2.0 Data Mining and Tools for Desktop Analysis of Risks

Abstract. Section 2 summarizes the tools applied to this desk-top analysis of risks associated with interbasin biota transfers regardless of the roles played by potential diversions of Missouri River source waters to Red River basin. Predicated on the outcomes of problem formulation, e.g., conceptual models, measurement and assessment endpoints were characterized and linked to the primary tool—habitat equivalency analysis—for evaluating consequences. Terminology critical to the evaluation of risks of biota transfer was characterized, including project-specific definitions for "introduction," "native," "alien," and "invasive." Data-mining techniques were applied to open literature searches initiated for compiling existing data and information on biota of concern. Outcomes of those searches are detailed and summarized. Potential pathways directly associated with engineered interbasin water diversions were considered as one of many competing pathways linked to human device(s) or natural events (i.e., those not linked to anthropogenic activities). A series of nested fault-probability trees (FPTs) were built to graphically illustrate the biota transfer process potentially captured by interbasin water diversions and competing pathways linked to anthropogenic or natural (not aided by human devices or activity) processes. Tools applied to the analysis of risks are characterized, including categorical and spatiotemporal tools employing traditional dot maps to characterize current distributions of biota of concern, and genetic algorithms focused on ecological-niche models to project potential distributions for these species.

2.0 Tools and Data-Mining Overview

Risk analysis may be conducted with various levels of effort and tap into a variety of data and information available to the analyst. From a technical perspective, there are three varieties of risk analysis that potentially serve natural resource managers in their efforts to manage risks in the face of various levels of uncertainty (see ASTM 2004; EPA 2003, 1998, 1992; Foran and Ferenc 1999). While terminology varies from author to author, the analysis of risks can be implemented through (1) "desktop" efforts reliant on available information (e.g., open-source, peer-reviewed technical literature) and existing data sources, (2) screening efforts that are implemented along a spectrum of designed or observational studies, and (3) comprehensive efforts that are generally phased interrelated studies resulting from previously completed desktop and screening level efforts. While these categories may be conventionally characterized as discrete forms, a desktop analysis may be implemented with various levels of effort, ranging from a "preassessment" activity that provides analysis and commentary sufficient to support decisions regarding the need for further study (e.g., NOAA 1997) to comprehensive studies that are variously implemented as data mining or integrated field-laboratory efforts involving designed studies to address environmental or engineering issues (see Downes et al. 2002; Doppelt et al. 1993; Margoluis and Salafsky 1998).

The current investigation is a focused desktop-level analysis that was designed to address questions fielded from Reclamation and Technical Team and reflected in the regional conceptual model in Section 1 (see Section 1, Figure 3). As a desktop analysis of risks, the current work was implemented through a comprehensive literature survey targeted on the list of biota of concern (Table 1) developed in collaboration with Reclamation and Technical Team. The literature survey yielded existing information, largely peer-reviewed literature and data compilations (see Appendix 3A and Appendix 3B), that was evaluated using available data analysis tools as briefly summarized in the following text and Appendix 4.

2.1 Terminology and Identification of Biota of Concern

For purposes of the present study focused on biological invasions potentially associated with interbasin water transfers, definitions of terms were critical to the analysis. Four terms in particular must be clearly understood: introduction, native species, alien species, and invasive species. We have followed the terminology of the National Invasive Species Council (NISC) as specified in Executive Order 13112 in this report (Office of the President, Executive Order 13112, 1999).

An “introduction” means the intentional or unintentional escape, release, dissemination, or placement of a species into an ecosystem as a result of human activity. “Native species” are those that, other than as a result of an introduction, historically occurred or currently occurs in a specific region. An “alien species” means, with respect to a particular ecosystem, any species, including its seeds, eggs, spores, or other biological material capable of propagating that species that is not native to that ecosystem. In contrast, our definition of “invasive species” follows as an alien species whose introduction does or is likely to cause economic or environmental harm or harm to human health. Although not alien or invasive species, a limited focus of the current analysis also fell on biota transfers between regions that merely reflected a movement of species across basin boundaries; that is, the species presently occurs in each region but the interbasin transfer of water expedites movements of subpopulations between regions. Although not invasions by definition, biota of concern in the present analysis included selected species that are present in each basin regardless of population levels and current distributions, and are potentially associated with adverse impacts on the receiving system.

In identifying biota of concern, species lists for Upper Missouri River and Red River basins were initially referenced with respect to the compilation of candidate species of concern. Then, given Technical Team input, species selected as biota of concern (Section 1, Table 1) were characterized as being widely characterized as invasive species (see http://www.invasivespecies.gov/; http://anstaskforce.gov/; http://www.nrcs.usda.gov/technical/invasive.html; http://plants.usda.gov/) and were considered as species likely to emigrate from Upper Missouri River basin to Red River basin. In addition to those invasive species identified as representatives of a larger pool of candidate species, selected species occurring in both Missouri River and Red River basins were also included as biota of concern, because interbasin water transfers were considered potential links between basins that might be associated with shifts in metapopulations, resulting in an increased occurrence of a species in the receiving basin, e.g., waterborne disease agents could potentially be transferred consequent to interbasin water diversion (see Embrey et al. 2002; Greenblatt and Spigelman 2003; http://www.phac-aspc.gc.ca/new_e.html; http://www.cdc.gov/mmwr/distrnds.html; http://wildfishsurvey.fws.gov/). Within the context of risk analysis and the process of developing a technical analysis suited to the risk-management needs of stakeholders, the selection of species as biota of concern was influenced by the emigrant's attributes of invasiveness (e.g., see Elton 1958; Ehrlich 1976; FAO 2002 for overview; Johnson and Carlton 1996; Moyle and Light 1996; Ricciardi and Rasmussen 1998; Kolar and Lodge 2002 for aquatic nuisance species; Goodwin et al 1998; Higgins et al. 1999; NRC 2002; Rejmánek and Richardson 1996; Rejmánek 2000 for invasive plants) and its role as a "representaive species" in the crafting of a risk characterization focused on an analysis that must be generalized to anticipate invasions by species that present unique species signatures, yet possess attributes that are common to invasive species as a general category of species of concern.

In combination with these attributes of invasiveness, inclusion of a species as biota of concern also focused on the likely ecological receptors, or those species in Red River basin, that would likely be adversely affected if a given species invaded from the Upper Missouri River basin. Any given biota of concern may adversely impact single or multiple species in their role as ecological receptors. The species identified as "likely to emigrate" were identified as biota of concern and their selection was influenced by those species "likely to be adversely affected" as ecological receptors adversely impacted as a consequence of invasion, e.g., unionid mussels were ecological receptors likely to be adversely affected by invasions by zebra mussel. Criteria for inclusion on this list of biota of concern were:

- Organisms identified as biota of concern were included on lists of invasive species previously compiled by the NISC or similar organizations (Invasive Species Specialist Groups, ISSG of the Global Invasive Species Program, GISP), with a particular emphasis on lists having geographic ties to the northern Great Plains and Great Lakes regions.
- Organisms identified as biota of concern were cited by Centers for Disease Control and Prevention and resource management agencies (e.g., US Fish & Wildlife Service, state agencies) as causative agents of waterborne disease in the states of the northern Great Plains and adjacent areas in both United States and Canada.
- Organisms identified as biota of concern have been reported as disease-causing agents in fish, wildlife, or domestic livestock and are directly linked to surface water pathways.
- Initial literature surveys suggested that these species might be supported by data previously published or available for analysis, which would enhance their role as species representative of the broader general focus on the wide range of invasive species that might be critical to evaluations of risks associated with biota transfers in the future.

In our current application, ecological receptors in the Red River basin were considered as those native species most likely adversely affected by a successful biological invasion. For example, zebra mussel (e.g., Johnson and Padilla 1996, Johnson and Carlton 1996, Johnson et al. 2001) are well characterized with respect to their competitive advantage over native bivalves. Similarly, the displacement of native species of mollusks and other aquatic invertebrates is clearly indicated subsequent to invasions by the New Zealand mudsnail (e.g., Richards et al. 2001; see http://www.esg.montana.edu/aim/mollusca/nzms/), and invasions by spiny water flea (e.g., Yan et al. 2001) have adverse effects on ecological receptors in the Great Lakes region where the species has adversely affected aquatic invertebrates and fishes.

Emerging diseases, although not a primary focus of the current effort, do potentially represent species invasions or exchanges of disease agents between the Upper Missouri River and Red River basins. For example, an emerging disease that would represent a species invasion for the Red River basin would be whirling disease caused by *Myxosoma cerebralis* (*Myxobolus cerebralis*). Salmonids (especially, strains of rainbow trout, *Oncorhynchus mykiss*) are potential ecological receptors adversely affected by a successful invasion of *M. cerebralis* (Noga 1996; Bartholomew and Wilson 2002). Similarly, while not invasive species by our definition, disease agents of fish, wildlife, and human populations represent biota potentially subject to interbasin transfer through water diversions from the Missouri River to the Red River basins (e.g., serotypes of *Escherichia coli* and various species of *Salmonella* as those relate to waterborne disease). Representative biota listed in Table 1 have been considered within the current investigations purview at the request of Reclamation and Technical Team, and screening of candidate species was limited. As detailed in the analysis of risks (Section 3) and the characterization of risks (Section 4), the economic consequences (Section 5) associated with invasions of these biota of concern and the adverse effects associated with the most likely impacted ecological receptors are considered in the text of this report, with the bulk of summarized technical materials associated with these species contained in appendices supporting the main body of the report.

2.2 Derivation of Initial Conceptual Model for Biological Invasions of Red River Basin from Upper Missouri River Basin

The iterative process characteristic of the current practice of risk assessment (see Section 1) yielded a draft conceptual model (Figure 6) initially developed from the regional conceptual model (see Section 1, Figure 3), and subsequently modified to yield an operational conceptual model given Reclamation and Technical Team inputs during problem formulation (Figure 7). The operational conceptual model that guided the work summarized in this current report incorporates

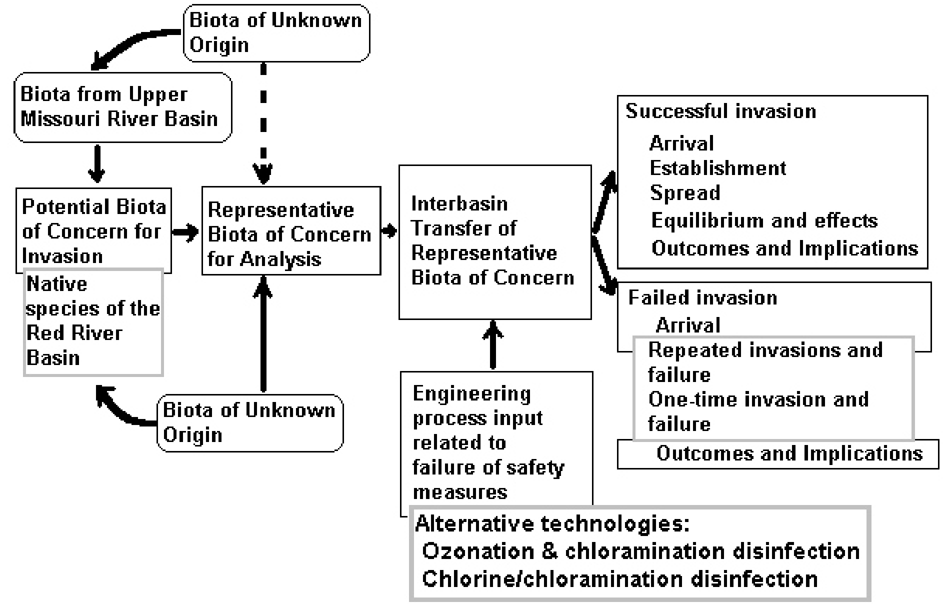


Figure 6. Preliminary conceptual model that guided early discussions with Reclamation and Technical Team to address the analysis, characterization, and interpretation of risks associated with biota transfers consequent to water diversions from HUC10 to HUC09.

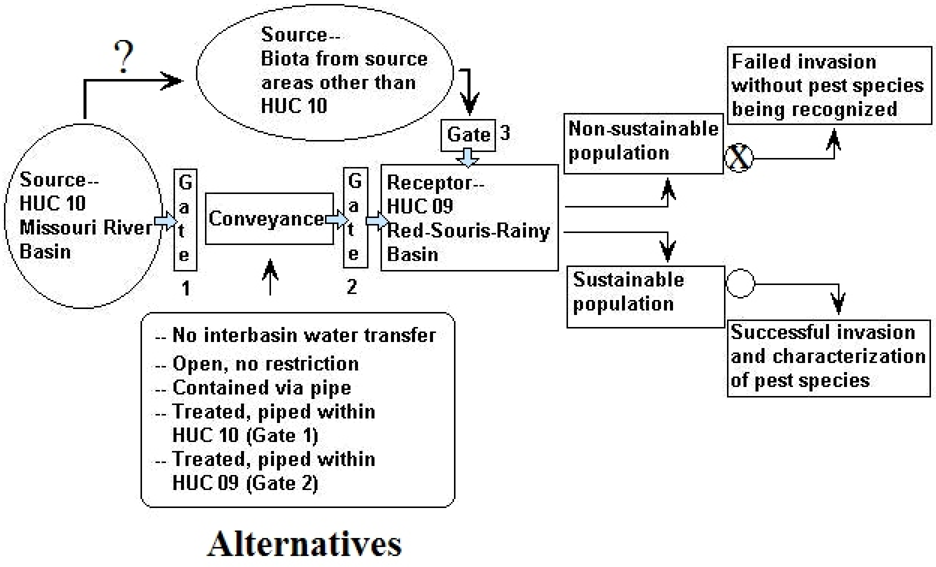


Figure 7. Revised conceptual model which became the operational linkage between the draft conceptual model developed early in the study with those elements of the project that were subsequently identified as critical to Reclamation and Technical Team needs for evaluating of risks and consequences (e.g., alternatives) potentially associated with interbasin water transfers.

sources of invasive species from the Upper Missouri River basin emigrating to the receiving Red River basin through various pathways, including those (1) directly reflecting interbasin water transfers reflected diversions from the Missouri River, (2) other invasions mediated by alternate routes of invasion dependent on human intervention (but not Garrison-related) pathways, or (3) invasions independent of anthropogenic activities.

