski, K., & Kløve, B. (2017). Quantifying spatial groundwater dependence in peatlands through a distributed isotope mass balance approach. Water Resources Research, 53(3), 2524–2541. https://doi.org/10.1002/2016WR019661	SW-GW interaction, Boreal, Peatlands, Modelling, Isotopes, Mapping GDEs	Finland
Klausmeyer, K., Howard, J., Keeler-Wolf, T., Davis-Fadtke, K., Hull, R., & Lyons, A. (2018). Mapping Indicators of Groundwater Dependent Ecosystems in California: Methods Report. The Nature Conservancy. https://www.groundwaterresourcehub.org/where-we-work/california/mapping-indicators-gdes/	Mapping GDEs, Arid	USA - California

Full Citation	Indicator	Location
Kløve, B., Ala-aho, P., Bertrand, G., Boukalova, Z., Ertürk, A., Goldscheider, N., Ilmonen, J., Karakaya, N., Kupfersberger, H., Kværner, J., Lundberg, A., Mileusnić, M., Moszczynska, A., Muotka, T., Preda, E., Rossi, P., Siergieiev, D., Šimek, J., Wachniew, P., … Widerlund, A. (2011). Groundwater dependent ecosystems. Part I: Hydroecological status and trends. Environmental Science & Policy, 14(7), 770–781. https://doi.org/10.1016/j.envsci.2011.04.002	GDEs, Boreal, Peatlands, Methods	Finland
LaRocque, A., & Leblon, B. (2022). (rep.). Remote Sensing of Ground-Water Dependent Ecosystems in the Oil Sands Area (Final Report, pp. 1–28). Fredericton, New Brunswick: University of New Brunswick.	GDE mapping, Machine learning, NDVI	Alberta Oil Sands Region
Lidberg, W., Nilsson, M., & Ågren, A. (2020). Using machine learning to generate high-resolution wet area maps for planning forest management: A study in a boreal forest landscape. Ambio, 49(2), 475–486. https://doi.org/10.1007/s13280-019-01196-9	Machine Learning, Wet area mapping, Boreal	Sweden
Manchuk, J. G., Birks, J. S., McClain, C. N., Bayegnak, G., Gibson, J. J., & Deutsch, C. V. (2021). Estimating Stable Measured Values and Detecting Anomalies in Groundwater Geochemistry Time Series Data Across the Athabasca Oil Sands Area, Canada. Natural Resources Research (New York, N.Y.), 30(2), 1755–1779. https://doi.org/10.1007/s11053-020-09801-5	SW-GW interaction, Modelling, geochemistry, Boreal	Alberta Oil Sands Region
Martínez-Santos, P., Díaz-Alcaide, S., De la Hera-Portillo, A., & Gómez-Escalonilla, V. (2021). Mapping groundwater-dependent ecosystems by means of multi-layer supervised classification. Journal of Hydrology (Amsterdam), 603, 126873-. https://doi.org/10.1016/j.jhydrol.2021.126873	GDE mapping, Machine Learning, Semi-arid, MLMapper	Spain
The Nature Conservancy and the Desert Research Institute. Global Groundwater Dependent Ecosystem Map, Version 1.2.0. https://kklausmeyer.users.earthengine.app/view/global-gde. June 2023	GDE mapping, Arid, Google Earth Engine	World

Full Citation	Indicator	Location
Pérez Hoyos, I., Krakauer, N., Khanbilvardi, R., & Armstrong, R. (2016). A Review of Advances in the Identification and Characterization of Groundwater Dependent Ecosystems Using Geospatial Technologies. Geosciences, 6(2), 17-. https://doi.org/10.3390/geosciences6020017	GDE mapping, Methods, Remote sensing Review	World
Rampheri, M. B., Dube, T., Dondofema, F., & Dalu, T. (2023a). Progress in the remote sensing of groundwater-dependent ecosystems in semi-arid environments. Physics and Chemistry of the Earth. Parts A/B/C, 130, 103359-. https://doi.org/10.1016/j.pce.2023.103359	GDE Mapping, Review, Remote sensing, Semi-arid	World
Rohde, M. M., Froend, R., & Howard, J. (2017). A Global Synthesis of Managing Groundwater Dependent Ecosystems Under Sustainable Groundwater Policy. Ground Water, 55(3), 293–301. https://doi.org/10.1111/gwat.12511	Mapping GDEs, Policy, Methods	World
Rohde, M. M., Biswas, T., Housman, I. W., Campbell, L. S., Klausmeyer, K. R., & Howard, J. K. (2021). A Machine Learning Approach to Predict Groundwater Levels in California Reveals Ecosystems at Risk. Frontiers in Earth Science (Lausanne), 9. https://doi.org/10.3389/feart.2021.784499	GDE Mapping, Machine Learning, GW monitoring, Modelling, Semi-arid	USA-California
Rosa, E., Larocque, M., Hatch, C. E., & Springer, A. E. (2023). Editorial: "Novel approaches for understanding groundwater dependent ecosystems in a changing environment." Frontiers in Earth Science (Lausanne), 11. https://doi.org/10.3389/feart.2023.1165061	GDE mapping, Review, Remote sensing, Modelling	World
Rossi, P. M., Ala-aho, P., Ronkanen, A.-K., & Kløve, B. (2012). Groundwater–surface water interaction between an esker aquifer and a drained fen. Journal of Hydrology (Amsterdam), 432–433, 52–60. https://doi.org/10.1016/j.jhydrol.2012.02.026	SW-GW interaction, Mapping, Modelling, Geochemistry, Boreal, Peatlands	Finland
Schmidt, A., Gibson, J. J., Santos, I. R., Schubert, M., Tattrie, K., & Weiss, H. (2010). The contribution of groundwater discharge to the overall water budget of two typical Boreal lakes in Alberta/Canada estimated from a radon mass balance. Hydrology and Earth System Sciences, 14(1), 79–89. https://doi.org/10.5194/hess-14-79-2010	Radon, Water balance, Boreal, Lakes, Isotopes, Modelling	Alberta Oil Sands Region

