Field data to predictions with statistics in between: linking Eastern waterfan observations to landscape attributes

Atlantic Canada Species At Risk Habitat

Modelling Seminar Presentation

Kellina L. Higgins, M.Sc., GeoFlora

Project team: Atlantic Canada Conservation Data Centre (ACCDC)

and Parks Canada (PC)

February 24th, 2023

Photo credits: Neil Vinson (Parks Canada) and Colin Chapman (ACCDC)

Report

Eastern Waterfan (*Peltigera hydrothyria*) occurrence and abundance analysis for Fundy National Park, New Brunswick. March 31, 2021. Report submitted by Kellina L. Higgins, Sean Blaney, Charity Robicheau & James Churchill, Atlantic Canada Conservation Data Centre to Parks Canada

Field crews: A.G. Belliveau, M. Baxley, C. Chapman, J. Klymko, D. Mazerolle, C. Porter, & N. Vinson

Description

Ecological field data can come in many forms: systematic experimental designs with set variables recorded (ecological studies), opportunistic observations (citizen science), surveys to detect species (botanical surveys) and more. Each form of data is valuable in its own way yet has it own limits. Sometimes data recorded for one purpose can be used in another context. Here, data collected with the objective of detecting populations of Eastern Waterfan (*Peltigera hydrothyria*) was used in predictive modelling to estimate the probability of its occurrence and its density across the landscape in and around Fundy National Park.

Methods: Field data was collected by the Atlantic Canada Conservation Data Centre (ACCDC) and Parks Canada supplemented by other occurrences contributed to ACCDC. In addition to observations of populations, GPS tracks representing surveys were used to estimate absence records. Landscape metrics derived from remote-sensing and photo-interpreted forest inventory data were used in the analysis to determine key habitat attributes : stream size, elevation, slope, stream aspect, canopy height, stand composition, stand age, crown closure, distance to roads, and distance to harvest cut-blocks. Conditional inference trees, logistic regression and multiple regression were used to analyze the data with the statistical software R in order to predict its presence and abundance.

Results: Waterfan was found in medium-size streams far from roads and harvest cut-blocks at higher elevations. Its presence was also influenced by the stand composition. Larger colonies were found in steeper sections of streams with northern and eastern flow orientations. The probability of water occurrence and its abundance was inferred based on these landscape attributes to identify potential waterfan hotpots in other streams.

Discussion: The field notes recording variables such as dominant canopy species, stream substrate, stream speed and percent canopy cover could not be included in any of the analyses due to inconsistencies in data collection between field crews and large data gaps. In addition, biases may have been introduced in the density estimates given that the methodology differed between field crews and by size of colonies. Thus, field data collection could be improved to adopt a common methodology to estimate population sizes and to record field attributes in order to better understand habitat requirements and develop more accurate predictive models.

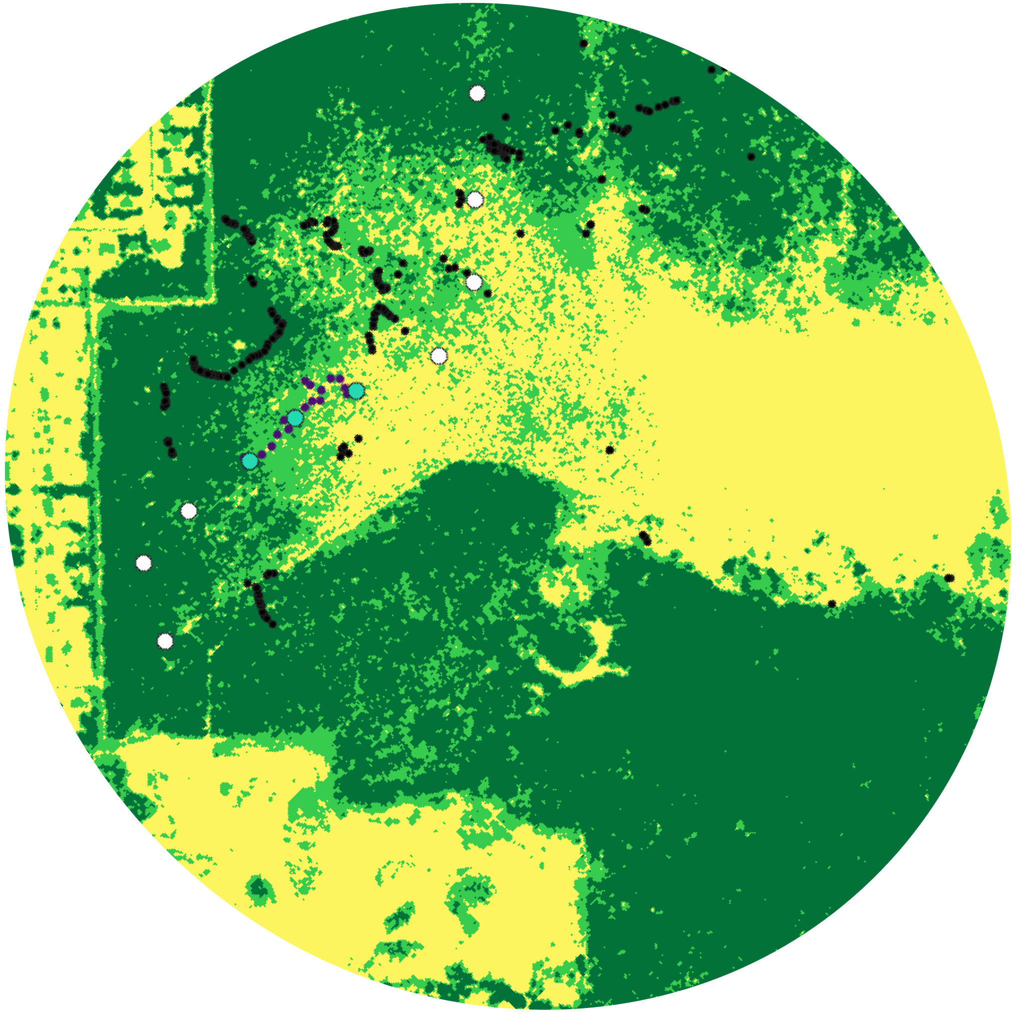
Themes

Types of field surveys	Data collection methods
Eastern waterfan surveys in and around Fundy National Park	Locally abundant rare species
Lanscape parameters influencing waterfan presence and density	Statistical analysis techniques • Conditional inference tree, logistic regression and multiple linear regression
Botanical survey notes describing habitat	Challenges using botanical survey notes for quantitative analysis
Optimization botanical field surveys	Learning from analysis challenges to build better models

Field data types

systematic experimental designs with set variables recorded	ecological studies
surveys to detect species	botanical surveys
opportunistic observations	citizen science

Data quality vs quantity



Field data types

consistency

systematic experimental designs with set variables recorded	ecological studies
surveys to detect species	botanical surveys
opportunistic observations	citizen science

more species records

Eastern Waterfan (*Peltigera hydrothyria*)