2.3 Assessment Endpoints

Within the risk assessment process, and in particular during problem formulation, discussions with risk managers yield guidance for technical support activities such as in this current effort. Within an ecological context, assessment endpoints are selected for risk assessment during problem formulation. In this current investigation, assessment endpoints were globally identified as valued ecosystem components to be protected, in this instance, populations of ecological receptors and habitats potentially adversely impacted by species invasions consequent to biota transfers linked to water diversions between basins. While a range of professional opinion is evident in the identification and characterization of assessment endpoints and valued ecosystem components, the identification of global assessment endpoints reflects the systems-level bearing of the Technical Team (see Appendix 17 for listing of Technical Team membership). Technical support developed subsequent to that input reflects biota of concern being viewed within an ecological context, with the primary focus for evaluating consequences resolving on habitat equivalency analysis (HEA; see Section 5). For developing risk scenarios that reflected complete or potentially complete exposure pathways identified during preliminary problem formulation, this initial analysis identified biota of concern that could be relatively easily linked to assessment endpoints captured through HEA that were ecologically relevant, measurable or predictable, and susceptible to biological stressors such as invasive species. By integrating an analysis of economic consequences using HEA at this early stage of technical support, our technical analysis should better support Reclamation and Technical Team’s needs related to environmental management and policy decisions. Given the economic analysis tool of choice—habitat equivalency analysis—the integration of global assessment endpoints that guided the technical analysis through the selected biota of concern as “measurement endpoints” (Suter 1993) closely linked to global assessment endpoints summarized in Table 2.

These global assessment endpoints reflected concerns conveyed by Reclamation and Technical Team, particularly as those regional concerns were captured by the selected biota of concern. Although the technical analysis focused on measurement endpoints at the species level of organization, assessment endpoints characterized in part through HEA were considered representatives of the much larger body of “receptors” potentially at risk to species invasions, including adverse affects at the community and ecosystem levels of organization, environmental (both ecological and public) health, and potentially measurable population-level effects. During the development of these assessment endpoints, ecological relevance was an important consideration in selecting appropriate ecological receptors representing these assessment endpoints. From ecological perspectives, relevant conditions considered in the process of identifying global assessment endpoints included:

- Effects associated with the absence of a species normally expected to occur
- Effects associated with reduction in population size
- Effects associated with altered in community structure
- Habitat degradation or loss
- Diminished or reduced ecological function

Table 2. Global assessment endpoints linking species-specific analysis on selected biota of concern with habitat equivalency analysis supporting analysis of potential economic consequences associated with interbasin water transfers

Population-level	Community-level	Ecosystem-level
♦ Extinction ♦ Abundance ♦ Yield or production ♦ Changed age or size-class (demographic structure) ♦ Disease occurrence (changes in mortality or morbidity)	♦ Market or sport value ♦ Recreational quality ♦ Habitat alteration to less useful or desired type ♦ Changed community structure	♦ Productive capacity

2.4 Measures of Adverse Effects

Measurements of adverse effects, traditionally identified as measurement endpoints (see Suter 1993; EPA 2003, 1998, 1992), were used to quantify adverse effects associated with completed pathways in evaluating alternatives and in characterizing risks likely to occur subsequent to a species invasion. The primary measures of adverse effects are captured by the selected biota of concern, their selection in part determined by the ecological receptors most likely adversely affected by a particular species invasion, e.g., zebra mussels displacing native unionid mussels, and tamarisk displacing native cottonwoods and willow in riparian zones. Good measurement endpoints are intended to correspond to or are predictive of the selected assessment endpoints with the conceptual model establishing links between assessment endpoints and measurement endpoints. For the present study, our focus on invasive species or species having otherwise adversely impacted receptors expedites the identification of measures of adverse effects, and their linkage to assessment endpoints ensures a technically founded transition between risk and economic consequence analysis.

2.5 Pathway Analysis

The pathways analysis conducted as part of the current study reflects guidance from NISC (2001; see http://www.invasivespecies.gov/), with the analysis of pathways focused on both spatial and temporal distributions of biota of concern in source areas of the Missouri River and receptors in the Red River basin most likely responsive to species invasions, if they occur. As part of problem formulation, a brief but illustrative listing of pathways in Figure 5 (see Section 1) clearly suggested the modes of transit that potentially serve the operational conceptual model of Figure 7. In part, the interbasin transfer scenarios that follow in Section 3 are straightforward derivatives of these graphical summary of issues voiced by Reclamation and Technical Team. Distinguishing between invasions linked to intentional transfers from those linked to unintentional and largely accidental transfers is critical to the analysis of risks, the characterization of uncertainty associated with that analysis, and the evaluation of economic consequences of biota transfers.

The root cause of biota transfers in the current investigation is focused on potential water diversions from the Missouri River to the Red River basin. But alternative and competing pathways were critically evaluated within the risk analysis, given the role that these alternate pathways play in linking, or potentially linking, the areas of concern and their intuitively high likelihood for serving as confounding factors in the analysis of risks. For example, intentional and unintentional importation of animals or plants (e.g., commercial- or transportation-linked dispersal mechanisms) represents important pathways for the introduction of biota by direct means or coincidentally, e.g., as hitchhiking organisms.

Pathways analysis is a critical element to the multifaceted process of dispersal regardless of whether anthropogenic or natural mechanisms (or combinations thereof) are the primary linkage between source areas and receiving areas. NISC (2001; see http://www.invasivespecies.gov/) devoted much time and effort to the characterization of pathways that is critical to our current investigation, and we presently rely on its lead in characterizing pathways that potentially mediate biota transfers, whether they were directly linked to interbasin water transfers or not. Similarly, we follow the lead of NISC (2001; see http://www.invasivespecies.gov/) with respect to its focus on uncertainty as an explicit component of the analysis that complements an evaluation of pathways, e.g., a pathway may be identified as being associated with “high risk” but characterized by significant uncertainties with respect to its role in any given invasion process.

Given the list of representative species as biota of concern (Section 1, Table 1) and the task of identifying and characterizing pathways as factors contributing significantly to invasion risks, the current investigation again followed the lead of NISC (2001; see http://www.invasivespecies.gov/) in evaluating pathways directly related to potential interbasin water diversions and competing pathways potentially confounding the characterization of risk. As suggested by NISC (2001; see http://www.invasivespecies.gov/), pathways were categorized "depending on whether they influenced the probability of introduction or the consequences of the introduction" and were evaluated in conjunction with an "explicit analysis of the degree of uncertainty in the data" available for analysis. Graphically, pathways considered in the current work are captured in Annex, Figure 1 through Annex, Figure 5, which are referred to as "fault-probability trees" (FPTs) and will guide the derivation of quantitative estimates of risk (see Section 3). Fault-probability trees represent a melding of tools common to systems analysis (e.g., analysis of ecosystems or engineering systems) and probability estimation (see Barlow 1998; Bartlett 1960; Bedford and Cooke 2001; Blischke and Parbhakar Murthy 2000; Borgelt and Kruse 2002; Fielding 1999; Haupt and Haupt 2004, 1998; Huzurbazar 2005; Li 1981; Peterson et al. 2002a-d; Peterson and Vieglais 2001; Puccia and Levins 1985; Salthe 1985; Shigesada and Kawasaki 1997; Stockwell and Noble 1991, 1992; Williamson 1996).

Following the lead of NISC (2001; see http://www.invasivespecies.gov/), the current investigation considered pathways as (1) the physical process (e.g., a device or mode of transit, a specific route or course of invasion) that links source and receiving areas and (2) an integrated concept that melds physical processes linking source and receiving areas with attributes of the transferred biota itself (e.g., invasive species may have reproductive mechanisms that favor their migration and establishment in a receiving area). As such, pathway analysis in the current investigation considered a number of attributes that influenced the probability estimates (either quantitative or qualitative) associated with particular biota of concern and the pathways that provide an avenue for dispersal of the species into the Red River basin. These attributes of pathways illustrate the integrated physicochemical and biological characteristics of the invasion process that were critical to the current analysis of risks:

- **Pathway magnitude**: the absolute or relative contribution of a given pathway as a linkage between Missouri River sources and receiving areas in the Red River basin. The magnitude of any particular pathway would reflect, for example, the numbers of species potentially relying on the pathway for transit; the propagule pressure associated with a particular pathway which captures the interrelated factors of inoculation strength (i.e., how propagules characterize each biota transfer or invasion event, also referred to as a "trial") and frequency (i.e., how often to events or trials occur); and the diversity of species that is carried by a particular pathway.

- **Propagule survival**: survival of propagules during transit or their viability and sustainability following transfer reflect the potentially variable ability of a transferred organism to successfully migrate from Missouri River source to the receiving Red River basin throughout the organism's life cycle, e.g., transfer may occur at any time during a species' life history (as adult or larval stages), and some development stages may be more amenable to invasiveness than others.

- **Varying detection probabilities**: a pathway may be characterized by different likelihoods (quantitative or qualitative) for a particular species transfer, e.g., a species may be difficult to manage in transit, or for observing a species during or following its transfer, e.g., a species “presence” or “absence” may be difficult to measure during the transfer process.

- **Habitat similarity between source and receiving areas**: biota transfers, and more likely, species invasions will be highly dependent upon habitat availability and habitat quality in the receiving basin, e.g., successful biota transfers may not be fully realized if habitats in the receiving areas of the Red River basin are not comparable to those in the Missouri River source areas.

- **Likelihoods of establishing a sustainable population and continued dispersal in receiving systems will vary**: regardless of the means of transfer between Missouri River sources and the receiving areas of the Red River basin, biological attributes of a transferred organism will influence its capacity to attain sustainable populations, e.g., whether the species continues to increase its distribution in the receiving system will depend in part on its reproductive fitness in the receiving area and other factors such as habitat interconnectness that might facilitate its spread within and beyond that immediately invaded region.

- **Difficulty in managing a species, if it becomes established**: biota transfers, particularly those that are truly invasions according to our current definitions, may be exacerbated if biological attributes of the invader confer robust invasiveness potential to the species, e.g., although potentially influenced by habitat, management technologies for controlling or potentially eliminating the species may lessen the risks associated with the species transfer.

Each of these attributes may serve as potential risk management tools to control species dispersal. Or if dispersal has already occurred, the species’ attributes may be critical in the control of spread and other mitigation issues, e.g., if propagules successfully breach basin boundaries through a completed pathway, do propagule’s arrive at the receiving environment in good health and potentially capable of establishing a sustainable population?

Life-history summaries, including presence data, current status, and distribution data for biota of concern, were collected as available from a variety of sources (see Appendix 3A and Appendix 3B). Species-specific data of particular interest in the analysis of aquatic biota included: trophic status, parental care (fishes and aquatic invertebrates only), maximum body size, size of species native range, physiological tolerance, distance from nearest native source, prior invasion success, fecundity, and propagule pressure. These attributes were assigned values in discrete categories based on a survey of current literature (see Appendix 3A and Appendix 3B). Only the variable “prior invasion success” was continuously distributed; otherwise, variables were categorical and ordinal measures which were selected over continuous measures given the relatively limited reliable quantitative data for the majority of biota of concern (see Appendix 4 regarding details of categorical data, measurement data, and related terminology used in Section 2). Categorical scores for each variable were considered as specified below:

- **Trophic status**: categories for trophic status were designated: (1) parasites, (2) carnivores and herbivores, (3) omnivores, (4) filter-feeders on zooplankton or phytoplankton, including detritivores, (5) autotrophs (primarily cyanobacteria and vascular plants) based on life-history characteristics of biota of concern.

- **Parental investment in offspring (fishes and aquatic invertebrates only)**: for fishes and aquatic invertebrates, parental care categories (scored 1–4) were based on species life histories and included: (1) livebearers including ovoviviparous fishes, (2) biota of concern confer protection of propagules, e.g., fish may protect their embryos or larvae, (3) modes of spawning reflect limitations in propagule release by parents; e.g., for fish, fertilized eggs may be hidden but no additional parental care is apparent, or for aquatic plants, seeds or vegetative reproductive structures have limited unaided dispersal potential, (4) water-column spawners with no parental care (e.g., mollusks and fishes that scatter their eggs in the environment).

- **Maximum adult size (fishes only)**: this category is the maximum length individuals achieved, scored as 3 for "large," 2 for "intermediate," and 1 for "small."

- **Size of native range**: georeferenced data were not available for all biota of concern; hence, using the available presence and distribution data, biota of concern were categorically assigned a score of 1 to 5 based on presence data in North America (and in the instance of carp, worldwide distribution). These categorical scores were: (1) range occupies <5% of one zoogeographic subregion, e.g., local endemics, (2) range occupies 5–50% of one zoogeographic subregion; 3) range occupies >50% of one zoogeographic subregion, (4) range occupies more than one zoogeographic subregion. If biota of concern presented sufficient data for characterizing their current and predicted distributions, these were also reflected in their categorical assignments.

- **Physiological tolerance**: this variable represents tolerance to changes in water quality (usually temperature, dissolved oxygen, turbidity, and salinity) or to extreme conditions in water quality, based on Halliwell et al. (1999) for aquatic vertebrates and extended for our current application, with the addition of an extremely tolerant category: (1) intolerant: biota have low physiological tolerance to changes or extremes in water quality. (2) moderately tolerant biota capable of living in water with moderately high variability in water quality, (3) tolerant biota capable of living in waters in which water quality often reaches their limits of physiological tolerance for short periods, (4) extremely tolerant biota capable of living in waters with water quality that excludes most other members of the taxon.

- **Distance from nearest native source**: because dispersal distances are infrequently characterized and reflect retrospective estimates, categories were used to characterize distances from source to receiving areas that historically have been observed for biota of concern: (1) <150 km, or within HUC or current distribution suggested movements across HUC boundaries are not likely events unless presence data suggest current distribution reflects occurrence near HUC boundary, (2) 150–1000 km, or within HUCs and diffusive dispersal not unlikely at HUC boundaries; 3) 1000–3000 km, or within North America outside above areas; 4) >3000 km or dispersal from other continents beyond HUC boundaries for either Missouri River or Red River basin.

- **Prior invasion success**: a category that reflects past performance of the biota of concern as an invasive species. This score ranged from 1 to 4 as defined: (1) invasion potential limited unless mitigating factors such as human-assisted introductions occur (e.g., introductions of species greater than 1000 km from current distribution), (2) invasion potential restricted due to limiting life-history attributes (e.g., species is physiologically limited by low temperatures) or limited means of self-dispersal (e.g., propagules such as seeds are not disperse by wind), (3) species is opportunistic given translocation to receiving areas is amply supported by physical or biological means (e.g., species may have limited means to self-disperse but is a highly capable “hitchhiker” and relatively tolerant of widely varying environmental conditions), (4) species is highly mobile and has a widespread distribution, but presents life-history attributes that require relatively invariant environmental conditions to assure sustainable populations develop in receiving areas, (5) invasion highly likely given the species life-history attributes (e.g., propagules easily dispersed and highly tolerant of widely varying environmental conditions).

- **Fecundity and propagule pressure**: propagule pressure is the number of propagules (at any stage of life-history) that historically represents sufficient numbers to successfully establish a self-sustaining population in areas being invaded. Categories ranged from 1 to 4 and represented logarithmically spaced estimates as single events or multiple events that are effectively linked in any given cohort’s common life history: (1) less than 10 individuals released, (2) greater than 10 but less than 100 individuals, (3) greater than 100 but less than 1000 individuals, (3) greater than 1000 but less than 10,000 individuals released, and (4) greater than 10000 individuals released. Categories were applied to this analysis, since measured counts of propagule pressure are infrequently recorded in the literature and when available are frequently rough estimates based on the most likely scenarios for the introduction.