Full Citation	Indicator	Location
Serov, P. A., & Kuginis, L. (2017). A groundwater ecosystem classification-the next steps. International Journal of Water, 11(4), 328-362.	GDE definition	World
van der Kamp, G., & Hayashi, M. (2009). Groundwater-wetland ecosystem interaction in the semiarid glaciated plains of North America. Hydrogeology Journal, 17(1), 203–214. https://doi.org/10.1007/s10040-008-0367-1	SW-GW interaction, Modelling, geochemistry, Prairies	Alberta
Walker, J., Almasi, I., Stoakes, F., Potma, K., & O'Keefe, J. (2017). Hypogenic karst beneath the Athabasca Oil Sands; implications for oil sands mining operations. Bulletin of Canadian Petroleum Geology, 65(1), 115–146. https://doi.org/10.2113/gscpgbull.65.1.115	Karst formations, stratigraphy	Alberta Oil Sands Region
Watts, C. L., Hatch, C. E., & Wicks, R. (2023). Mapping groundwater discharge seeps by thermal UAS imaging on a wetland restoration site. Frontiers in Environmental Science, 10. https://doi.org/10.3389/fenvs.2022.946565	GDE mapping, thermal, wetlands	USA - Massachusetts
Wells, C. M., & Price, J. S. (2015). The hydrogeologic connectivity of a low-flow saline-spring fen peatland within the Athabasca oil sands region, Canada. Hydrogeology Journal, 23(8), 1799–1816. https://doi.org/10.1007/s10040-015-1301-y	SW-GW interaction, Peatlands, Boreal, Geochemistry	Alberta Oil Sands Region

Appendix C. Biological Indicator Literature Review

Table C.1. List of papers included in the biological literature review. Papers are sorted by indicator category, then by lead author. Papers not included in the text of the literature review but that provide supporting information for aquatic, terrestrial, and subterranean ecosystems are listed at the bottom.

Full Citation	Indicator Category	Location
Manuscripts explicitly included in Aquatic GDE Literature Review		
Cooper, D. J., & Andrus, R. E. (1994). Patterns of vegetation and water chemistry in peatlands of the west-central Wind River Range, Wyoming, U.S.A. Canadian Journal of Botany, 72(11), 1586–1597. https://doi.org/10.1139/b94-196	Vegetation	Wyoming, USA
Elmes, M. C., Davidson, S. J., & Price, J. S. (2021). Ecohydrological interactions in a boreal fen–swamp complex, Alberta, Canada. Ecohydrology, 14(7), e2335. https://doi.org/10.1002/eco.2335	Vegetation	Poplar Fen, Alberta
Glaser, P. H., Siegel, D. I., Reeve, A. S., Janssens, J. A., & Janecky, D. R. (2004). Tectonic drivers for vegetation patterning and landscape evolution in the Albany River region of the Hudson Bay Lowlands. Journal of Ecology, 92(6), 1054–1070. https://doi.org/10.1111/j.0022-0477.2004.00930.x	Vegetation	Hudson Bay Lowlands, Canada
Jeglum, J. K. (1991). Definition of trophic classes in wooded peatlands by means of vegetation types and plant indicators. Annales Botanici Fennici, 28(3), 175–192. https://www.jstor.org/stable/23725328	Vegetation	Ontario, Canada
Kuglerová, L., Dynesius, M., Laudon, H., & Jansson, R. (2016). Relationships Between Plant Assemblages and Water Flow Across a Boreal Forest Landscape: A Comparison of Liverworts, Mosses, and Vascular Plants. Ecosystems, 19(1), 170–184. https://doi.org/10.1007/s10021-015-9927-0	Vegetation	Boreal Sweden
Larocque, M., Ferlatte, M., Pellerin, S., Cloutier, V., Munger, J. L., Paniconi, C., & Quillet, A. (2016). Chemical and botanical indicators of groundwater inflow to Sphagnum-dominated peatlands. Ecological Indicators, 64, 142–151. https://doi.org/10.1016/j.ecolind.2015.12.012	Vegetation	Quebec, Canada

Full Citation	Indicator Category	Location
Munger, J., Pellerin, S., Larocque, M., & Ferlatte, M. (2014). Espèces végétales indicatrices des échanges d'eau entre tourbière et aquifère [Plant species indicative of water exchange between bog and aquifer]. Le Naturaliste canadien, 138(1), 4–12. https://doi.org/10.7202/1021038ar	Vegetation	Quebec, Canada
Vitt, D. H., & Chee, W.-L. (1990). The relationships of vegetation to surface water chemistry and peat chemistry in fens of Alberta, Canada. Vegetatio, 89(2), 87–106. https://doi.org/10.1007/BF00032163	Vegetation	Alberta, Canada
Vitt, D. H., House, M., & Glaeser, L. (2022). The response of vegetation to chemical and hydrological gradients at a patterned rich fen in northern Alberta, Canada. Journal of Hydrology: Regional Studies, 40, 101038. https://doi.org/10.1016/j.ejrh.2022.101038	Vegetation	Alberta, Canada
Wells, C. M., & Price, J. S. (2015). A hydrologic assessment of a saline-spring fen in the Athabasca oil sands region, Alberta, Canada – a potential analogue for oil sands reclamation. Hydrological Processes, 29(20), 4533–4548. https://doi.org/10.1002/hyp.10518	Vegetation	Alberta, Canada
Lehosmaa, K., Jyväsjärvi, J., Ilmonen, J., Rossi, P. M., Paasivirta, L., & Muotka, T. (2018). Groundwater contamination and land drainage induce divergent responses in boreal spring ecosystems. Science of The Total Environment, 639, 100–109. https://doi.org/10.1016/j.scitotenv.2018.05.126	Vegetation; Microbes; Macroinvertebrates	Finland
Febria, C. M., Beddoes, P., Fulthorpe, R. R., & Williams, D. D. (2012). Bacterial community dynamics in the hyporheic zone of an intermittent stream. The ISME Journal, 6(5), 1078–1088. https://doi.org/10.1038/ismej.2011.173	Microbes	Ontario, Canada
Groult, B., St-Jean, V., & Lazar, C. S. (2023). Linking Groundwater to Surface Discharge Ecosystems: Archaeal, Bacterial, and Eukaryotic Community Diversity and Structure in Quebec (Canada). Microorganisms, 11(7), 1674. https://doi.org/10.3390/microorganisms11071674	Microbes	Quebec, Canada
Fraser, B. G., & Williams, D. D. (1998). Seasonal Boundary Dynamics of a Groundwater/ Surface-Water Ecotone. Ecology, 79(6), 2019–2031. https://doi.org/10.1890/0012-9658(1998)079[2019:SBDOAG]2.0.CO;2	Macroinvertebrates	Ontario, Canada

