



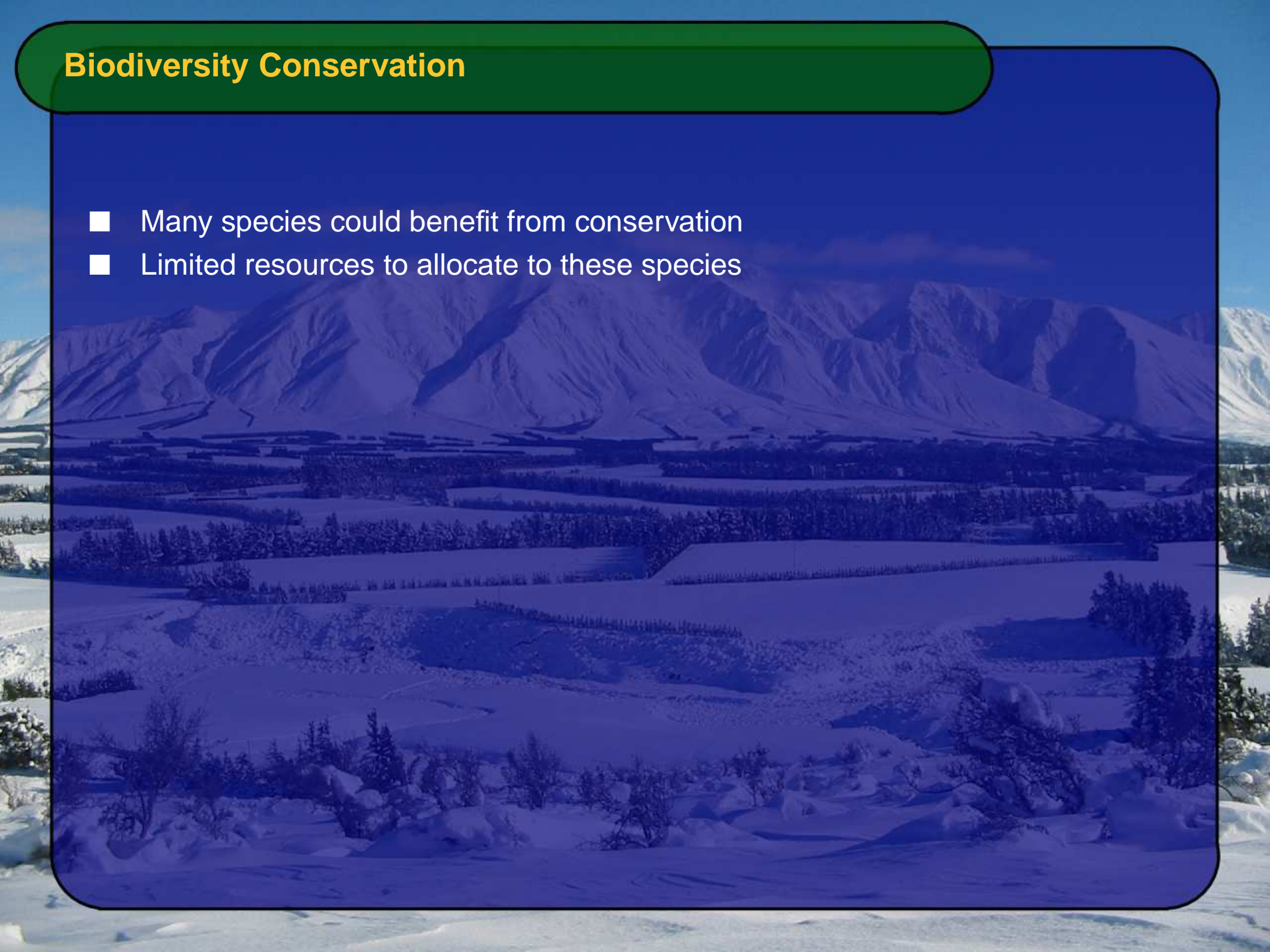
Phylogenetics and Biodiversity Conservation

Klaas Hartmann, University of Canterbury (New Zealand)

