Canadian Journal of Fisheries and Aquatic Sciences

M. cerebralis establishment and spread: A graphical synthesis

Journal:	<i>Canadian Journal of Fisheries and Aquatic Sciences</i>
Manuscript ID	cjfas-2020-0352.R2
Manuscript Type:	Perspective
Date Submitted by the Author:	22-Jul-2021
Complete List of Authors:	Ramazi, Pouria; University of Alberta Fischer, Samuel; University of Alberta Alexander, Julie; Oregon State University James, Clayton; Alberta Environment and Parks, Fish and Wildlife Paul, Andrew; Alberta Ministry of Environment and Parks, Fish and Wildlife Policy Greiner, Russell; University of Alberta, Department of Mathematical and Statistical Sciences Lewis, Mark; University of Alberta, Biological Sciences and Mathematical and Statistical Sciences
Keyword:	whirling disease, spread dynamics, graphical knowledge representation, <i>Myxobolus cerebralis</i>
Is the invited manuscript for consideration in a Special Issue? :	Not applicable (regular submission)

SCHOLARONE™
Manuscripts

M. cerebralis establishment and spread: A graphical
synthesis

^aPouria Ramazi   ^{b,j}Samuel M. Fischer   ^dJulie Alexander

^{e,f}Clayton T. James   ^gAndrew Paul   ^{c,h}Russell Greiner

^{b,i}Mark A. Lewis

^aDepartment of Mathematics and Statistics, Brock University, St. Catharines, ON, L2S 3A1,
Canada

^bDepartment of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB,
T6G 2G1 Canada

^cDepartment of Computing Science, University of Alberta, Edmonton, AB, T6G 2E8 Canada

^dDepartment of Microbiology, Oregon State University, Corvallis, OR 97330 US

^eAlberta Environment and Parks, Edmonton, AB, T5J 4G8, Canada

^f(present address) Fisheries and Oceans Canada, Government of Canada, Edmonton, AB, T6X
0J4 Canada

^gFish and Wildlife, Alberta Environment and Parks, Cochrane, Alberta, T4C 1A5, Canada

^hAlberta Machine Intelligence Institute, Edmonton, AB, T5J 3B1 Canada

¹⁷ ⁱDepartment of Biological Sciences, University of Alberta, Edmonton, AB, T6G 2E9 Canada

¹⁸ ^jHelmholtz Center for Environmental Research – UFZ, Department of Ecological Modelling,

¹⁹ Permoserstrasse 15, 04318 Leipzig, Germany.

Draft

Abstract

M. cerebralis is the parasite causing whirling disease, which has dramatic ecological impacts due to its potential to cause high mortality in salmonids. The large-scale efforts, necessary to underpin an effective surveillance program, have practical and economic constraints. There is, hence, a clear need for models that can predict the parasite spread. Model development, however, often heavily depends on knowing influential variables and governing mechanisms. We have developed a graphical model for the establishment and spread of *M. cerebralis* by synthesizing experts’ opinion and empirical studies. First, we conducted a series of workshops with experts to identify variables believed to impact the establishment and spread of the parasite *M. cerebralis* and visualized their interactions via a directed acyclic graph. Then we refined the graph by incorporating empirical findings from the literature. The final graph’s nodes correspond to variables whose considerable impact on *M. cerebralis* establishment and spread is either supported by empirical data or confirmed by experts, and the graph’s directed edges represent direct causality or strong correlation. This graphical model facilitates communication and education of whirling disease and provides an empirically driven framework for constructing future models, especially Bayesian networks.

Keywords

Myxobolus cerebralis; whirling disease; spread dynamics; graphical knowledge representation; risk assessment; directed acyclic graph; Bayesian networks.

1 Introduction

Myxobolus cerebralis (Myxozoa, Bivalvulida: Myxobolidae) (Hofer, 1903) is a parasite that infects salmonids, including trout and whitefish, resulting in *whirling disease*. The parasite invades the fish via the skin and consumes cartilage, which can cause skeletal deformities and the characteristic fish “whirling” instead of normal swimming behavior (Hedrick and

El-Matbouli, 2002; DuBey et al., 2007). Whirling disease was described in Europe (Halliday, 1976) in 1903, established in the US in the 1950s-2000, and was first detected in Canada in 2016 in the province Alberta (Alberta Environment and Parks, 2017). Up to 90% of infected juvenile fish may die (Elwell et al., 2010), resulting in economic (Turner et al., 2014) and ecological costs (Kerans and Zale, 2002). Infection is preventable in fish culture settings by disrupting parasite transmission (e.g., breaking the parasite's life cycle by inactivation of parasite spores by ultraviolet (UV) irradiation (Hedrick et al., 2008), chlorination, or heating (Wagner, 2002)), but management in natural ecosystems is limited to mitigating the risk of parasite exposure and establishment because parasite elimination is impractical (Ayre et al., 2014), if not impossible. Modeling the spread and establishment of the parasite can inform management by revealing areas of high risk, allowing these areas to be prioritized for management actions.

The parasite has a complex lifecycle involving two obligate susceptible hosts including a salmonid fish and a tubificid oligochaete worm *Tubifex tubifex* (Gilbert and Granath Jr, 2003; Hedrick and El-Matbouli, 2002), and two waterborne spore stages, myxospores and actinospores (triacinomyxons, or TAMs) (Wolf and Markiw, 1984). Susceptibility varies among salmonids, and the development and severity of clinical signs including myxospore formation (necessary for parasite transmission) depend on salmonid age and size and exposure conditions (dose and environmental conditions (Hoffman and Putz, 1969; O'Grodnick, 1979; Hedrick et al., 1999, 2001b,a, 2003; Downing et al., 2002; Ryce et al., 2004, 2005)). Myxospores are released when the fish decomposes or is eaten (Hedrick and El-Matbouli, 2002), or for some fish, such as Brown Trout *Salmo trutta*, while still alive (Nehring et al., 2002). Myxospores settle out in the stream sediments, where they can be ingested by *T. tubifex*. After 60–90 days, infected *T. tubifex* release the neutrally buoyant TAM stage into the water column with feces, where they can go on to infect new fish hosts (Hedrick and El-Matbouli, 2002), completing the lifecycle.

The typical construction of models for disease/parasite establishment and spread relies

on first, identifying the covariates, and second, understanding their cumulative effects on the target variable. Researchers have used logistic regressions to identify influential covariates (Schisler and Bergersen, 2002), yet the complex disease dynamics are often de-emphasized by these additive models, which are unable to model complex interactions amongst the covariates. Others have employed mechanistic models to understand the spread dynamics of *M. cerebralis* and other myxozoan parasites, but with few covariates: ordinary differential equations with constant environmental covariates (Turner et al., 2014) and partial differential equation (PDE)s with the two non-constant covariates temperature and stream discharge (Schakau et al., 2019). In addition to simplifying the computations and analysis, limiting the choice of covariates is often inevitable in mechanistic models as the exact mechanisms by which the covariates impact the target variable are unknown. However, for *M. cerebralis*, many experiments have provided evidence for correlations between parasite establishment/spread and potential covariates (Krueger, 2002; Shirakashi and El-Matbouli, 2010; Kaeser and Sharpe, 2006a). Although typically insufficient to build predictive models, these results provide the essentials for understanding the underlying processes. Based on the correlations and both descriptive and mechanistic knowledge from the literature, experts can intuitively develop a causal understanding of the establishment and spread of the parasite, based on which qualitative and quantitative models can be constructed. Communicating the obtained “understanding”, however, can be daunting and often tedious.

Researchers have described their “beliefs” about processes via *directed acyclic graphs* (DAG), whose nodes represent the target variable and covariates, and arcs represent direct causal relationships (Wu et al., 2018; Bode et al., 2017; Eklöf et al., 2013; Herring et al., 2015). Such DAGs provide a framework for evaluating whether individual covariates impact the response variable directly, and if not, what the intermediate variables are.

Defining the seemingly trivial notion of “*A* causes *B*” has been long and controversially discussed by scholars. The two most used definitions are Granger causality (Granger, 1969) and Pearl’s causality Pearl (2009). Viewing “*A* causes *B*” as “ $A \rightarrow B$ ” leads naturally

to using DAGs to represent casual connections among multiple variables, which has since been exploited and rigorously characterized in recent decades [Greenland et al. \(1999\)](#); [Pearl \(2009\)](#); [Pearl and Mackenzie \(2018\)](#). However, the fairly complicated notions as well as often infeasible interventions have hindered the broad use of the rigorously defined frameworks. In many cases, the use of DAGs has been limited to representing the causal relationships among the variables, in only an intuitive sense.

Besides being graphical, and hence, easy to interpret, a DAG can be readily used to construct a *Bayesian network (BN)*, which is a *probabilistic graphical model* that can be fitted with data to make predictions (see Section 4.3). [Ayre et al.](#) have modeled the risk of establishment of *M. cerebralis* at a given management unit via a three-layer tree-structured BN with six covariates, categorized into three *latent* (*i.e.*, unobserved) variables: worm host environment, connectivity (to other infected management units), and (co-occurrence with and spawning of) fish host habitat. The model provides a simple explanation of a subset of the covariates involved in the parasite spread. Also, [Bartholomew et al.](#) have constructed a partially-directed graph to assess the risk of *M. cerebralis* establishment. However, a more detailed model is needed to elaborately incorporate the role of myxospore dissemination and spread (hereafter termed ‘propagule pressure’), *e.g.*, angler movement and fish movement, and other covariates such as landscape, stream discharge, and water chemistry.

Our objective was to synthesize literature findings and experts’ opinions on the spread dynamics of *M. cerebralis* in any given river system and to visualize them via a DAG. To build the DAG, we conducted three workshops to compile experts’ beliefs and merged these with empirical findings extracted from the literature. We identified the variables that have a considerable impact on the settlement of *M. cerebralis* and connected almost any pair of variables whenever one had an immediate impact on the other. The resulting DAG indicates the direct and indirect mechanisms by which the covariates affect parasite establishment, as well as the intermediate variables necessary to understand the indirect mechanisms. The graph compiles the current knowledge of the “causal relationships” between the variables

potentially involved in the spread of *M. cerebralis*, and hence, can inform management and provide the basis for future model development.

DAGs are well suited for modeling the complex lifecycle of *M. cerebralis*. Because the lifecycle involves two different susceptible hosts, salmonid fish and worms *Tubifex tubifex*, and two distinct waterborne spore stages, *myxospores* and *triacinomyxon* (TAM)s (Gilbert and Granath Jr, 2003; Hedrick and El-Matbouli, 2002; Wolf and Markiw, 1984), each stage can be compartmentalized and linked via a series of nodes. Moreover, nodes can be created specifically for myxospore production in infected fish hosts that are distinct from those releasing myxospore to the environment. This characteristic of DAGs may be particularly relevant for *M. cerebralis* because of the variability associated with parasite transmission and success in different phases of the lifecycle. For example, myxospores are primarily released from juvenile fish host tissue and settle into the sediment when the fish dies and decomposes or is eaten (Hedrick and El-Matbouli, 2002); however, they may also be released from live, clinically asymptomatic, adult fish, such as Brown Trout *Salmo trutta* (Nehring et al., 2002). In contrast, the myxospores must be ingested by *T. tubifex* before the TAM stage can be released, after around 60–90 days at appropriate temperatures (Kerans et al., 2005). Infected worm hosts then release TAMs into the water column along with faeces, and TAMs in turn, may infect fish hosts upon contact with the skin (Hedrick and El-Matbouli, 2002). DAGs possess the flexibility to carefully illustrate these complex dynamics.