Full Citation	Indicator Category	Location
Williams, D. D. (1993). Nutrient and flow vector dynamics at the hyporheic/groundwater interface and their effects on the interstitial fauna. Hydrobiologia, 251(1), 185–198. https://doi.org/10.1007/BF00007178	Macroinvertebrates	Ontario, Canada
Bousfield, E. L., & Holsinger, J. R. (1981). A second new subterranean amphipod crustacean of the genus Stygobromus (Crangonyctidae) from Alberta, Canada. Canadian Journal of Zoology, 59(9), 1827–1830. https://doi.org/10.1139/z81-250	Stygofauna	Alberta, Canada
Dole-Olivier, M.-J., Creuzé des Châtelliers, M., Galassi, D. M. P., Lafont, M., Mermillod-Blondin, F., Paran, F., Graillot, D., Gaur, S., & Marmonier, P. (2022). Drivers of functional diversity in the hyporheic zone of a large river. Science of The Total Environment, 843, 156985. https://doi.org/10.1016/j.scitotenv.2022.156985	Stygofauna	France
Graillot, D., Paran, F., Bornette, G., Marmonier, P., Piscart, C., & Cadilhac, L. (2014). Coupling groundwater modeling and biological indicators for identifying river/aquifer exchanges. SpringerPlus, 3, 68. https://doi.org/10.1186/2193-1801-3-68	Stygofauna	Switzerland
Holsinger, J. R. (1980). Stygobromus canadensis, a new subterranean amphipod crustacean (Crangonyctidae) from Canada, with remarks on Wisconsin refugia. Canadian Journal of Zoology, 58(2), 290–297. https://doi.org/10.1139/z80-034	Stygofauna	Alberta, Canada
Hose, G. C., Chariton, A. A., Daam, M. A., Di Lorenzo, T., Galassi, D. M. P., Halse, S. A., Reboleira, A. S. P. S., Robertson, A. L., Schmidt, S. I., & Korbel, K. L. (2022). Invertebrate traits, diversity and the vulnerability of groundwater ecosystems. Functional Ecology, 36(9), 2200–2214. https://doi.org/10.1111/1365-2435.14125	Stygofauna	N/A
Saccò, M., Guzik, M. T., van der Heyde, M., Nevill, P., Cooper, S. J. B., Austin, A. D., Coates, P. J., Allentoft, M. E., & White, N. E. (2022). eDNA in subterranean ecosystems: Applications, technical aspects, and future prospects. Science of The Total Environment, 820, 153223. https://doi.org/10.1016/j.scitotenv.2022.153223	Stygofauna	N/A
Ayotte, J. B., Parker, K. L., Arocena, J. M., & Gillingham, M. P. (2006). Chemical composition of lick soils: Functions of soil ingestion by four ungulate species. Journal of Mammalogy, 87(5), 878–888. https://doi.org/10.1644/06-MAMM-A-055R1.1	Mammals	British Columbia, Canada

Full Citation	Indicator Category	Location
Worker, S. B., Kielland, K., & Barboza, P. S. (2015). Effects of geophagy on food intake, body mass, and nutrient dynamics of snowshoe hares (Lepus americanus). Canadian Journal of Zoology, 93(4), 323–329. https://doi.org/10.1139/cjz-2014-0237	Mammals	Alaska, USA
Land, E., & Peters, C. N. (2023). Groundwater impacts on stream biodiversity and communities: A review. Journal of Freshwater Ecology, 38(1), 2260801. https://doi.org/10.1080/02705060.2023.2260801	Background	N/A
Hancock, P. J., Boulton, A. J., & Humphreys, W. F. (2005). Aquifers and hyporheic zones: Towards an ecological understanding of groundwater. Hydrogeology Journal, 13(1), 98–111. https://doi.org/10.1007/s10040-004-0421-6	Background	N/A
Ruff, S. E., Humez, P., De Angelis, I. H., Diao, M., Nightingale, M., Cho, S., Connors, L., Kuloyo, O. O., Seltzer, A., Bowman, S., Wankel, S. D., McClain, C. N., Mayer, B., & Strous, M. (2023). Hydrogen and dark oxygen drive microbial productivity in diverse groundwater ecosystems. Nature Communications, 14(1), 3194. https://doi.org/10.1038/s41467-023-38523-4	Background; Microbes	Alberta, Canada
Driscoll, K. P., Smith, D. M., Warren, S. D., & Finch, D. M. (2019). Riparian ecosystems of the Salmon-Challis National Forest: An assessment of current conditions in relation to the natural range of variability (RMRS-GTR-394; General Technical Report, p. 190). USDA Forest Service. https://doi.org/10.2737/RMRS-GTR-394	Other; Natural Range of Variablity in Riparian Ecosystems	Idaho, USA
Ionescu, D., Bizic, M., Karnatak, R., Musseau, C. L., Onandia, G., Kasada, M., Berger, S. A., Nejstgaard, J. C., Ryo, M., Lischeid, G., Gessner, M. O., Wollrab, S., & Grossart, H.-P. (2022). From microbes to mammals: Pond biodiversity homogenization across different land-use types in an agricultural landscape. Ecological Monographs, 92(3), e1523. https://doi.org/10.1002/ecm.1523	Other; Biotic Homogenization	Germany
Manuscripts providing supporting information for aquatic, terrestrial, and subterranean ecosystems		
Graup, L. J., Tague, C. L., Harpold, A. A., & Krogh, S. A. (2022). Subsurface Lateral Flows Buffer Riparian Water Stress Against Snow Drought. Journal of Geophysical Research: Biogeosciences, 127(12), e2022JG006980. https://doi.org/10.1029/2022JG006980	Terrestrial	California, USA