January 31, 2007

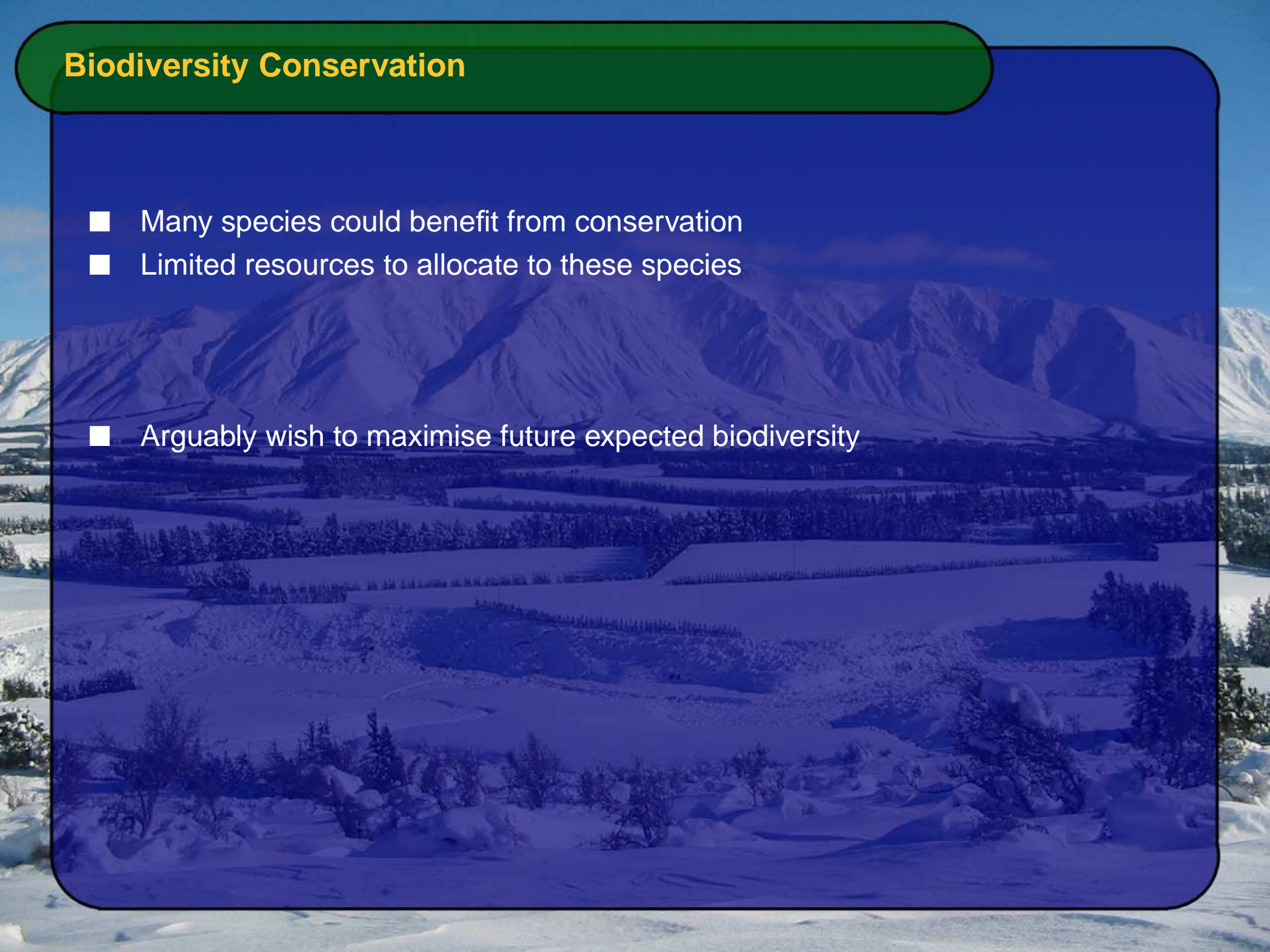
Biodiversity Conservation

- Many species could benefit from conservation
- Limited resources to allocate to these species



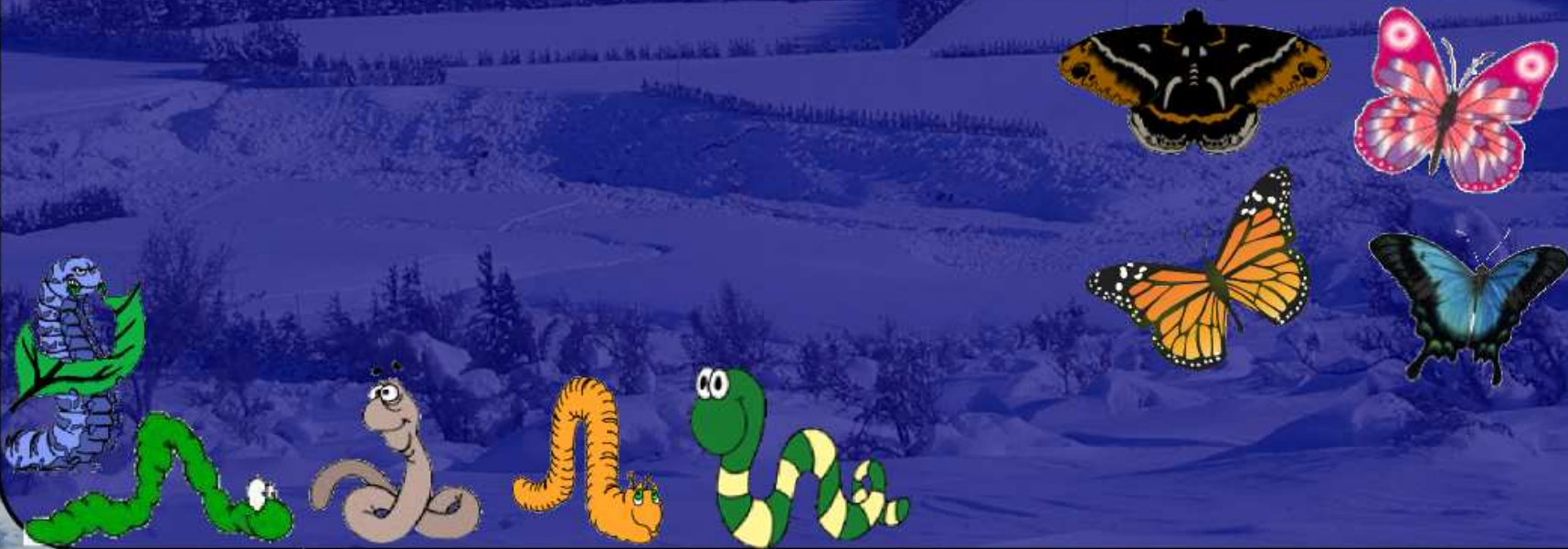
Biodiversity Conservation

- Many species could benefit from conservation
- Limited resources to allocate to these species
- Arguably wish to maximise future expected biodiversity