COSEWIC - Threatened

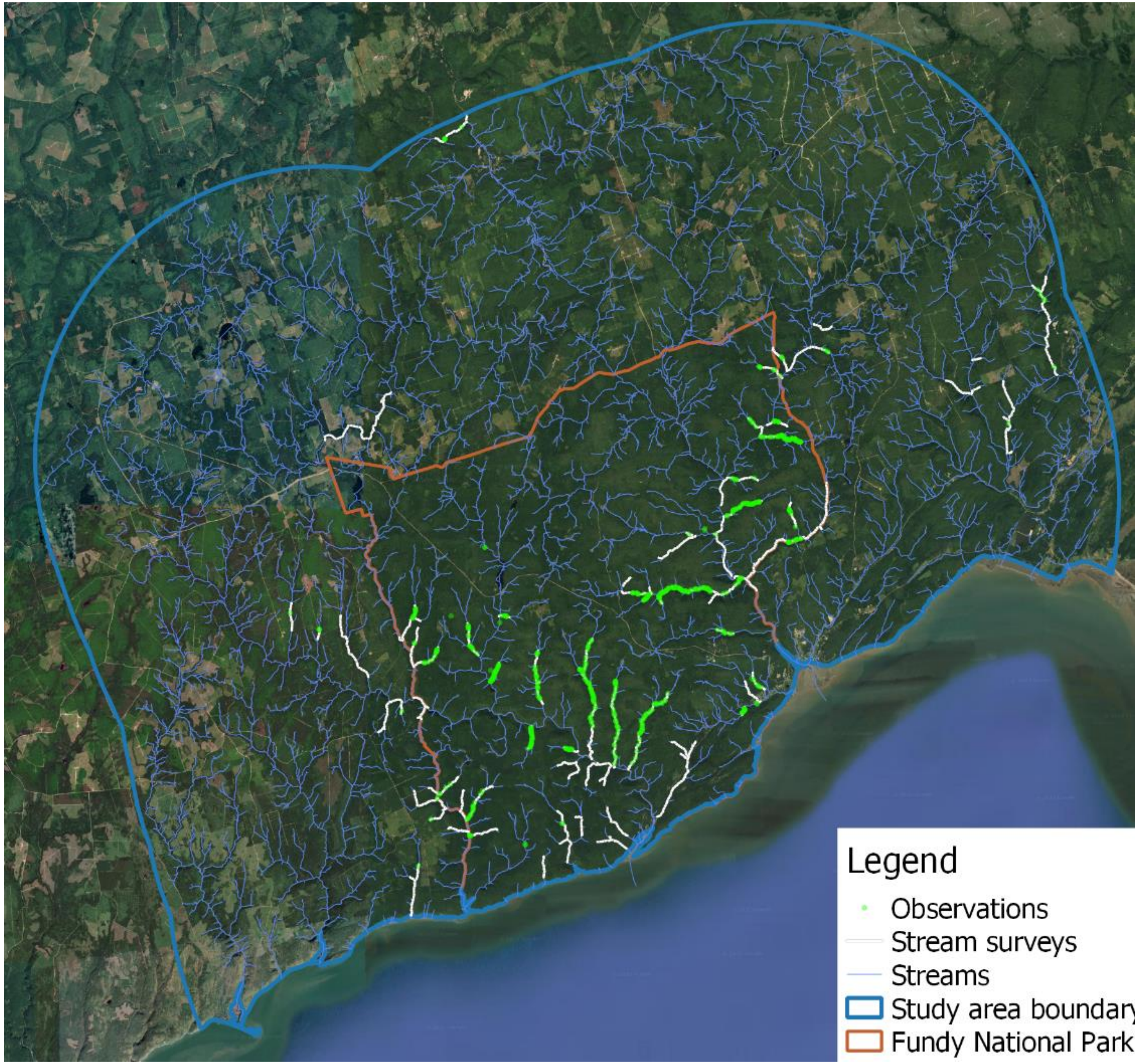
Habitat

- Small streams
- Covers rocks
- Low sedimentation
- Shaded areas

Botanical field surveys

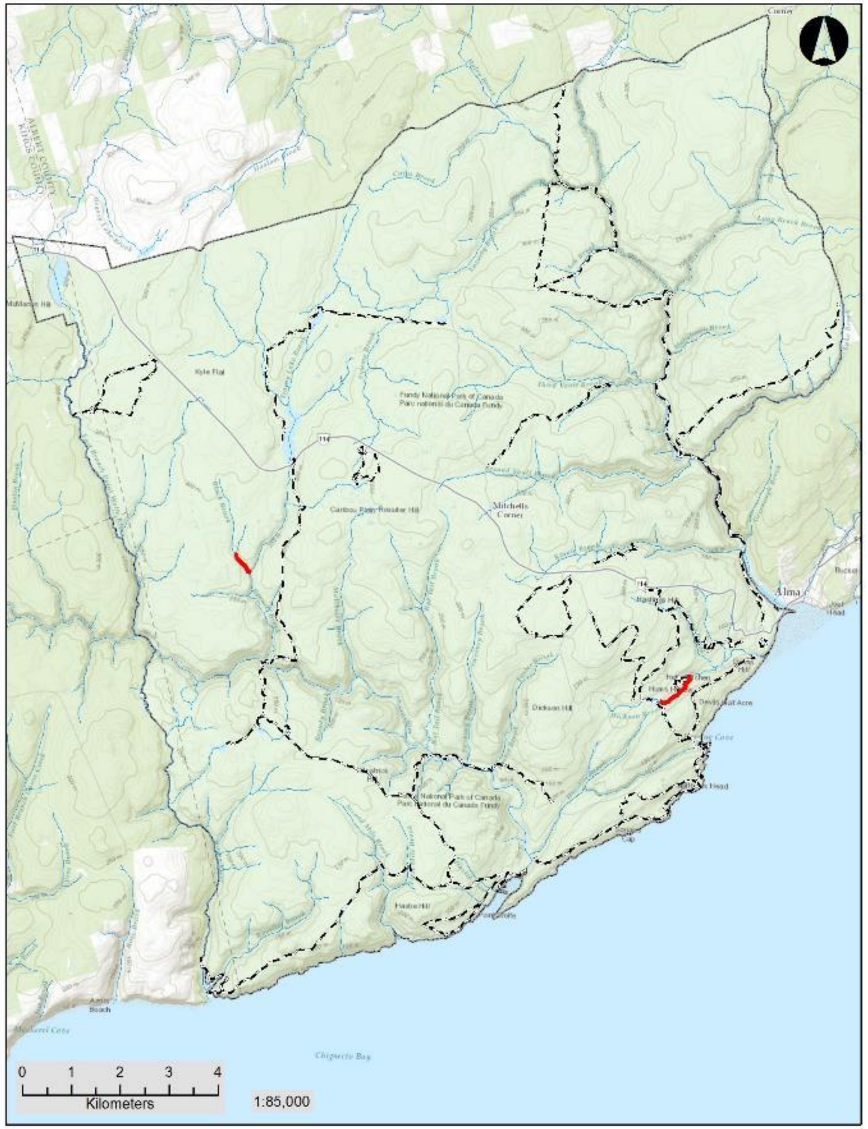
- Walk the brooks
- Estimated surface area of colonies
- Extrapolation for streams with abundant waterfan
- Field seasons 2019-2020

Eastern Waterfan observations (green points) in study area within Fundy National Park: 750 records

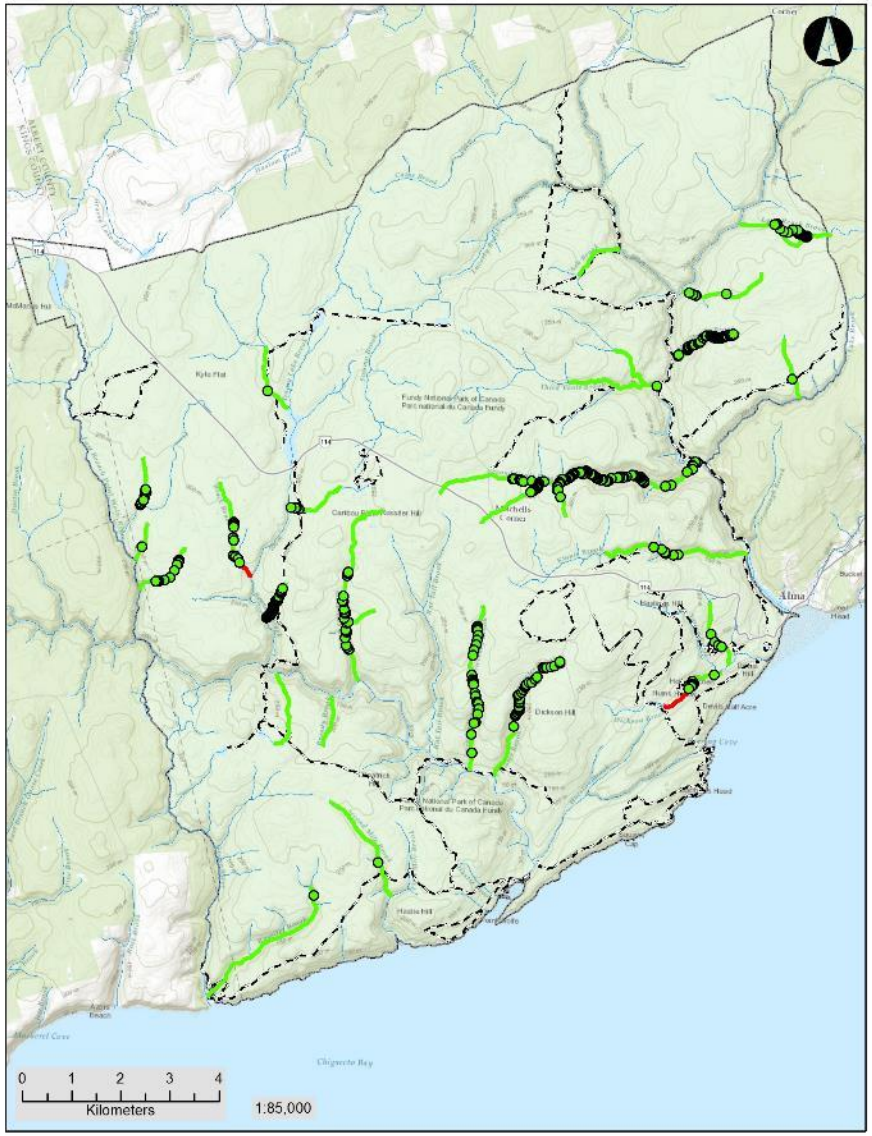


Extent in Fundy NP expanded based on field surveys in 2019 and 2020 by ACCDC and PC