Full Citation	Indicator Category	Location
Lilles, E. B., Purdy, B. G., Chang, S. X., & Macdonald, S. E. (2010). Soil and groundwater characteristics of saline sites supporting boreal mixedwood forests in northern Alberta. Canadian Journal of Soil Science, 90(1), 1–14. https://doi.org/10.4141/CJSS08040	Terrestrial	Alberta, Canada
Tai, X., Mackay, D. S., Sperry, J. S., Brooks, P., Anderegg, W. R. L., Flanagan, L. B., Rood, S. B., & Hopkinson, C. (2018). Distributed Plant Hydraulic and Hydrological Modeling to Understand the Susceptibility of Riparian Woodland Trees to Drought-Induced Mortality. Water Resources Research, 54(7), 4901–4915. https://doi.org/10.1029/2018WR022801	Terrestrial	Alberta, Canada
Zimmerman, O. R., Pearce, D. W., Woodman, S. G., Rood, S. B., & Flanagan, L. B. (2023). Increasing contribution of alluvial groundwater to riparian cottonwood forest water use through warm and dry summers. Agricultural and Forest Meteorology, 329, 109292. https://doi.org/10.1016/j.agrformet.2022.109292	Terrestrial	Alberta, Canada
Power, G., Brown, R. S., & Imhof, J. G. (1999). Groundwater and fish—Insights from northern North America. Hydrological Processes, 13(3), 401–422. https://doi.org/10.1002/(SICI)1099-1085(19990228)13:3<401::AID-HYP746>3.0.CO;2-A	Fish	North America
Martinez, A., Anicic, N., Calvaruso, S., Sanchez, N., Puppieni, L., Sforzi, T., Zaupa, S., Alvarez, F., Brankovits, D., Gąsiorowski, L., Gerovasileiou, V., Gonzalez, B., Humphreys, W., Iliffe, T., Worsaae, K., Bailly, N., & Fontaneto, D. (2018). A new insight into the Stygofauna Mundi: Assembling a global dataset for aquatic fauna in subterranean environments. ARPHA Conference Abstracts, 1, e29514. https://doi.org/10.3897/aca.1.e29514	Stygofauna	N/A
Chiloane, C., Dube, T., & Shoko, C. (2022). Impacts of groundwater and climate variability on terrestrial groundwater dependent ecosystems: A review of geospatial assessment approaches and challenges and possible future research directions. Geocarto International, 37(23), 6755–6779. https://doi.org/10.1080/10106049.2021.1948108	Background, Terrestrial	N/A
Goodbrand, A., Westbrook, C. J., & van der Kamp, G. (2019). Hydrological functions of a peatland in a Boreal Plains catchment. Hydrological Processes, 33(4), 562–574. https://doi.org/10.1002/hyp.13343	Other; Hydrological Functioning of Peatlands	Saskatch-ewan, Canada

Full Citation	Indicator Category	Location
Miller, C. A., Benscoter, B. W., & Turetsky, M. R. (2015). The effect of long-term drying associated with experimental drainage and road construction on vegetation composition and productivity in boreal fens. Wetlands Ecology and Management, 23(5), 845–854. https://doi.org/10.1007/s11273-015-9423-5	Other; Peatland Drying	Alberta, Canada
Rhode, M. M., Saito, L., & Smith, R. (2020). Groundwater Thresholds for Ecosystems: A Guide for Practitioners (p. 37). Global Groundwater Group (G3), The Nature Conservancy. https://www.scienceforconservation.org/products/groundwater-thresholds-for-ecosystems/	Other; Groundwater Thresholds	N/A
Rooney, R. C., & Bayley, S. E. (2011). Setting reclamation targets and evaluating progress: Submersed aquatic vegetation in natural and post-oil sands mining wetlands in Alberta, Canada. Ecological Engineering, 37(4), 569–579. https://doi.org/10.1016/j.ecoleng.2010.11.032	Other; Wetland Condition	Alberta, Canada

Appendix D. Data Compilation

Table D.1. Full list of datasets compiled to support GDE mapping in the oil sands region.

Categories	Data Name
Climate	Climate data (Data gap)
Geography	Annual Unit Runoff
Geography	Digital Elevation Model (DEM) - Advanced Land Observing Satellite (ALOS)
Geography	Flow Accumulation - ALOS Derived
Geography	Flow Direction - ALOS Derived
Geography	HUC 8, 10
Geography	Slope - ALOS Derived
Geography	Topographic Wetness Index (TWI)- ALOS Derived
Geology	AGS DIG_2023_0017 (Modelled Surfaces of Quaternary Units NAOS)
Geology	AGS REP_99 (Paleogeography,Evaporite Karstification, and Salt Cavern Potential)
Geology	Bedrock (Map 600)
Geology	DIG_2023_0017 (Modelled Surfaces of Quaternary Units NAOS)
Geology	MAP_632; DIG_2022_0031
Geology	Permafrost presence
Geology	Permeability - derived from geological materials
Geology	Quaternary Unit Picks in the North Athabasca Oil Sands (NAOS) Region
Geology	Subterranean data (Data gap)
Geology	surficial geological maps (bedrock)
Geology	surficial geological maps (bedrock) - updated - Maps 618-621
Groundwater	Aquanty: Depth to water tables, exchange fluxes, groundwater seepage along the Athabasca River