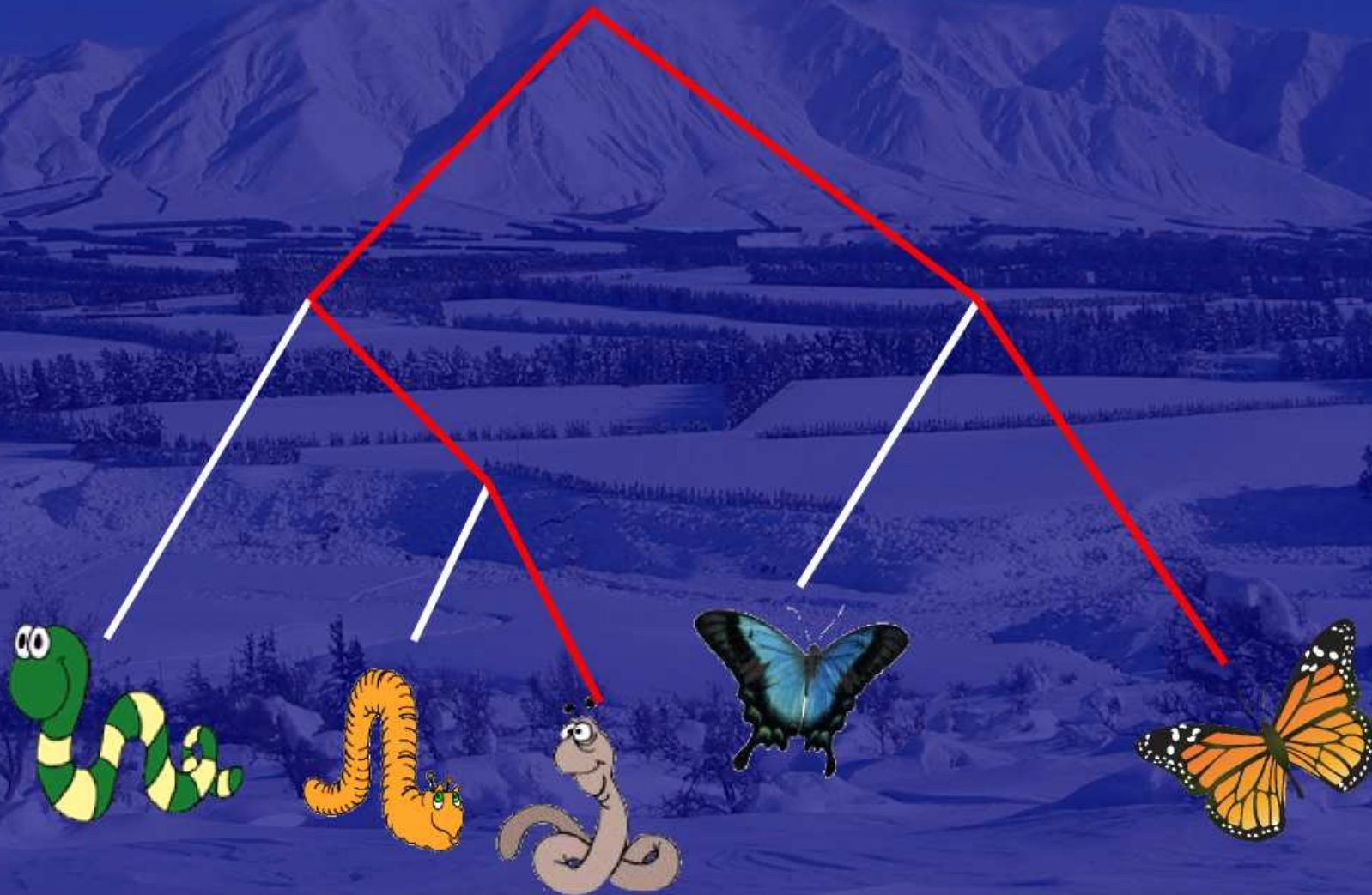
Biodiversity Conservation

- Many species could benefit from conservation
- Limited resources to allocate to these species
- Arguably wish to maximise future expected biodiversity