2 Methods

The initial phase of the project was to construct a conceptual pathway for *M. cerebralis* spread in order to identify the underlying covariates and mechanisms. To accomplish this we conducted three workshops with the following attendees from Alberta Environment (AE), University of Alberta (UA), and Oregon State University (OSU):

- Aquaculture: Janelle Sloychuk (aquaculture specialist, AE), Trish Kelley (aquaculture

specialist, AE);

- Data scientist/management: Alicia Kennedy (GIS analyst, AE), Chad Sherburne (resource data biologist, AE), Laurie Gallagher (regional issues manager, AE);
- Fisheries biologists: Andrew Paul (population ecologist and environmental modeller, AE), Bev Larson (fish disease specialist, AE), Clayton James (fisheries biologist, AE), Dave Park (director fisheries management policy, AE), Laura MacPherson (fish stock assessment biologist, AE), Marie Veillard (fisheries biologist, AE), Michael Sullivan (fisheries scientist, AE);
- Invertebrate ecologists: Julie Alexander (invertebrate ecologist, OSU);
- Modelers: Mark A. Lewis (mathematical biologist, UA), Pouria Ramazi (modeller, applied mathematician, UA), Russell Greiner (machine-learning expert, UA).

Collectively, this expert team gradually constructed a DAG representing the conceptual pathways during the workshops. The first workshop focused on host and parasite habitat, resulting in a rough initial DAG that covered most of the major covariate groups, including hydrology, weather, and propagule pressure. However, some details, especially those on the underlying mechanisms of the propagule pressure, remained untouched. In the second workshop, we clarified some unclear terms used to describe the nodes, such as *flashiness* and modified some links to improve the representation of the causal relationships. We also incorporated the parasite and fish pressure and human movement variables into the graph. The final workshop focused on further clarifications and detailed discussions on the components of propagule pressure.

We then used the literature on whirling disease to improve the graph created during the three workshops. We focused on empirical studies identifying the driving factors of the parasite establishment and spread. The literature supported many of the nodes and links. We modified the remaining unsupported and included new nodes and links to better explain those covariates and mechanisms that were not discussed in detail during the meetings.

177 During the workshops and literature-based refinements, we went through the following
178 procedure to develop the graph: (0) Consider an arbitrary aquatic network, partitioned
179 into several segments of roughly the same size with water flowing unidirectionally – *i.e.*,
180 downstream – from one section to another. (1) Consider an arbitrary segment and model
181 the status of *M. cerebralis* establishment by the response variable *parasite establishment*.
182 (2) Model all major variables that affect parasite establishment as individual nodes, and link
183 each of them to the parasite establishment node. These variables are either observable (*i.e.*,
184 well-defined and measurable), such as host worm density, or latent (*i.e.*, abstract and not
185 measurable), such as joint propagule pressure. (3) Repeat Part 2 for each of the secondary-
186 variables mentioned above and so on until we reach an observable node that is either often
187 readily available to the managers or estimable by well-established models in the literature.
188 So, in the end, we obtain a set of nodes $\mathcal{V}$ comprising the response, observable, and latent
189 variables. In the following section, we describe each node, explain how to measure it, list its
190 parenting nodes, and explain their influences on each other.

191 To obtain the arc set $\mathcal{E}$, we link a node i to a node j if there is scientific evidence or expert
192 intuition that variable i causes or has a considerable impact on variable j . However, in rare
193 situations, we may link two variables just because they are strongly correlated, and one may
194 predict the other. For example, there is a mathematical relationship between discharge, yearly
195 precipitation, and watershed size, allowing any of them to be computed by the other two. In
196 such instances, we arrange the variables by linking the easily measurable or predictable ones
197 to the less accessible. For our purpose, accessible variables are those that are easily measured
198 and likely to be widely available, *e.g.*, temperature, measured by temperature loggers, as
199 opposed to variables that are less easily measured, *e.g.*, parasite spore densities, and thus
200 less likely to be widely available. Generally, if two variables represented similar quantities, we
201 selected the one with the best mechanistic justification or, if the variables seemed similarly
202 valid, the most accessible one. To simplify the network, we avoid linking of a variable A
203 directly to another variable C if there exists a path from A to C , *e.g.*, $A \rightarrow B \rightarrow C$, that

already explains the effect of A on C . For example, stream velocity affects or partially predicts flashiness (defined as frequency and rapidity of short-term changes in discharge (Baker et al., 2004)), which in turn affects host fish density, and hence, there is no direct link from stream velocity to host fish density. However, stream velocity and flashiness have different impacts on myxospore success, and hence, they are both linked to it. We constrain the arcs to be directional and prohibit directed cycles, resulting in the DAG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$.

3 Results

The final graph obtained by synthesizing literature findings and expert opinions is displayed in Figure 1. Brief descriptions of the nodes (variables) and information on how they can be measured are given in Table 1. In what follows, we describe the graph roughly from bottom to top via Figures 2 to 4.

Table 1: Variable description and measurement/data accessibility.

	Variable	Description	Measurement / Computation
Parasite occurrence	Infected fish density	The density of infected salmonid fish.	May be measured by testing caught fish using quantitative polymerase chain reaction (qPCR) (Fytilis et al., 2013).
	Infected worm density	The density of infected <i>T. tubifex</i> .	May be measured by sampling <i>T. tubifex</i> worms in the river sediment and subsequently testing worms for the parasite using qPCR (Alexander et al., 2011; Barry and Hanington, 2021).
	Parasite establishment	The presence and ability of <i>M. cerebralis</i> to complete its life cycle.	It is difficult to track the full life cycle of the parasite. Presence of infected fish and worms may be used as a proxy. Although challenging, the detection of myxospores in sediment Barry and Hanington (2021) and TAMs in water Lukins et al. (2007); Nehring and Thompson (2003a); James et al. (in press) is also possible (Gates and others, 2007).
Propagule pressure	Joint propagule pressure	A measure for the number of parasites from external sources introduced at the section of interest.	This is a latent variable, and hence, not measurable.

Table 1: Variable description and measurement/data accessibility.

Variable	Description	Measurement / Computation
Infected fish pressure	The number of infected fish migrating to or through a river section.	Difficult to measure directly. May instead be estimated via fish movement models incorporating the parasite establishment states of other sites in the river system.
Local parasite pressure	A measure for the number of infected worms and free-floating parasites from sources introduced to the section of interest via the stream flow.	This is a latent variable, and hence, not measurable. A score can be defined based on the parasite establishment states of neighboring river sections.
Flood-induced propagule pressure	A measure for the number of parasites and infected hosts introduced to the section of interest via flood-induced water connections, <i>e.g.</i> , when a fish pond gets flooded and infected fish escape to the river section.	This is a latent variable, and hence, not measurable.
Inland navigation	A measure for the number of parasites introduced via vessels used for recreation and transportation of goods.	May be estimated using on-site surveys or, if available, high-quality data on inland navigation (Mari et al., 2011). Although vessel traffic may be easily observable, the volume of transported propagules is difficult to determine.
Overland propagule transport	A measure for the number of parasites introduced via recreational and commercial overland traffic, <i>e.g.</i> , when fish are transported to stock fish ponds.	Surveys reveal the number of anglers accessing a river section after visiting a infected site and the transportation pathways of potentially infected material to fish hatcheries and stocked fishing ponds. However, surveys can rarely cover all sites, and models may be required to estimate the incoming traffic at sites for which no data are available (Bossenbroek et al., 2001; Potapov et al., 2010; Muirhead and MacIsaac, 2011; Drake and Mandrak, 2014).
Habitat and hosts TAM success	The number/density of TAM that survive and infect fish.	May be determined by the myxospore burden (amount of myxospores in fish). TAM density may be measured using water filtration methods (Thompson and Nehring, 2000; James et al., in press; Lukins et al., 2007).

Table 1: Variable description and measurement/data accessibility.

	Variable	Description	Measurement / Computation
	Host fish density by age and species	(vector-valued) The density of susceptible salmonid fish in different age and species classes (Elwell et al., 2010).	May be determined via fish sampling.
	Host worm density	The density of <i>T. tubifex</i> .	May be determined via sediment sampling Alexander et al. (2011).
	Myxospore success	The number/density of myxospores that survive and eventually turn into TAM.	May be estimated by the amount of TAM in the water column. (Thompson and Nehring, 2000; James et al., in press; Lukins et al., 2007)
Water chemistry	Conductivity	The ability of water to transmit electrical flow. High conductivity indicates high concentration of dissolved electrolytes.	May be measured using conductivity meters.
	Water pH	The water pH value; serves as a measure for the acidity of the water.	May be measured with pH meters. In sufficiently deep lakes, stratification can result in differing pH values in the top and bottom water layers.
	Dissolved oxygen	The concentration of dissolved oxygen.	May be measured using a dissolved oxygen meter and sensor.
	Nutrient enrichment	Content of organic matter and inorganic nutrients in the water column.	Organic and inorganic nutrients can be measured through standardized methods for water quality sampling (ccm, 2011; Hellawell, 1986).
Landscape	Agricultural and aquacultural facilities	The proximity to agriculture production facilities, e.g., feedlots, that potentially affect the water quality	The locations of facilities of concern may be determined by means of detailed surveys and spatial databases.
	Vegetation and land use	The predominant vegetation and land use in the surrounding area; captured either by a categorical variable indicating the types forest, grassland, pasture land, arable land, etc., or by a vector indicating the percentage of the different types (Kaesler and Sharpe, 2006a).	General data on the vegetation are often readily available on large scales. Information on land use may be obtained from government records.

Table 1: Variable description and measurement/data accessibility.

Variable	Description	Measurement / Computation	
Soil	The predominant soil composition in the surrounding area; characterized by the chemical properties, grain size, water repellency, etc., (Schoeneberger et al., 2012).	May be measured by soil sampling (Schoeneberger et al., 2012). However, characteristics important to hydrology can be difficult to measure and may be estimated via models based on other characteristics (Boorman et al., 1995). Also, although often extrapolated, soil data are available in many areas due to their wide use.	
Climate	Degree days	The cumulative daily water temperature exceeding a certain threshold over a year.	Measurement requires continuous temperature measurements at different sites. May be estimated by the widely available air temperature measurements.
	Yearly precipitation	The per-area precipitation averaged over a year.	May be computed using the widely available weather data.
Hydrology / river morphology	River sediment	The size distribution of particles in the sediment of the waterbody.	May be measured by extracting, drying, sieving, and weighing sediment samples from waterbodies. The content of organic matter may be determined by the weight loss of a sediment sample after incineration (Kaesler and Sharpe, 2006b; McGinnis and Kerans, 2013).
	Flashiness	The temporal variation of the stream velocity or discharge; indicates how strongly and rapidly a stream is affected by weather events such as precipitation.	Typically computed based on repeated stream velocity and discharge measurements (Baker et al., 2004; Allan and Ibañez Castillo, 2009), which requires continuous sampling effort.
	Stream velocity	The velocity of water current; that is, the distance that a weightless particle in the stream would pass per time unit.	Measured with a variety of methods differing in accuracy and required effort. Since the water current is not constant throughout the river cross section, stream velocity is typically measured at several locations along the cross section and at specific depths (Whiting, 2016).
	Hydraulic radius	The ratio of a channel’s cross-sectional area to its wetted perimeter, the length of the river bed and bank in contact with water taken along a cross section (Gordon et al., 2004).	May be calculated directly from measurements taken from the river bed on-site. Measurements can be extrapolated onto the whole river network via geostatistical methods using the stream width (Shahrokhnia et al., 2004).

Table 1: Variable description and measurement/data accessibility.

Variable	Description	Measurement / Computation
Channel roughness	Also called Manning's n (Gauckler, 1867; Manning et al., 1890). A coefficient capturing a river's resistance to flow due to factors such as surface roughness, vegetation and channel irregularity (Chaudhry, 2007).	Can be estimated directly from on-site measurements of discharge, hydraulic radius and slope. Indirect estimation is possible using stream classification, geology, substrate size, channel characteristics, river stage, vegetation or some combination (Leopold, 1994; Gordon et al., 2004)
Stream slope	The slope of the river section.	May be calculated by dividing the difference in elevation between the beginning and the end of the river section by the river's length in that section. Elevation data may be taken from remote sensing data sets or measured on-site.
Stream width	The width of the stream.	May be measured on-site or computed from mapping data. Varies in predictable patterns with discharge (Leopold, 1994).
Discharge	The volume of water entering and leaving the river section over time.	May be measured by portable devices, which are often based on simultaneous measurements of the stream velocity and cross sectional area. If the flow velocity cannot be measured, the slope of the stream may be used as a proxy. May also be measured by built infrastructures, resulting in relatively effortless and accurate measurements (Whiting, 2016). Alternately, less accurate and precise estimates of discharge can be made by adjusting discharge measured in nearby watersheds by watershed area and potentially other covariates (Leopold, 1994).
Watershed area	The area of the catchment region from which water will flow into the river section.	Data on the watershed area are often readily available.