2013:
2 brooks COSEWIC

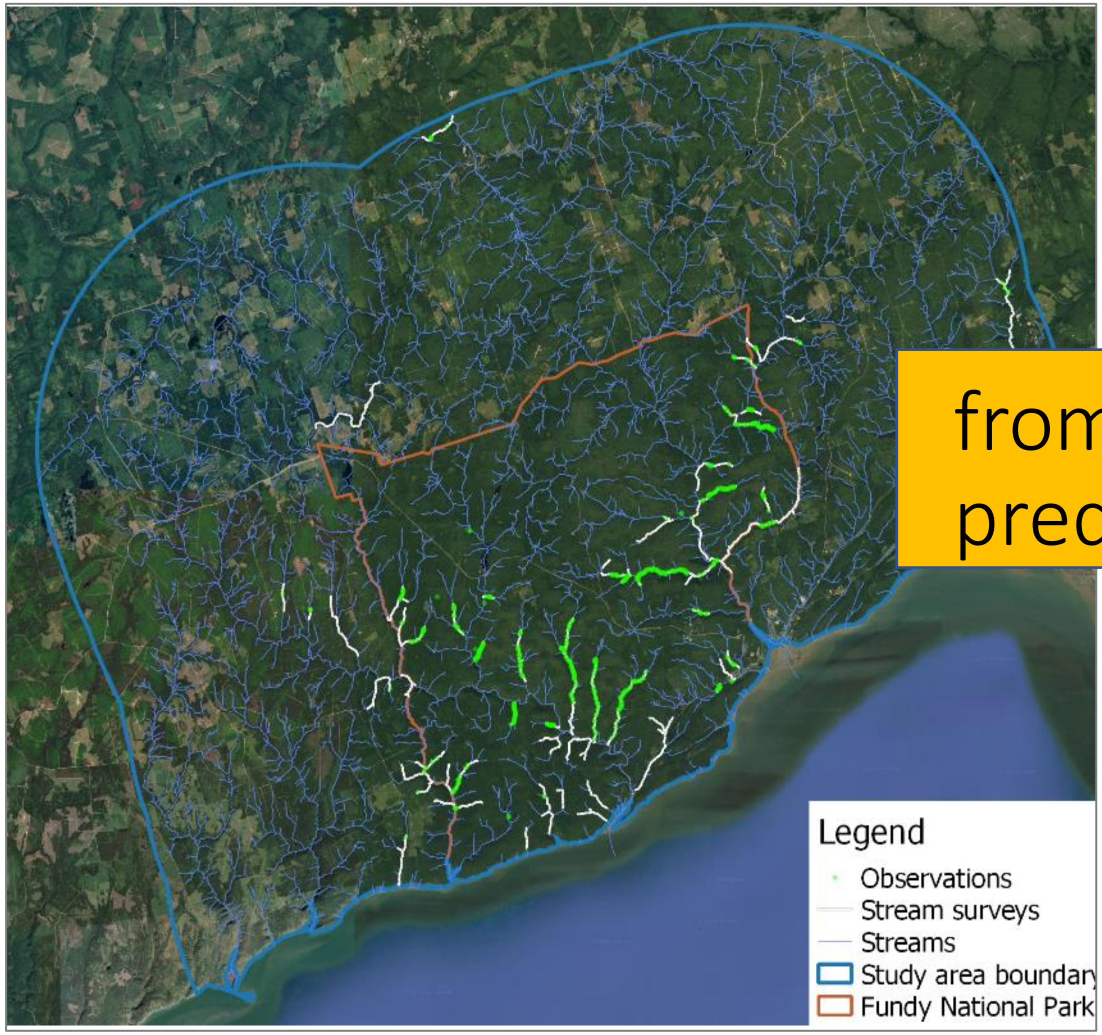


2021:
28 brooks in FNP

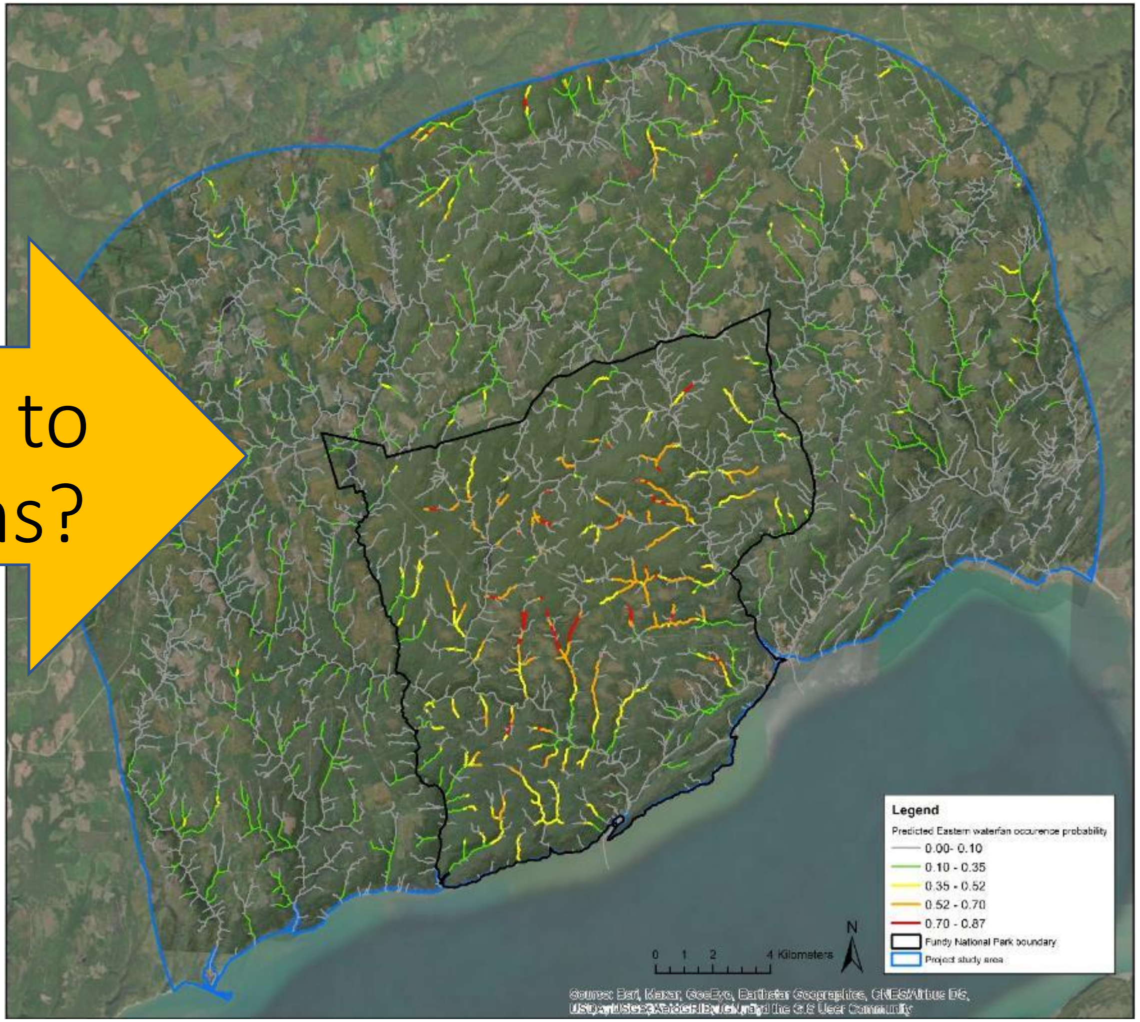


Surveys detected several massive populations within Fundy National Park!

Next steps: Could we use field records to make generalizations about waterfan habitat at the small-scale? Could we use observations and survey routes to make predictions about potential waterfan hotpots elsewhere in the park?



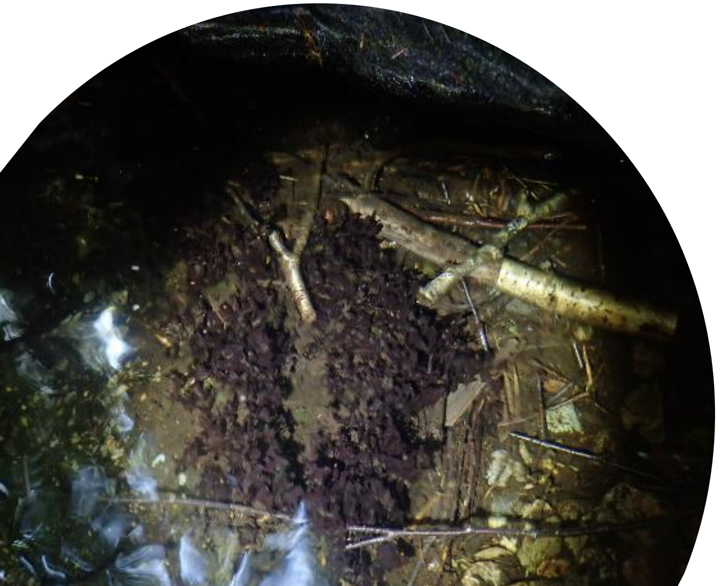
from data to predictions?

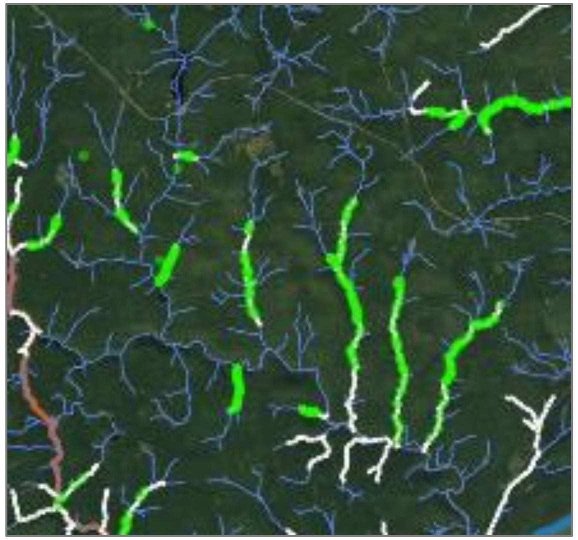


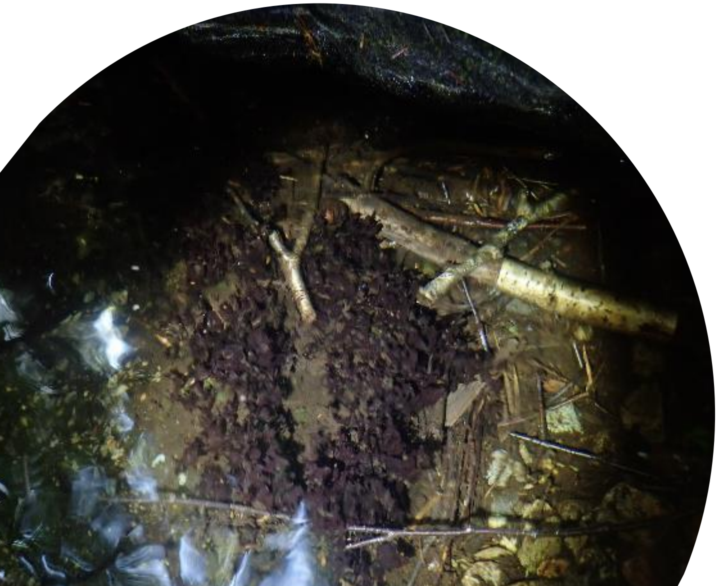
Estimating waterfan abundance along linear features

Botanical field surveys

- Walk the brooks
- Estimated surface area of colonies
- Extrapolation for streams with abundant waterfan
- Field seasons 2019-2020



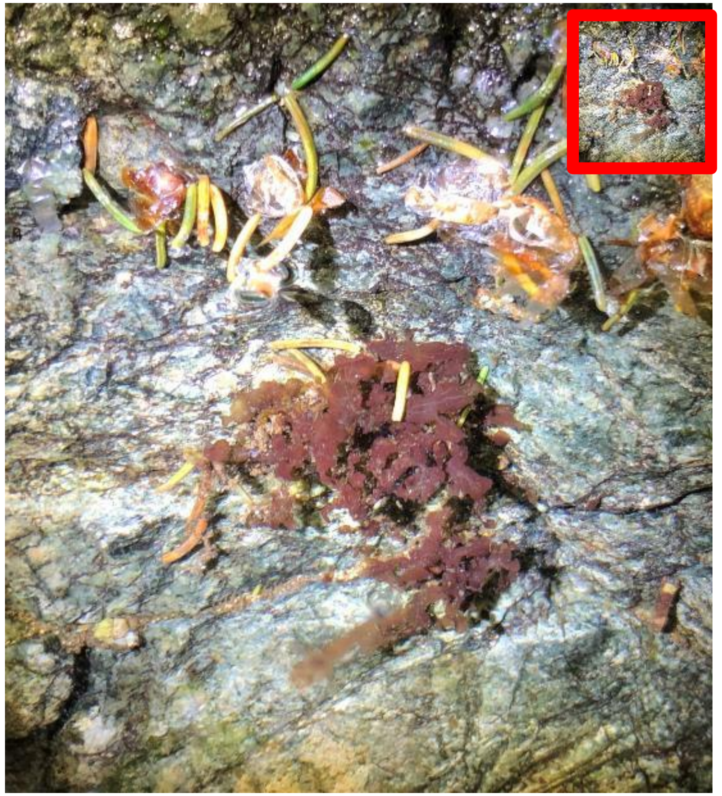




Botanical field surveys

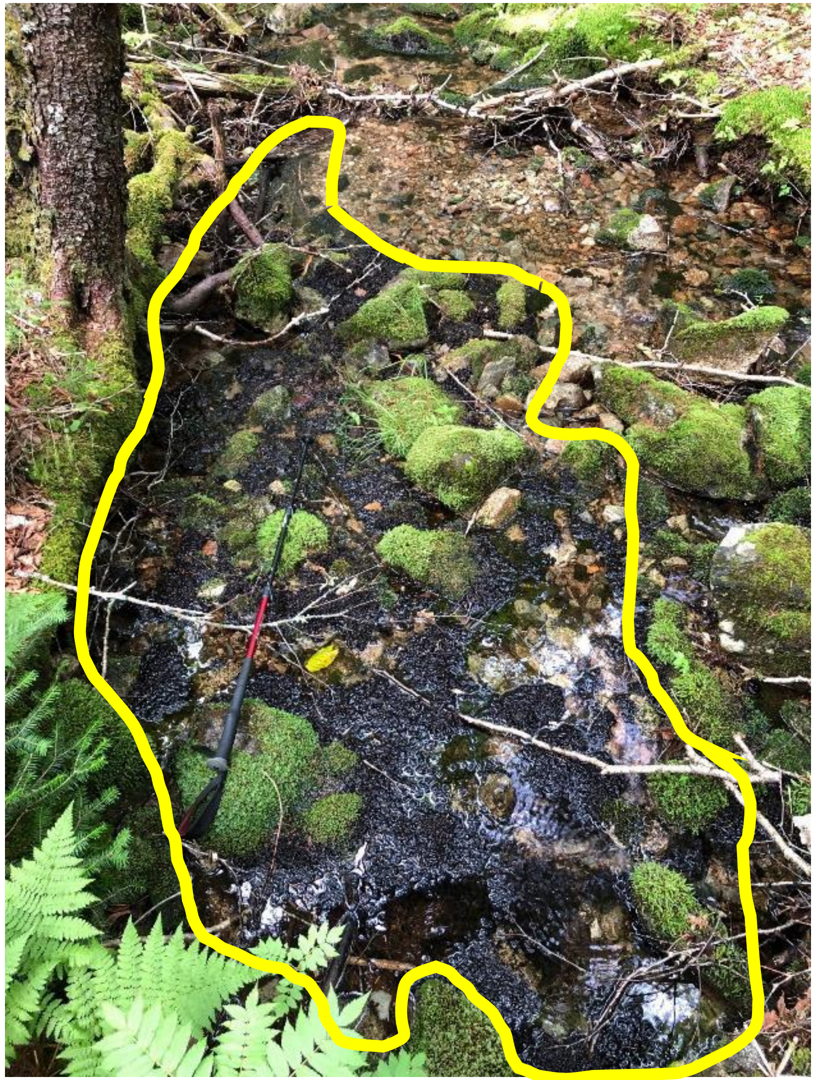
- Walk the brooks
- Estimated surface area of colonies
- Extrapolation for streams with abundant waterfan
- Field seasons 2019-2020

Estimating waterfan colony size?