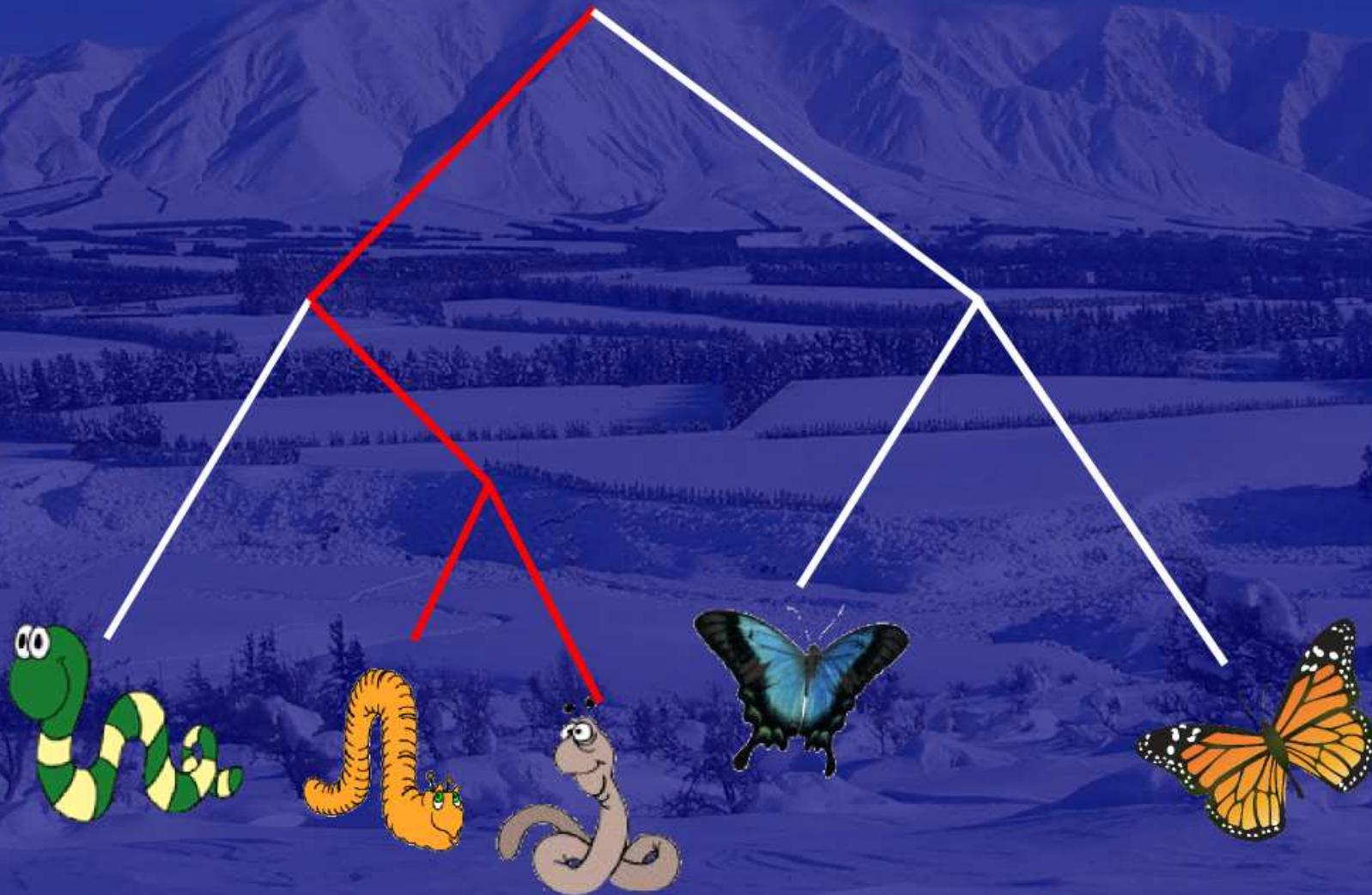
Phylogenetic Diversity

- Phylogenetic diversity as a measure of biodiversity



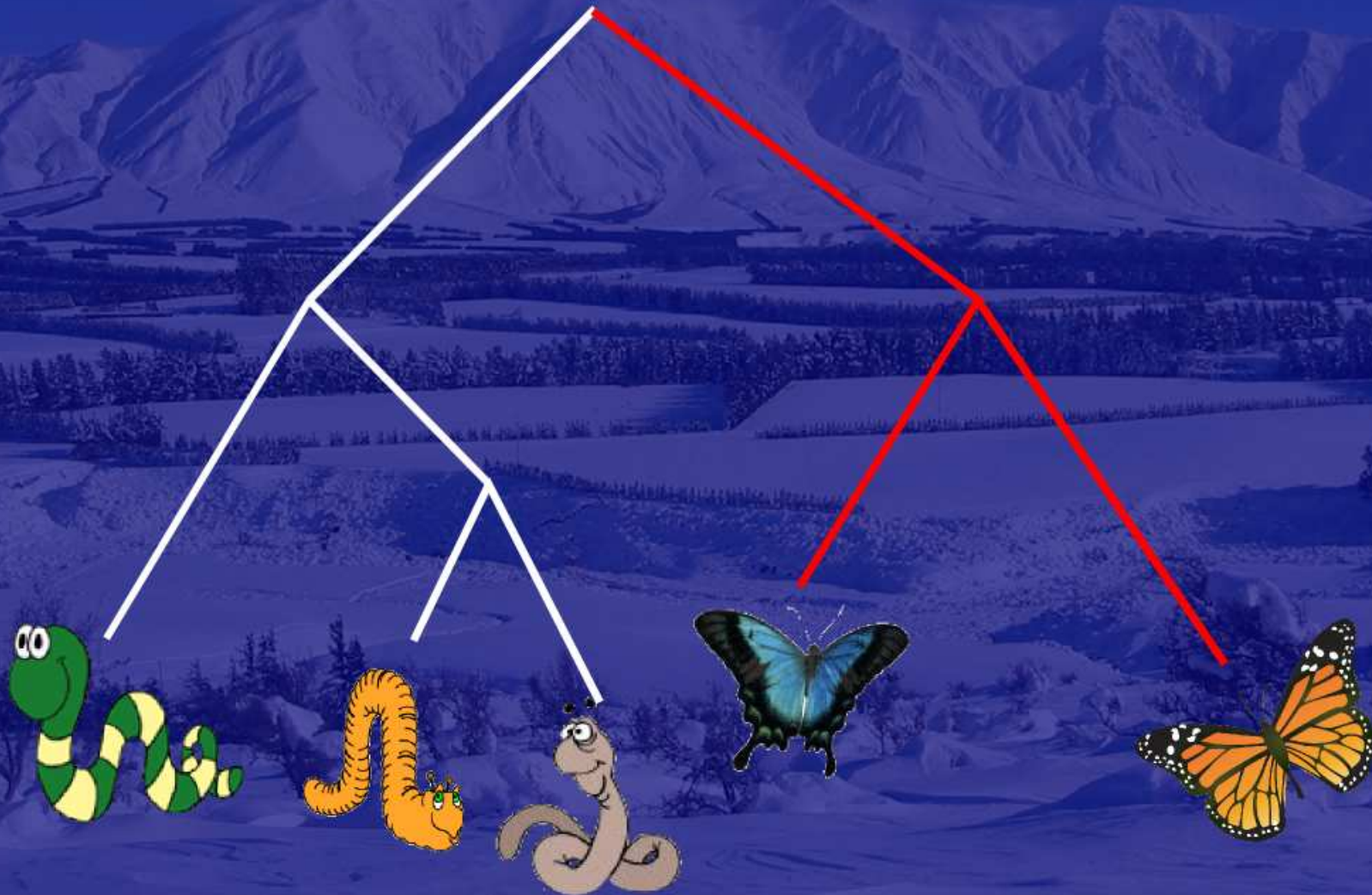
Phylogenetic Diversity

- Phylogenetic diversity as a measure of biodiversity



Phylogenetic Diversity

- Phylogenetic diversity as a measure of biodiversity



Mike Steel's Greedy Optimisation

To obtain the set of species, S , that has greatest PD:

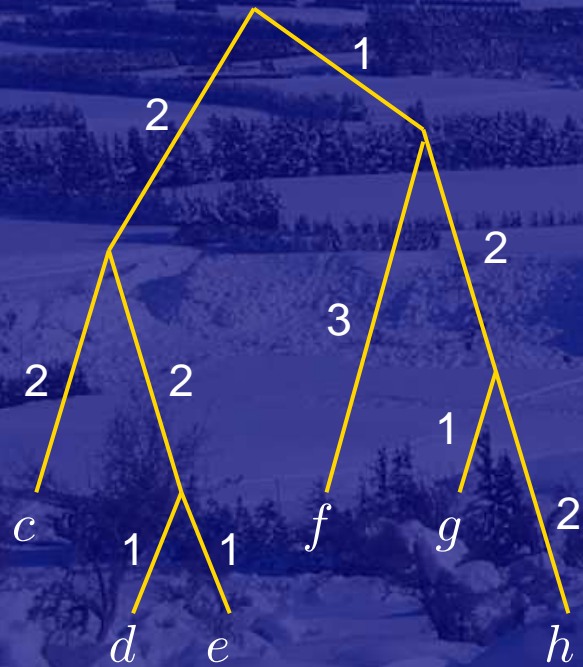
- Begin with the empty set $S = \emptyset$
- Sequentially add the species that gives the greatest PD increase to S
- Stop when S is the desired size