In the current investigation, we have incorporated a categorical analysis for characterizing risk and used a rank-categorical approach (see Campbell and Kreisch 2003; Agresti 2002; NISC 2001) to evaluate each of the biota of concern listed in Table 1 by adapting a set of categorical assignments suggested by Campbell and Kreisch (2003), Kolar and Lodge (2002) and Marchetti et al (2004) among others (Ashton and Mitchell 1989; Rejmánek and Richardson 1996; Rejmánek 2000, e.g., for invasive plants) to evaluate those attributes identified for each biota of concern.

As suggested by Marchetti et al. (2004), we used the following general analytical approach based on Burnham and Anderson (2002) to assess the influence of species characteristics on three stages of the invasion process. First, we gathered existing data and information on species presence, characteristics, spread, and abundance (see Appendix 3A and Appendix 3B), then completed a spatial analysis that was focused on predicting the potential distribution of each biota of concern. Depending on the available life history, past records of invasiveness, and the current and predicted distributions of biota of concern, we then characterized risk as a function of the categorical evaluation of potential risk and predicted distributions of biota of concern.

Quantitatively, two techniques were applied to the analysis and characterization of risk in Section 3 and Section 4, respectively. Various information measures such as Akaike's Information Criterion (see Arndt 2001) and statistical tools such as chi-square analysis of goodness of fit (see Zar 1999) are applicable to identifying "best fit" spatial projections of potential species' distribution (see Appendix 4). In our current work, the latter tool was applied to our spatial analysis. When data were sufficient, categorical and spatial analyses were completed for quantitative evaluations of risk; however, not all biota of concern presented sufficient data to warrant quantitative analysis, and risks were characterized qualitatively through a narrative characterization (see Section 4).

Categorical assignments of risks were also applied to estimates of risk derived from the evaluation of simple FPTs (Figure 8 through Figure 12) that captured the stepwise invasion process that practically linked pathways analysis with our operational conceptual model (Figure 7). The melding of categorical and quantitative estimates of risk assured that differences in data availability from one biota of concern to another would not adversely influence characterization of risks associated with biota of concern but would reinforce the importance of uncertainty in the interpretation of risk and the development of risk management practices available to manage data insufficiency.

2.6 Tools to Characterize Risk and Economic Consequences

The primary task that characterizes this objective of the current study focuses on the derivation of estimates of risk and the consequences potentially associated with those risks. Although oversimplified for our present purposes (and highly dependent on data sufficient for implementation), in general the analysis of risks considered conditional probabilities that describe the invasion of the Red River basin by any species originating from the Upper Missouri basin as

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum\limits_{j=1} P(B|A_j)P(A_j)}$$

where the event, A*i*, is predicated on B repeatedly over space-time. Such tools are commonly applied to engineering systems (see Bedford and Cook 2001, Serrano 2001) and biological systems, including species invasions (Hayes 1996, Levin 1989, Williamson 1989, Williamson 1996, Paine et al. 1998). Although evaluation of risks conceptually hinged on evaluating conditional probabilities, in our current evaluation, simple probability evaluations were the tool of choice (see Appendix 4). For example, FPTs in Annex, Figure 1 through Annex, Figure 5 reflect the "simple" stepwise constructions supporting a simple probability analysis of the biota transfer and invasion process. Biologically, the generalized top event might be "successful biological invasion of Red River basin by biota originating from Upper Missouri River basin." Here, the success of invasion would be predicated on prior independent events occurring, e.g., (1) "biota transfer successfully completed," (2) "invasive species established a reproductive population," and (3) "a reproductive population attains sustainable numbers to compete against indigenous species," with each prior event amenable to decomposition and more comprehensive characterization as data allow (pathways may be incomplete, biota transfer from source area may not lead to establishment of invasive species population given failure to find suitable habitats or hosts, etc.. in the target area). Ultimately, the statements of probability of invasive species established in the target area (Red River basin) would be developed for each of the biota of concern identified in the conceptual model(s).

Given the conceptual models developed early in discussions and finally revised as the current investigations operation conceptual model to guide scenario development (Figure 7), the development of a series of FPTs reflects a graphical summary of these conditioned events that were the focus of quantitative risk estimation (see Section 3 and Section 4 for analysis and characterization of risks, respectively). Here, we briefly describe the graphic representations of FPTs and characterize the calculations that yield quantitative risk estimates summarized in Section 3. The reader is referred to Appendix 4 for greater detail on the preliminary derivations of probability estimates for biota transfers as depicted in Annex, Figure 1 through Annex, Figure 5. Each FPT captures one pathway or a series of interrelated pathways that potentially link biota from the Missouri River basin with the Red River basin. For example, FPT 1 is the primary depiction of pathways directly linked to water diversions proposed under the DWRA, and subsequent fault-probability trees (Annex, Figure 1 thorugh Annex, Figure 5) reflect alternative transfer and invasion pathways.

Consequence analysis was fully integrated with the analysis of risk, and as identified during problem formulation (see Section 1), the analysis of economic consequences relied upon tools commonly applied to the evaluation of restoration and compensatory costs in Natural Resource Damage Assessments (NRDA). Invasion of habitats by nonnative species and the subsequent displacement of native species from those habitats clearly suggest that HEA is appropriate as a tool to evaluate economic consequences potentially associated with interbasin biota transfers. Section 5 summarizes the HEA process which focused on the costs of compensatory measures that offset the loss of wildlife habitat function associated with invasive species. Outputs of risk characterization provide inputs to the HEA process, and the subsequent integration of risk and costs provide insights to the consequences associated with biota transfers as captured by the supporting conceptual model. Also, tools available to the natural resource and environmental economist have been applied to resource valuation, including those tools applied to nonmarket valuation and characterization of nonuse, option, and existence values.

Confidence in the conclusions of risk characterization may be increased by using several lines of evidence to interpret and compare risk estimates, including an evaluation of the relevance of evidence to the assessment endpoints, the relevance of evidence to the conceptual model, the sufficiency and quality of existing data, the strength of cause and effect relationships noted in comparative studies, and the relative uncertainty associated with each line of evidence and the concordance (or absence of concordance) across various lines of evidence.

2.6.1 Focus on ecological adversity. Risk characterization discusses whether ecological receptors are exposed to invasive species capable of causing adverse effects to the overall ecosystem or to the particular valued species within that ecosystem (assessment endpoint), including a focus on whether ecological receptors may be adversely affected in the future (see Section 1; EPA 1992, EPA 1998, Suter 1993, Minnesota Sea Grant/Michigan Sea Grant, 2001).

The nature and intensity of effects were evaluated to distinguish adverse effects from effects occurring within the normal pattern of variability. Spatial and temporal scales were also considered in assessing adverse effects. The spatial dimension involves both the extent and pattern of adverse effects, as well as the context of the effects within the ecosystem. Factors to consider include the absolute area affected, the extent of sensitive habitats affected compared with a larger area of interest, and the current and future land and water use within the ecosystem. The temporal scale of adverse effects for ecosystems can vary from short term (e.g., seconds to minutes to days for altered photosynthesis yielding advantages to invasives for establishing sustainable populations) to long term (e.g., decades to centuries for adverse effects reflected in changes in biodiversity). Risk assessors should recognize that the time scale of adverse effects operate within the context of multiple natural time scales. For example, visible changes in the productivity of an aquatic system may not become evident for many years after initial biological invasion. The potential for recovery of a system was also considered in assessing ecological adversity. Recovery is the rate and extent of return of a population or community to a condition that existed before the introduction of invasive species. Examples include reestablishment of a species to a specified density or recolonization during recovery following removal of a biological invader.

2.6.2 Uncertainty analysis. A discussion of uncertainties or the lack of relevant information is a necessary part in an even-handed characterization of risks associated with a biological invasion. Sources of uncertainty contribute to possible overestimated or underestimated ecological risks. The objective of uncertainty analysis was to describe and quantify, where possible, what is known and not known about effects and linkage between source and receiving areas. Uncertainty analysis increases assessment credibility by quantitatively or categorically characterizing the magnitude of uncertainties and their relationship to risk characterization.

Uncertainties may be addressed and their effects minimized for any risk assessment, with the results of uncertainty analysis being used to identify data gaps and direct data collection activities. For the evaluation of biota transfers and the biological invasions subsequent to water transfers between the Upper Missouri basin and Red River, species distributions will be critical data to the risk analysis. Additionally, and as available, the risk analysis will depend on data that reflect a quantitative basis for evaluating the transfer and establishment of invasive species, the spread and development of equilibrium populations of invasive species and the effects and potential implications of invasive species. For example, although data may not be sufficient for each biota of concern, demographic data related to life-table analysis would ideally be applied to the analysis wherein survivorship and maternity functions and reproductive rates would be considered as a basis of analysis. Ecologically, habitat data may be critical to the analysis (e.g., habitats not sufficient to sustain an invading species) as would potential environmental or engineering data that suggest limitations to successful invasions (e.g., ambient temperature extremes or water treatment may limit success). Similarly, data critical to a fully developed consequence analysis would encompass biological data (e.g., species distributions, functions key to life-table analysis) and economic data essential to an analysis of the impacts of invasive species and the determination of compensatory measures sufficient to offset those impacts.

The methodological approach applied in the current analysis was observational and relied on existing data or information in the form of peer-reviewed literature or government documents. Analytically and statistically as possible, these encountered data were reviewed for data quality. When possible, primary data sources were used in developing the risk and consequence analysis. Tools selected for the analysis reflected the contingencies predicated by available data, and as such, those tools commonly applied to encountered data analysis (see Appendix 4). In the absence of primary data, peer-reviewed literature from open sources was relied upon, as well as government documents that have met data quality objectives for the project reports being reviewed (see data quality discussions for the Heinz Center (2002) reports such as http://www.heinzctr.org/ecosystems/report.html, and related technical appendices available at http://www.heinzctr.org/ecosystems/pdf_files/sotne_tech_notes.pdf, and http://www.heinzctr.org/ecosystems/fr_water_technotes/fr_water_non_ntv_spec.shtml).

2.6.3 Data quality evaluation. The derivation of probabilities for interbasin transfers of biota of concern will only be as good as the data used in their calculation, which necessarily means the characterization of risks must be completed in parallel with an evaluation of data quantity and data quality. As a source of uncertainty, data quality and quantity are critical to the interpretation of species invasion probabilities. Also, uncertainty will vary from one species to the next, depending on the available data; hence, risks dependent on probability estimates have been characterized in light of their associated uncertainty. For this investigation, existing literature and available data used as inputs for the derivation of quantitative estimates of risk were obtained from the peer-reviewed literature, from open-source data compilations having documented quality assurance practices, and from USGS-reviewed data compilations. In addition, as available, other Department of the Interior (DOI) sources (e.g., US Fish and Wildlife Service, National Park Service) provided data and available literature with documented quality assurance practices (see http://www.epa.gov/nerlesd1/gqc/courses/qa.htm; http://geography.usgs.gov/standards/).

2.7 Implementing Data Search and Retrieval

Natural resource managers frequently seek technical support for developing practices and policy. The work summarized in this report reflects the technical findings for such a request focused on issues related to biota transfers potentially resulting from water diversions from the Missouri River to the Red River basin. As reflected in Section 1, risk analysis has found increasing application for crafting adaptive resource management practices wherein technical inputs and outputs—realized or potential—to a managed system (e.g., wildlife refuge, agriculture lands) are considered within a “what-if” context focused on potential outcomes that likely influence practices and policy proactively (see Gunderson et al. 1995; Holling 1978; Jensen and Bourgeron 2001; Walters 1986). Regardless of landscape setting, whether it is solely spatial or spatiotemporal, an initial evaluation of risks involved in various management practices available to the manager may be solicited for guidance on which of many management practices might be applied to the specific circumstances. Often, a “desktop” risk analysis is the first step in the characterization of risks and the evaluation of and its dependence on existing data and information directly or indirectly linked to the questions identified in problem formulation.

2.7.1 Encountered data and data mining. Desktop analysis requires working with data and existing information; hence, the analytical tools are more observational than experimental and rely on data-mining search and compilation (see Chen 2001; Wolkenhauer 2001). As such, collection of data in desktop studies is similar to preliminary field investigations in ecological studies, since data in ecological and environmental studies, especially reconnaissance level efforts, are often not collected via an experimental or sampling design but through observational studies. The primary tool in data collection for desktop analysis is data mining, a discipline lying at the interface of statistics, database technology, pattern recognition, and machine learning. Data mining is focused on the secondary analysis of data extracted from the existing literature (e.g., previously published material, compiled databases) in order to characterize relationships among variables typical of new questions that may be linked to these existing sources of information. Data mining relies on an inductive process and is primarily concerned with secondary data analysis.

2.7.1.1 Literature Search: Collection of existing data and information to evaluate risks associated with biota transfers between Missouri River and Red River basins. The main database providers used included Cambridge Scientific Abstracts (CSA) and OCLC FirstSearch. Databases searched in CSA included Aquatic Sciences and Fisheries Abstracts, Biological Sciences, Environmental Sciences and Pollution Management; and to a lesser extent, AquaLine, Water Resources Abstracts, GeoRef, Biology Digest, Conference Papers Index, Medline and Toxline. Databases in OCLC FirstSearch that were searched included Agricola, ArticleFirst, BasicBiosis, Dissertations, GeoBase, and WorldCat. BioAgIndex, Electronic Collections Online, PapersFirst, and Proceedings. Ingenta database provider was also used for some searches.

Search terms. Searches for identifed biota of concern used the scientific name (at genus or species level), and common names, if applicable. Depending on the number of citations found, additional search terms were added. Terms used would refer to the distribution and spread of the species, its life history and habitat, and its interaction with other species. In some cases, for example, the bacteria, the focus was on the natural occurrence of the biota and risk assessment. Citations related to detection and control measures were generally included. When there were an overwhelming number of citations, the search in some databases was limited to more recent references (within the last 10 years).

2.7.1.2 Search outcomes. Existing literature and data collected from the literature search was dominated by "effects data" derived from past studies—observational and experimental—focused on the effects that a particular biota of concern had on a receiving system and "pathways data" which reflected available literature resources focused on the spatiotemporal linkages between biota of concern and their geographic distributions. Graphically, Figure 8 through Figure 13 summarize citation counts for biota of concern and reflect a relatively wide range in literature and existing data available for the current data-mining effort focused on the biota transfer questions identified during problem formulation.