The graph can be used to predict in each river section, whether *M. cerebralis* is established, *i.e.*, is present and able to complete its life cycle. Increased densities of infected salmonids and oligochaete hosts indicate parasite establishment (Figure 2). If no on-site host samples are available, environmental variables may be used to predict *M. cerebralis* establishment.

Figure 1: The directed acyclic graph obtained from the literature and experts’ opinions on the spread of *M. cerebralis*. To compact the network representation, we use *plate models* (Koller and Friedman, 2009) to encode repeated structures, denoted by the red rectangles on the top left. Call the largest sub-graph whose nodes and arcs are all inside the rectangle, the *plate*, and the node that is out of the rectangle but linked to a node inside the rectangle, the *outside node*. For example, in the leftmost rectangle (plate), the plate consists of the nodes *infected fish density (past)* and *distance (river)* linked to *incoming infected fish*, and the outside node is *infected fish pressure*. The plate model encodes the following graph: N number of repeated inside sub-graphs that are all linked to the outside node, where N is the number at the upper-right corner of the rectangle. Back to the example, where N is the number of *river sections*, the plate model represents a graph, where the sub-graph repeated in the plate is repeated as many times as the number of *river sections*, and the node *incoming infected fish* in each plate is linked to the outside node *infected fish pressure*, which can also be seen as an “aggregation node”.

Figure 2: **Parasite occurrence.** The directed acyclic graph indicates two groups of variables that directly influence or “cause” the parasite occurrence: (i) propagule pressure and (ii) parasite hosts and success.

Successful establishment requires the introduction of *M. cerebralis* propagules from external sources, as well as potential hosts along with proper environmental conditions, as explained below.

3.1 Propagule pressure

The node joint-propagule-pressure summarizes the variety of pathways via which *M. cerebralis* can be introduced to a site (Figure 3). The node represents a numeric score, which is higher as more parasites are introduced from external sources. This could happen directly via introduction of parasite spores or indirectly via introduction of infected hosts. To account for the differences in propagules introduced via different vectors - and thus their ability to establish a population - we compute the propagule pressure score as a weighted sum of propagule counts from different sources and life stages. The weights representing establishment potential could be estimated based on empirical data on establishment success.

Propagules may be distributed locally with the downstream flow, or, on larger scales, along with introduced infected fish. Long-distance and cross-watershed dispersal could occur via recreational and commercial anthropogenic movement (Modin, 1998; Bartholomew et al., 2005b), including both overland traffic and inland navigation. Flooding events that connect

Figure 3: **Propagule pressure.** See the caption of Figure 1 for an explanation on the plate notation indicated by the red rectangles.

separated water sections, such as fish ponds, to the river network allow parasites and infected hosts to be dispersed between unconnected waterbodies. Due to the time lag between the initial introduction of propagules and establishment of a parasite population, the spreading events must have occurred before the establishment time point.

The node *local parasite pressure* captures the possibility that the stream transports TAMs and infected worms from other river sections to the site of interest. Although *T. tubifex* can migrate short (1 km/yr *e.g.*, (Cassie, 1971)) upstream distances, local parasite pressure is directed mainly downstream. Moreover, since TAMs and host worms are unlikely to be carried over long distances by the stream, local pressure can be limited to adjacent upstream river-sections only. The pressure depends on the density of infected hosts and parasite habitat quality in the adjacent river sections. Host density is a rough indicator of the number of released myxospores and TAMs, and habitat quality may indicate the number of propagules produced per host and whether they survive long enough to reach a neighboring river section.

Infected fish can spread *M. cerebralis* over intermediate distances (10–200 km) in both

upstream and downstream directions, *e.g.*, 180 km for Mountain Whitefish (Northcote, 1997; Benjamin et al., 2014). The number of fish migrating between two sites decreases with their distance and may depend on the number of fish at the origin and the fish habitat quality at the destination site. Therefore, the number of incoming infected fish may be estimated by the weighted sum of infected fish in other river sections. The weights incorporate the number of hosts at the site of interest, as well as the length, shape, direction, and velocity of the connecting stream.

Overland propagule transport can occur via human recreational and commercial overland traffic. Recreational traffic includes anglers or recreational boaters carrying infested watercraft, gear, fish, or live wells (Bartholomew et al., 2005b). Parasite distribution may occur via commercial transportation if infected fish are introduced to stocked fishing ponds or fish hatcheries or if fish stocks are fed with products made from infected fish (Modin, 1998; Bartholomew et al., 2005b). The number of potential anthropogenic parasite vectors (“agents”) travelling between two sites depends on the “attractiveness” of the origin and destination and the distance between the two (Bossenbroek et al., 2001; Potapov et al., 2010; Muirhead and MacIsaac, 2011; Drake and Mandrak, 2014). The attractiveness of a river section may be computed based on the number of visiting agents, which in turn can be estimated based on covariates such as the river-section size, proximity to dense human populations, and available infrastructure (Fischer et al., 2020). Since the factors driving attractiveness may be different for recreational and commercial traffic, it may be necessary to consider these transporting mechanisms separately. In addition to the vector traffic, the densities of infected worm and fish at the origin must also be taken into account.

Inland navigation also facilitates parasite dispersal and is considered a major vector for several aquatic invasive species (Galil et al., 2007; Panov et al., 2009) via hull fouling and ballast water transport. Inland navigation has the potential to disperse *M. cerebralis* as well. However, inland navigation requires sufficiently large and deep rivers, and hence, may be significant for only certain systems.

As with anthropogenic overland transport, inland navigation may be estimated based on the attractiveness of boat launches, measured, for example, by the tonnage of goods or number of people passing through them (Mari et al., 2011) and the waterway distance between the locations. Since propagules are taken up and released not only at destinations but also in intermediate river sections, the propagule pressure due to inland navigation may be driven by the traffic passing through a river section rather than just the number of vessels arriving at particular destination points.

Natural parasite dispersion between unconnected waterbodies can occur both via wildlife such as waterfowl (Taylor and Lott, 1978) and during flooding events. Though flood events facilitating cross-contamination may be rare, flood-induced propagule pressure may be significant in permitting parasites to spread between unconnected river systems or escape from infected stocked fish ponds. The chances of establishing flood-induced water-connections may be assessed via a geo-spatial analysis in combination with precipitation data. The number of dispersing propagules may be estimated from the density of infected worms and fish in the donor ponds and the corresponding parasite habitat quality.

3.2 Parasite hosts and success

Suitable environmental conditions for myxospores and TAMs and the presence of susceptible hosts are prerequisites for the establishment of *M. cerebralis*. Both the myxospore and TAM stages are directly influenced by environmental stressors while they are waterborne and indirectly influenced by environmental stressors while they are within the hosts (Figure 4). These factors, hence, significantly affect the parasite establishment and spread likelihood.

Success for a myxospore requires production in an infected fish followed by release to the sediments and surviving long enough to infect a compatible worm host. Suitable environmental conditions including water temperature (degree days), stream velocity, and flashiness affect myxospore success. Optimal thermal environments are characterized by low maximum, high minimum, and low mean stream temperatures (Krueger, 2002), with the optimal value

Figure 4: **Parasite host and success.** The directed acyclic graph indicates four groups of variables that directly influence or “cause” the parasite hosts and success: (i) landscape, (ii) water chemistry, (iii) climate, and (iv) hydrology and river morphology.

of 15 °C (El-Matbouli et al., 1999). However, myxospores tolerate freezing at –20 °C for at least three months (El-Matbouli and Hoffmann, 1991) and can remain viable following a variety of thermal extremes (Wagner et al., 2003). Myxospores also tolerate extreme stream velocities; however, they proliferates more in slow currents (Hallett and Bartholomew, 2008; Krueger, 2002). Stream flashiness is another variable that may prevent the establishment of *M. cerebralis*, because myxospores are passive spores that cannot swim or burrow to avoid being displaced. When infected host fish decompose, myxospores are deposited in the sediment. Consequently, myxospores may be flushed away, reducing their probability of ingestion by a host worm in flashy streams.

Success for a TAM requires production in an infected worm host, followed by release to the water column and surviving long enough to infect a compatible fish host. TAM success also depends suitable environmental conditions, including degree days, stream velocity, and water pH and conductivit. TAMs are produced at water temperatures from 5 °C to 30 °C (El-Matbouli et al., 1999) and require 1,320 degree-days at 15 °C and 1,456 degree-days at 8 °C (Kerans et al., 2005) to mature within their worm hosts. TAM release duration decreases with increased temperature (Blazer et al., 2003). TAMs can remain viable in the water column for up to 15 days at temperatures between 6 °C and 15 °C (El-Matbouli et al., 1999), but only up to three days at temperatures above 19 °C (Markiw, 1992). In addition, TAM release rate appears to be optimal at moderate water temperatures (Krueger, 2002). Studies have reported low and high parasite prevalence among *Cutthroat Trout* (*Oncorhynchus clarki*) at 9.5 °C and 12 °C, respectively (de la Hoz Franco and Budy, 2004), and maximum prevalence in *Rainbow Trout* at 13 °C (Baldwin et al., 2000).

Water chemistry also affects TAM viability and transmission, but the relationships may be confounded by multiple environmental variables. For example, water conductivity is correlated with the prevalence of infection in sentinel fish in multiple field studies (Sandell et al., 2001; Krueger et al., 2006), suggesting TAM viability or transmission increases with conductivity. However, 20-30% salinity for at least 6 hours can kill TAMs (Smith et al.,

2002), and salinity is typically measured by conductivity. TAM viability is also affected by water pH: the further pH moves from neutral, the more TAM viability decreases (Smith et al., 2002). No significant relationship has been described between dissolved oxygen and TAM viability (Smith et al., 2002) or fish infection (Schisler et al., 2000).

In addition to environmental conditions necessary to the survival and success of *M. cerebralis* in its waterborne or out-host life stages, the parasite requires sufficient densities of both susceptible salmonid and worm hosts (oligochaete) to complete its life cycle.

The density of the compatible oligochaete host *T. tubifex* determines how effectively myxospores eventually develop into TAMs. The host worm density depends on river sediment composition, nutrient enrichment of the water, water oxygen concentration, and water temperature. In general, *T. tubifex* tolerates a wide range of environmental conditions as long as certain minimal habitat requirements are satisfied and can achieve extremely high densities when conditions are optimal (Alexander et al., 2011).

T. tubifex is tolerant of poor water quality conditions (low oxygen concentration and high nutrient enrichment). This gives the species a significant advantage over its competitors under unfavorable habitat conditions. As a consequence, *T. tubifex* has been observed at high densities in organically polluted/enriched habitats, and at comparably low densities in habitats having good water quality, where its competitors dominate (Kaesler and Sharpe, 2006b, 2008). The competition strength is, however, hypothesized to be highly temperature dependent and govern *T. tubifex* densities only in sufficiently warm rivers (water temperatures above 10 °C for more than 14-20 weeks annually (Kaesler and Sharpe, 2006a)). Therefore, there may be a strong interaction between the covariates nutrient enrichment and degree days. *T. tubifex* and oligochaete competitors have been found to have low densities in confined rivers (Anderson, 2011) and steep slope ($\geq 4\%$) rivers (Kaesler and Sharpe, 2006a), which typically have coarse-grain sediment. Several studies suggest that the species prefers fine silt and clay sediment (Lazim and Learner, 1987; Sauter and Güde, 1996).