Categories	Data Name
Groundwater	Base of Groundwater Protection Data (estimated elevation for the base of the deepest formation that is likely to contain nonsaline groundwater)
Groundwater	DIG_2014_0025 (Springs) [Locations, Chemistry]
Groundwater	Distribution of Hydraulic Head in the Peace River / Viking / Bow Island Hydrostratigraphic Unit
Groundwater	Higher resolution groundwater level (Data gap)
Groundwater	Hydraulic head (Data gap)
Groundwater	Kisters Surface and Groundwater data
Groundwater	Map 593 (Distribution of Total Dissolved Solids in the Peace River / Viking / Bow Island Hydrostratigraphic Unit)
Groundwater	Map 594 (Distribution of Hydraulic Head in the Peace River / Viking / Bow Island Hydrostratigraphic Unit)
Groundwater	Map 596 (Distribution of Total Dissolved Solids in the Grand Rapids Hydrostratigraphic Unit)
Groundwater	Map 597 (Distribution of Hydraulic Head in the Grand Rapids Hydrostratigraphic Unit)
Groundwater	Map 612 (Distribution of Total Dissolved Solids in the McMurray Hydrostratigraphic Unit)
Groundwater	Map 613 (Distribution of Hydraulic Head in the McMurray Hydrostratigraphic Unit)
Groundwater	Operators/EIAs GW Chemistry
Groundwater	Operators/EIAs Water Levels
Groundwater	Spring Compilation (AGS)
Groundwater	Spring Compilation (InnoTech)
Groundwater	Thalwegs
Groundwater and Surface water	2022 Water Use Data
Groundwater and Surface water	EarthFX data (Data gap)
Human Impact Data	Data on land use changes (Data gap)
Landcover	Biological data (Data gap)

Categories	Data Name
Landcover	Eco_AB_10TM - this has 4 scales of ecosite from Agriculture Canada
Landcover	FinalEBM - Harvest, Fire, PreHarvest, SitePrep and SiteEquip layers (ABMI)
Landcover	Forest Fire Polygons
Landcover	Higher resolution thermal data (Data gap)
Landcover	Soil Landscapes of Canada
River and Lake Surveys	Isotope Sampling (ISO ABMI)
River and Lake Surveys	RAMP Hydrometric Monitoring Locations
River and Lake Surveys	RAMP Water Quality Monitoring Locations
River Surveys	Electromagnetic (EM31) Surveys (InnoTech/Advisian)
River Surveys	Water quality/LTRN
River Surveys	WSC/RAMP Stream Gauging
Surface Water	Kisters - 2022-23-osm-wetland-monitoring-surface-water-quality
Surface Water	Surface water/ groundwater interaction (Data gap)
Various	Temporal resolution data (Data gap)
Wetland surveys	OSM Wetland Inventory Pilot Area (ABMI/DUC)