Mike Steel's Greedy Optimisation

To obtain the set of species, S , that has greatest PD:

- Begin with the empty set $S = \emptyset$
- Sequentially add the species that gives the greatest PD increase to S
- Stop when S is the desired size

$$S = \emptyset, PD(S) = 0$$



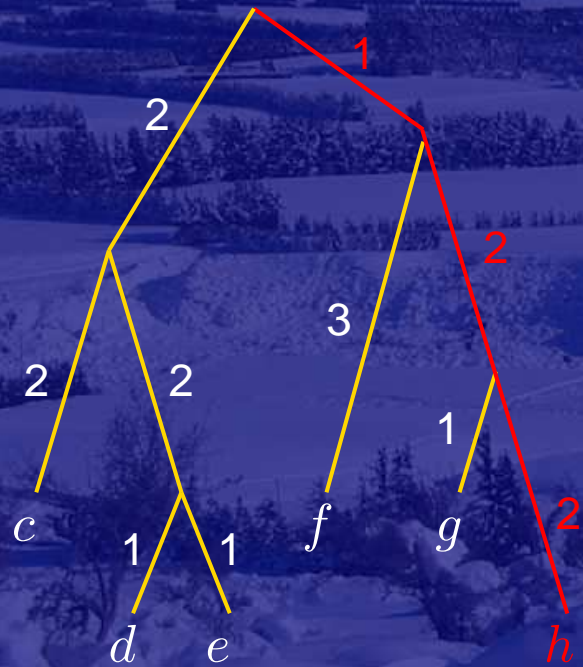
Species	Contribution to $PD(S)$
c	4
d	5
e	5
f	4
g	4
h	5

Mike Steel's Greedy Optimisation

To obtain the set of species, S , that has greatest PD:

- Begin with the empty set $S = \emptyset$
- Sequentially add the species that gives the greatest PD increase to S
- Stop when S is the desired size

$$S = \{h\}, PD(S) = 5$$



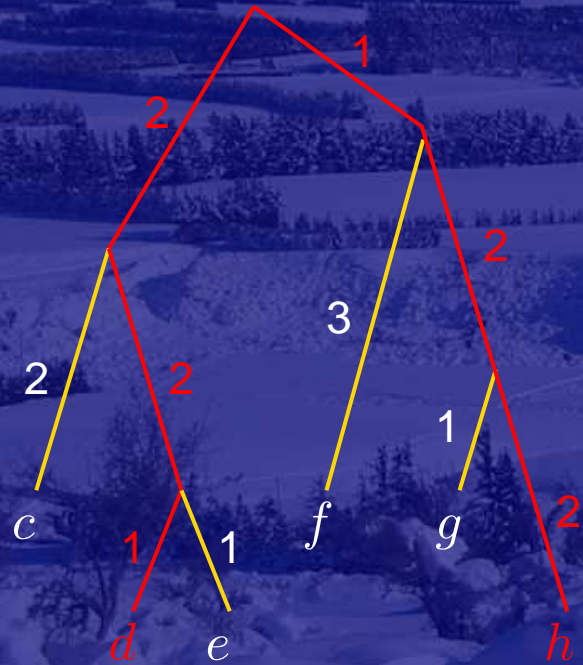
Species	Contribution to $PD(S)$
c	4
d	5
e	5
f	3
g	1

Mike Steel's Greedy Optimisation

To obtain the set of species, S , that has greatest PD:

- Begin with the empty set $S = \emptyset$
- Sequentially add the species that gives the greatest PD increase to S
- Stop when S is the desired size

$S = \{h, d\}, PD(S) = 10$



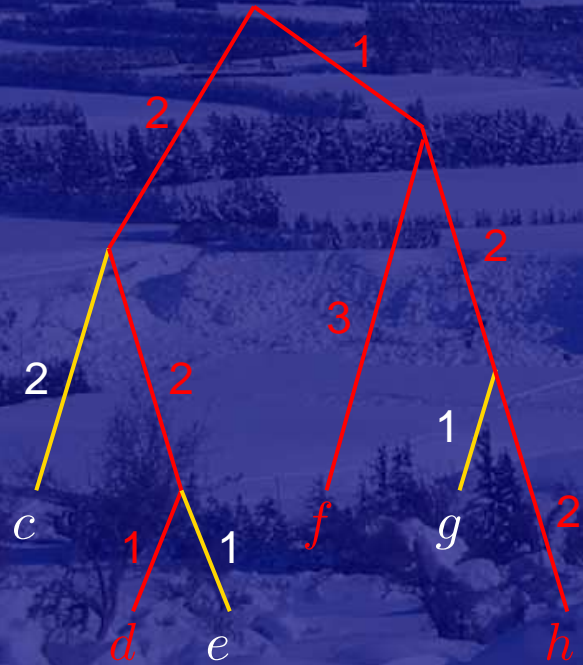
Species	Contribution to $PD(S)$
c	2
e	1
f	3
g	1

Mike Steel's Greedy Optimisation

To obtain the set of species, S , that has greatest PD:

- Begin with the empty set $S = \emptyset$
- Sequentially add the species that gives the greatest PD increase to S
- Stop when S is the desired size

$$S = \{h, d, f\}, PD(S) = 13$$

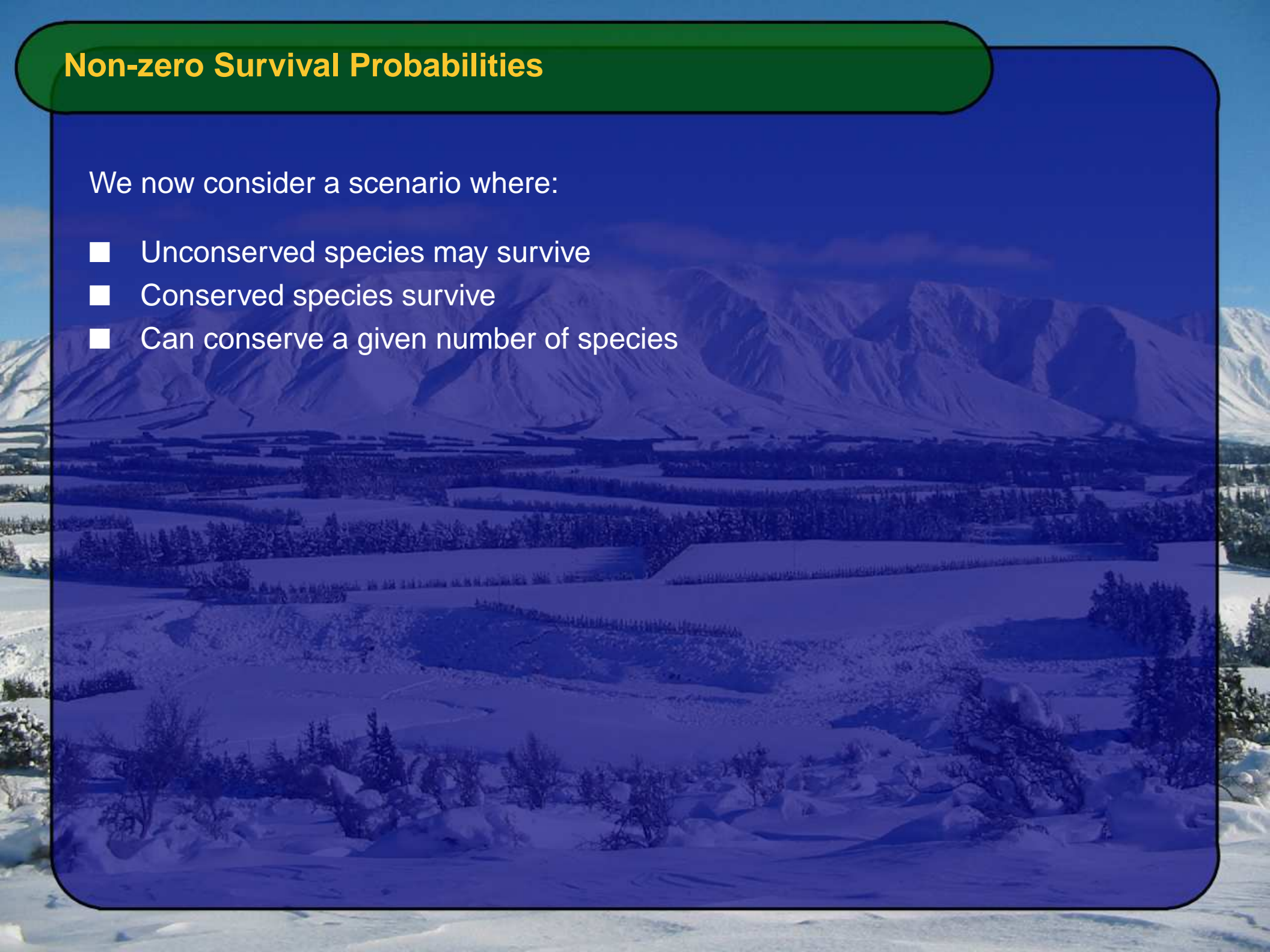


Species	Contribution to $PD(S)$
c	2
e	1
g	1

Non-zero Survival Probabilities

We now consider a scenario where:

- Unconserved species may survive
- Conserved species survive
- Can conserve a given number of species



Non-zero Survival Probabilities

We now consider a scenario where:

- Unconserved species may survive
- Conserved species survive
- Can conserve a given number of species

Solutions obtained using a greedy algorithm:

- Rescale branch lengths
- Apply Mike Steel's greedy algorithm to rescaled branch lengths

Hartmann, K. and M. Steel. 2006. Phylogenetic diversity from combinatorics to ecology.

Rescaling Branch Lengths

$$\lambda'_i := \lambda_i p(i) = \lambda_i \left(\prod_{j \in C_i} (1 - a_j) \right)$$

λ_i Length of i th edge

$p(i)$ Probability that i th edge will not survive without conservation

C_i Set of children of i th edge

a_j Survival probability of species j

Rescaling Branch Lengths

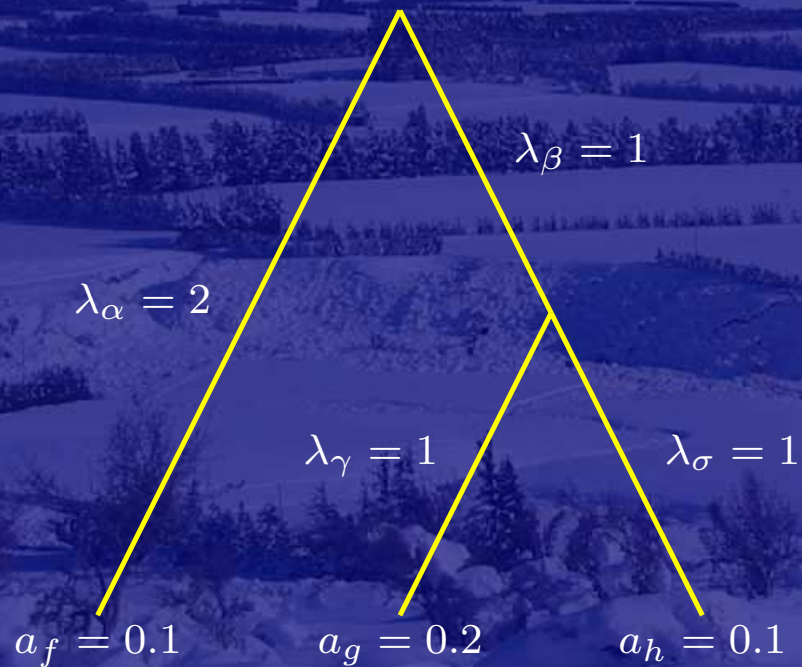
$$\lambda'_i := \lambda_i p(i) = \lambda_i \left(\prod_{j \in C_i} (1 - a_j) \right)$$