The species and age distribution of salmonid fish determine their susceptibility to *M.*

cerebralis (Elwell et al., 2010; Ryce et al., 2005), and hence, the ability of *M. cerebralis*. Therefore, it is imperative to model salmonid fish density as a vector-valued variable containing the density of susceptible salmonid fish by age and species class. To account for temporal variation, we may consider the average species densities in the productive season of *M. cerebralis*. In Colorado, TAM production was reported highest during late spring to late fall with water temperature ranging from 10 to 15 °C (Nehring et al., 2003; James et al., in press).

Several factors affect salmonid density (Bjornn and Reiser, 1991; Warren et al., 2015). However, since salmonid species differ in behavior and habitat preferences, it is difficult to identify predictors that are generally applicable (Warren et al., 2015). Furthermore, different factors are important at different life stages. Salmonid species have temperature tolerances ranging from just above 0 °C to 29.8 °C with preferred ranges variable among species (Bjornn and Reiser, 1991), *e.g.*, 10 °C–16 °C for bull trout, which overlaps almost perfectly with the ideal reproduction temperature-range of *M. cerebralis* (El-Matbouli et al., 1999), but close to 20 °C for rainbow trout (Scott and Crossman, 1973). Salmonids also require a relatively high oxygen concentration, with distress symptoms starting at oxygen concentrations below 6 mg/L (Bjornn and Reiser, 1991; Armstrong et al., 2003). The oxygen concentration, in turn, can be strongly affected by nutrient enrichment, as the decomposition of organic matter can use a significant amount of oxygen (Hellawell, 1986). Water pH values outside the range of 4.5–9 are likely harmful to salmonids (Crisp, 1996). The sediment type can have an effect via silt particles in the water, decreasing fish growth rates (Sigler et al., 1984). For spawning, salmonid species require minimal water depths of 6–30 cm, stream velocities between 0.11 m/s and 1.09 m/s, and temperatures between 2.2 °C and 20 °C (incubation: 4.4 °C–14.4 °C) (Bjornn and Reiser, 1991; Armstrong et al., 2003). Embryo survival is reduced in the presence of fine sediment (>1 mm size) (Armstrong et al., 2003; Greig et al., 2005). Furthermore, acidic water (pH 5.8–6.5) has been found to inhibit upstream migration and nest-digging of females (Ikuta et al., 2003).

3.3 Relationships between environmental covariates

Not every environmental variable affecting *M. cerebralis* establishment may be available at all sites of interest. Therefore, it is useful to know how these variables are interrelated and affected by other covariates that are easier to measure and more likely to be available.

Nutrient enrichment is often driven by the amount of fertilizer and contents of liquid manure that are flushed into the river. Therefore, nutrient enrichment may be estimated by near-by vegetation and land use. In addition to pollutants introduced from fields and pasture areas, agricultural, industrial, and municipal facilities in close proximity to a waterbody can be significant sources of organic pollution. This includes fish hatchery discharges (Kaesler and Sharpe, 2006a), animal farms (Kaesler and Sharpe, 2008), sewage effluents, or other facilities in close river proximity contributing significant nutrient inputs to the stream. A weighting of different sources of pollutants can be conducted based on distance to the stream and empirical findings on the respective effects of different facility types (Hellawell, 1986).

River sediment composition depends on the soil, vegetation, and land use in the drainage area and on the water flow (Schumm et al., 2020). Intensive agricultural land use can result in soil erosion, and thus, a higher inflow of sediments to the river system. Furthermore, the amount of flushed sediment depends on the vegetation close to the river bed and in the drainage area (Allan and Ibañez Castillo, 2009). In addition, land use and vegetation affect the content of organic matter in the sediment.

The volume of introduced sediment depends also on the hydrological properties of the river. The faster and more variable the water current and the steeper the slope, the larger and more particles will be introduced to the river from the river bank (Allan and Ibañez Castillo, 2009). Furthermore, sediment inflow from upstream river sections increases with discharge. However, sediment composition depends not only on the amount of introduced particles but also the settlement of these particles. Since faster rivers also transport more particles than slow rivers, fine-grain sediment may settle at slower downstream river sections. However, due to the complex interactions of the different covariates, intricate modelling approaches

(White, 2005) are often necessary to accurately estimate sediment composition in rivers.

Stream velocity can be computed from the hydraulic radius, stream slope, and channel roughness, using Manning's equation Chaudhry (2007). The hydraulic radius can be estimated via discharge and stream width. Discharge may be estimated by considering the yearly precipitation as a source for water and the area of the watershed, where the precipitation could be collected Touazi et al. (2004). Alternatively, the discharge could be computed as the sum of the discharge of upstream river segments. This, however, requires discharge to be known at first-order streams Allan and Castillo (2007), which in turn could be estimated via the method described above. As discharge, hydraulic radius, and stream velocity are often temporally and spatially variable within a river, we may consider average values over space and time.

High flashiness streams are typically characterized via short phases of strongly increased discharge. However, some measures of flashiness include normalization factors, making it independent of discharge. Flashiness has been found to decrease with increasing watershed area (Baker et al., 2004; Holko et al., 2011). Furthermore, flashiness is positively correlated with imperviousness of the soil, which in turn can be affected by human land use (Roy et al., 2005; Allan and Ibañez Castillo, 2009). Increased agricultural land use and forest cover in the watershed have been found to be negatively correlated to flashiness (Holko et al., 2011).

4 Discussion

Whirling disease can dramatically increase the mortality of some salmonid species, *e.g.*, it was responsible for the near-to-complete collapse of rainbow trout in Colorado and Montana (Nehring and Thompson, 2003b), and there is now evidence of whirling disease impacting populations of *Rainbow Trout* in Alberta James et al. (in press). Although effects on recreational fishing and tourism may be negligible in some areas (Elwell et al., 2010), the overall economic, ecological and social costs can be significant (Ben-David et al., 2016; Turner et al.,

2014). Successful management requires a thorough spatial assessment of the parasite establishment risk. By synthesizing expert opinions and a wide body of literature, we have developed a graph displaying a detailed and clear overview of the many factors currently believed to affect *M. cerebralis* establishment and the interplay of these factors. Consequently, the graph facilitates both our understanding and the management of whirling disease.

Our DAG provides an easy-access visualization of the existing results on and experts' understanding of the establishment and spread of *M. cerebralis*. It additionally takes the key first step to identifying data requirements for quantitative analysis, building the corresponding BN, examining expert's understandings in the form of hypothesis testing, and making probabilistic risk assessments.

4.1 Significance

Early detection of invasive species can increase the outcome of control and mitigation strategies, lowering the impact and encountered costs (Finnoff et al., 2010; Blackwood et al., 2010; Epanchin-Niell and Wilen, 2012). However, continuous monitoring of large river systems is costly and resource intensive. Therefore, models can help managers to focus monitoring efforts by assessing invasion risk and guiding management efforts to the locations where they are most effective.

A particular challenge for risk assessment models is the diversity of factors potentially impacting the ability of the invader – in this case, *M. cerebralis* – to establish in a new location. As a result, insights from a variety of scientific fields are required to understand and predict the parasite invasion, establishment, and spread. Earlier models for *M. cerebralis* spread either focused on specific mechanisms, thereby neglecting important environmental variables and human-mediated dispersal (Ayre et al., 2014), or provided rather general frameworks for risk assessment without accounting for many insights from empirical research (Bartholomew et al., 2005a).

Our approach addresses both issues by merging empirical results from a large body of

literature from different research areas to form a single comprehensive model for spatial prediction of *M. cerebralis* establishment. By being specific about mechanistic links, the model provides a framework into which future empirical findings can be embedded. The graph also facilitates the formulation of hypotheses that may be tested to improve our understanding of *M. cerebralis* spread.

In particular, the impact of propagule pressure on parasite spread is fairly quantified, thanks to the ongoing research on human, specifically angler, movement models [Erlander and Stewart \(1990\)](#); [Leung et al. \(2004\)](#); [Bossenbroek et al. \(2007\)](#); [Mari et al. \(2011\)](#). The effect of parasite-establishment-(past) is also captured by mechanistic models [Turner et al. \(2014\)](#); [Schakau et al. \(2019\)](#); however, more research is required to quantify the relationship between the other two parasite-occurrence covariates, *i.e.*, infected-fish-density and infected-worm density, with the target variable parasite-establishment. The importance of the habitat-and-host variables in parasite establishment is confirmed in many studies [Elwell et al. \(2010\)](#); [Ryce et al. \(2005\)](#); [Alexander et al. \(2011\)](#); [Gilbert and Granath Jr \(2001\)](#), yet a model capturing the governing mechanisms is missing. The effect of climate covariates on the other nodes in the DAG is well-studied [Krueger \(2002\)](#); [Kerans et al. \(2005\)](#); [Touazi et al. \(2004\)](#). The role of water-chemistry nodes in parasite establishment is perhaps the most blurred, as there have been few related studies [Sandell et al. \(2001\)](#); [Smith et al. \(2002\)](#) with many open questions. There is a rich body of literature on the hydrology-and-river-morphology variables [Chaudhry \(2007\)](#); [Gordon et al. \(2004\)](#). Moreover, the effects of the two variables stream-velocity and discharge on the habitat-and-host nodes have been quantified using partial differential equations in [Schakau et al. \(2019\)](#). However, characterizing the co-effect of flashiness, river-sediment, stream-velocity, and stream-slop remains concealed. Similarly, more work is required to quantify the exact effect of landscape variables on other connected variables. Out of all these variables, understanding the underlying mechanisms of those directly connected to the target node parasite-establishment is likely to be more influential in modeling the parasite spread and establishment process.

Our DAG and the reviewed empirical results also provide the essentials for building future mechanistic, statistical, and machine-learning models. Modellers can start from the target variable *parasite establishment* and model the mechanism via which it is impacted by its parenting nodes. Modellers may repeat this process for the nodes higher in the hierarchy and stop at any variable for which data is available in the area of interest since then there is no need to estimate that variable using its predictors in the upper hierarchy.

The information provided in this synthesis largely summarizes the necessary variables needed to estimate the potential establishment and spread of *M. cerebralis* or similar aquatic invasive species into novel environments. As such, the conceptual information presented herein would primarily appeal to not yet established jurisdictions where the risk of parasite introduction is high. For example, British Columbia currently have no known detections of the parasite to date but has susceptible fish populations in close proximity to *M. cerebralis* positive watersheds in Alberta and have recently implemented a surveillance program (REF, 2018). This would represent a scenario whereby our DAG may help inform sampling efforts in areas where a higher probability of detecting the parasite exists.

On the other hand, for locations such as Alberta, in which the distribution of the parasite is generally known but there exists a risk of whirling disease outbreak and loss to fish populations (James et al., in press), variables (nodes) more specific to disease outbreak (rather than parasite spread and establishment) should be considered in the predictive models. Such variables include: (i) a high abundance of the aquatic worm host, (ii) the presence of vulnerable juvenile fish hosts, and (iii) water temperatures ideal for the development of the parasite (Elwell et al., 2010; James et al., in press).

4.2 Graph construction

The graph reveals the interaction hierarchy of the variables impacting *M. cerebralis* establishment. Thereby, direct effects (or “causalities”) are represented by links, and indirect effects via paths of links. Although many links in the graph represent direct causal relationships,

the exact mechanism behind some links may be unknown, and for some relationships there is no causal direction. For example, discharge and velocity relate via a simple mathematical equation: $\text{discharge} = \text{velocity} \times \text{stream width} \times \text{stream depth}$. However, discharge and velocity are not causally related, *i.e.*, neither velocity nor discharge “causes” the other. Hence, the link from discharge to velocity is not causal but indicates dependence or correlation.