Appendix E. Additional Results Figures

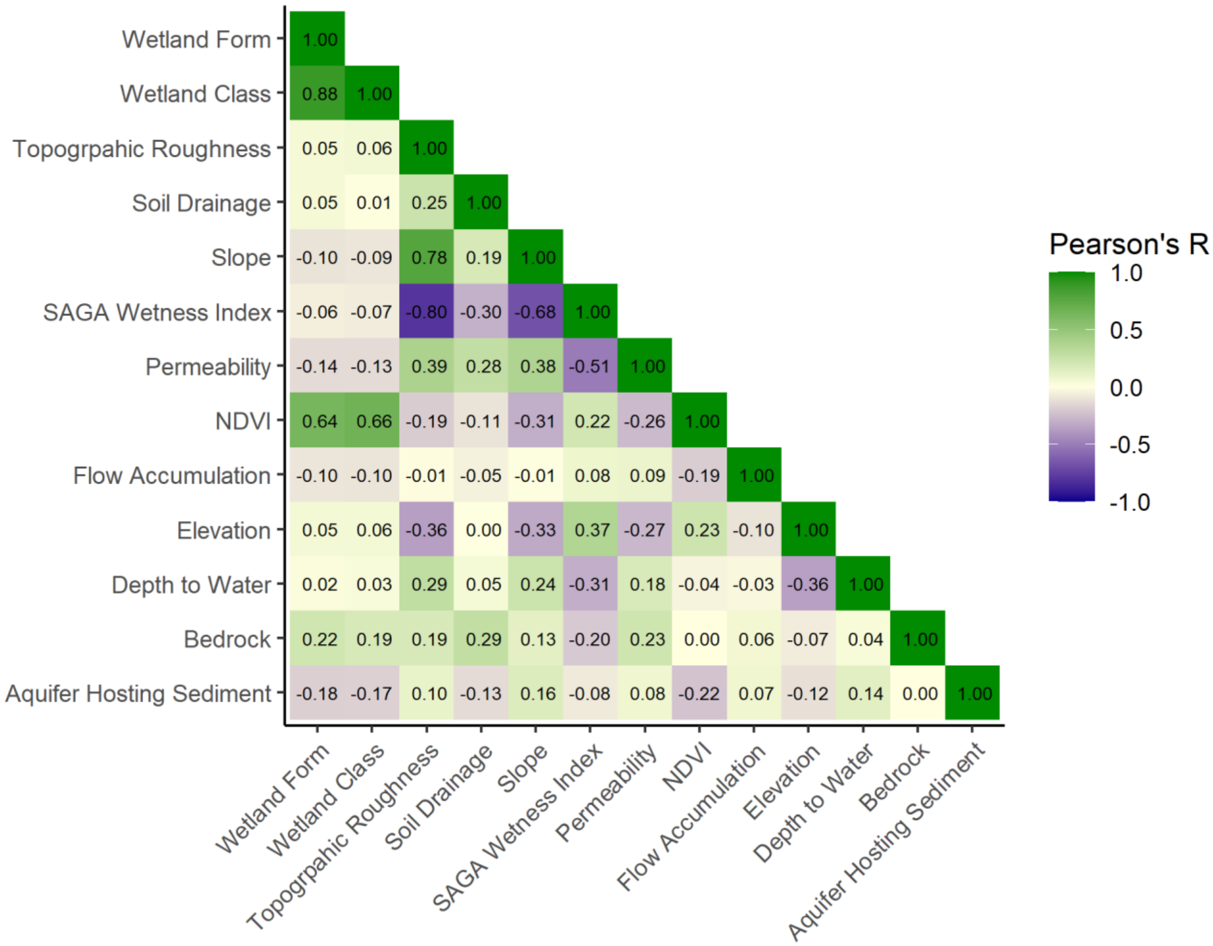

Figure E.1. Results of pairwise correlation analysis of initial 13 explanatory variables to be used in the models. Numbers in the plot show the Pearson correlation coefficient. The explanatory variables used were: aquifer hosting sediment, bedrock, depth to water, elevation, flow accumulation, normalized difference vegetation index (NDVI), permeability, SAGA wetness index, slope, soil drainage, topographic roughness, wetland class, and wetland form.

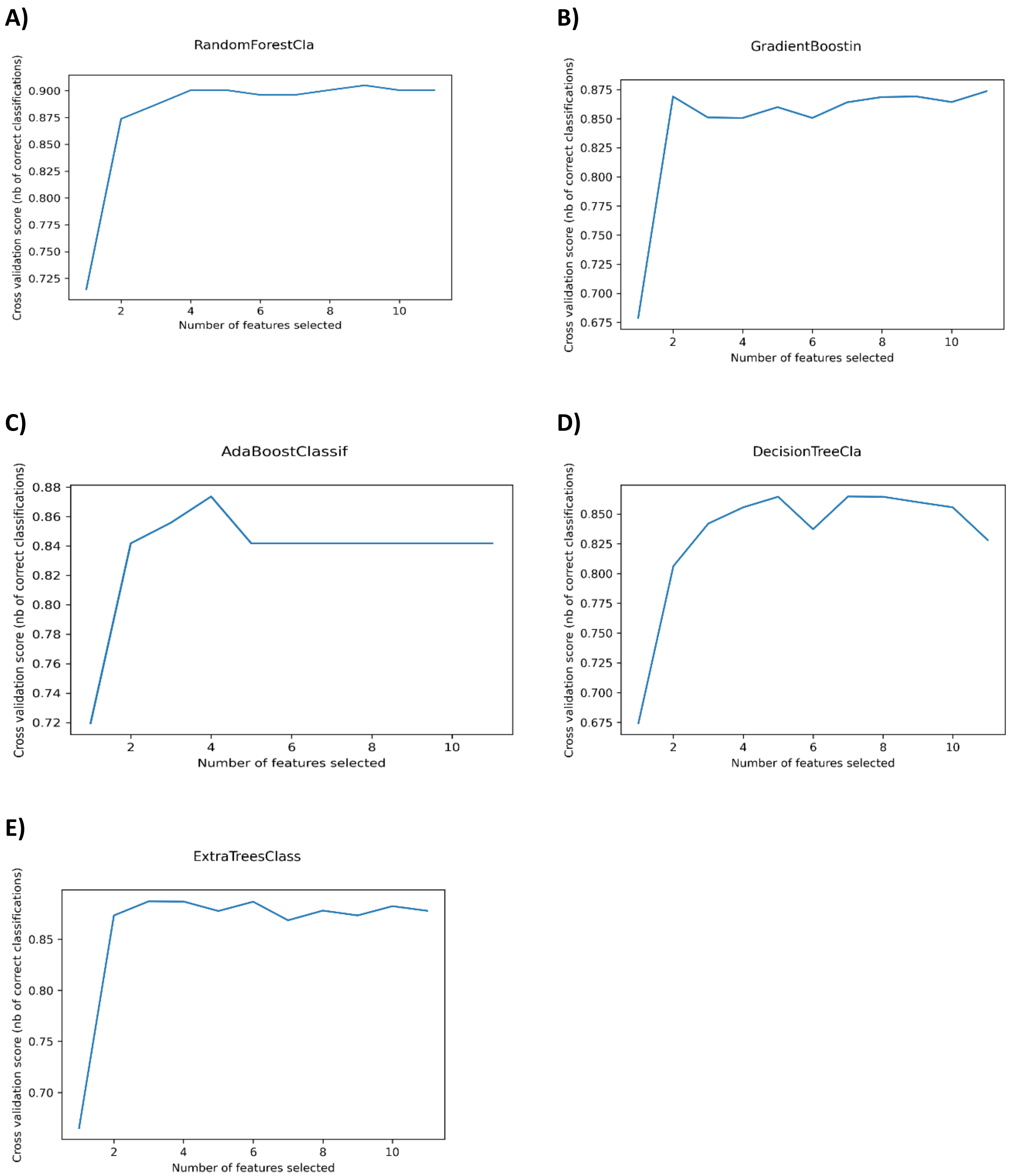

Figure E.2. Outcome of recursive feature reduction for the top five models showing optimization of explanatory features.