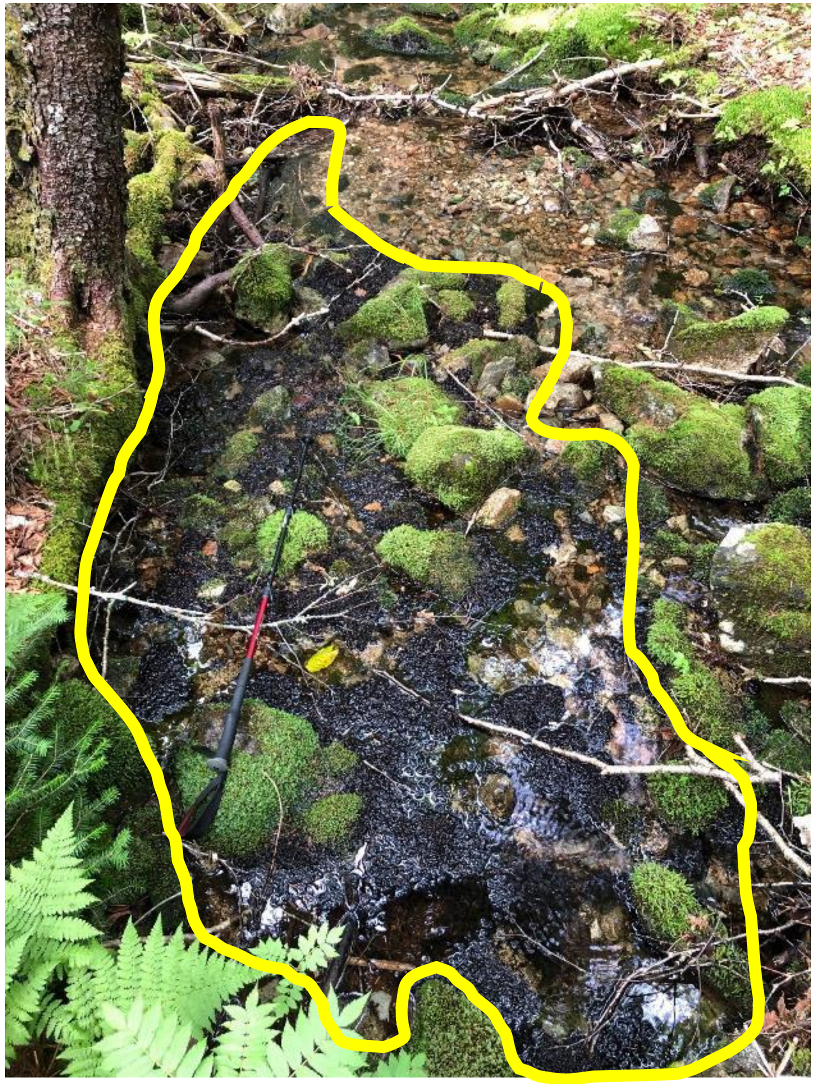
Estimating waterfan colony size

1 thalli, ~2750cm2



1 stream section, ~15 000cm2

Estimating waterfan surface : interpretation



Field notes:

carpeted up stream for 1-2 m; 1 thallus golfball-sized, 5 thalli baseball-sized, 1 thallus basketball-sized.

Summarizing observations along streams

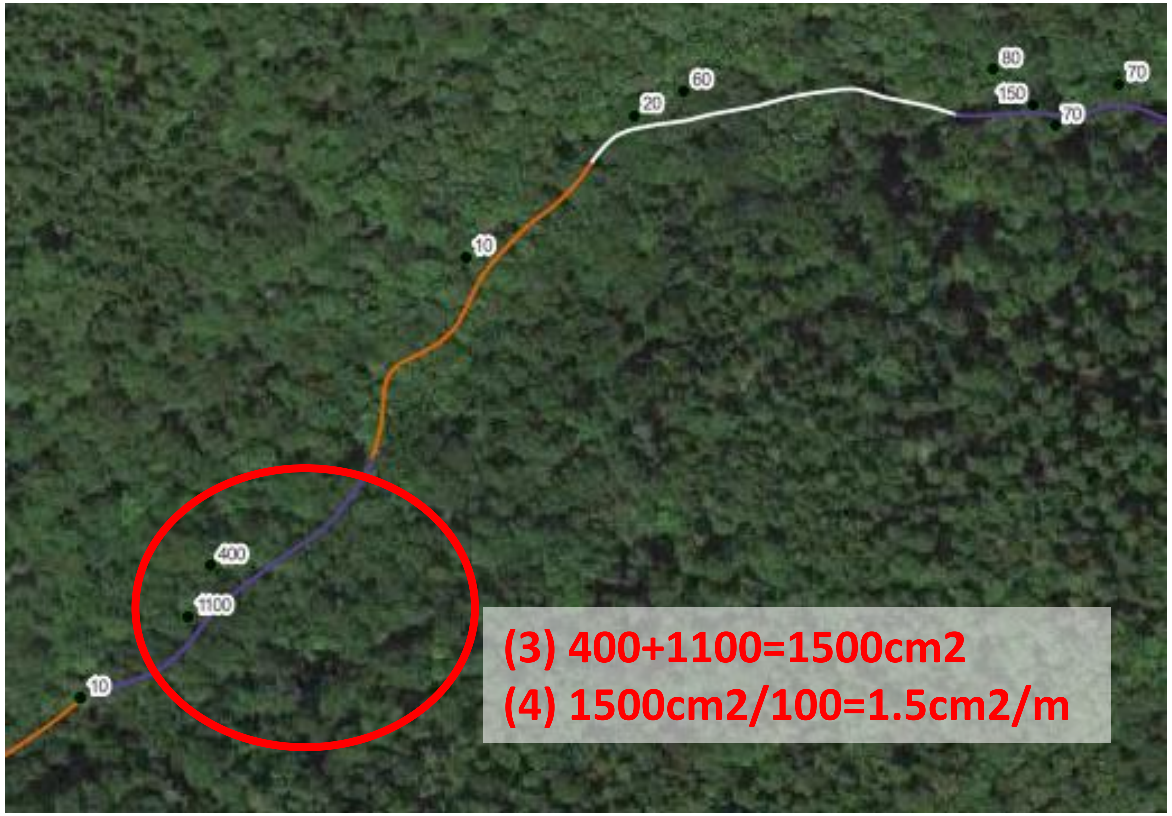
- each observation as data point?
- number of observations per area/length?
  - 1 observation = 1 continuous colony (may cover >5-10m)
  - 5 observations = 5 separate thalli within 5-10m
  - large colonies are underestimated
- waterfan area (cm2) by length?
  - need area provided by field crews

Waterfan density along streams : cm2/m

1. Split stream into sections of <100m
2. Calculate (estimate) total area per observation
3. Sum of cover over stream section
4. Total waterfan area (cm2)/stream section length

Waterfan density along streams : cm2/m

1. Split stream into sections of <100m
2. Calculate (estimate) total area per observation
3. Sum of cover over stream section
4. Total waterfan area (cm2)/stream section length



Analysis challenges

- Species distribution tied to stream networks
  - Linear distribution
  - Influence downstream only (dispersal limited along network)
- Differentiation of individuals unclear
- Small organism dependent on small-scale habitat features

Landscape habitat features analysed

Input datasets

Dataset	Type	Source
Digital Elevation Model (DEM)	1 m raster	Province of New Brunswick
Canopy Height Model (CHM)	1 m raster	Province of New Brunswick
DEM-derived stream network	Vector	Atlantic Canada Conservation Data Centre
Provincial hydrographic network	Vector	Province of New Brunswick
Fundy National Park boundary	Vector	Canadian Parks and Conserved Areas Database
Forest inventory	Vector	Province of New Brunswick
Eastern Waterfan observations	Vector	Atlantic Canada Conservation Data Centre
Eastern Waterfan absence data	Vector	Atlantic Canada Conservation Data Centre

Landscape attributes : remote-sensing derived

Dataset	Type	Attribute
Digital Elevation Model (DEM)	1 m raster	Elevation (m) – 50m buffer each side stream
		Stream slope (radians) – change in elevation over 100m stream section
		Stream aspect : direction of stream flow • Aspect related to North-South (-1 to 1) • Aspect related to East-West (-1 to 1)
		Stream network derived from DEM
Canopy Height Model (CHM)	1 m raster	Canopy height (m) – 50m buffer each side stream

Landscape attributes : photo-interpreted

Dataset	Type	Attribute
Forest inventory	Vector	Distance to roads (m) – distance from stream to polygon classified as transportation corridor
		Distance to harvest (m) – distance from stream to polygon classified as harvested within the last 40 years
		Stand age : young, immature, mature and overmature
		Stand composition defined by canopy groups: 25 dominant-codominant species pairs were grouped into 8 categories
		Crown cover : 0-50%, 50-70% and 70-100%
		Land cover : Forest, Wetland and Open water (main classes from multiple Primary Land Use)
Prov. hydrographic	Vector	Stream order : 0 to 5 (small to large)

Statistical techniques and output

Analysis steps in R statistical software

1. Preliminary analysis to select landscape type to include

2. Identifying attributes determining the likelihood of Eastern Waterfan presence
   - Conditional inference tree analysis
   - Logistic regression