There is not a unique choice of variables to explain parasite establishment dynamics, even when they have a simple life cycle. The interactions between environmental and biotic variables are often highly complex and involve numerous intermediate factors. For example, vegetation and land use can affect river sediment via different mechanisms, including the erosion of soil that can then be flushed into water bodies. To reduce the graph complexity, we focused on the variables and interactions deemed most influential and neglected latent intermediate variables if they had a single predictor. However, if the intermediate variables were already included in the graph, we used them to represent mechanisms more accurately. For example, a positive correlation has been reported between sentinel rainbow trout infection and conductivity (that can serve as a proxy for nutrient enrichment) in (Sandell et al., 2001). This correlation may be explained via the positive effect of nutrient enrichment on host worm density, which increases TAM density, parasite establishment, and infected fish density, in order. Therefore, we did not include a direct link between nutrient enrichment (or conductivity) and fish infection prevalence in the graph.

In our effort to synthesize literature findings, we needed to consider potential intermediate variables for correlations and unify results by merging similar variables considered in different studies. Some mechanisms may be similarly well described via different environmental predictors. For example, maximal stream temperature is strongly correlated with degree days, and either of the two variables could be used to model a temperature-driven mechanism. Similarly, changes in conductivity can be used as a predictor for water pollution via nutrient enrichment (Fondriest Environmental Inc., 2014). However, because there are multiple mechanisms that impact conductivity, we chose to focus on nutrient enrichment

instead.

We excluded variables with little influence, such as the density of non-susceptible host fish, *e.g.*, longnose suckers (*Catostomus catostomus*) and carp. Although TAMs can still attach to these “non-target” fish, thereby reducing the density of TAMs available to infect target hosts (Kallert et al., 2009), there are typically still enough TAMs to complete the parasite life-cycle.

We have constructed the graph under the assumption that variables take value in the same fixed time interval, say over a year. Some variables such as annual precipitation and degree days are defined on this basis; others such as flashiness and parasite establishment may be defined at a specific time of the year, say the end of August. One may average out some of the variables over the whole year or the time interval of interest. For example, stream velocity may be averaged over the year, or just over the months that are believed to have maximal effect on the establishment of the parasite. Some variables may have been measured occasionally, and hence, at different time intervals than other variables. For example, it might be that water pH is measured at year 2010, whereas the parasite establishment is studied at 2015. This is typically not a major issue as long as the spatial pattern of the variable remains unchanged over the years.

4.3 Extension to Bayesian networks

In addition to its qualitative use, our DAG can be used to assess the risk of parasite establishment quantitatively if the network is parameterized as a BN. If each node v in the DAG is understood as a random variable, we can parameterize its dependency on its parents by the *conditional probability distribution (CPD)*. More specifically, associated with each node v a random variable X_v is defined, which can be either discrete or continuous. For example, the discrete random variable $X_{\text{discharge}}$ can be either Low or High to indicate the discharge level. Similarly, corresponding to the parents of this node, the binary random variables $X_{\text{watershed-area}}$ and $X_{\text{yearly-precipitation}}$ are defined. Being affected by its parental nodes, the

$X_{\text{watershed-area}}$	$X_{\text{yearly-precipitation}}$	$X_{\text{discharge}}$	
		Low	High
Low	Low	0.9	0.1
Low	High	0.5	0.5
High	Low	0.4	0.6
High	High	0.2	0.8

Table 2: Conditional probability distribution (CPD) of discharge based on watershed-area and yearly-precipitation. The numbers are for illustrative purposes only and indicate the probability of discharge given the levels of watershed-area and yearly-precipitation. The numbers at each row add up to 1.0 as they correspond to a probability distribution.

probability distribution of discharge depends on that of its parents. This results in the CPD $P(X_{\text{discharge}}|X_{\text{watershed-area}}, X_{\text{yearly-precipitation}})$ which can be presented as Table 2. In general, the CPD of each random variable X_v is written as $P(X_v|X_{\text{Pa}_v})$, where X_{Pa_v} indicates the random variables corresponding to the parents of node v .

One may either manually specify the CPDs or learn them from data (see (Marcot et al., 2006) for guidelines). The network $\mathcal{G}$ together with the CPDs associated with each node results in a BN that factorizes the joint probability distribution of all of the variables $X_{\mathcal{V}}$ as follows (Pearl, 1988; Koller and Friedman, 2009):

$$P(X_{\mathcal{V}}) = \prod_{v \in \mathcal{V}} P(X_v|X_{\text{Pa}_v}).$$

The resulting BN can then be used to compute the chances of parasite establishment in a region of interest, given the specified values of some of the covariates.

The BN also encodes a precise set of probabilistic interdependencies among the variables. Any two nodes that are linked by a path of arcs without a “v-structure”, *i.e.*, two arcs pointing towards each other ($a \rightarrow b \leftarrow c$), are probabilistically dependent. For example, TAM-success and infected-worm-density are probabilistically dependent as they are linked by the directed path TAM-success $\rightarrow$ parasite-establishment $\rightarrow$ infected-worm-density in Figure 1. So are TAM-success and myxospore-success as they are linked by the path TAM-success $\leftarrow$

degree-days $\leftarrow$ myxospore-success. However, conductivity and water-pH are probabilistically independent as they are only linked by the path conductivity $\rightarrow$ TAM-success $\leftarrow$ water-pH, which is a v-structure. There are many other more-subtle ways to read off independencies (and its complement, dependencies), which can be used to help understand the connections. For example, two variables may become conditionally independent given another variable. One may refer to the notion of d-separation for identifying all conditional independencies between the variables in a BN (Pearl, 1988). Such (conditional) independencies can then be tested based on collected data, and if violated, suggests ways to produce a more accurate DAG structure.

Compared to other mechanistic, statistical, and even most machine-learned models, BNs have the advantage of handling missing values – *e.g.*, Figure 1 shows more than 50 nodes; a standard model would not be able to estimate the chance of parasite-establishment unless we provide all values of these nodes (or at least, those of parasite-establishment’s *Markov blanket*, which are its parents, its children, and the “co-parents” of those children). However, as this BN is a probability distribution, one can simply marginalize out the unspecified variables to obtain the probability of parasite establishment.

Moreover, some parts of the BN, for example the sub-graph consisting of the node propagule pressure and its parents and connecting links, can be modelled separately by models such as gravity models (Muirhead and MacIsaac, 2011; Potapov et al., 2011) and then be connected to the remaining of the BN. This allows us to mechanistically model those parts with known mechanisms and estimate the rest from data. Finally, control can also be integrated into BNs via the introduction of *decision nodes*, representing management actions that can be taken at all or some of the partitioned regions, and *utility nodes*, indicating the associated costs and benefits (Koller and Friedman, 2009). The resulting BN is also referred to as a *Bayesian decision network* or *influence diagram* (Nyberg et al., 2006).

Instead of manually constructing the DAG, which produced the model in this paper, one may learn it directly from data. That is, given the variables and an existing dataset, a BN-

learning algorithm will attempt to produce the BN that best fits the data, with respect to some scoring function, *e.g.*, log-likelihood (Scutari, 2009; Koller and Friedman, 2009; Ramazi et al., 2021b). Although such data-driven BN structures are a good fit to the data and may suggest connections beyond the state-of-the-art knowledge, they may not make perfect ecological sense, especially if one attempts to read the links causally. Moreover, manually-constructed BNs often suffer less from overfitting. Therefore, depending on the goal, one may use either the data-driven or manually-constructed structure. The first is typically better at making predictions whereas the second may be better for interpreting the already known scientific hypotheses. Clearly, if a manually-constructed BN encompasses accurately all causal relationships and underlying mechanisms, and for each instance to predict, we have values for all of the covariates, then we expect it to outperform other structures, including the data-driven ones. However, this is rarely the case in real world scenarios, motivating the parallel development of machine-learning approaches for prediction purposes.

4.4 Limitations

Although we have collected empirical results confirming the included links, our graph does not yield quantitative results. It describes only which variable affects which other, but does not reveal how. For example, while it claims that water chemistry, river sediment, and nutrient enrichment affect host worm density, it does not provide the exact effecting mechanism. The mechanisms could be modelled either by using machine-learning algorithms Ramazi et al. (accepted, 2021a), based on data, or mechanistically with, for example, logistic regressions or PDEs (Schakau et al., 2019), based on the literature (*cf.* the references that we provided for each node). In particular, *evolutionary game theory* provides a framework to model human behavior in terms of “cooperation” or “defection”, that is, limiting or accelerating disease spread by following or violating decontamination policies Nowak and Sigmund (2005); Govaert et al. (2017, 2021); Riehl et al. (2018).

The presented graph is directed acyclic, and does not model the full life cycle of the

parasite. This is because a DAG does not allow for modelling reciprocal relationships between the variables. For example, high fish density enhances parasite establishment; however, the establishment of the parasite may decrease fish density. We have not modelled the second, reciprocal effect as we have already modelled the first, unreciprocal. Nevertheless, a temporal extension of the model allows capturing these effects. For example, if parasite establishment and fish density are present at years t and $t + 1$ in the graph, then we can link fish density at year t to parasite establishment at year $t + 1$, and similarly, parasite establishment at year t to fish density at year $t + 1$. Such temporal connections lead to the construction of *temporal Bayesian networks* (Koller and Friedman, 2009).

Besides the limitations due to the acyclic structure, DAGs may not easily capture highly dynamic processes on a mechanistic level, and other approaches such as differential equations may be better suited for this task (Schakau et al., 2019). On a time-aggregated and phenomenological level, however, graphical models can yield accurate results (Koller and Friedman, 2009), and the option to include a variety of many environmental variables in the model may offset mechanistic deficiencies.

Finally, we have conducted a thorough literature review to collect empirical knowledge about mechanisms deemed significant by experts and described in the literature. However, it is difficult to oversee the whole diversity of relevant scientific studies, and future research may suggest new or, even different, predictor variables and interactions. Nonetheless, the presented DAG provides a framework that can be extended to accommodate such results in a relatively easy, modular fashion.

Acknowledgments

Thank you to Alberta Environment & Parks (AEP) for funding this research (grant number RES0037031) and providing the data and for having the technical discussions. Thank you to the Greiner and Lewis Research Groups for helpful feedback. M.A.L. is also grateful for

the support through the NSERC Discovery and the Canada Research Chair programs. R.G. is grateful for funding from NSERC Discovery, and Alberta Machine Intelligence Institute.

References

government of british columbia.

Protocols manual for water quality sampling in Canada, 2011. Canadian Council of Ministers of the Environment Winnipeg, Canada.

Freshwater Fisheries Society of BC. <https://www.gofishbc.com/Blog/Conservation/Whirling-Disease.aspx>, 2018.

Alberta Environment and Parks. Whirling disease program report (2017) Edmonton, AB. <https://www.alberta.ca/release.cfm?xID=46262A00CB125-CCE0-5AC7-9815E058B20176B3>, 2017.

Julie D. Alexander, Billie L. Kerans, Todd M. Koel, and Charlotte Rasmussen. Context-specific parasitism in tubifex tubifex in geothermally influenced stream reaches in yellow-stone national park. *Journal of the North American Benthological Society*, 30(3):853–867, 2011.

J David Allan and María M Castillo. *Stream ecology: structure and function of running waters*. Springer Science & Business Media, 2007.

J. David Allan and María M. Ibañez Castillo. *Stream ecology: structure and function of running waters*. Springer, Dordrecht, 2. ed., reprinted edition, 2009. ISBN 978-1-4020-5583-6 978-1-4020-5582-9. OCLC: 837704805.

James E. Anderson. The gravity model. *Annual Review of Economics*, 3 (1):133–160, September 2011. ISSN 1941-1383, 1941-1391. doi: 10.1146/

annurev-economics-111809-125114. URL <http://www.annualreviews.org/doi/10.1146/annurev-economics-111809-125114>.

J.D Armstrong, P.S Kemp, G.J.A Kennedy, M Ladle, and N.J Milner. Habitat requirements of Atlantic salmon and brown trout in rivers and streams. *Fisheries Research*, 62(2): 143–170, May 2003. ISSN 01657836. doi: 10.1016/S0165-7836(02)00160-1. URL <https://linkinghub.elsevier.com/retrieve/pii/S0165783602001601>.