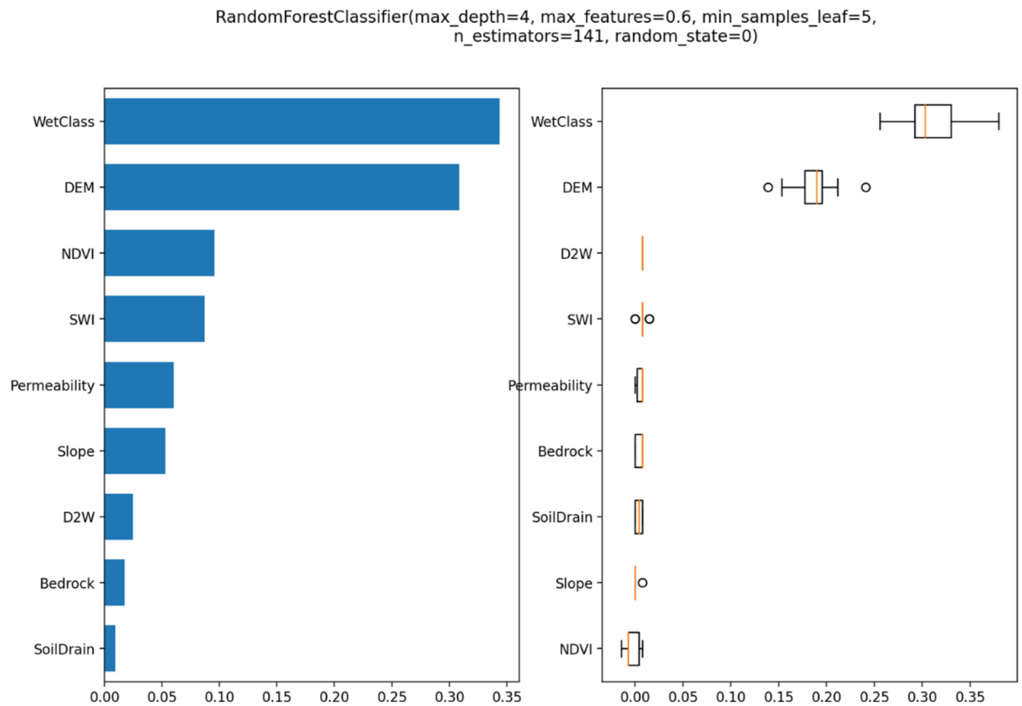

Figure E.3. *Random forest classifier weighted feature importance normalized to sum of 1 (left). Permutation features importance (right).*

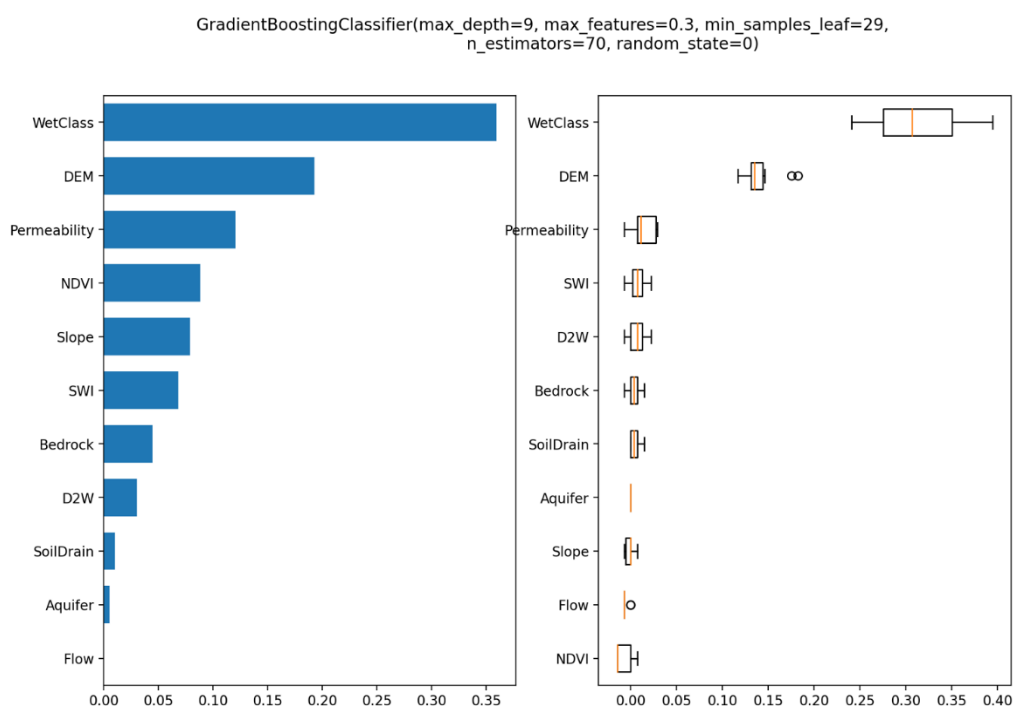

Figure E.4. *Gradient boosting classifier weighted feature importance normalized to sum of 1 (left). Permutation feature importance (right).*