2.8 Implementing Data Analysis

The following overview provides a summary of analytical tools and the underlying assumptions that are associated with these tools for the data analysis detailed in Section 3 and the characterization of risks in Section 4. Please refer to Aven (2003), Barlow (1998), Blischke and

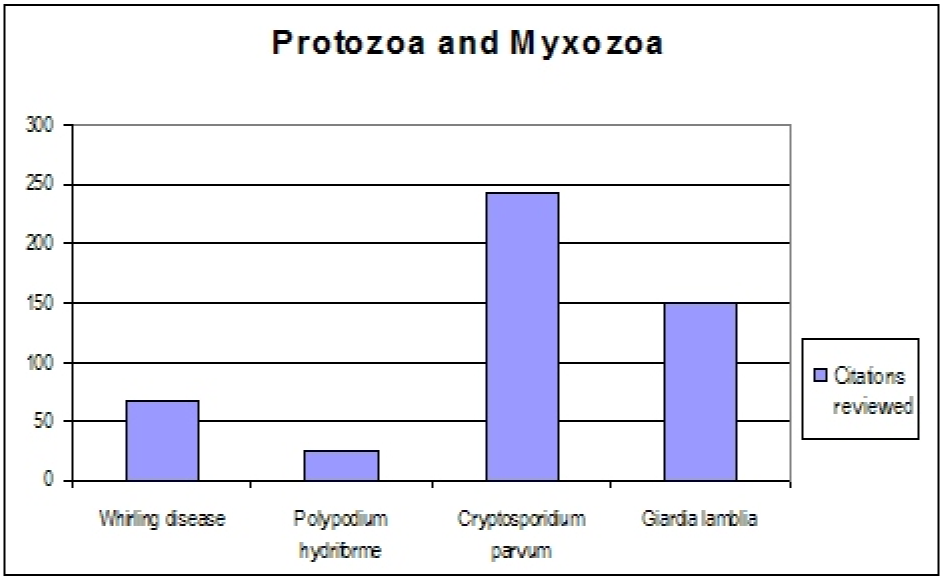


Figure 8. Summary numbers of literature citations collected and subsequently compiled for protozoa and myxozoa (see Appendix 3B).

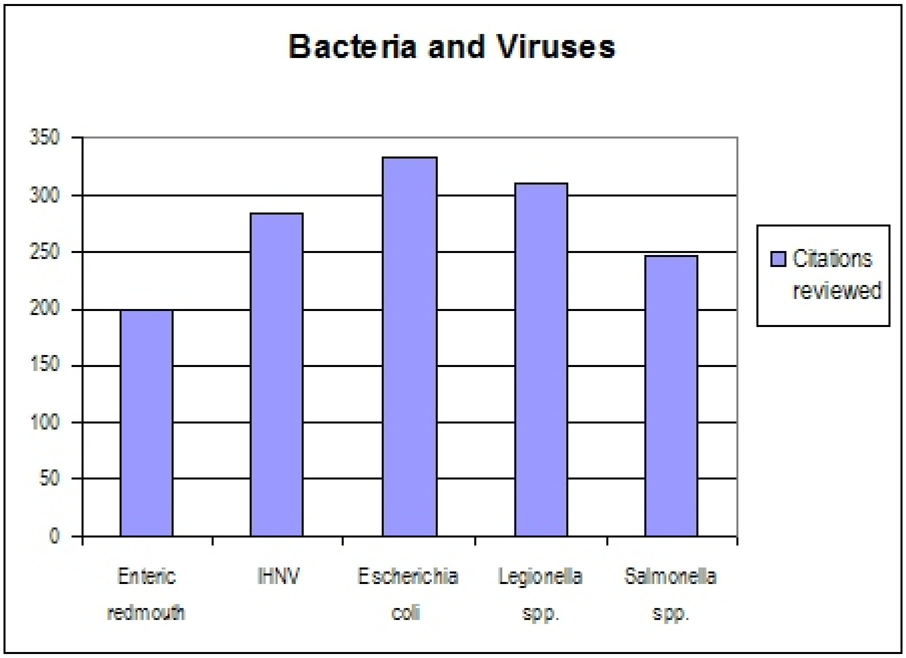


Figure 9. Summary numbers of literature citations collected and subsequently compiled for bacteria and viruses (see Appendix 3B).

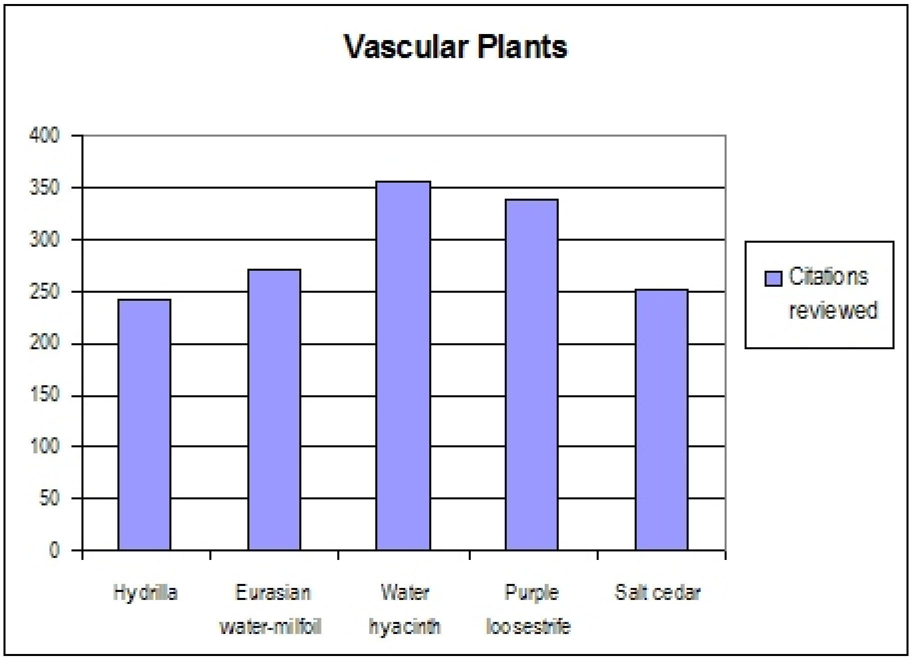


Figure 10. Summary numbers of literature citations collected and subsequently compiled for aquatic vascular plants (see Appendix 3A).

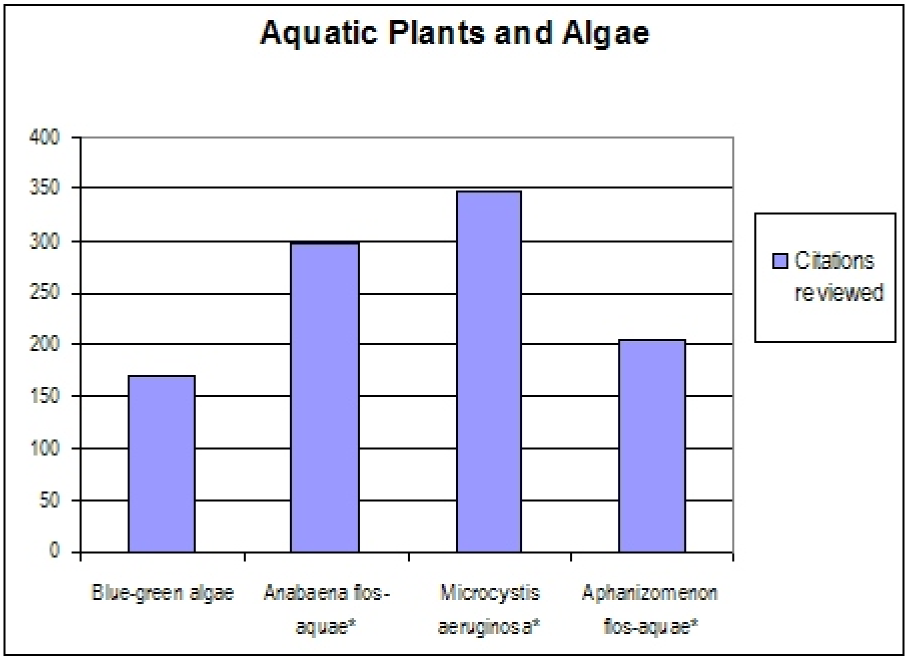


Figure 11. Summary numbers of literature citations collected and subsequently compiled for cyanobacteria (see Appendix 3B).

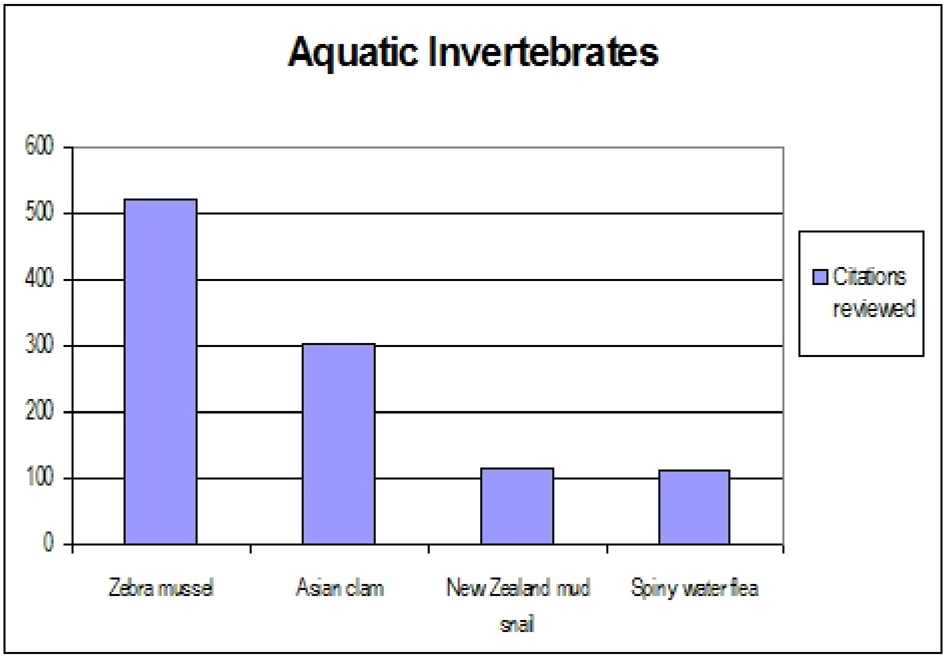


Figure 12. Summary numbers of literature citations collected and subsequently compiled for aquatic invertebrates (see Appendix 3A).

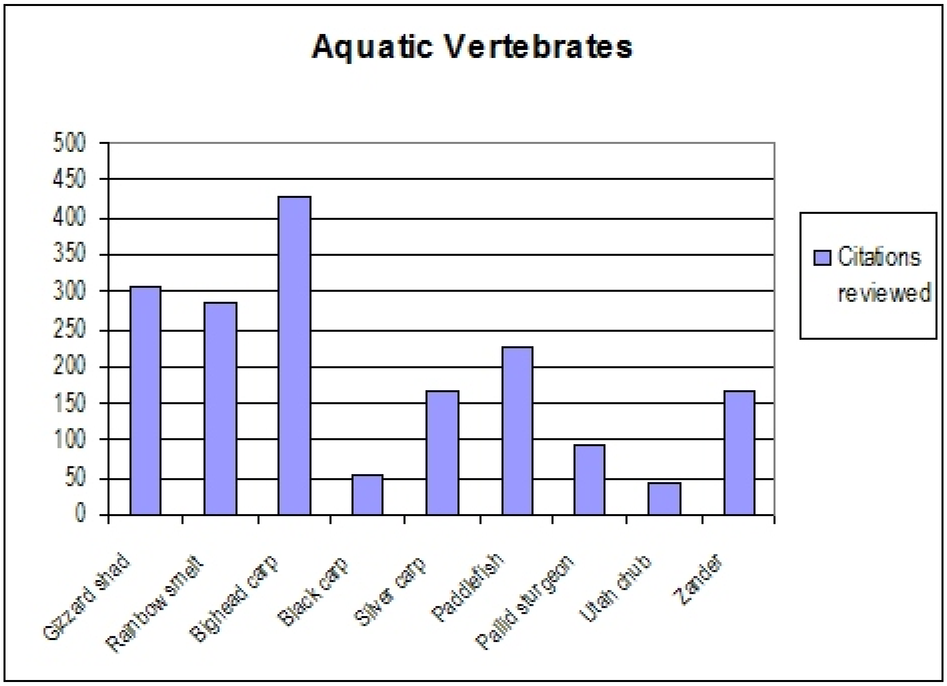


Figure 13. Summary numbers of literature citations collected and subsequently compiled for aquatic vertebrates (i.e., fishes; see Appendix 3A).

Parbhakar Murthy (2000), Borgelt and Kruse (2002), Huzurbazar (2005), and Appendix 4 as derived in part from NIST/SEMATECH (2004; http://www.itl.nist.gov/div898/handbook/) for background information for supporting data analysis.

2.8.1 Tree diagrams and pathway analysis.

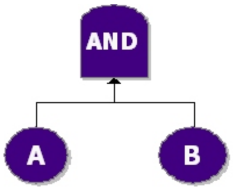
Tree diagrams are used to represent probability problems involving multiple events, generally where the events are sequential and independent. The tree diagram consists of a number of branches, where each event in a flow of events has branches characteristic of potential outcomes of trials (see Burgman 2005 for application of these techniques in conservation practice). Probabilities, then, are associated with each of the branches in a flow of events. Construction of a tree diagram continues until the process is fully characterized. In the present work, tree diagrams occur as graphic tools to evaluate (1) probabilities of serial, independent events and (2) potential fault trees associated with a preliminary analysis of control system reliability.

Fault trees are symbolic analytical logic techniques that can be applied to analyze system failure or reliability and related characteristics. A fault-tree diagram follows a structure that represents a graphical model of the pathways within a system that can lead to a potential loss or failure, in our application, a species invasion. The pathways interconnect intermediary events and conditions using standard logic symbols (e.g., AND-gate, OR-gate). Fault-tree diagrams consist of gates and events connected with lines. The AND and OR gates are the two most commonly used gates in a fault tree wherein gates are used to consider two events (called “input events”) that can lead to another event (called the “output event” or “outcome”). If the occurrence of either input event causes the output event to occur, then these input events are connected using an OR gate. Alternatively, if both input events must occur in order for the output event to occur, then they are connected by an AND gate.

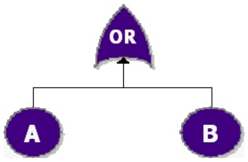
2.8.1.1 Basic gates.

Gates are the logic symbols that interconnect contributory events and conditions in a fault-tree diagram. The AND, OR gates, and Voting OR gates are the basic types of gates in classical fault-tree analysis that pertain to our present investigation.