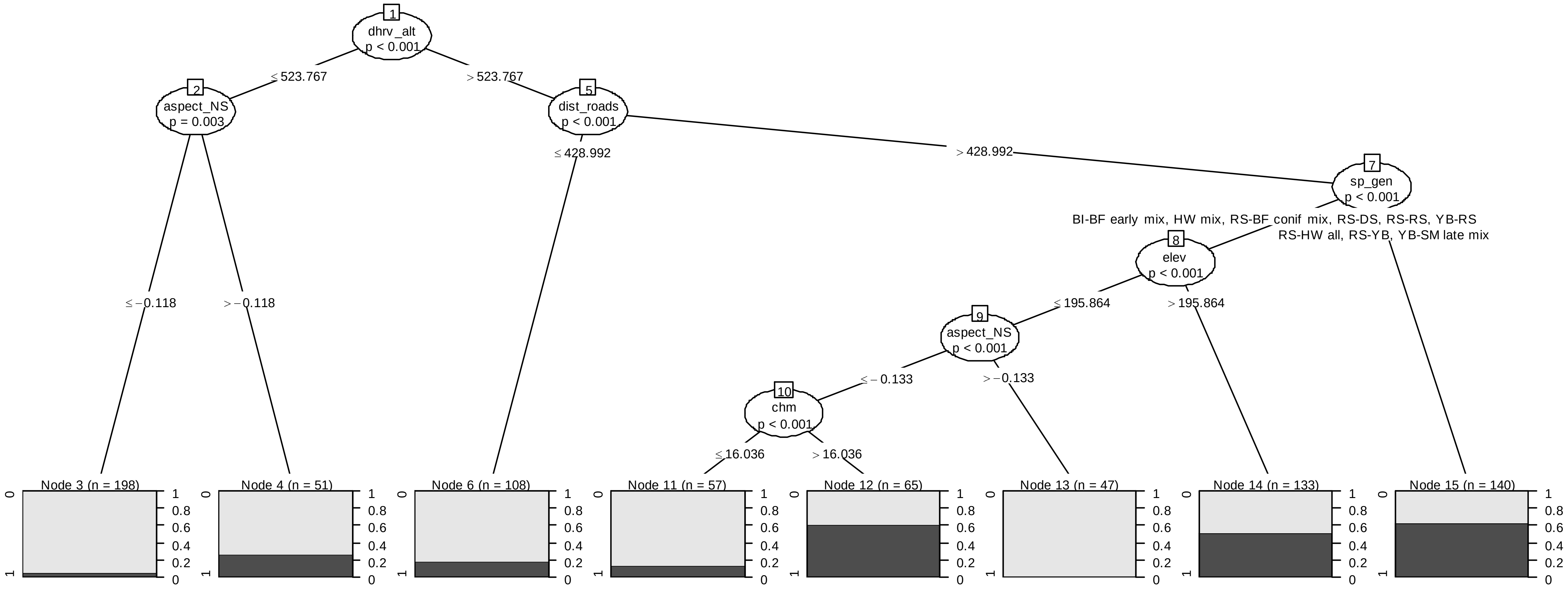
3. Identifying attributes contributing to Eastern Waterfan density (based on records with measurements of thallus or patch area)
   - Multiple regression

Preliminary categorical analysis

land cover / stream order	forest	wetland (WL)	open water (WB)	all types
0 (very small streams)	0/55 (0%)	0/2 (0%)	NA	0%
1	167/457 (37%)	0/6 (0%)	NA	36%
2	91/351 (26%)	0/12 (0%)	0/1 (0%)	25%
3	3/40 (8%)	0/37 (0%)	0/23 (0%)	3%
4 (large rivers)	NA	NA	0/55 (0%)	0%
all stream sizes	29%	0%	0%	

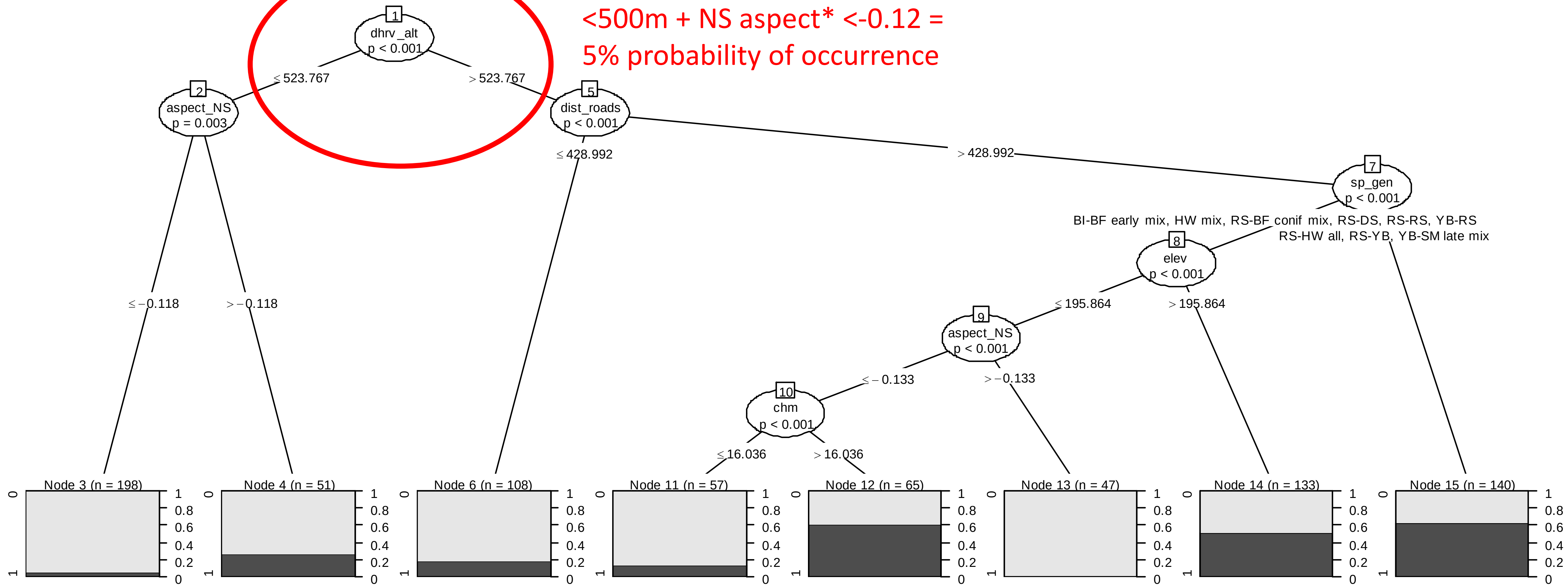
Analysis focused ONLY on medium-sized streams in forest

Conditional inference tree analysis



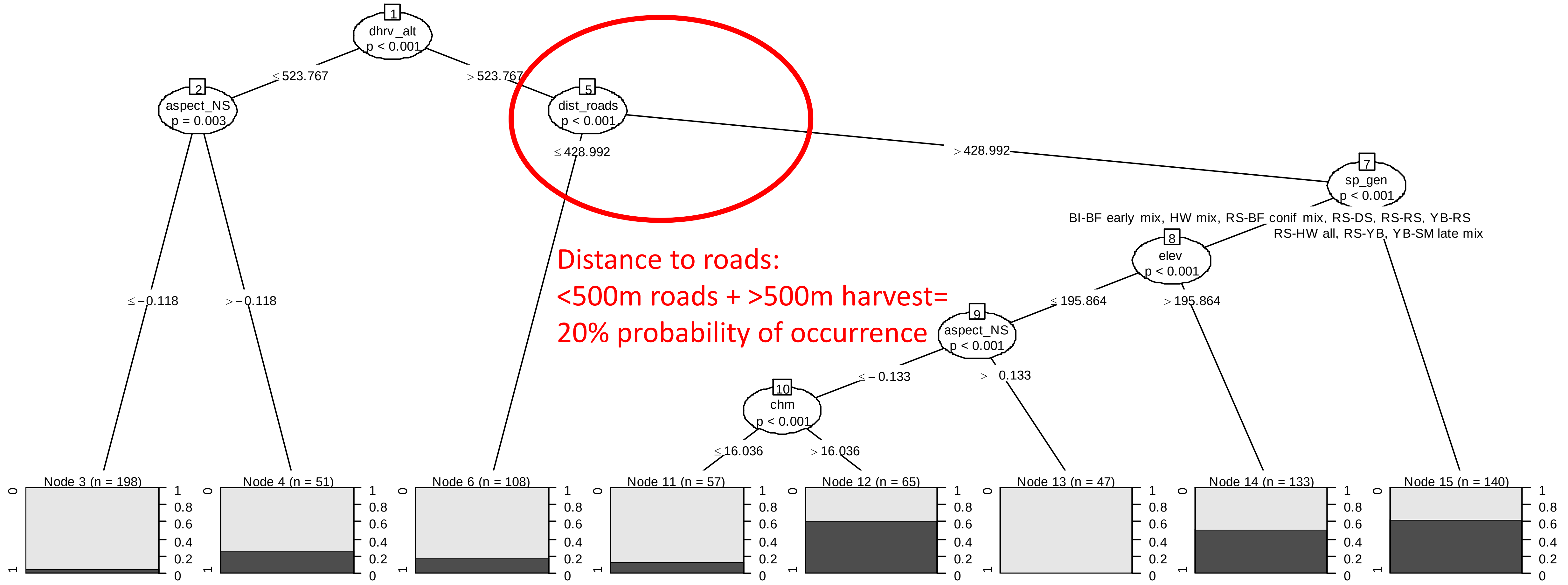
Conditional inference tree analysis

Distance to harvest cutblocks:
<500m + NS aspect* <-0.12 =
5% probability of occurrence



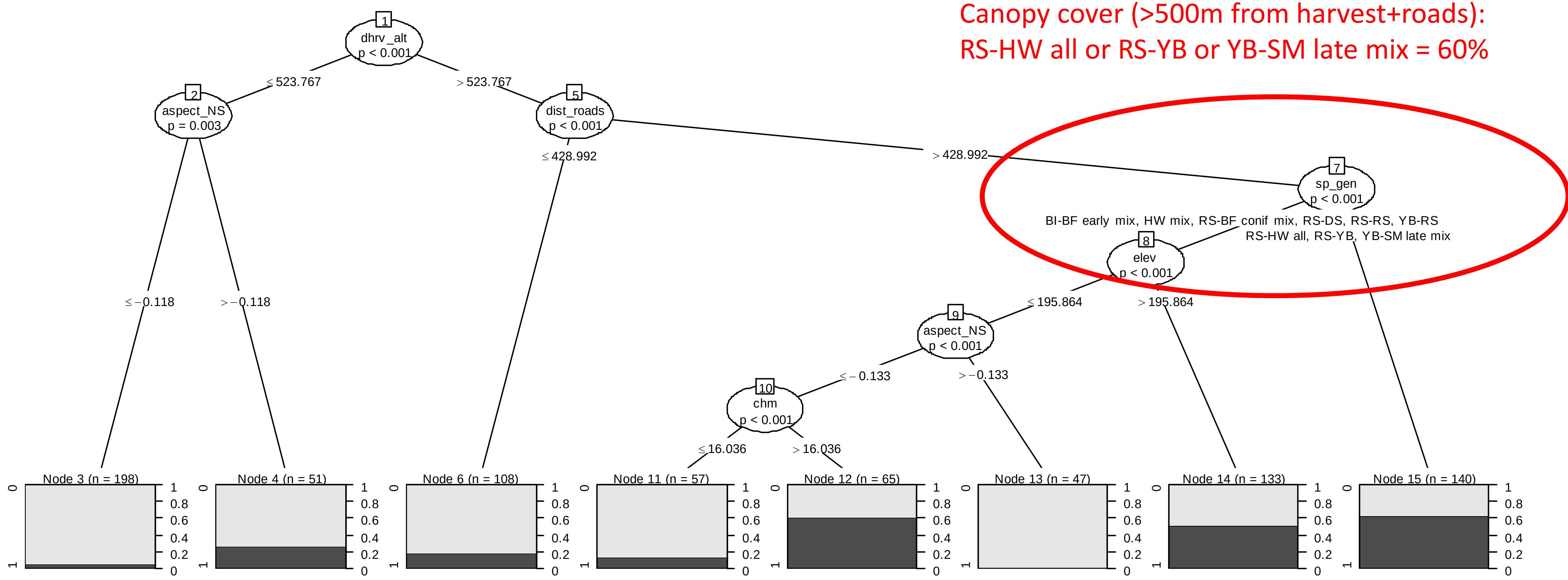
*NS aspect: orientation with respect to North-South (where -1 is due South and 1 is due North)