Kimberley Kolb Ayre, Colleen A Caldwell, Jonah Stinson, and Wayne G Landis. Analysis of regional scale risk of whirling disease in populations of colorado and rio grande cutthroat trout using a bayesian belief network model. *Risk Analysis*, 34(9):1589–1605, 2014.

David B. Baker, R. Peter Richards, Timothy T. Loftus, and Jack W. Kramer. A new flashiness index: characteristics and applications to midwestern rivers and streams. *Journal of the American Water Resources Association*, 40(2):503–522, April 2004. ISSN 1093-474X, 1752-1688. doi: 10.1111/j.1752-1688.2004.tb01046.x. URL <http://doi.wiley.com/10.1111/j.1752-1688.2004.tb01046.x>.

Thomas J Baldwin, E Richard Vincent, Ronald M Silflow, and Danielle Stanek. *Myxobolus cerebralis* infection in rainbow trout (*Oncorhynchus mykiss*) and brown trout (*Salmo trutta*) exposed under natural stream conditions. *Journal of Veterinary Diagnostic Investigation*, 12(4):312–321, 2000.

Veillard M. James C.T. Brummelhuis L. Pila E.A. Turnbull A. Oploo A.O. Han X. Barry, D.E. and P.C. Hanington. qpcr-based environmental monitoring of the whirling disease-causing parasite myxobolus cerebralis and phylogenetic analysis of the tubificid hosts in alberta, canada. *Diseases of Aquatic Organisms*, 2021. doi: <https://doi.org/10.3354/dao03608>.

Jerri L Bartholomew, Billie L Kerans, Ronald P Hedrick, Stuart C Macdiarmid, and James R Winton. A risk assessment based approach for the management of whirling disease. *Reviews in Fisheries Science*, 13(4):205–230, 2005a.

- 709 Jerri L. Bartholomew, Billie L. Kerans, Ronald P. Hedrick, Stuart C. Macdiarmid, and
710 James R. Winton. A Risk Assessment Based Approach for the Management of Whirling
711 Disease. *Reviews in Fisheries Science*, 13(4):205–230, October 2005b. ISSN 1064-1262,
712 1547-6553. doi: 10.1080/10641260500326842. URL [https://www.tandfonline.com/doi/
713 full/10.1080/10641260500326842](https://www.tandfonline.com/doi/full/10.1080/10641260500326842).
- 714 Jonathan Ben-David, Stephen D Atkinson, Yulia Pollak, Gilad Yossifon, Uri Shavit, Jerri L
715 Bartholomew, and Tamar Lotan. Myxozoan polar tubules display structural and functional
716 variation. *Parasites & vectors*, 9(1):549, 2016.
- 717 Joseph R Benjamin, Lisa A Wetzel, Kyle D Martens, Kimberly Larsen, and Patrick J Con-
718 nolly. Spatio-temporal variability in movement, age, and growth of mountain whitefish
719 (*Prosopium williamsoni*) in a river network based upon pit tagging and otolith chemistry.
720 *Canadian Journal of Fisheries and Aquatic Sciences*, 71(1):131–140, 2014.
- 721 TCj Bjornn and DW Reiser. Habitat requirements of salmonids in streams. *American*
722 *Fisheries Society Special Publication*, 19(837):138, 1991.
- 723 Julie Blackwood, Alan Hastings, and Christopher Costello. Cost-effective management
724 of invasive species using linear-quadratic control. *Ecological Economics*, 69(3):519–527,
725 January 2010. ISSN 09218009. doi: 10.1016/j.ecolecon.2009.08.029. URL [https:
726 //linkinghub.elsevier.com/retrieve/pii/S0921800909003590](https://linkinghub.elsevier.com/retrieve/pii/S0921800909003590).
- 727 Vicki S Blazer, Thomas B Waldrop, William B Schill, Christine L Densmore, and David
728 Smith. Effects of water temperature and substrate type on spore production and release
729 in eastern *Tubifex tubifex* worms infected with *Myxobolus cerebralis*. *The Journal of para-
730 sitology*, pages 21–26, 2003.
- 731 Michael Bode, Christopher M Baker, Joe Benshemesh, Tim Burnard, Libby Rumpff, Cindy E
732 Hauser, José J Lahoz-Monfort, and Brendan A Wintle. Revealing beliefs: using ensemble

ecosystem modelling to extrapolate expert beliefs to novel ecological scenarios. *Methods in Ecology and Evolution*, 8(8):1012–1021, 2017.

D. B Boorman, J. M Hollis, and A Lilly. Hydrology of soil types: a hydrologically- based classification of the soils of the United Kingdom. Technical Report 126, Institute of Hydrology, Wallingford, UK, 1995. OCLC: 912955904.

Jonathan M. Bossenbroek, Clifford E. Kraft, and Jeffrey C. Nekola. Prediction of long-distance dispersal using gravity models: zebra mussel invasion of inland lakes. *Ecological Applications*, 11(6):1778–1788, December 2001. ISSN 1051-0761. doi: 10.1890/1051-0761(2001)011[1778:POLDDU]2.0.CO;2. URL [http://doi.wiley.com/10.1890/1051-0761\(2001\)011\[1778:POLDDU\]2.0.CO;2](http://doi.wiley.com/10.1890/1051-0761(2001)011[1778:POLDDU]2.0.CO;2).

Jonathan M. Bossenbroek, Ladd E. Johnson, Brett Peters, and David M. Lodge. Forecasting the expansion of zebra mussels in the United States. *Conservation Biology*, 21(3):800–810, June 2007. ISSN 0888-8892, 1523-1739. doi: 10.1111/j.1523-1739.2006.00614.x. URL <http://doi.wiley.com/10.1111/j.1523-1739.2006.00614.x>.

R Morrison Cassie. Elliott, jm 1971. some methods for the statistical analysis of samples of benthic invertebrates. sci. publ. 25. freshwater biol. ass., ambleside, westmorland, uk 144 p.£ 0.95. *Limnology and Oceanography*, 16(3):596–597, 1971.

M Hanif Chaudhry. *Open-channel flow*. Springer Science & Business Media, 2007.

D. T. Crisp. Environmental requirements of common riverine European salmonid fish species in fresh water with particular reference to physical and chemical aspects. *Hydrobiologia*, 323(3):201–221, May 1996. ISSN 0018-8158, 1573-5117. doi: 10.1007/BF00007847. URL <http://link.springer.com/10.1007/BF00007847>.

Ernesto de la Hoz Franco and Phaedra Budy. Linking environmental heterogeneity to the distribution and prevalence of *Myxobolus cerebralis*: a comparison across sites in a northern utah watershed. *Transactions of the American Fisheries Society*, 133(5):1176–1189, 2004.

- 758 Daniel C Downing, Thomas E McMahon, Billie L Kerans, and E Richard Vincent. Relation
759 of spawning and rearing life history of rainbow trout and susceptibility to myxobolus
760 cerebralis infection in the madison river, montana. *Journal of Aquatic Animal Health*, 14
761 (3):191–203, 2002.
- 762 D. Andrew R. Drake and Nicholas E. Mandrak. Bycatch, bait, anglers, and roads: quan-
763 tifying vector activity and propagule introduction risk across lake ecosystems. *Ecological*
764 *Applications*, 24(4):877–894, June 2014. ISSN 1051-0761. doi: 10.1890/13-0541.1. URL
765 <http://doi.wiley.com/10.1890/13-0541.1>.
- 766 Robert J DuBey, Colleen A Caldwell, and William R Gould. Relative susceptibility and
767 effects on performance of rio grande cutthroat trout and rainbow trout challenged with
768 *Myxobolus cerebralis*. *Transactions of the American Fisheries Society*, 136(5):1406–1414,
769 2007.
- 770 Anna Eklöf, Si Tang, and Stefano Allesina. Secondary extinctions in food webs: a bayesian
771 network approach. *Methods in Ecology and Evolution*, 4(8):760–770, 2013.
- 772 M El-Matbouli and RW Hoffmann. Effects of freezing, aging, and passage through the
773 alimentary canal of predatory animals on the viability of *Myxobolus cerebralis* spores.
774 *Journal of Aquatic Animal Health*, 3(4):260–262, 1991.
- 775 M El-Matbouli, TS McDowell, DB Antonio, KB Andree, and RP Hedrick. Effect of wa-
776 ter temperature on the development, release and survival of the triactinomyxon stage of
777 *Myxobolus cerebralis* in its oligochaete host. *International Journal for Parasitology*, 29(4):
778 627–641, 1999.
- 779 Leah C Steinbach Elwell, Kajsa Eagle Stromberg, Eileen KN Ryce, and Jerri L Bartholomew.
780 Whirling disease in the united states: A summary of progress in research and management.
781 In *Proceedings of the Wild Trout X Symposium, West Yellowstone, Montana*, page 203,
782 2010.

Rebecca S. Epanchin-Niell and James E. Wilen. Optimal spatial control of biological invasions. *Journal of Environmental Economics and Management*, 63(2):260–270, March 2012. ISSN 00950696. doi: 10.1016/j.jeem.2011.10.003. URL <https://linkinghub.elsevier.com/retrieve/pii/S0095069611001392>.

Sven Erlander and Neil F Stewart. *The gravity model in transportation analysis: theory and extensions*, volume 3. Vsp, 1990.

David Finnoff, Alexei Potapov, and Mark A. Lewis. Control and the management of a spreading invader. *Resource and Energy Economics*, 32(4):534–550, November 2010. ISSN 09287655. doi: 10.1016/j.reseneeco.2010.04.003. URL <http://linkinghub.elsevier.com/retrieve/pii/S0928765510000230>.

Samuel M. Fischer, Martina Beck, Leif-Matthias Herborg, and Mark A. Lewis. A hybrid gravity and route choice model to assess vector traffic in large-scale road networks. *Royal Society Open Science*, pages 1–26, May 2020.

Fondriest Environmental Inc. Conductivity, Salinity & Total Dissolved Solids, March 2014. URL <https://www.fondriest.com/environmental-measurements/parameters/water-quality/conductivity-salinity-tds/>. Library Catalog: www.fondriest.com.

Nikolaos Fytilis, Donna M Rizzo, Ryan D Lamb, Billie L Kerans, and Lori Stevens. Using real-time pcr and bayesian analysis to distinguish susceptible tubificid taxa important in the transmission of *myxobolus cerebralis*, the cause of salmonid whirling disease. *International Journal for Parasitology*, 43(6):493–501, 2013.

Bella S. Galil, Stefan Nehring, and Vadim Panov. Waterways as Invasion Highways – Impact of Climate Change and Globalization. In Wolfgang Nentwig, editor, *Biological Invasions*, volume 193, pages 59–74. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. ISBN 978-3-540-77375-7 978-3-540-36920-2. doi: 10.1007/978-3-540-36920-2_5. URL http://link.springer.com/10.1007/978-3-540-36920-2_5. Series Title: Ecological Studies.