λ_i Length of i th edge

$p(i)$ Probability that i th edge will not survive without conservation

C_i Set of children of i th edge

a_j Survival probability of species j



Rescaling Branch Lengths

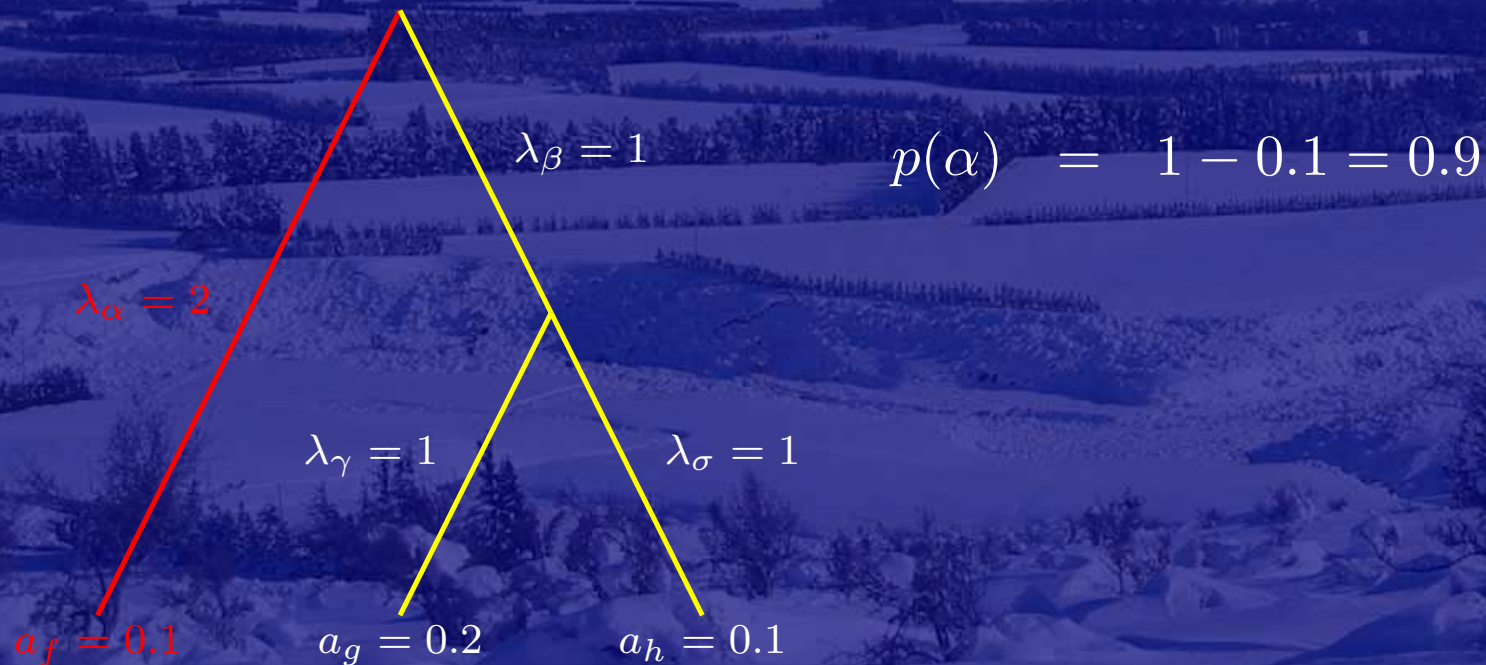
$$\lambda'_i := \lambda_i p(i) = \lambda_i \left(\prod_{j \in C_i} (1 - a_j) \right)$$

λ_i Length of i th edge

$p(i)$ Probability that i th edge will not survive without conservation

C_i Set of children of i th edge

a_j Survival probability of species j



Rescaling Branch Lengths

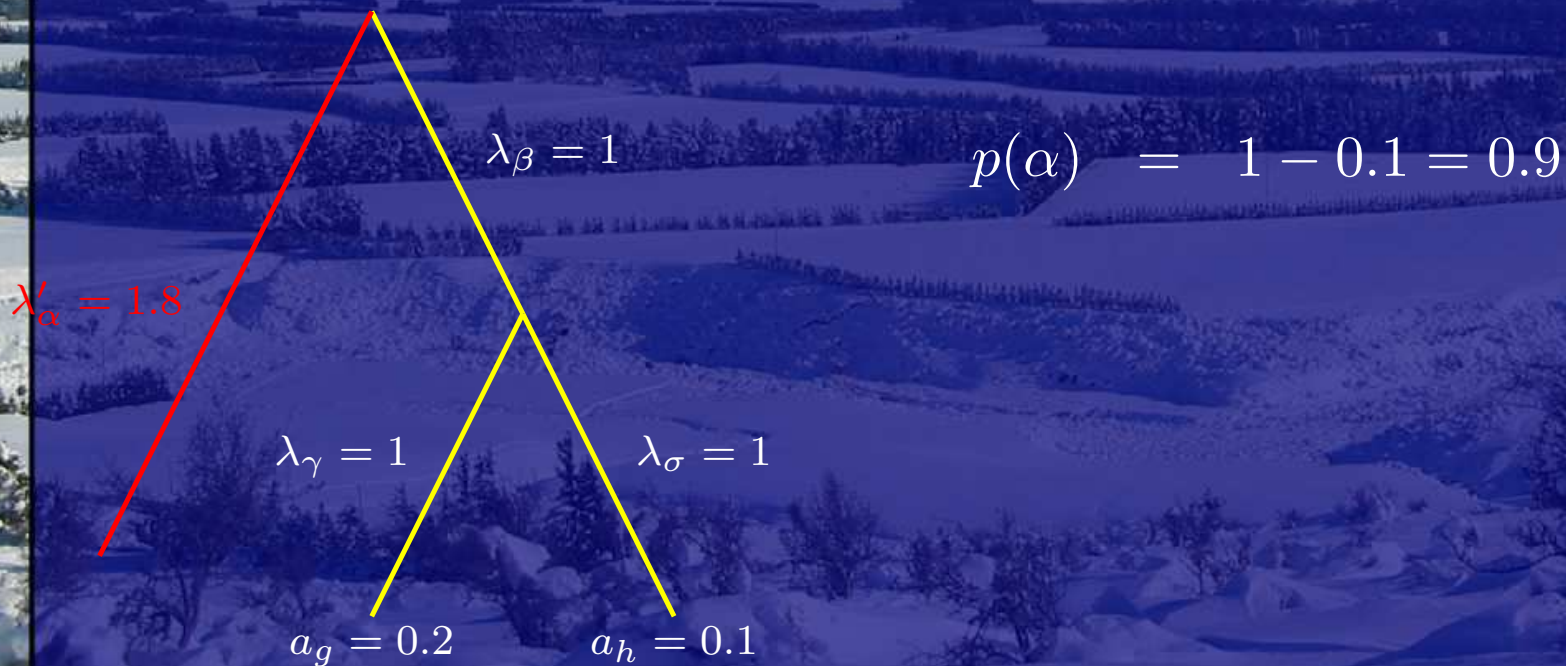
$$\lambda'_i := \lambda_i p(i) = \lambda_i \left(\prod_{j \in C_i} (1 - a_j) \right)$$