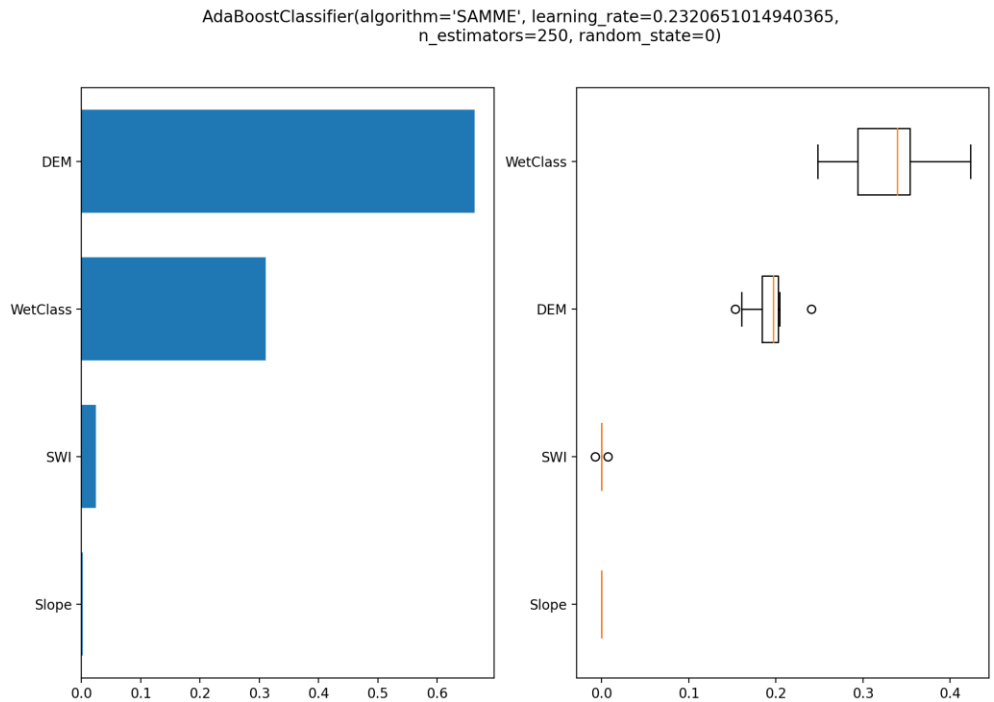

Figure E.5. *AdaBoost classifier weighted feature importance normalized to sum of 1 (left). Permutation feature importance (right).*

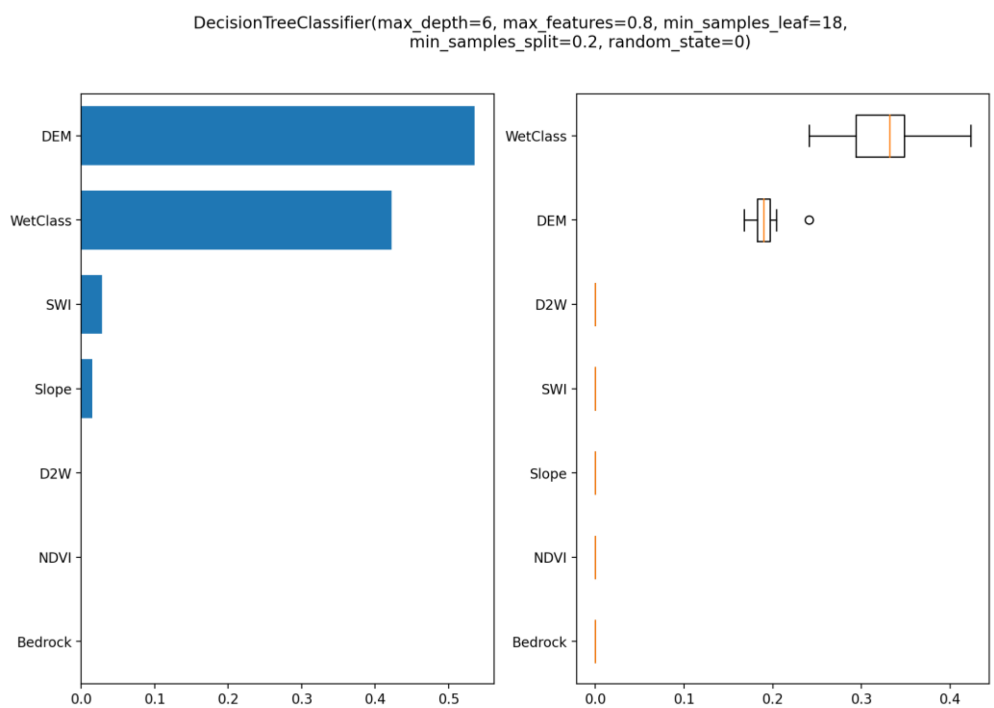

Figure E.6. *Decision Tree classifier weighted feature importance normalized to sum of 1 (left). Permutation feature importance (right).*

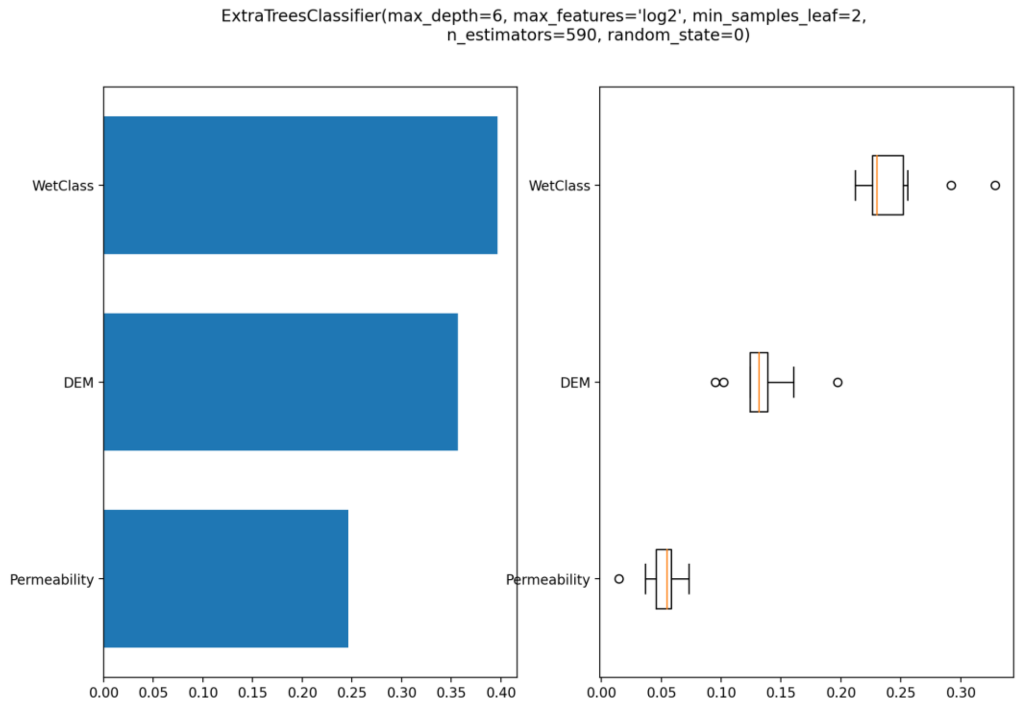

Figure E.7. Extra Tree classifier feature importance. Weighted feature importance normalized to sum of 1 (left). Permutation feature importance.