- 808 Kiza Kristine Gates and others. *Myxospore detection in soil and angler movement in south-*
809 *western Montana: implications for whirling disease transport*. PhD Thesis, Montana State
810 University-Bozeman, College of Letters & Science, 2007.
- 811 Ph Gauckler. *Etudes Théoriques et Pratiques sur l'Ecoulement et le Mouvement des Eaux*.
812 Gauthier-Villars, 1867.
- 813 Michael A Gilbert and Willard O Granath Jr. Persistent infection of myxobolus cerebralis,
814 the causative agent of salmonid whirling disease, in tubifex tubifex. *Journal of Parasitology*,
815 87(1):101–107, 2001.
- 816 Michael A Gilbert and Willard O Granath Jr. Whirling disease of salmonid fish: life cycle,
817 biology, and disease. *Journal of Parasitology*, 89(4):658–667, 2003.
- 818 Nancy D Gordon, Thomas A McMahon, Brian L Finlayson, Christopher J Gippel, and Rory J
819 Nathan. *Stream hydrology: an introduction for ecologists*. John Wiley and Sons, 2004.
- 820 Alain Govaert, Pouria Ramazi, and Ming Cao. Convergence of imitation dynamics for public
821 goods games on networks. In *2017 IEEE 56th Annual Conference on Decision and Control*
822 *(CDC)*, pages 4982–4987. IEEE, 2017.
- 823 Alain Govaert, Pouria Ramazi, and Ming Cao. Rationality, imitation and rational imitation
824 in spatial public goods games. *IEEE Transactions on Control of Network Systems*, 2021.
- 825 Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral
826 methods. *Econometrica: journal of the Econometric Society*, pages 424–438, 1969.
- 827 Sander Greenland, Judea Pearl, and James M Robins. Causal diagrams for epidemiologic
828 research. *Epidemiology*, pages 37–48, 1999.
- 829 S.M. Greig, D.A. Sear, and P.A. Carling. The impact of fine sediment accumulation on
830 the survival of incubating salmon progeny: Implications for sediment management. *Sci-*
831 *ence of The Total Environment*, 344(1-3):241–258, May 2005. ISSN 00489697. doi:

10.1016/j.scitotenv.2005.02.010. URL <https://linkinghub.elsevier.com/retrieve/pii/S004896970500118X>.

Sascha L Hallett and Jerri L Bartholomew. Effects of water flow on the infection dynamics of *Myxobolus cerebralis*. *Parasitology*, 135(3):371–384, 2008.

MM Halliday. The biology of *myxosoma cerebralis*: the causative organism of whirling disease of salmonids. *Journal of Fish Biology*, 9(4):339–357, 1976.

Ronald P Hedrick and Mansour El-Matbouli. Recent advances with taxonomy, life cycle, and development of *Myxobolus cerebralis* in the fish and oligochaete hosts. In *American Fisheries Society Symposium*, pages 45–54. American Fisheries Society, 2002.

Ronald P Hedrick, Terry S McDowell, Kaveramma Mukkatira, Marios P Georgiadis, and Elizabeth MacConnell. Susceptibility of selected inland salmonids to experimentally induced infections with myxobolus cerebralis, the causative agent of whirling disease. *Journal of Aquatic Animal Health*, 11(4):330–339, 1999.

Ronald P Hedrick, Terry S McDowell, Kaveramma Mukkatira, Marios P Georgiadis, and Elizabeth MacConnell. Salmonids resistant to ceratomyxa shasta are susceptible to experimentally induced infections with myxobolus cerebralis. *Journal of Aquatic Animal Health*, 13(1):35–42, 2001a.

Ronald P Hedrick, Terry S McDowell, Kaveramma Mukkatira, Marios P Georgiadis, and Elizabeth MacConnell. Susceptibility of three species of anadromous salmonids to experimentally induced infections with myxobolus cerebralis, the causative agent of whirling disease. *Journal of Aquatic Animal Health*, 13(1):43–50, 2001b.

Ronald P Hedrick, Terry S McDowell, Gary D Marty, Geoffrey T Fosgate, Kaveramma Mukkatira, Karin Myklebust, and Mansour El-Matbouli. Susceptibility of two strains of rainbow trout (one with suspected resistance to whirling disease) to myxobolus cerebralis infection. *Diseases of Aquatic Organisms*, 55(1):37–44, 2003.

- 857 Ronald P Hedrick, Terry S McDowell, Kaveramma Mukkatira, Elizabeth MacConnell, and
858 Brian Petri. Effects of freezing, drying, ultraviolet irradiation, chlorine, and quaternary
859 ammonium treatments on the infectivity of myxospores of *Myxobolus cerebralis* for *Tubifex*
860 *tubifex*. *Journal of Aquatic Animal Health*, 20(2):116–125, 2008.
- 861 J. M. Hellawell. The Effects of Organic Enrichment. In *Biological Indicators of Freshwa-*
862 *ter Pollution and Environmental Management*, pages 155–211. Springer Netherlands, Dor-
863 drecht, 1986. ISBN 978-94-010-8417-8 978-94-009-4315-5. doi: 10.1007/978-94-009-4315-5_
864 6. URL http://link.springer.com/10.1007/978-94-009-4315-5_6.
- 865 Carlie E Herring, Jonah Stinson, and Wayne G Landis. Evaluating nonindigenous species
866 management in a bayesian networks derived relative risk framework for padilla bay, wa,
867 usa. *Integrated environmental assessment and management*, 11(4):640–652, 2015.
- 868 B Hofer. Über die drehkrankheit der regenbogenforelle. *Allgemeine Fischerei-Zeitung*, 28(1):
869 7–8, 1903.
- 870 Glenn L Hoffman and Robert E Putz. Host susceptibility and the effect of aging, freezing,
871 heat, and chemicals on spores of myxosoma cerebralis. *The Progressive Fish-Culturist*, 31
872 (1):35–37, 1969.
- 873 Ladislav Holko, Juraj Parajka, Zdeno Kostka, Peter Škoda, and Günter Blöschl. Flashiness of
874 mountain streams in Slovakia and Austria. *Journal of Hydrology*, 405(3-4):392–401, August
875 2011. ISSN 00221694. doi: 10.1016/j.jhydrol.2011.05.038. URL <https://linkinghub.elsevier.com/retrieve/pii/S0022169411003660>.
- 877 K. Ikuta, Y. Suzuki, and S. Kitamura. Effects of low pH on the reproductive behavior of
878 salmonid fishes. *Fish Physiology and Biochemistry*, 28(1-4):407–410, 2003. ISSN 0920-
879 1742. doi: 10.1023/B:FISH.0000030607.75565.74. URL <http://link.springer.com/10.1023/B:FISH.0000030607.75565.74>.
- 880

C.T. James, M.F. Veillard, A.M. Martens, E.A. Pila, A. Turnbull, P. Hanington, A. Luek, J. Alexander, and R.B. Nehring. Whirling disease in the Crowsnest River: An emerging threat to wild salmonids in Alberta. *Canadian Journal of Fisheries and Aquatic Sciences*, in press.

Adam J Kaeser and William E Sharpe. Patterns of distribution and abundance of *Tubifex tubifex* and other aquatic oligochaetes in *Myxobolus cerebralis* enzootic areas in pennsylvania. *Journal of Aquatic Animal Health*, 18(1):64–78, 2006a.

Adam J. Kaeser and William E. Sharpe. Patterns of Distribution and Abundance of *Tubifex tubifex* and Other Aquatic Oligochaetes in *Myxobolus cerebralis* Enzootic Areas in Pennsylvania. *Journal of Aquatic Animal Health*, 18(1):64–78, March 2006b. ISSN 0899-7659, 1548-8667. doi: 10.1577/H05-016.1. URL <http://doi.wiley.com/10.1577/H05-016.1>.

Adam J. Kaeser and William E. Sharpe. The Ecology of *Tubifex tubifex* in two *Myxobolus cerebralis* Enzootic Streams in Pennsylvania. *Journal of Freshwater Ecology*, 23(4):575–588, December 2008. ISSN 0270-5060, 2156-6941. doi: 10.1080/02705060.2008.9664245. URL <http://www.tandfonline.com/doi/abs/10.1080/02705060.2008.9664245>.

Dennis M Kallert, Edit Eszterbauer, Daniel Grabner, and Mansour El-Matbouli. In vivo exposure of susceptible and non-susceptible fish species to *Myxobolus cerebralis* actinospores reveals non-specific invasion behaviour. *Diseases of aquatic organisms*, 84(2):123–130, 2009.

Billie L Kerans and Alexander V Zale. The ecology of myxobolus cerebralis. In *American Fisheries Society Symposium*, pages 145–166. American Fisheries Society, 2002.

Billie L Kerans, Richard I Stevens, and JC Lemmon. Water temperature affects a host–parasite interaction: *Tubifex tubifex* and *Myxobolus cerebralis*. *Journal of Aquatic Animal Health*, 17(3):216–221, 2005.

Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT Press, 2009.

- 906 Rebecca C Krueger, BL Kerans, E Richard Vincent, and Charlotte Rasmussen. Risk of
907 myxobolus cerebralis infection to rainbow trout in the madison river, montana, usa. *Eco-*
908 *logical Applications*, 16(2):770–783, 2006.
- 909 Rebecca Caroline Krueger. *Correlations among environmental features, Myxobolus cerebralis*
910 *infection prevalence in oligochaetes, and salmonid infection risk in the Madison River,*
911 *Montana*. PhD thesis, Montana State University-Bozeman, College of Letters & Science,
912 2002.
- 913 M. N. Lazim and M. A. Learner. The influence of sediment composition and leaf litter on
914 the distribution of tubificid worms (Oligochaeta): A field and laboratory study. *Oecologia*,
915 72(1):131–136, April 1987. ISSN 0029-8549, 1432-1939. doi: 10.1007/BF00385057. URL
916 <http://link.springer.com/10.1007/BF00385057>.
- 917 Luna Bergere Leopold. *A View of the River*. Harvard University Press, 1994.
- 918 Brian Leung, John M. Drake, and David M. Lodge. Predicting invasions: Propagule pressure
919 and the gravity of Allee effects. *Ecology*, 85(6):1651–1660, June 2004. ISSN 0012-9658.
920 doi: 10.1890/02-0571. URL <http://www.esajournals.org/doi/abs/10.1890/02-0571>.
- 921 Halcyon J Lukins, Alexander V Zale, and Frederic T Barrows. A packed-bed filtration system
922 for collection of myxobolus cerebralis triactinomyxons. *Journal of aquatic animal health*,
923 19(4):234–241, 2007.
- 924 Robert Manning, John Purser Griffith, T F Pigot, and Leveson Francis Vernon-Harcourt.
925 *On the flow of water in open channels and pipes*. 1890.
- 926 Bruce G Marcot, J Douglas Steventon, Glenn D Sutherland, and Robert K McCann. Guide-
927 lines for developing and updating bayesian belief networks applied to ecological modeling
928 and conservation. *Canadian Journal of Forest Research*, 36(12):3063–3074, 2006.
- 929 L. Mari, E. Bertuzzo, R. Casagrandi, M. Gatto, S. A. Levin, I. Rodriguez-Iturbe, and A. Ri-
930 naldo. Hydrologic controls and anthropogenic drivers of the zebra mussel invasion of the

Mississippi-Missouri river system. *Water Resources Research*, 47(3):1–16, March 2011.
ISSN 00431397. doi: 10.1029/2010WR009920. URL <http://doi.wiley.com/10.1029/2010WR009920>.

Maria E Markiw. Experimentally induced whirling disease ii. determination of longevity of the infective triactinomyxon stage of *Myxobolus cerebralis* by vital staining. *Journal of Aquatic Animal Health*, 4(1):44–47, 1992.

Stephanie McGinnis and Billie L Kerans. Land use and host community characteristics as predictors of disease risk. *Landscape Ecology*, 28(1):29–44, 2013.

John Modin. Whirling disease in California: a review of its history, distribution, and impacts, 1965–1997. *Journal of Aquatic Animal Health*, 10(2):132–142, 1998. Publisher: Wiley Online Library.

Jim R. Muirhead and Hugh J. MacIsaac. Evaluation of stochastic gravity model selection for use in estimating non-indigenous species dispersal and establishment. *Biological Invasions*, 13(11):2445–2458, November 2011. ISSN 1387-3547, 1573-1464. doi: 10.1007/s10530-011-0070-3. URL <http://link.springer.com/10.1007/s10530-011-0070-3>.

R B Nehring and K. G. Thompson. *Whirling Disease Investigations: Federal Aid Project F-237-R10*. Colorado Division of Wildlife Job Progress Report, Fort Collins, CO, 2003a.