Conditional inference tree analysis



Conditional inference tree analysis

Canopy cover (>500m from harvest+roads):
RS-HW all or RS-YB or YB-SM late mix = 60%



Canopy groups (example)

Too few stream sections covered by Red Spruce-Birch to include as its own category (RS-BI) : combined with generic Red Spruce-Hardwood

RS-YB and YB-RS very abundant so each gets its own category in analysis

	Red spruce with hardwood (RS-HW all)	Red spruce with yellow birch (RS-YB)	Yellow birch with red spruce (YB-RS)	Yellow birch with sugar maple (YB-SM late mix)
RS-BI	15			
RS-HW	20			
RS-SM	1			
RS-TH	64			
SM-RS	3			
YB-BF	4			
RS-BI	15			
RS-HW	20			
RS-YB		142		
YB-RS			93	
SM-SM				1
SM-YB				6
YB-RM				1
YB-SM				34
YB-YB				5

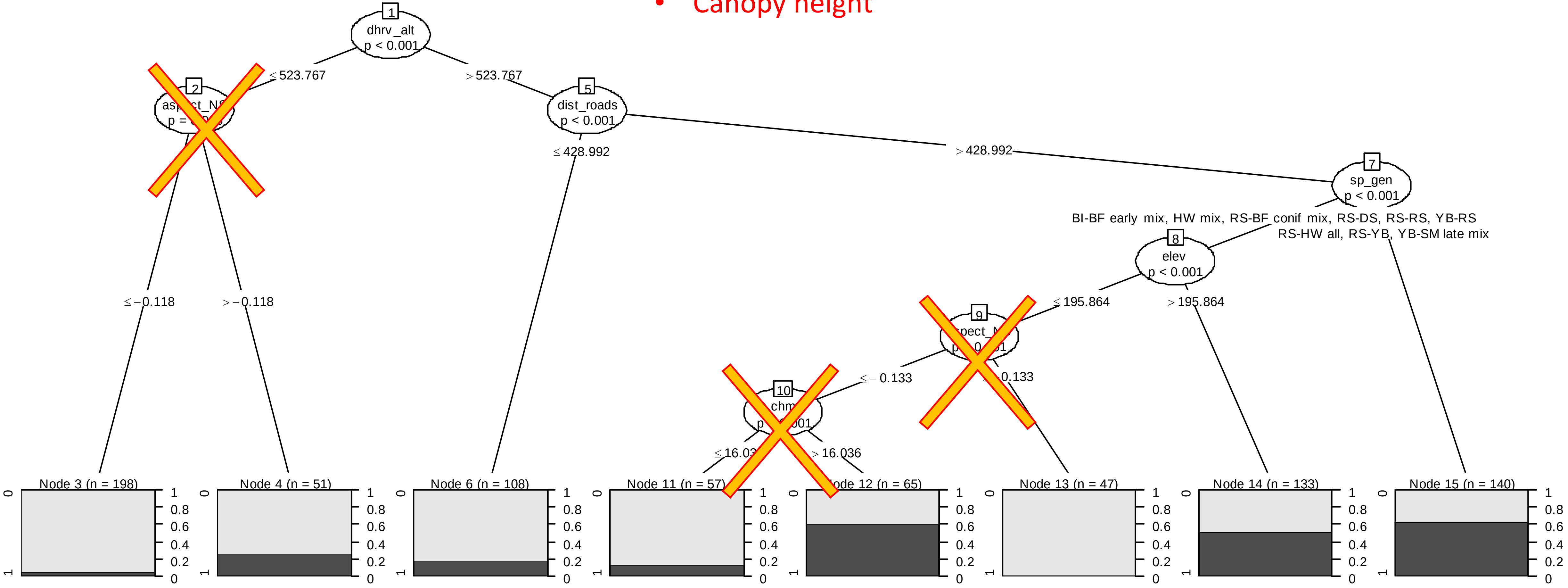
Logistic regression output

The group with highest probability of waterfan occurrence has highest increase for each canopy type here

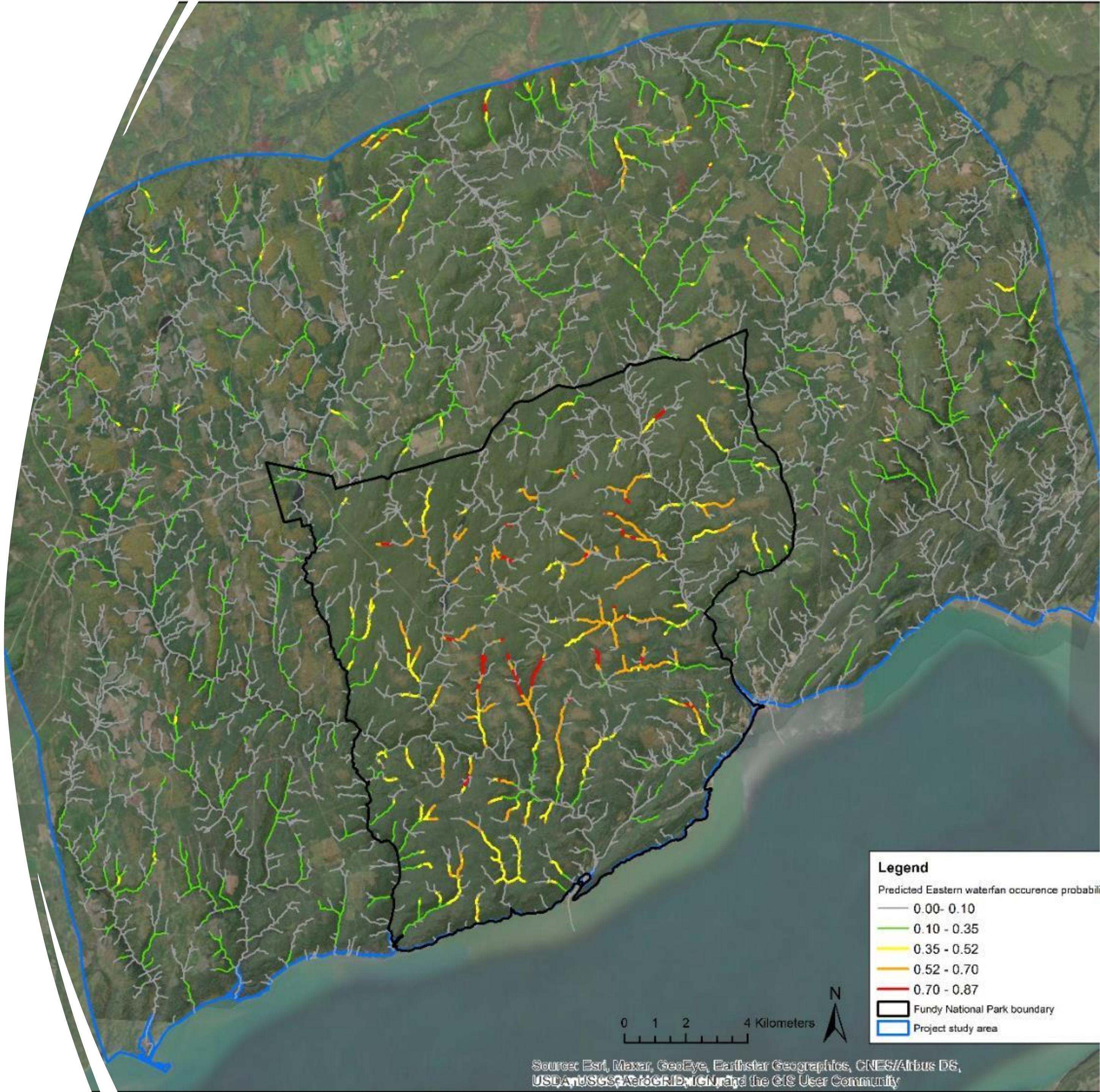
	coefficient (log-odds)	p	Lowest value	Highest value	Times increase
(Intercept) BF-BI early mix	-4.498	<0.0001			
HW mix (n=25)	-0.204	0.7596	0	1	0.8
RS-BF conif mix (n=109)	-0.193	0.7159	0	1	0.8
RS-DS (n=119)	0.340	0.5016	0	1	1.4
RS-RS (n=161)	-0.302	0.5451	0	1	0.7
RS-YB (n=119)	0.171	0.7354	0	1	1.2
YB-RS (n=86)	0.239	0.6463	0	1	1.3
RS-HW all (n=96)	0.677	0.1812	0	1	2.0
YB-SM late mix (n=45)	1.260	0.0249	0	1	3.5
Distance to harvest (m), maximum 2500m	0.001	<0.0001	50	2500	3.9
Distance to roads (m), maximum 500m	0.003	0.0001	50	500	4.0
Elevation (m)	0.007	<0.0001	25	325	8.2