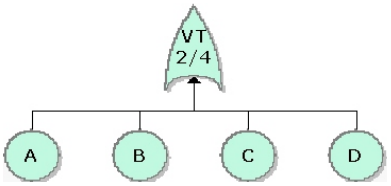
AND Gate. In an AND gate, the output event occurs if all input events occur. In system reliability terms, this implies that all components must fail (input) in order for the system to fail (output) and is conventionally illustrated as:



OR Gate. In an OR gate, the output event occurs if at least one of the input events occurs. In system reliability terms, this implies that if any component fails (input) then the system will fail (output). The illustration below shows a simple fault-tree diagram in which either A or B must occur in order for the output event to occur. In this diagram, the two events are connected to an OR gate. If the output event is system failure and the two input events are component failures, then this fault tree indicates that the failure of A or B causes the system to fail.



Voting OR Gate. In a Voting OR gate, the output event occurs if one or more of the input events occur. In system reliability terms, this implies that if any k-out-of-n components fail (input) then the system will fail (output), e.g., for a sexually reproducing invasive species, at least one of each sex must successfully breach barriers between source and receiving areas. Convention would illustrate a fault-tree diagram with a 2-out-of-4 Voting OR gate as indicated below where the system will fail if any two of its input components fail. The fault tree looks at "k-out-of-n" failures for the system failure.



Combining Basic Gates. Gates within a fault-tree analysis can also be combined to create more complex representations. Regardless of its complexity, a fault-tree diagram is always drawn in a top-down manner and with its lowest item being a basic event block.

Pathway analysis for the current investigation was completed in conjunction with fault-probability tree diagrams and summarized invasion scenarios developed from the conceptual model(s) in problem formulation (see Section 1 and operation conceptual model, Figure 7). Reliability analysis, especially for infrastructures such as those envisioned for implementing interbasin water transfers, has been folded into the pathways analysis, since it is a key component reflected in the developing estimates of risk and evaluation the risk reduction potential of alternative control system technologies. By doing so, alternative controls technologies have also been considered within the context of risk reduction in Section 4.

2.8.1.2 Fault-probability trees and reliability analysis. A graphic depiction of pathways that potentially link biota from source areas in the Missouri River basin with their potential distribution expansions to the Red River basin are captured in Annex, Figure 1 through Annex, Figure 5. While relatively simple in construction, these FPTs represent an integration of tools from systems analysis that afford us with an opportunity to incorporate alternative control systems into our risk analysis. The reliability of any given control system of interest to Reclamation and Technical Team, be that chemical treatment of source waters (e.g., chlorination, chloramination), use of various water filtration technologies (e.g., microfiltration, ultrafiltration), or combined technologies can be considered within the context of risk reduction when these control systems are contrasted to water transfers in the absence or failure of control systems (see Section 4). As such, the interrelationships between probability characterizations for biota transfer and systems reliability can be graphically summarized for considering various scenarios in the current analysis.

For example, any engineered control system reflects a range in performance that varies with time where a system's quality reflects a snapshot of that system at "time zero," or T_0, or the system's start-up. Reliability, however, is dynamic and captures a system's day-to-day operation. For example, system defects, e.g., in a water treatment system, at time zero likely express mistakes in manufacture of component parts or construction of the system, while reliability defects reflect the aging process of the system and its component parts. To describe reliability defects, a probability model that describes the fraction of failures over time or the system's life distribution model should be known because accurate prediction and control of reliability play important roles in helping a system to meet its performance goals. See Appendix 4 for additional background on reliability analysis as integrated into the analysis of risks for the current investigation.

2.9 Categorical Data and its Analysis: Tools Used in Current Analysis of Encountered Data

A brief review of Appendix 4 provides background on categorical data and measurement data, and in this section we briefly review the primary tools brought to the analysis of risks currently under investigation.

2.9.1 Logistic regression. Logistic regression is instrumental in the analysis of risk, including the development of predicted species distributions for biota of concern (see Hosmer and Lemeshow 2000; see also §4.3.2). As in simple linear regression, logistic regression gives each regressor a coefficient b_1 which measures the regressor's independent contribution to variations in the dependent variable. In evaluating dependent variables that are binary (that is, they result from a Bernoulli process wherein outcomes can only take values of 0 and 1), logistic regression yields estimates of probability (p) of a dependent variable being 1 rather than 0 based upon our knowledge of independent variables, i.e., we are interested in the probability of a biota transfer consequent to a water diversion being successful (value = 1) or not successful (value = 0). Logistic regression requires making a logistic transformation of p, commonly referred to as the logit of p. Logit(p) is the log (to base e) of the odds or likelihood ratio that the dependent variable is 1, and in symbols is defined as

$$\text{logit}(p) = \log(p/(1-p)).$$

Values of p can only range from 0 to 1, but logit (p) can range from negative infinity to positive infinity with the logit scale being symmetrical around the logit 0.5 (Table 3).

Table 3. The relationship between probability of success (p) and logit(p).

p	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99
logit(p)	-0.847	-0.405	0	0.405	0.847	1.386	2.197	2.944	4.595

As indicated by values in this summary table, the differences between extreme probabilities are spread out over a larger interval than mid-range values (e.g., differences of logits between success rates of 0.95 and 0.99 are much greater than logits between 0.5 and 0.7); hence, the logit scale is approximately linear in mid-range and logarithmic at extreme values.

Logistic regression involves fitting an equation of the form to the empirical data available to the analyst; that is,

$$\text{logit}(p) = a + b_1x_1 + b_2x_2 + b_3x_3 + \ldots$$

Logistic regression finds a “best-fit” equation using a maximum likelihood method in contrast to the simple least-squares methods of linear regression. Hence, the goodness of fit and overall significance statistics used in logistic regression are different from those used in linear regression. Log likelihoods are critical in evaluating logistic regression wherein probability associates with a specified hypothesis, e.g., the null hypothesis that all coefficients in the regression equation take the value zero. Calculations of such a likelihood of observing the exact data we actually did observe under this hypothesis generally yields a very small number. The values are usually transformed to the natural logarithm; hence, a value of log likelihood. Log likelihoods are always negative, since probabilities are nearly always less than one.

2.9.2 Monte Carlo analysis, Markov chain-Monte Carlo analysis, and bootstrap resampling methods. Monte Carlo, Markov Chain-Monte Carlo (MCMC), and bootstrap resampling methods (see Davison and Hinkley 1997; Efron 1982; Gamerman 1997; Manly 1991) were applicable to characterizing quantitative estimates of risk of biota transfers between Missouri River source and Red River basin receiving systems. For example, MCMC is an algorithm that generates a Markov chain of random samples (i.e., each sample is conditionally dependent on the preceding sample) whose stationary distribution is the likelihood function. In contrast, bootstrap resampling relies on assumptions that empirical samples are adequate as representations of populations and that characterizations of the underlying population may be attained by simply resampling that empirical sample set that is characterized by a statistic, β^{hat}, based on a sample of size, T. In bootstrap resampling we assume that the empirical sample of size T approximates the entire sampling distribution of β^{hat} by investigating the variation of β^{hat} over a large number of pseudosamples obtained by resampling the same data. For the resampling, we assume the original sample was drawn from a single population and that samples were independent, and a Monte Carlo simulation is used on the available sample values which are drawn randomly with replacement. No underlying assumptions of normality are required in bootstrap resampling. The bootstrap is completed via a large number of resamples (e.g., 1000) of size T from the original sample created, with each element of the original sample having the same probability ($1/T$) of being in a sample. The initial idea behind bootstrapping was that a relative frequency distribution of β^{hat}s calculated from the resamples can be a good approximation to its sampling distribution.

2.9.3 Spatial and temporal analysis. Our analysis of risks is largely based on a “snapshot” of species distributions in ecological space-time, yet our intent is projecting future distributions of biota of concern given the dynamic process of species dispersal. In the current investigation the temporal and spatial attributes of species distributions are summarized primarily through dot maps and other qualitative thematic maps (e.g., county maps) for displaying point data and regional count data, respectively. Potential species distributions are predicted using a technique increasingly common in conservation biology and in studies focused on assessing biological diversity (e.g., Peterson and Vieglais 2001; Peterson et al. 2002d).

2.9.3.1 Dot maps. Qualitative and quantitative thematic maps have a long history in the analysis of species distributions and epidemiology (see Lawson et al. 1999; Lawson 2001), especially when data are collected as a result of a point process linked to some geographic feature of the landscape (e.g., some natural or political features that can be projected on a map). Qualitative thematic maps form a variety of maps that differ from general or topographic maps, since they specify distributions of categorical data or nominal data (see Appendix 4) in their simplest forms, e.g. different types of natural features such as soils and vegetation or different types of human cultural attributes such as religions and occupations.

Dot maps are one of the oldest but simplest quantitative thematic maps used to summarize spatial point data. While dot maps provide a quick visual display of point data across some geographic region, it does not project estimates of actual numbers or areal patterns. Simple dot maps, however, do provide a good illustration of distribution and form the basis for the spatial analysis of geographically referenced point data concerned with spatial patterns of species or disease occurrence. In the current investigation dot maps are used to summarize spatial point data when locations of species occurrence are known, or regional count data when location information is limited and the numbers of species occurrences are aggregated for distinctive regions, most often county or state records.

2.9.3.2 Genetic algorithm for rule-set prediction (GARP) and Desktop GARP. There are numerous tools that project spatial point data, e.g., as evident from a simple dot map, to a continuous species distribution where interpolations have been made to account for areas between dots to yield continuously distributed projections (see Cressie 1993; Daly and Gani 1999; Elliot et al. 2000; Lawson et al. 1999; Lawson 2001; Scott et al. 2002). Alternatively, projections of spatial point data may assume disjunct collections of continuous distributions that result from point data that occur as clusters on a simple dot map. In the current investigation, potential species distributions were predicted by using contemporary tools—genetic algorithms and machine learning—increasingly used in conservation biology (see Fielding 1999; Sait and Youssef 1999; Spall 2003; Peterson and Vieglais 2001; Peterson et al. 2002a-d). While simple dot maps are examples of one of our oldest tools for evaluating species distributions, these contemporary tools reflect the development of computer-based analysis of spatial processes and their applications to biological and ecological questions.

Genetic algorithms (GAs) constitute one class of artificial intelligence applications and were inspired by models of genetics and evolution (Holland 1975). They have been applied to various problems not amenable to traditional computational methods because the search space of all possible solutions is too large to search exhaustively in a reasonable amount of time (Stockwell and Noble 1992). Genetic Algorithm for Rule-Set Prediction (GARP) is an expert system, machine-learning approach to predictive modeling (Stockwell and Peters 1999). In general, GA is a machine-learning method (e.g., decision trees, genetic algorithms, rules, and neural nets; Haupt and Haupt 1998, 2004; Fielding 1998), which includes relatively recent techniques previously impractical to implement due to an absence in available computing resources. With the increasing availability of computers sufficient to the prerequisites of GARP and other machine-learning techniques, the range of machine-dependent analytical tools applicable to the addressing problems of natural resource managers has markedly increased, e.g., in evaluating issues related to biological diversity, future land and water use, and the evaluation of alternative environmental management practices (see Scott et al. 2002; Scott et al. 1993). GARP has proven especially successful in predicting species' potential distributions under a wide variety of situations (see Peterson and Cohoon, 1999; Peterson et al. 1999, 2001, 2002a–d; Peterson, 2001; Feria and Peterson 2002; Stockwell and Peters 1999). Chen and Peterson (2000), Peterson and Vieglais (2001), and Anderson et al. (2002) provide general explanations of the GARP modeling process and interpretation of potential distributions; see Stockwell and Noble (1992) and Stockwell and Peters (1999) for technical details.

Developing predicted distributions, as either projections of past distribution or estimates of future species distribution expansion, is an increasingly applied tool in biogeography, evolution, ecology, conservation, invasive-species management, and disease prevention (see contributed papers in Scott el al. 2002; Peterson et al. 1999, 2000; Peterson and Vieglais 2001; Anderson et al. 2002). As illustrated by these and other peer-reviewed publications, these approaches combine presence data with environmental variables (both biotic and abiotic factors, including, e.g., temperature, precipitation, elevation, geology, and vegetation) to create a model of the species' distribution based on physical habitat requirements associated with environmental variables. Georeferenced presence data occur as latitude and longitude coordinates for confirmed localities that are characterized by vouchered museum or herbarium specimens or observational accounts (as captures or multiple-observer sightings) collected during designed or opportunistic biological surveys. While presence data are relatively well documented, absence data are infrequently recorded and not necessary for implementation of GARP (Stockwell and Peters 1999; Anderson et al. 2002).

Simple stated, GARP views presence data relative to existing digital maps of biological and physical habitat attributes. Environmental variables typically examined in such modeling efforts encompass include relatively few measures of ecological niche (see often-cited Hutchinson 1957), but available digital environmental coverages provide many variables that commonly influence any given species distributions. Given the analysis of presence data predicated on existing biological and physical habitat attributes, the resulting model is subsequently projected onto a map of the study region, here, the northern Great Plains of North America, wherein the potential distribution of the species of concern is characterized within the context of previously unoccupied space (see Chen and Peterson 2000; Peterson and Vieglais 2001). Models are generally based on the species' fundamental niche (Hutchinson 1957; MacArthur 1968; Wiens 1989), so the spatial projections represent potential areas occupied by a species. In areas currently unoccupied, closely related species may already occur, or the projections may represent potential areas of suitable habitat to which the species has failed to disperse or in which it has previously occupied the area but had been extripated in the near or relatively distant past. Overprediction resulting from the niche-based models, however, allows for comparisons of potential and realized distributions, which may benefit natural resource managers faced with great numbers of potentially invasive species that may be effectively prioritized relative to their varying potentials to successfully invade unoccupied regions (Peterson et al. 1999; Peterson and Vieglais 2001; Anderson et al. 2002).

For the current investigation regarding risks potentially associated with biota transfers between Missouri River and Red River basins, GARP was selected as the predictive tool applied to the evaluation of species distributions potentially expressed by biota of concern when georeferenced data were sufficient to the projection. DesktopGarp is a desktop version of the GARP algorithm and is available as a software package for biodiversity and ecologic research that allows the user to predict and analyze wild species distributions. The genetic algorithm in GARP (http://biodi.sdsc.edu/ and http://beta.lifernapper.org/ desktopgarp/) is a means of generating a set of rules. Each rule, which is a model in itself, is an "if-then" statement used for making inferences about the values of the variable of interest. The sets of rules developed by GARP are more accurately described as inferential models rather than mathematical models, since inferential models are more closely related to logic than mathematics and the basic process is logical inference rather than calculation (see http://biodi.sdsc.edu/ and http://beta.lifernapper.org/

desktopgarp/). GARP presently uses four types of rules: atomic, BIOCLIM rules, range rules, and logit rules.