R Barry Nehring and Kevin G Thompson. Whirling disease risk assessment: the colorado perspective. In *Proceedings of the Whirling Disease Symposium*, pages 31–32. Citeseer, 2003b.

R Barry Nehring, Kevin G Thompson, Karen A Taurman, and David L Shuler. Laboratory studies indicating that living brown trout *Salmo trutta* expel viable *Myxobolus cerebralis* myxospores. In *American Fisheries Society Symposium*, pages 125–134. American Fisheries Society, 2002.

- 955 R Barry Nehring, Kevin G Thompson, David L Shuler, and Terence M James. Using sedi-
956 ment core samples to examine the spatial distribution of *Myxobolus cerebralis* actinospore
957 production in windy gap reservoir, colorado. *North American Journal of Fisheries Man-*
958 *agement*, 23(2):376–384, 2003.
- 959 T G Northcote. Potamodromy in salmonidae—living and moving in the fast lane. *North*
960 *American Journal of Fisheries Management*, 17(4):1029–1045, 1997.
- 961 Martin A Nowak and Karl Sigmund. Evolution of indirect reciprocity. *Nature*, 437(7063):
962 1291–1298, 2005.
- 963 J Brian Nyberg, Bruce G Marcot, and Randy Sulyma. Using bayesian belief networks in
964 adaptive management. *Canadian Journal of Forest Research*, 36(12):3104–3116, 2006.
- 965 Joseph J O’Grodnick. Susceptibility of various salmonids to whirling disease (myxosoma
966 *cerebralis*). *Transactions of the American Fisheries Society*, 108(2):187–190, 1979.
- 967 Vadim E Panov, Boris Alexandrov, Kestutis Arbačiauskas, Rosa Binimelis, Gordon H Copp,
968 Michal Grabowski, Frances Lucy, Rob SEW Leuven, Stefan Nehring, Momir Paunović,
969 Vitaliy Semenchenko, and Mikhail O Son. Assessing the Risks of Aquatic Species In-
970 vasions via European Inland Waterways: From Concepts to Environmental Indicators.
971 *Integrated Environmental Assessment and Management*, 5(1):110, 2009. ISSN 1551-3777,
972 1551-3793. doi: 10.1897/IEAM{_}2008-034.1. URL [http://doi.wiley.com/10.1897/](http://doi.wiley.com/10.1897/IEAM_2008-034.1)
973 [IEAM_2008-034.1](http://doi.wiley.com/10.1897/IEAM_2008-034.1).
- 974 Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*.
975 Morgan Kaufmann, 1988.
- 976 Judea Pearl. *Causality*. Cambridge university press, 2009.
- 977 Judea Pearl and Dana Mackenzie. *The book of why: the new science of cause and effect*.
978 Basic books, 2018.

Alex Potapov, Jim R. Muirhead, Subhash R. Lele, and Mark A. Lewis. Stochastic gravity models for modeling lake invasions. *Ecological Modelling*, 222(4):964–972, July 2010. ISSN 03043800. doi: 10.1016/j.ecolmodel.2010.07.024. URL <http://linkinghub.elsevier.com/retrieve/pii/S0304380010003820>.

Alex Potapov, Jim R Muirhead, Subhash R Lele, and Mark A Lewis. Stochastic gravity models for modeling lake invasions. *Ecological Modelling*, 222(4):964–972, 2011.

Pouria Ramazi, Arezoo Haratian, Maryam Meghdadi, Arash Mari Oriyad, Mark A Lewis, Zeinab Maleki, Roberto Vega, Hao Wang, David S Wishart, and Russell Greiner. Accurate long-range forecasting of COVID-19 mortality in the USA. *Scientific Reports*, 11(1):1–11, 2021a.

Pouria Ramazi, Mélodie Kunegel-Lion, Russell Greiner, and Mark A Lewis. Exploiting the full potential of bayesian networks in predictive ecology. *Methods in Ecology and Evolution*, 12(1):135–149, 2021b.

Pouria Ramazi, Mélodie Kunegel-Lion, Russell Greiner, and Mark A. Lewis. Predicting insect outbreaks using machine learning: A mountain pine beetle case study. *Ecology and Evolution*, accepted.

James Riehl, Pouria Ramazi, and Ming Cao. A survey on the analysis and control of evolutionary matrix games. *Annual Reviews in Control*, 45:87–106, 2018.

A. H. Roy, M. C. Freeman, B. J. Freeman, S. J. Wenger, W. E. Ensign, and J. L. Meyer. Investigating hydrologic alteration as a mechanism of fish assemblage shifts in urbanizing streams. *Journal of the North American Benthological Society*, 24(3):656–678, September 2005. ISSN 0887-3593, 1937-237X. doi: 10.1899/04-022.1. URL <https://www.journals.uchicago.edu/doi/10.1899/04-022.1>.

Eileen KN Ryce, Alexander V Zale, and Elizabeth MacConnell. Effects of fish age and

- 1003 parasite dose on the development of whirling disease in rainbow trout. *Diseases of Aquatic*
1004 *Organisms*, 59(3):225–233, 2004.
- 1005 Eileen KN Ryce, Alexander V Zale, Elizabeth MacConnell, and Mark Nelson. Effects of
1006 fish age versus size on the development of whirling disease in rainbow trout. *Diseases of*
1007 *Aquatic Organisms*, 63(1):69–76, 2005.
- 1008 Todd A Sandell, Harriet V Lorz, Donald G Stevens, and Jerri L Bartholomew. Dynamics of
1009 myxobolus cerebralis in the lostine river, oregon: implications for resident and anadromous
1010 salmonids. *Journal of Aquatic Animal Health*, 13(2):142–150, 2001.
- 1011 Georg Sauter and Hans Güde. Influence of grain size on the distribution of tubificid
1012 oligochaete species. *Hydrobiologia*, 334(1-3):97–101, October 1996. ISSN 0018-8158,
1013 1573-5117. doi: 10.1007/BF00017358. URL [http://link.springer.com/10.1007/](http://link.springer.com/10.1007/BF00017358)
1014 [BF00017358](http://link.springer.com/10.1007/BF00017358).
- 1015 Vanessa Schakau, Frank M Hilker, and Mark A Lewis. Fish disease dynamics in changing
1016 rivers: Salmonid ceratomyxosis in the klamath river. *Ecological Complexity*, 40:100776,
1017 2019.
- 1018 George Schisler and E.P. Bergersen. Evaluation of risk of high elevation colorado waters to
1019 the establishment of *Myxobolus cerebralis*. *American Fisheries Society Symposium*, pages
1020 33–41, 2002.
- 1021 George J Schisler, Eric P Bergersen, and Peter G Walker. Effects of multiple stressors on
1022 morbidity and mortality of fingerling rainbow trout infected with *Myxobolus cerebralis*.
1023 *Transactions of the American Fisheries Society*, 129(3):859–865, 2000.
- 1024 Philip J Schoeneberger, Douglas A Wysocki, Ellis C Benham, and Soil Survey Staff. *Field*
1025 *book for describing and sampling soils*. Natural Resources Conservation Service, National
1026 Soil Survey Center, Lincoln, NE, 2012.

- 1027 Stanley A. Schumm, Dale F. Ritter, George Harry Dury, and Lawrence K. Lustig. River -
1028 Sediment yield and sediment load, February 2020. URL [https://www.britannica.com/
1029 science/river](https://www.britannica.com/science/river). Library Catalog: www.britannica.com.
- 1030 WB Scott and EJ Crossman. Fisheries research board of canada bulletin no. 184, fresh water
1031 fishes of canada. *Fish. Res. Board Can. Bull*, 966, 1973.
- 1032 Marco Scutari. Learning bayesian networks with the bnlearn r package. *arXiv preprint
1033 arXiv:0908.3817*, 2009.
- 1034 M A Shahrokhnia, Ali Reza Sepaskhah, and M Javan. Estimation of hydraulic parameters for
1035 karoon river by cokriging and residual kriging. *Iranian Journal of Science & Technology,
1036 Transaction B*, 28(B1):153–163, 2004.
- 1037 Sho Shirakashi and Mansour El-Matbouli. Effect of cadmium on the susceptibility of *Tubifex
1038 tubifex* to *Myxobolus cerebralis* (myxozoa), the causative agent of whirling disease. *Diseases
1039 of aquatic organisms*, 89(1):63–70, 2010.
- 1040 John W Sigler, TC Bjornn, and Fred H Everest. Effects of chronic turbidity on density and
1041 growth of steelheads and coho salmon. *Transactions of the American Fisheries Society*,
1042 113(2):142–150, 1984. Publisher: Wiley Online Library.
- 1043 MARK A Smith, Eric J Wagner, and AMy HowA. The effect of water characteristics on via-
1044 bility of the *Myxobolus cerebralis* actinospore. In *American Fisheries Society Symposium*,
1045 pages 227–238. American Fisheries Society, 2002.
- 1046 Robert L. Taylor and Marley Lott. Transmission of salmonid whirling disease by birds fed
1047 trout infected with *Myxosoma cerebralis*. *The Journal of protozoology*, 25(1):105–106,
1048 1978.
- 1049 Kevin G Thompson and R Barry Nehring. A simple technique used to filter and quantify the
1050 actinospore of *Myxobolus cerebralis* and determine its seasonal abundance in the Colorado
1051 River. *Journal of Aquatic Animal Health*, 12(4):316–323, 2000.

- 1052 Mustapha Touazi, Jean Pierre Laborde, and Najat Bhiry. Modelling rainfall-discharge at a
1053 mean inter-yearly scale in northern algeria. *Journal of Hydrology*, 296(1-4):179–191, 2004.
- 1054 Kimbra G Turner, Matthew J Smith, and Benjamin J Ridenhour. Whirling disease dynamics:
1055 an analysis of intervention strategies. *Preventive veterinary medicine*, 113(4):457–468,
1056 2014.
- 1057 E.J. Wagner. Whirling disease prevention, control, and management: A review. *American*
1058 *Fisheries Society Symposium*, 2002:217–225, 01 2002.
- 1059 Eric J Wagner, Mark Smith, Ronney Arndt, and Donald W Roberts. Physical and chem-
1060 ical effects on viability of the myxobolus cerebralis triactinomyxon. *Diseases of aquatic*
1061 *organisms*, 53(2):133–142, 2003.
- 1062 M. Warren, M. J. Dunbar, and C. Smith. River flow as a determinant of salmonid
1063 distribution and abundance: a review. *Environmental Biology of Fishes*, 98(6):1695–
1064 1717, June 2015. ISSN 0378-1909, 1573-5133. doi: 10.1007/s10641-015-0376-6. URL
1065 <http://link.springer.com/10.1007/s10641-015-0376-6>.
- 1066 Sue White. Sediment yield prediction and modelling. *Hydrological Processes*, 19(15):3053–
1067 3057, October 2005. ISSN 0885-6087, 1099-1085. doi: 10.1002/hyp.6003. URL <http://doi.wiley.com/10.1002/hyp.6003>.
- 1069 Peter J. Whiting. Flow measurement and characterization. In G. Mathias Kondolf and
1070 Hervé Piégay, editors, *Tools in Fluvial Geomorphology*, pages 260–277. Wiley, 1 edition,
1071 April 2016. ISBN 978-0-470-68405-4 978-1-118-64855-1. doi: 10.1002/9781118648551.ch12.
1072 URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118648551.ch12>.
- 1073 Ken Wolf and Maria E Markiw. Biology contravenes taxonomy in the myxozoa: new discov-
1074 eries show alternation of invertebrate and vertebrate hosts. *Science*, 225(4669):1449–1452,
1075 1984.

1076 Paul Pao-Yen Wu, Kathryn McMahon, Michael A Rasheed, Gary A Kendrick, Paul H York,
1077 Kathryn Chartrand, M Julian Caley, and Kerrie Mengersen. Managing seagrass resilience
1078 under cumulative dredging affecting light: Predicting risk using dynamic bayesian net-
1079 works. *Journal of Applied Ecology*, 55(3):1339–1350, 2018.

Draft