Variables removed by logistic regression

- Aspect
- Canopy height



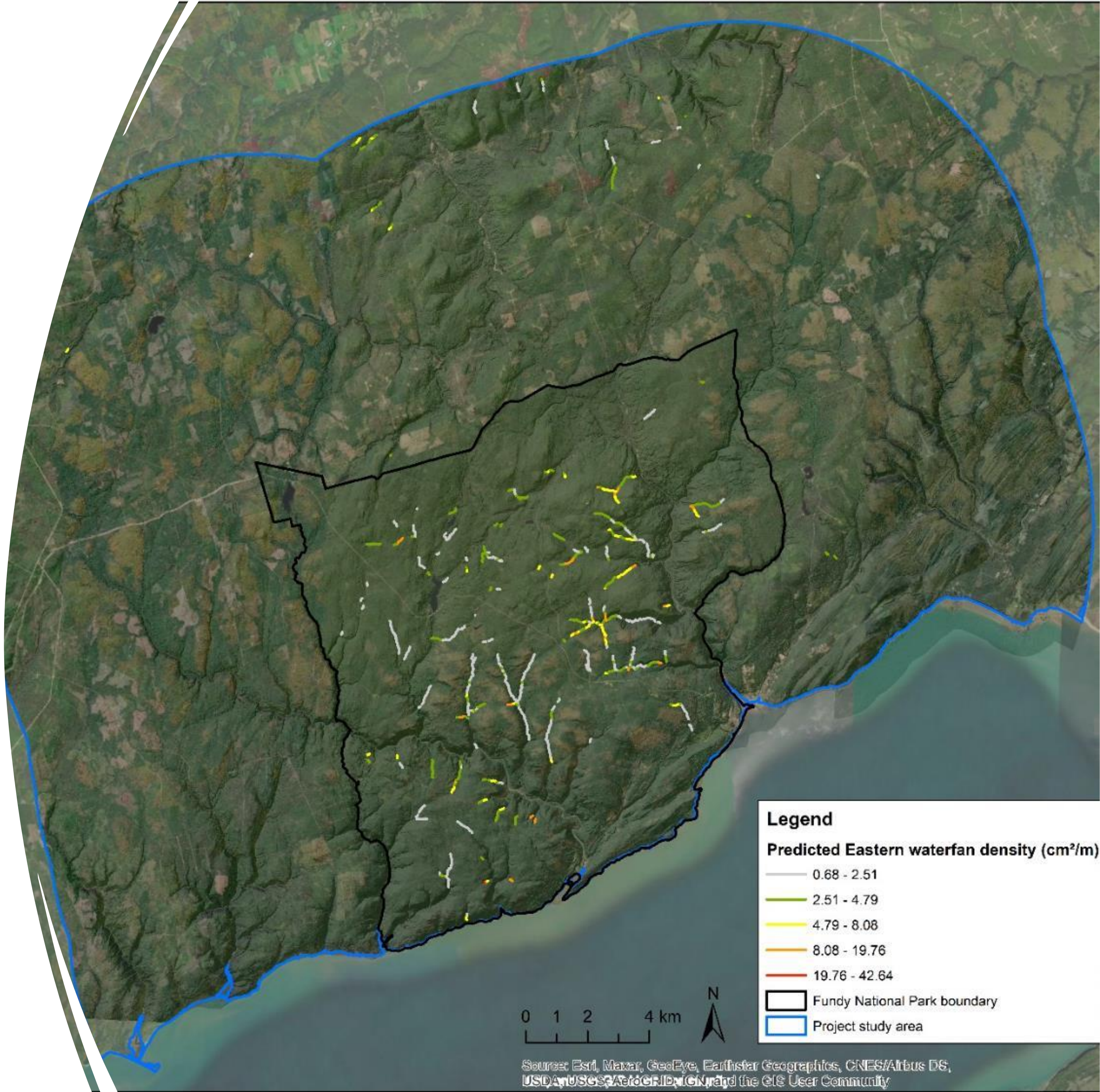
Waterfan presence probability



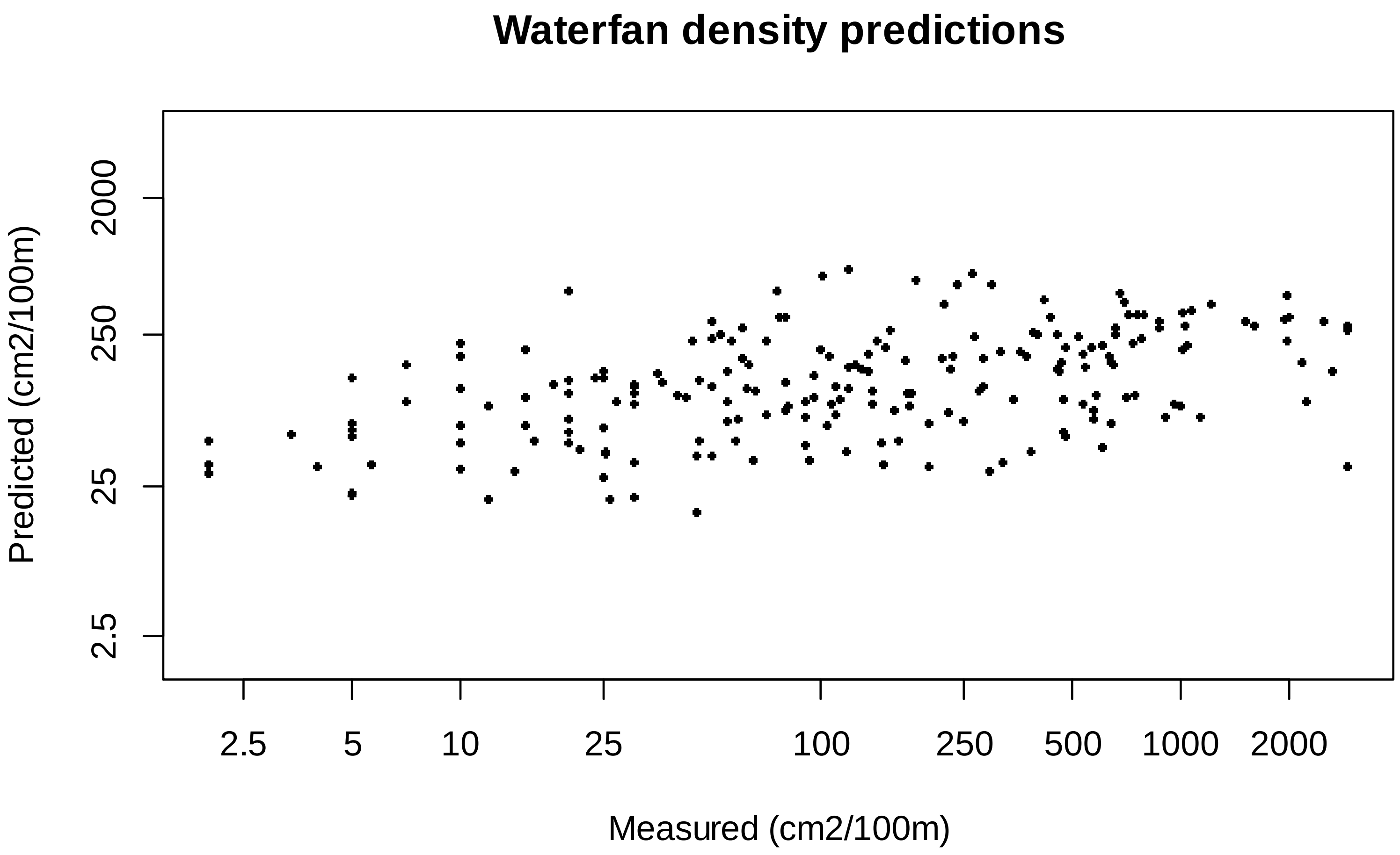
Multiple linear regression results

	slope (log)	p	Slope interpretation	lowest	highest
(Intercept) : sp = RS-BF (n=24)	-1.110	0.06	0.32cm2/m is the reference value (RS-BF, on a flat slope, 0m elevation and centred at NS/EW orientation		
RS-DS (n=26)	0.413	0.345	RS-DS has 50% more than RS-BF	0	1
RS-HW all (n=39)	0.275	0.492	RS-HW all has 32% more than RS-BF	0	1
RS-RS (n=32)	0.230	0.585	RS-RS has 26% more than than RS-BF	0	1
RS-YB (n=26)	1.020	0.022	RS-YB has 170% (2.7 times) more than RS-BF	0	1
YB-RS (n=35)	0.352	0.397	YB-RS has 42% more than RS-BF	0	1
YB-SM late mix (n=26)	0.158	0.743	YB-SM mix has 17% more than RS-BF	0	1
Slope (radians)	7.862	0.001	+15 degrees in slope increases by 610% (by 7.1 times)	0	0.25
Elevation (m)	0.004	0.028	+100m in elevation increases by 42%	25	325
Aspect relative to N (from -1 to 1)	0.511	0.004	angle towards North (relative to centre) increases by 67%	-1	1
Aspect relative to E (from -1 to 1)	0.388	0.026	angle towards East (relative to centre) increases by 48%	-1	1

Predicted waterfan density



Evaluating model strength



Challenges using botanical survey notes for quantitative analysis

Stream speed, substrate, colony size

Analysis challenges

- Species distribution tied to stream networks
  - Linear distribution
  - Influence downstream only (dispersal limited along network)
- Differentiation of individuals unclear
- Small organism dependent on small-scale habitat features
- Evident distribution pattern : species found primarily within park

Describe stream flow!

Convert field notes to habitat variables : stream speed

Stream speed	
speed class (0 to 5)	description
0	pools (includes waterfall pool or base of waterfall), backwater
1	trickle, very slow
2	slow, quiet
3	brook or stream with no indication, flowing, medium
4	fast-flowing, gentle waterfall
5	very fast, cascading, waterfall

356 records approximate stream speed, only 126 use detailed descriptions

field notes	stream speed class assigned
far end of pool half emersed	0
quiet side bare rock half submerged	2
bedrock **isolated pool far side of waterfall**	0
top of mossy rocks **very slow stream**	1
upstream facing **very slow stream**, and submerged on bedrock facing downstream	1
submerged in shallow bedrock pools of and **below very low water waterfall**	1
bedrock in **shallow side pool**	0
still braid of very slow stream	1
rock at stream edge facing into **somewhat sheltered side pool**	0
submerged on large rocks side **below small waterfall**	1
downstream and top in **side pool**	0
on rock or wet bryophytes in **fast-flowing bouldery and stony brook**	4
on rock or wet bryophytes and waterfall bryophyte in **fast-flowing** bedrock-laden and stony brook	4
cascading brook dominated by boulders and stones	5
In cascade	5

Scale matters!

Stream speed	
speed class (0 to 5)	description
0	pools (includes waterfall pool or base of waterfall), backwater
1	trickle, very slow
2	slow, quiet
3	brook or stream with no indication, flowing, medium
4	fast-flowing, gentle waterfall
5	very fast, cascading, waterfall

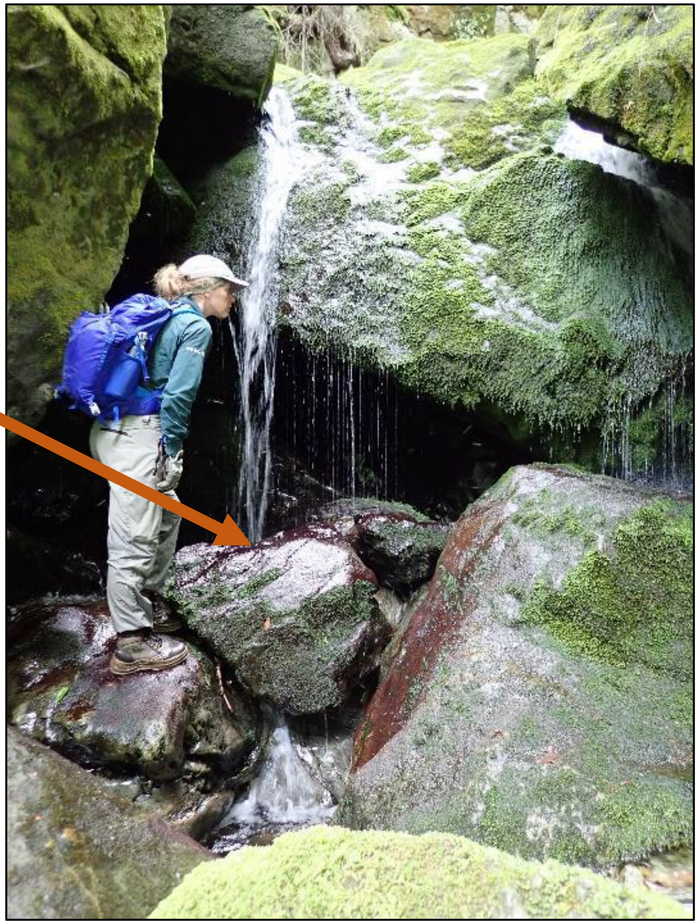
Scale matters!