Atomic rules are the simplest form of rule which uses only values of variables in the precondition of the rule. An atomic rule is most useful for categorical variables such as geological attributes, soil types or geographic features. The atomic rule precondition expresses values of variables, e.g., for categorical variables X and Y: If X = x and Y = y then ... . On the other hand, BIOCLIM rules are based on the form of model used in the BIOCLIM program used by Nix (1986) for predicting the range of a species from the environmental tolerances of a species. The BIOCLIM program develops a model through enclosing the range of the environmental values of the data points where a species occurs in a statistically defined envelope, typically the 95 percentile range. The environmental envelope defined by this range encloses 95% of the data points where the species occurs. The distribution of the species is predicted at those points that fall within that environmental envelope, and absence is predicted outside those points. In GARP, then, a BIOCLIM rule is generated by developing a statistically defined envelope as briefly indicated above. A number of differences between these rules exist:

- The range of the variables is not restricted to climate; variables other than climatic variables can be used.
- The rules are evaluated for significance and accuracy after formation of the range from the percentile range. This permits an accurate estimate of the quality of the rule, which does not necessarily have a relationship to the percentile value.
- The rule can predict presence or absence of a species, but not both.
- The negation of a BIOCLIM rule can be used to predict presence or absence.

The range rule in GARP is a generalization of the BIOCLIM rule. In a range rule a number of variables may be regarded as irrelevant, which is useful when the response of a species has the form of environmental tolerances or limitations. As such, a range rule expresses intervals for each of the variables in the precondition (e.g., If $X \geq x1$ and $X \leq x2$ then ...) which lends these rules good for predicting the distribution of a species with a range of tolerance to climatic variables.

Another family of rules called logit rules is useful when species respond to the environment through environmental gradients. Logit rules are an adaptation of logistic regression models to rules, as previously characterized. As noted, logistic regression is a form of regression equation where the output is transformed into a probability. For example the logistic regression gives the output probability p where:

$$y = a_0 + a_1x_1 + a_2x_2 + \ldots + a_nx_n$$

and,

$$= \qquad p \quad \frac{1}{1+e^{-y}} \, .$$

A logistic rule, in the manner of all "if then" rules, appends an outcome (such as presence or absence) to the logistic regression and uses the probability output from the regression to determine if the rule should be applied. Thus a logistic rule produces an outcome Y with a precondition X as the rule, "If p(X) is greater than some parameter, then Y; else rule not applied." A logistic regression, then, is a form of regression equation where the output is transformed into a probability *P* that determines if a rule should be applied.

GARP reduces error in predicted distributions by both significance and predictive accuracy (Stockwell and Peters 1999). The algorithm is largely successful in doing so without overfitting which is especially important when models are based on occurrence data compiled without a fixed study design (Peterson and Cohoon 1999). Because of stochastic elements in the algorithm (such as mutation and crossing over; Holland 1975; Stockwell and Noble 1992), however, no unique solution is produced; indeed, the underdetermination of the system yields multiple solutions holding the same value for the optimization criterion. Hence, the variability among resulting models (typical of most machine-learning problems) requires careful examination of possible sources of error in order to select the most predictive models (see http://beta.lifemapper.org/desktopgarp/).

Overview of GARP implementation. GARP searches for nonrandom associations between environmental characteristics of localities of known occurrence versus those of the overall study region. It works in an iterative process of rule selection, evaluation, testing, and incorporation, or rejection to produce a heterogeneous rule-set characterizing the species' ecological requirements (Peterson et al., 1999). First, a method is chosen from a set of possibilities (e.g., logistic regression, bioclimatic rules), and it is applied to the data. Then, a rule is developed and predictive accuracy is evaluated via training points intrinsically resampled from both the known distribution and from the study region as a whole (Stockwell and Peters 1999). The change in predictive accuracy from one iteration to the next is used to evaluate whether a particular rule should be incorporated into the model (rule-set). As implemented here, the algorithm runs either 2500 iterations or until addition of new rules has no appreciable effect on the intrinsic accuracy measure (convergence). The final rule-set, or ecological-niche model, is then projected onto a digital map as the species' potential geographic distribution, exported as an ASCII raster grid, and imported into ArcView 8.3 (ESRI Corporation, Redlands, California) using the Spatial Analyst Extension for visualization.

The base environmental data comprise a variety of geographic coverages (digitized maps). For our current implementation, we used 21 environmental coverages having a pixel size of 0.04°

x 0.04° (about 4.5 km x 4.5 km). These coverages consisted of georeferenced attributes of elevation, slope, aspect soil conditions, geological ages, geomorphology, coarse potential vegetation zones, and a series of coverages for solar radiation, temperature, and precipitation. For the latter three, separate coverages representing upper and lower bounds of isopleth intervals were included (for mean annual solar radiation, mean annual temperature, mean monthly temperature in January and July, mean annual precipitation, and mean monthly precipitation in January and July). Collection localities were randomly divided into training and test data sets (50% each) for each species. We generated 20 models for each species using their respective training sets; then, the same training set is used to create each of the 20 models for a species. Test points were withheld completely from GARP's model building and internal evaluation process and were used only for evaluating final models in the "best subsets" implementation (see Anderson et al. 2003).

For each model, the elements of the confusion matrix were collected then calculated values of the correct classification rate ((a + d)/(a + b + c + d)), the intrinsic omission error (c/(a + c)) and the intrinsic commission index (b/(b + d)) (Appendix 5). In some models, GARP failed to predict every pixel as either present or absent, and such pixels were categorized as "no data" in the resultant map and reclassified as predicted absence in further geographic analyses. These unpredicted pixels do not enter into the confusion matrix. Following the "best subsets" procedure as detailed in Anderson et al. (2003), a one-tailed χ^2-statistic was applied to test data to evaluate the significance of each model against a null hypothesis of no relationship between the prediction and the test data points. In a best subsets procedure, we test whether test points fell into areas predicted present more often than expected at random, given the overall proportion of pixels predicted present versus predicted absent for that species (modified from Peterson et al. 1999). The χ^2-value represented our extrinsic composite measure of overall model performance (including contributions of both omission and commission; see Anderson et al. 2002, 2003).

As detailed by Anderson et al. (2003), we considered the potential variability of GARP outputs and generated a suite of 20 models to predict the potential distribution rather than a single model, then we subsequently queried those outputs for a consistent signal among those model outputs. The equal-weight approaches, e.g., of Anderson et al. (2002) suggested that summing the 20 models for each species (value of 1 for a pixel of predicted presence; value of 0 for predicted absence) would yield a composite map comprised of pixels with values ranging from 0 to *n*, where *n* represents the number of models of the best subset that predicted the species' presence in the pixel. For visualization of these results, maps showing various thresholds of concordance among models are presented by quartiles. In outputting the prediction of the "best subsets" composite model, a color code was used for display and ease of interpretation. Red was used for predicted presences with 75%–100% of pixel overlay of composite map when overlays were in agreement, orange for 50%–75% agreement of pixel overlay of composite map, pale green for 25%–50% agreement of pixel overlay of composite maps, and green for 0–25% agreement of pixel overlay of composite maps.

Textual output was produced by GARP, including the confusion matrix and measures of predictive accuracy. The confusion matrix tabulates the fraction of results both observed in the data and predicted by the model. The overall predictive accuracy is the accuracy of the model as a whole on an independent test set (see Appendix 5).

2.10 Cited References and Bibliography

Agresti, A., 2002, Categorical analysis, Second edition, John Wiley & Sons, Inc., New York, 710pp.

American Society for Testing and Materials, 2004, Guide for assessing the hazard of a material to aquatic organisms and their uses, E1023, Annual Book of Standards, Volume 11.05, Biological effects and environmental fate, American Society for Testing and Materials, West Conshohocken PA.

Anderson, R. P., Lew, D. & Peterson, A. T., 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models, Ecological Modelling, 162: 211-232.

Anderson, R. P., M. Laverde, and A. T. Peterson, 2002, Geographical distributions of spiny pocket mice in South America: Insights from predictive models. Global Ecology and Biogeography 11:131-141.

Arndt, C., 2001, Information measures, Springer, Berlin, 547pp.

Ashton, P.J. and D.S. Mitchell, 1989, Aquatic plants: Patterns and modes of invasion, attributes of invading species and assessment of control programmes. In Drake, J.A., H.A. Mooney, F. di Castri, R.H. Groves, F.J. Kruger, M. Rejmánek and M. Williamson (Eds.). Biological invasions: A global perspective. John Wiley & Sons, Ltd., New York. Pp. 111-154.

Aven, T., 2003, Foundations of risk analysis, John Wiley & Sons, Ltd., Chichester, UK, 190pp.

Barlow, R.E., 1998, Engineering reliability, American Statistical Society, Alexandria, Virginia, and Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 199pp.

Bartholomew, J.L., and J.C. Wilson (eds.), 2002, Whirling disease: Reviews and current topics, American Fisheries Society Symposium 29, American Fisheries Society, Bethesda, Maryland, 247pp.

Bartlett, M.S., 1960, Stochastic population models in ecology and epidemiology, Methuen & Co., Ltd., London, UK, 90pp.

Bedford, T., and R. Cooke, 2001, Probabilistic risk analysis, Cambridge University Press, Cambridge, UK, 393pp.

Blischke, W.R., and D.N. Parbhakar Murthy, 2000, Reliability, John Wiley & Sons, Inc., New York, 812pp.

Borgelt, C., and R. Kruse, 2002, Graphical models, John Wiley & Sons, Ltd., Chichester, UK, 358pp.

Burgman, M., 2005, Risks and decisions for conservation and environmental management, Cambridge University Press, Cambridge, UK, 488pp.

Burnham, K.P., and D.R. Anderson, 2002, Model selection and multimodel inference, Second edition, Spring-Verlag, Inc., New York, 488pp.

Campbell, F, and P. Kreisch, 2003, Invasive species pathway team: Final report, National Invasive Species Council, Washington, D.C., 25pp.

Chen, G., and A. T. Peterson, 2000, A new technique for predicting distributions of terrestrial vertebrates using inferential modeling, Zoological Research 21:231-237.

Chen, X., 2001, Data mining and uncertain reasoning, John Wiley & Sons, Inc., New York, 370pp.

Cressie, N., 1993, Statistics for Spatial Data (rev. ed.). Wiley, New York, 900pp.

Daley, D.J., and J. Gani, 1999, Epidemic modeling, Cambridge University Press, Cambridge, UK, 213pp.

Davison, A.C., and D.V. Hinkley, 1997, Bootstrap methods and their applications, Cambridge University Press, Cambridge, UK, 582pp.

Doppelt, B., M. Scurlock, C. Frissell, and J. Karr, 1993, Entering the watershed, Island Press, Washington, D.C., 462pp.

Downes, B.J., L.A. Barmuta, P.G. Fairweather, D.P. Faith, M.J. Keough, P.S. Lake, B.D. Mapstone, and G.P. Quinn, 2002, Monitoring ecological impacts, Cambridge University Press, Cambridge, UK, 434pp.

Efron, B., 1982, The jackknife, the bootstrap, and other resampling plans, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 92pp.

Ehrlich P.R., 1976, Which animal will invade? In: Ecology of Biological Invasions of North America and Hawaii (Eds H.A. Mooney & J.A. Drake), pp. 79–95, Springer-Verlag, New York.

Elliot, P., J. Wakefield, N. Best, and D. Briggs, 2000, Spatial epidemiology, Methods and applications, Oxford University Press, Oxford, UK, 475pp.

Elton, C.S., 1958, The ecology of invasions by plants and animals. The University of Chicago Press, Chicago, Illinois, 181pp.

Embrey, M.A., R.T. Parkin, and J.M Balbus, 2002, Handbook of CCL microbes in drinking water, American Water Works Association, Denver, Colorado, 436pp.

Environmental Protestion Agency (EPA), 2003, Framework for Cumulative Risk Assessment, EPA/630/P-02/001F, Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, D.C.

EPA, 1998, Guidelines for Ecological Risk Assessment. EPA/630/R-95/002F, Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, D.C.

EPA, 1992, Framework for ecological risk assessment. EPA/630/R-92/001, Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, D.C.

FAO, 2002, Evaluation of introductions. http://www.fao.org/docrep/x5628/x5628e06.htm.

Feria, T. P., and A. T. Peterson, 2002, Using point occurrence data and inferential algorithms to predict local communities of birds, Diversity and Distributions 8:49-56.

Fielding, A., 1999, Machine learning methods for ecological applications, Kluwer Academic Publishers, Boston, Massachusetts, 261pp.

Foran, J.A. and S.A. Ferenc (eds.), 1999, Multiple stressors in ecological risk and impact assessment, SETAC Press, Pensacola, Florida, 100pp.

Gamerman, D., 1997, Markov chain Monte Carlo, Chapman & Hall/CRC, Boca Raton, Florida, 245pp.

Goodwin, B. J., A. J. McAllister and L. Fahrig, 1998, Predicting invasiveness of plant species based on biological information, Conservation Biology 13(2):422-426.

Greenblatt, C., and M. Spigelman (eds.), 2003, Emerging pathogens, Oxford University Press, Oxford, UK, 250pp.

Gunderson, L.H., C.S. Holling, and S.S. Light (eds.), 1995, Barriers and bridges to the renewal of ecosystems and institutions, Columbia University Press, New York, 587pp.

Halliwell, D. B., R. W. Langdon, R. A. Daniels, J. P. Kurtenbach, and R. A. Jacobson, 1999, Classification of freshwater fish species of the Northeastern United States for use in the development of indices of biological integrity with regional applications. Pages 301–338 in T. P. Simon (ed.), Assessing the sustainability and biological integrity of water resources using fish communities, CRC Press, New York.

Hayes, K.R., 1998, Bayesian statistical inference in ecological risk assessment. Technical Report Number 17, Centre for research on introduced marine pests. Australia.

Haupt, R.L., and S.E. Haupt, 2004, Practical genetic algorithms, Second Edition, Wiley-Interscience, A John Wiley & Sons, Inc., Publication, New York, 253pp.

Haupt, R.L., and S.E. Haupt, 1998, Practical genetic algorithms, Wiley-Interscience, A John Wiley & Sons, Inc. Publication, New York, 177pp.

Heinz Center, 2002, The state of the nation's ecosystems, The H. John Heinz Center for Science, Economics, and the Environment, published by Cambridge University Press, Cambridge, UK, 270pp.

Higgins, S. I., D. M. Richarson, R. M. Cowling and T.H. Trinder-Smith, 1999, Predicting the landscape-scale distribution of alien plants and their threat to plant diversity. Conservation Biology 13(2):303-313.

Holland, J.H., 1975, Adaptation in natural and artificial systems, The MIT Press, Cambridge, Massachusetts, 211pp.

Holling, C.S. (ed.), 1978, Adaptive environmental assessment and management, International Institute for Applied Systems Analysis, published by John Wiley & Sons, Inc., New York, 377pp.

Hosmer, D.W., and S. Lemeshow, 2000, Applied logistic regression, John Wiley & Sons, Inc., New York, 375pp

Hutchinson, G.E., 1957, Concluding remarks, In Cold Spring Harbour symposium on quantitative biology, Vol. 22, pp. 415-427.

Huzurbazar, A.V., 2005, Flowgraph models for multistate time-to-event data, John Wiley & Sons, Inc., New York, 270pp.

Jensen, M.E., and P.S. Bourgeron (eds.), 2001, A guidebook for integrated ecological assessments, Springer-Verlag, Inc., New York, 536pp.

Johnson, L.E., A. Ricciardi, and J.T. Carlton, 2001, Overland dispersal of aquatic invasive species: A risk assessment of transient recreational boating, Ecological Applications 11:1789-1799.

Johnson, L.E. and J.T. Carlton, 1996, Post-establishment spread in large-scale invasions: dispersal mechanisms of the zebra mussel (*Dreissena polymorpha*), Ecology 77:1686-1690.

Johnson, L.E. and D.K. Padilla, 1996, Geographic spread of exotic species: ecological lessons and opportunities from the invasion of the zebra mussel (*Dreissena polymorpha*), Biological Conservation 78:22-33.

Kolar, C.S., and D. M. Lodge, 2002, Ecological predictions and risk assessment for alien fishes in North America, Science, 298:1233-1236.

Lawson, A., Biggeri, A., Böhning, D., Lesaffre, E., Viel, J-f. and Bertollini, R., 1999, Disease mapping and Risk Assessment for Public Health. Wiley, Chichester.

Lawson, A.B., 2001, Statistical methods in spatial epidemiology, John Wiley & Sons, Ltd., Chichester, UK, 277pp.

Levin, S.A., 1989, Analysis of risk for invasions and control programs. In Drake, J.A., H.A. Mooney, F. di Castri, R.H. Groves, F.J. Kruger, M. Rejmánek and M. Williamson (Eds.). Biological invasions: A global perspective. John Wiley & Sons, Ltd., New York. Pp. 425-435.

Li, H.W., 1981, Ecological analysis of species introductions into aquatic systems. Transactions of American Fisheries Society 110:772-782.

MacArthur, R.H., 1968, The theory of the niche. In Population Biology and Evolution (R.C. Lewontin, ed.), Syracuse University Press, Syracuse, New York, pp. 159-176.

MacKenzie, S.H., 1996, Integrated resource planning and management, Island Press, Washington, D.C., 243pp.

Manly, B.F.J., 1991, Randomization, bootstrap, and Monte Carlo methods in biology, Second Edition, Chapman & Hall, London, UK, 399pp.

Marchetti,M.P., Moyle, P.B., and Levine, R, 2004, Alien fishes in California watersheds: characteristics of successful and failed invaders, Ecological Applications, 14(2):587–596

Margoluis, R., and N. Salafsky, 1998, Measures of success: Designing, managing, and monitoring conservation and development projects, Island Press, Washington, D.C., 362pp.

Minnesota Sea Grant/Michigan Sea Grant, 2001, Hazard analysis and critical control point analysis. MN SG-F11/MSG-00-400. University of Minnesota, Duluth MN.

Moyle, P.B. and T. Light, 1996, Biological invasions of fresh water: Empirical rules and assembly theory. Biological Conservation 78:149-161.

National Oceanic and Atmospheric Administration (NOAA), 1997, Natural Resource Damage Assessment Guidance Document: Scaling Compensatory Restoration Actions (Oil Pollution Act of 1990). Damage Assessment and Restoration Program. Silver Spring, MD.

NRC, 2002. Predicting Invasions of Nonindigenous Plants and Plant Pests. National Academy Press, Washington, D.C. 194p.

National Invasive Species Council (NISC), 2001, Meeting the Invasive Species Challenge: National Invasive Species Management Plan. 80 pp.

Nix, H.A., 1986, A biogeographic analysis of Australian elapid snakes. In: Atlas of Elapid Snakes of Australia. (Ed.) R. Longmore, pp. 4-15. Australian Flora and Fauna Series Number 7. Australian Government Publishing Service: Canberra.

NIST/SEMATECH, 2004, e-Handbook of Statistical Methods, available through World-wide Web http://www.itl.nist.gov/div898/handbook/, last accessed November 11, 2004.

Noga, E.J., 1996, Fish disease, Mosby, St. Louis, Missouri, 367pp.

Office of the President, 1999, Executive Order 13112, February 3, 1999, Established Invasive Species Council and specified its duties.

Paine, R.T., M.J. Tegner, and E.A. Johnson, 1998, Compounded perturbations yield ecological surprises. Ecosystems 1:535-545.

Peterson, A. T., L. G. Ball, and K. C. Cohoon, 2002a, Predicting distributions of tropical birds. Ibis 144:e27-e32.

Peterson, A. T., M. A. Ortega-Huerta, J. Bartley, V. Sanchez-Cordero, J. Soberon, R. H. Buddemeier, and D. R. B. Stockwell, 2002b, Future projections for Mexican faunas under global climate change scenarios. NATURE 416:626-629.

Peterson, A. T., V. Sanchez-Cordero, C. B. Beard, and J. M. Ramsey, 2002c, Ecologic niche modeling and potential reservoirs for Chagas disease, Mexico. Emerging Infectious Diseases 8:662-667.

Peterson, A. T., D. R. B. Stockwell, and D. A. Kluza, 2002d, Distributional prediction based on ecological niche modeling of primary occurrence data. Pages 617-623 in J. M. Scott, P. J. Heglund, and M. L. Morrison (eds.), Predicting Species Occurrences: Issues of Scale and Accuracy. Island Press, Washington, D.C.

Peterson, A. T., 2001, Predicting species' geographic distributions based on ecological niche modeling. Condor 103:599-605.

Peterson, A. T., and D. A. Vieglais, 2001, Predicting species invasions using ecological niche modeling. BioScience 51:363-371.

Peterson, A. T., V. Sanchez-Cordero, J. Soberon, J. Bartley, R. H. Buddemeier, and A. G. Navarro-Siguenza, 2001, Effects of global climate change on geographic distributions of Mexican Cracidae. Ecological Modelling 144:21-30.

Peterson, A. T., and K. C. Cohoon, 1999, Sensitivity of distributional prediction algorithms to geographic data completeness. Ecological Modelling 117:159-164.

Peterson, A. T., J. Soberon, and V. Sanchez-Cordero, 1999, Conservatism of ecological niches in evolutionary time. Science 285:1265-1267.

Puccia, C.J., and R. Levins, 1985, Qualitative modeling of complex systems, Harvard University Press, Cambridge, Massachusetts, 259pp.

Rejmánek, M. and D. D. Richardson, 1996, What attributes make some plant species more invasive? Ecology 77(6):1655-1660.

Rejmánek, M., 2000, Invasive plants: approaches and predictions. Austral Ecology 25:497-506.

Ricciardi, A., and J.B. Rasmussen, 1998, Predicting the identity and impact of future biological invaders: A priority for aquatic resource management. Canadian Journal of Fisheries and Aquatic Sciences 55:1759-1765.

Richards, D. C., Cazier, L.D., and G. T. Lester, 2001, Spatial distribution of three snail species including the invader Potamopyrgus antipodarum in a freshwater spring. Western North American Naturalist. 61(3): 375-380.

Sait, S.M., and H. Youssef, 1999, Iterative computer algorithms with applications in engineering, The Institute of Electrical and Electronic Engineers, Inc., The Computer Society, Los Alamitos, California, 387pp.

Salthe, S.N., 1985, Evolving hierarchical systems, Columbia University Press, New York, 343pp.

Scott, J.M., P.J. Heglund, M.L. Morrison, J.B Haufler, M.G. Raphael, W.A. Wall, and F.B. Samson (eds.), 2002, Predicting species occurrences, Issues of accuracy and scale, Island Press, Washington, D.C., 868pp.

Scott, J.M., David F., Csuti B., Noss R., Butterfield B., Groves C., Anderson H., Caicco S., D'erchia F., Edwards Jr. T.C., Ulliman J., and Wright R.G., 1993, Gap Analysis: A geographic approach to protection of biological diversity. Wildlife Monographs, No 123. The Wildlife Society, Inc. 1993. Blacksburg, VA 24061.

Serrano, S., 2001, Engineering uncertainty and risk analysis, HydroScience, Inc., Lexington, Kentucky, 456pp.

Shigesada, N. and K. Kawasaki, 1997, Biological invasions: Theory and practice. Oxford University Press, Oxford UK.

Spall, J.C., 2003, Introduction to stochastic search and optimization, John Wiley & Sons, Inc., New York, 595pp.

Stockwell, D. R. B., and D. P. Peters, 1999, The GARP modelling system: Problems and solutions to automated spatial prediction. International Journal of Geographic Information Systems 13:143-158.

Stockwell, D. R. B., and I. R. Noble, 1992, Induction of sets of rules from animal distribution data: A robust and informative method of analysis, Mathematics and Computers in Simulation 33:385-390.

Stockwell, D.R.B. and I.R. Noble., 1991, Induction of sets of rules from animal distribution data: a robust and informative method of data analysis. Mathematics and Computers in Simulation. 32:249-254.

Suter, G.W., 1993, Ecological risk assessment. Lewis Publishers, Boca Raton, Florida, 538pp.

Walters, C., 1986, Adaptive management of renewable resources, republished by The Blackburn Press, 2001, Caldwell, New Jersey, 374pp.

Weins, J.A., 1986, Spatial scale and temporal variation in studies of shrubsteppe birds In Community Ecology (eds Diamond, J. and T.J. Case), pp154-172, Harper & Row Pub. Inc., New York.

Williamson, M. 1989. Mathematical models of invasion. In Drake, J.A., H.A. Mooney, F. di Castri, R.H. Groves, F.J. Kruger, M. Rejmánek and M. Williamson (Eds.). Biological invasions: A global perspective. John Wiley & Sons, Ltd., New York. Pp. 329-350.

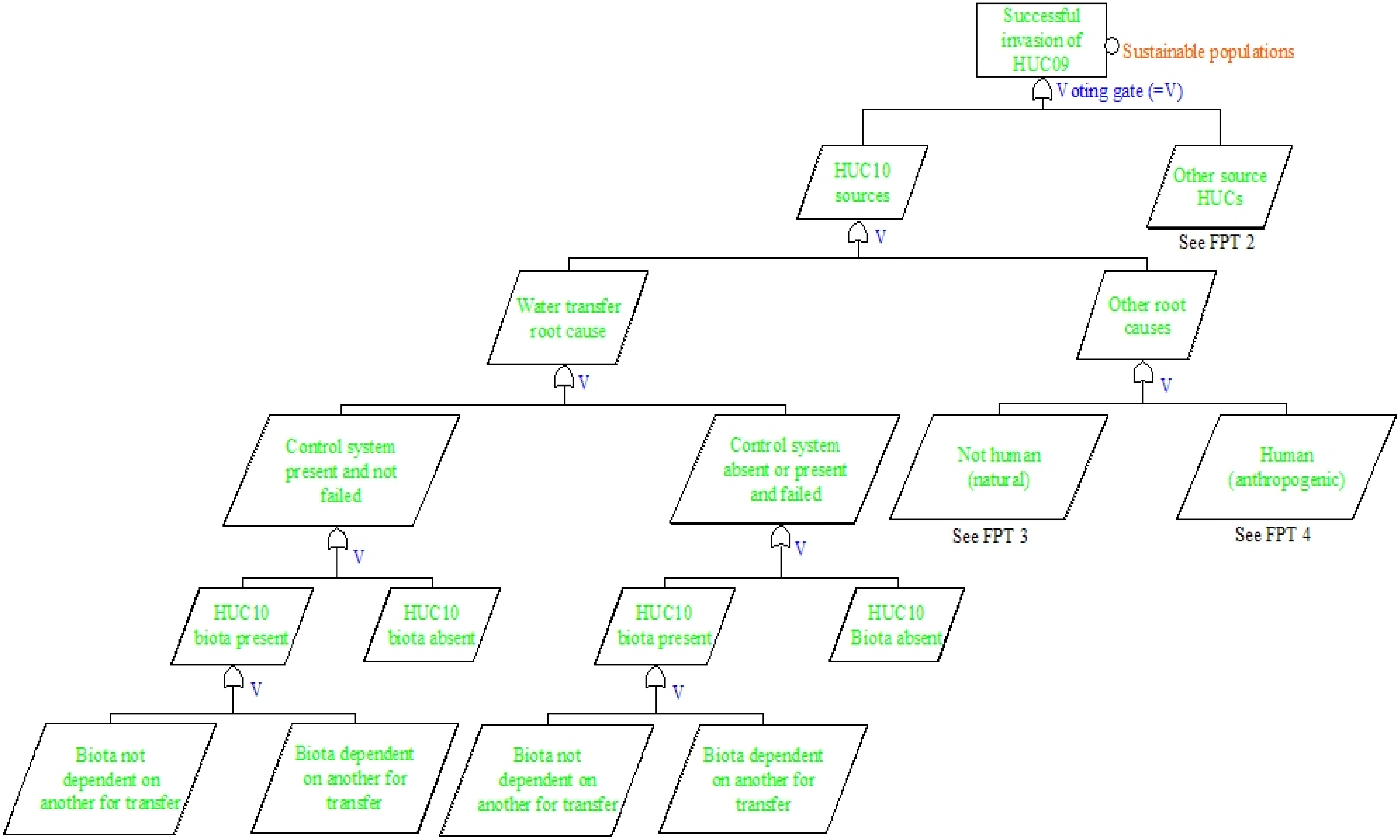
Williamson, M. 1996. Biological invasions. Population and Community Biology Series 15, Chapman & Hall. London, UK, 244pp.

Wolkenhauer, O., 2001, Data engineering, John Wiley & Sons, Inc., New York, 263pp.