Stream speed	
speed class (0 to 5)	description
0	pools (includes waterfall pool or base of waterfall), backwater
1	trickle, very slow
2	slow, quiet
3	brook or stream with no indication, flowing, medium
4	fast-flowing, gentle waterfall
5	very fast, cascading, waterfall

field notes	**stream speed class assigned**
far end of pool half emersed	0
rock at stream edge facing into **somewhat sheltered side pool**	0
submerged on large rocks side **below small waterfall**	1
quiet side bare rock half submerged	2
flowing stream	3
on rock or wet bryophytes and waterfall bryophyte in **fast-flowing** bedrock-laden and stony brook	4
cascading brook dominated by boulders and stones	5

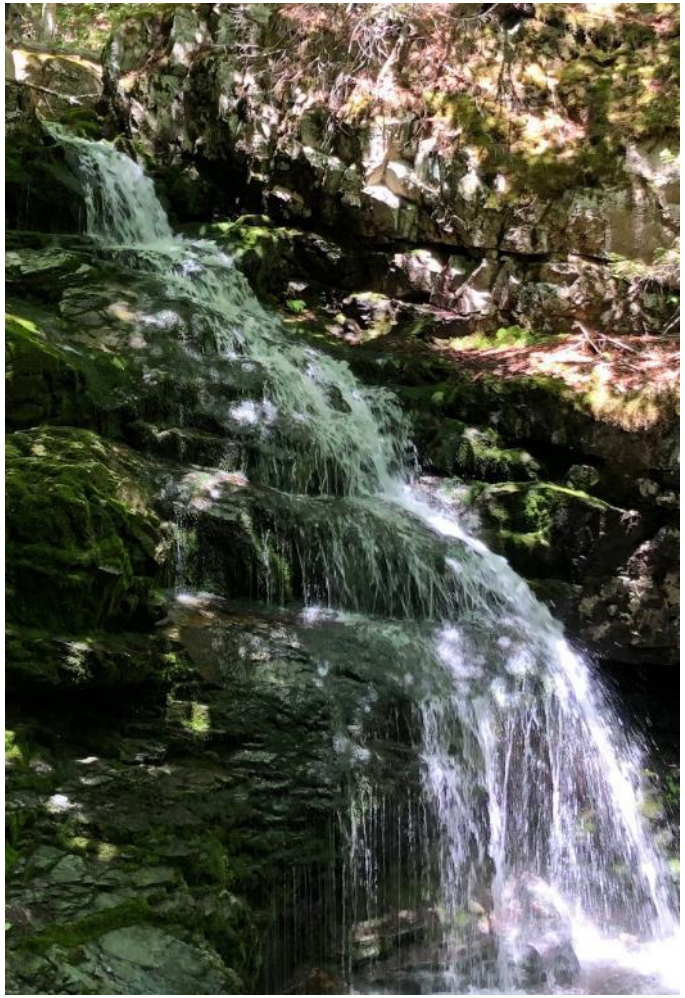
Microhabitat challenges

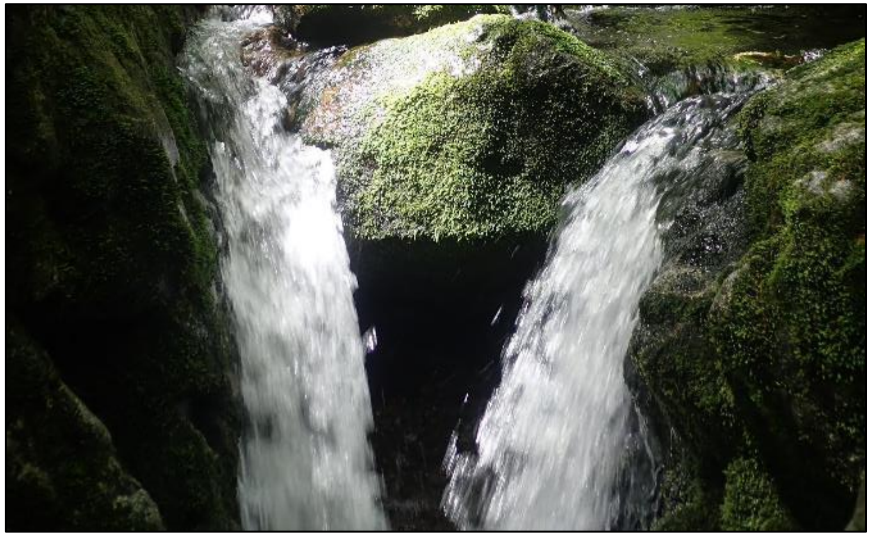
« Waterfan usually found at the base of waterfalls » -- Neil Vinson (field crew)



a) Define zones to evaluate
    1. Stream speed 1m upstream of thalli (population)
    2. Stream speed directly at position of thalli
    3. Stream speed 1m downstream of thalli
    4. Stream speed over 5-10m stream section from thalli : overall speed as an average and all speeds that cover at least 1m length
b) Use pre-defined classes to estimate stream speed
c) Include sections without waterfan at random/set distances (compare waterfan presence to general stream characteristics)

Describe substrate?





Convert field notes to habitat variables : rock substrate size and form

Rock substrate size and form	
very small	gravel, coarse sand
small	cobbles, large cobble, large gravel
medium	stones, rocks, bare rock
large	boulders, large rocks
bedrock	bedrock, outcrop, conglomerate rock, boulder/outcrop, ledge

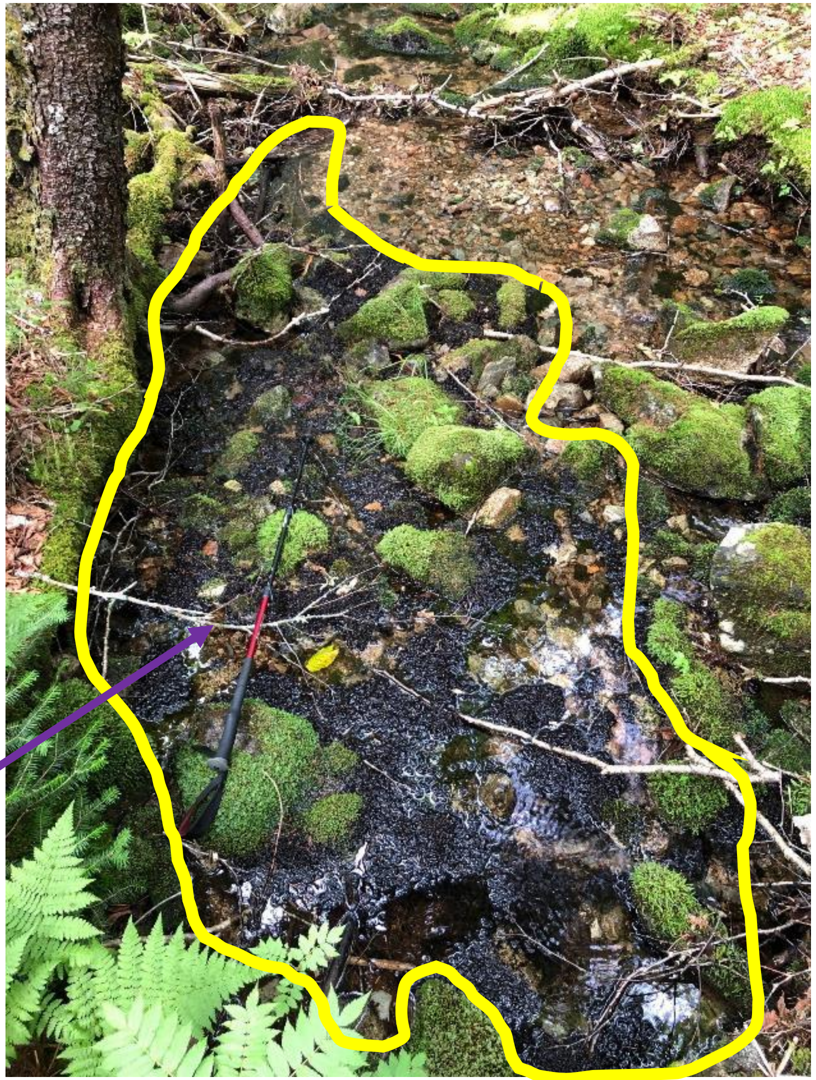
field notes	substrate class assigned
bare rock above water but below high water	medium
in water with **boulders** and pools	large
quiet brook edge **on boulder** beneath water level	large
bedrock at stream's edge	bedrock
brook with **large cobble substrate**	small
large gravel, side pool	small
semi-rounded bedrock	bedrock
Trickle of a stream with **bedrock, boulders and coarse sand**	very small, large, bedrock

Challenges of varying colony size

1 thalli, ~2750cm2

Easily count and measure area covered by small easily bounded colonies

- Difficult to distinguish colonies when it carpets streambed
- Length of stream (m) easier metric than area (cm2)
- Fatigue when multiple colonies encountered in a row



1 stream section, ~15 000cm2

Challenges of varying colony size : example

Descr.: thalli covering a total at least 1150 cm2

Descr.: thalli covering a total of 5700 cm2; stopped counting, but Peltigera dense and continuous past here.

Descr.: thalli covering a total of 80 cm2; begins becoming sparse here, not seen for a bit.

Ideas for survey improvements: adapt to small organisms!

Create field sheets with suggested metrics

- List of substrate classes by size and type
- List of stream speeds to record

Record both local and surrounding habitat features

- Quiet brook; thallus found at in still pool at base of waterfall
- Sugar maple forest; colony located at base of yellow birch
- Vertical rock face in spruce forest; plants growing on thin (<3cm) layer organic soil on ledge

Record both cover, individuals and reference area

- Colony covering 40cm2; 3 distinct patches
- 30 individuals recorded over 10m stream length; 1 individual=10cm2

Be mindful of variation in colony size and provide numeric estimates

Field data types : drawbacks

consistency

systematic experimental designs with set variables recorded

ecological studies

- Randomized locations may not include rare species
- Training and planning effort is high

surveys to detect species

botanical surveys

- No information on habitat unsuitable for species
- Estimates on species population size lack consistency
- Description of habitat incomplete or at different scales between observers or field survey locations

opportunistic observations

citizen science

- Only species presence records
- No habitat variables

more species records

Field data types: advantages

consistency

systematic experimental designs with set variables recorded

ecological studies

- Systemic covariables recorded allow comparison between points
- Consistent quantitative measures
- Randomized locations include unsuitable habitat

surveys to detect species

botanical surveys

- Search effort to detect multiple populations
- Field notes record habitat attributes

opportunistic observations

citizen science

- Most observers involved
- Add observations in context of other work or recreation

more species records

Optimizing botanical surveys

systematic experimental designs
with set variables recorded

- Use existing survey data to create of list systematic covariables to record to allow comparison between points
- Prioritize quantitative measures and avoid extrapolation
- Include some randomized locations to describe unsuitable habitat

consistency

surveys to detect species

- Search effort to detect multiple populations
- Field notes record habitat attributes

botanical surveys that use ecological study techniques

more species records