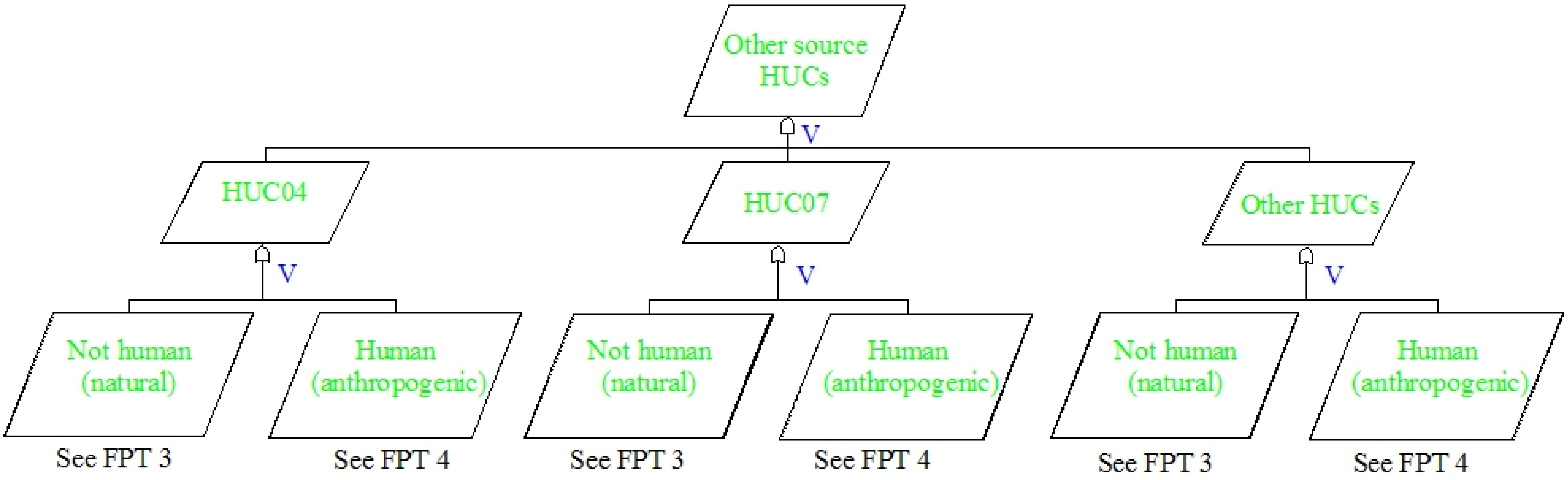
Yan, N.D., A. Blukacz, W.G. Sprules, P.K. Kindy, D. Hackett, R.E. Girard, and B.J. Clark, 2001. Changes in Zooplankton and the Phenology of the Spiny Water Flea, Bythotrephes, Following its Invasion of Harp Lake, Canadian Journal of Fisheries and Aquatic Sciences 58(12): 2341-2350.

Zar, J.H., 1999, Biostatistical analysis. Fourth Edition. Prentice-Hall. Upper Saddle River, New Jersey, 663pp plus appendices and index.

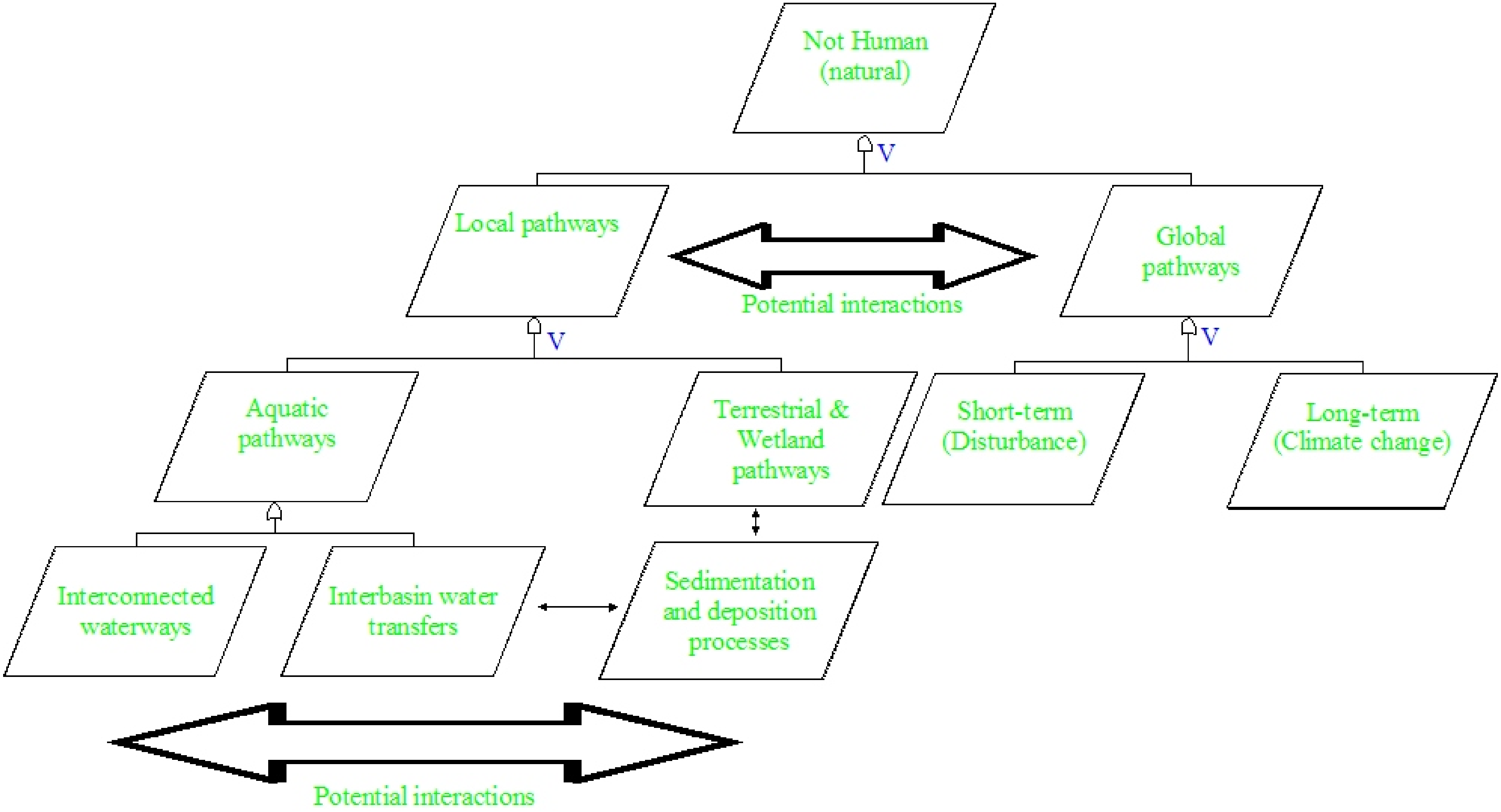
Annex, Figure 1. Fault-probability tree (FPT) 1


Successful invasion of HUC09
Sustainable populations
Voting gate (=V)
HUC10 sources
Other source HUCs
See FPT 2
V
Water transfer root cause
Other root causes
V
V
Control system present and not failed
Control system absent or present and failed
Not human (natural)
See FPT 3
Human (anthropogenic)
See FPT 4
V
V
HUC10 biota present
HUC10 biota absent
HUC10 biota present
HUC10 Biota absent
V
V
Biota not dependent on another for transfer
Biota dependent on another for transfer
Biota not dependent on another for transfer
Biota dependent on another for transfer

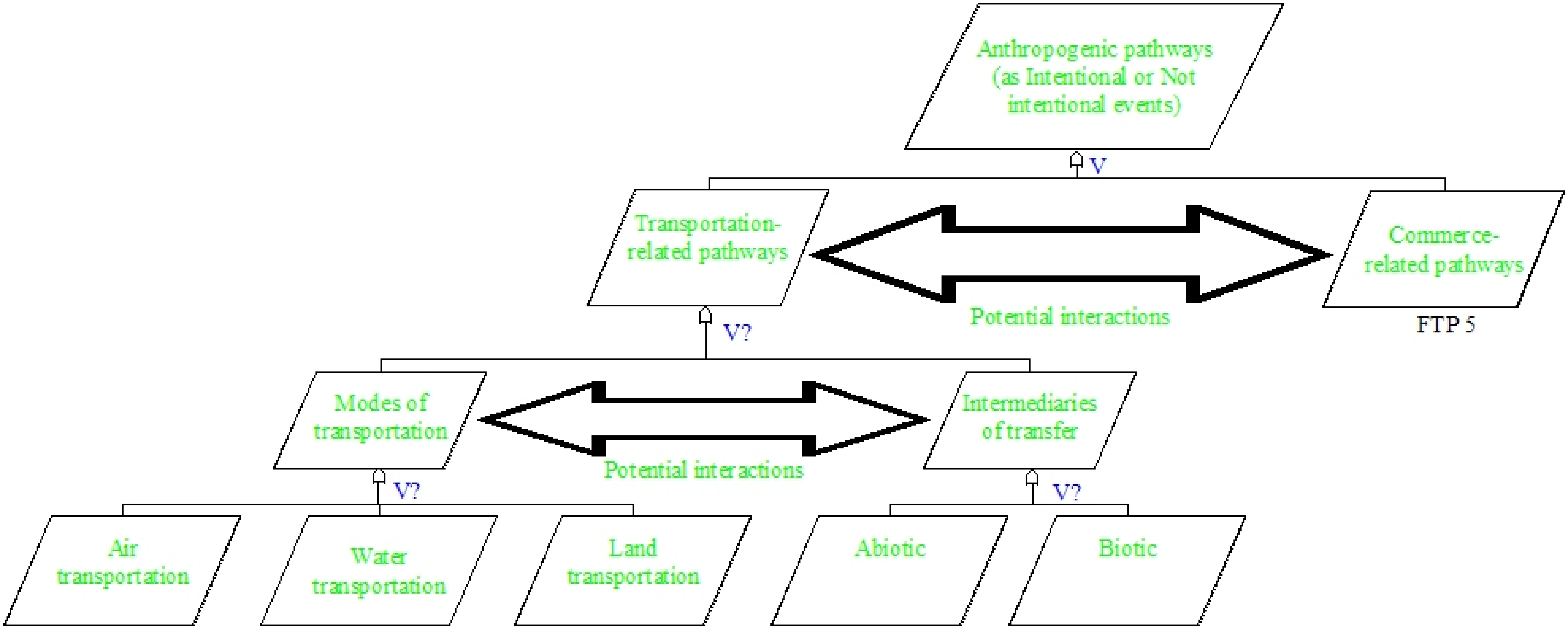
Annex, Figure 2. FPT 2


Other source HUCs
V
HUC04
HUC07
Other HUCs
V
V
V
Not human (natural)
Human (anthropogenic)
Not human (natural)
Human (anthropogenic)
Not human (natural)
Human (anthropogenic)
See FPT 3
See FPT 4
See FPT 3
See FPT 4
See FPT 3
See FPT 4

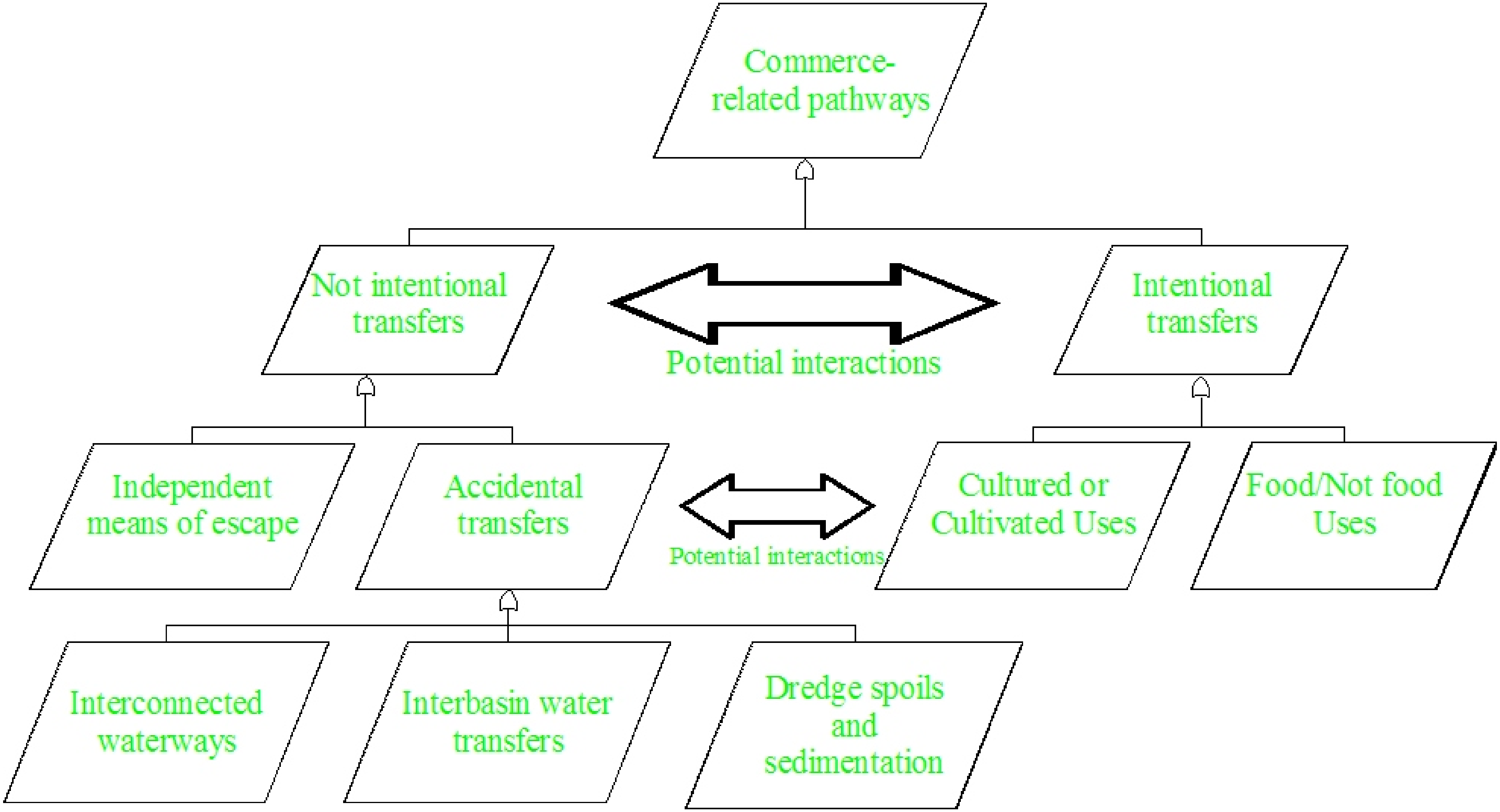
Annex, Figure 3. FPT 3


Not Human
(natural)
V
Local pathways
Potential interactions
Global
pathways
V
V
Aquatic
pathways
Terrestrial &
Wetland
pathways
Short-term
(Disturbance)
Long-term
(Climate change)
Interconnected
waterways
Interbasin water
transfers
Sedimentation
and deposition
processes
Potential interactions

Annex, Figure 4. FPT 4


Anthropogenic pathways (as Intentional or Not intentional events)
V
Transportation-related pathways
Potential interactions
Commerce-related pathways
FTP 5
V?
Modes of transportation
Potential interactions
Intermediaries of transfer
V?
V?
Air transportation
Water transportation
Land transportation
Abiotic
Biotic

Annex, Figure 5. FPT 5


Commerce-related pathways
Not intentional transfers
Potential interactions
Intentional transfers
Independent means of escape
Accidental transfers
Potential interactions
Cultured or Cultivated Uses
Food/Not food Uses
Interconnected waterways
Interbasin water transfers
Dredge spoils and sedimentation