λ_i Length of i th edge

$p(i)$ Probability that i th edge will not survive without conservation

C_i Set of children of i th edge

a_j Survival probability of species j



Rescaling Branch Lengths

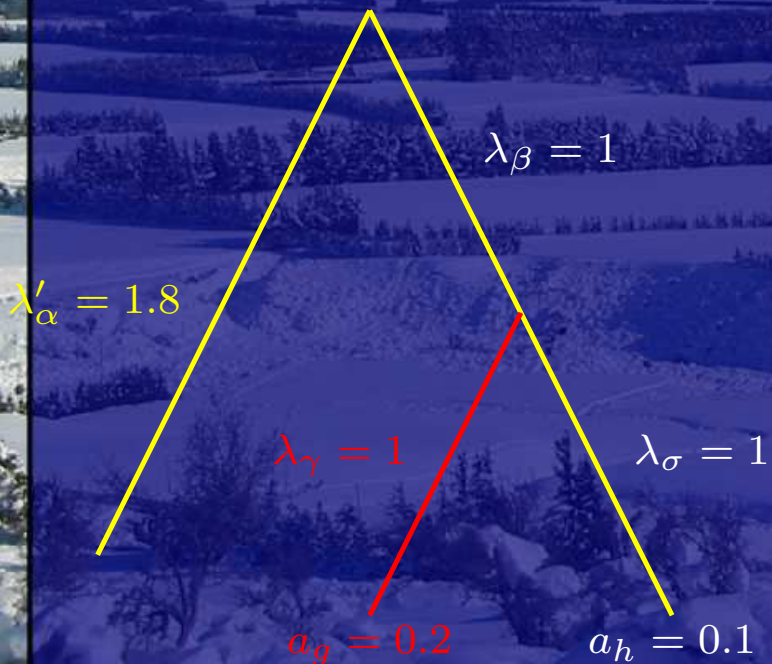
$$\lambda'_i := \lambda_i p(i) = \lambda_i \left(\prod_{j \in C_i} (1 - a_j) \right)$$

λ_i Length of i th edge

$p(i)$ Probability that i th edge will not survive without conservation

C_i Set of children of i th edge

a_j Survival probability of species j



$$p(\alpha) = 1 - 0.1 = 0.9$$

$$p(\gamma) = 1 - 0.2 = 0.8$$

Rescaling Branch Lengths

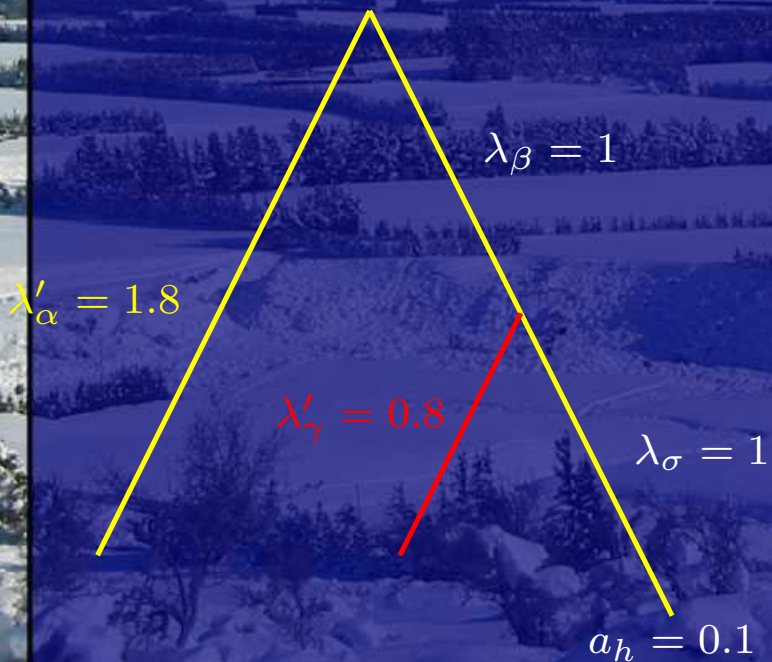
$$\lambda'_i := \lambda_i p(i) = \lambda_i \left(\prod_{j \in C_i} (1 - a_j) \right)$$

λ_i Length of i th edge

$p(i)$ Probability that i th edge will not survive without conservation

C_i Set of children of i th edge

a_j Survival probability of species j



$$p(\alpha) = 1 - 0.1 = 0.9$$

$$p(\gamma) = 1 - 0.2 = 0.8$$

Rescaling Branch Lengths

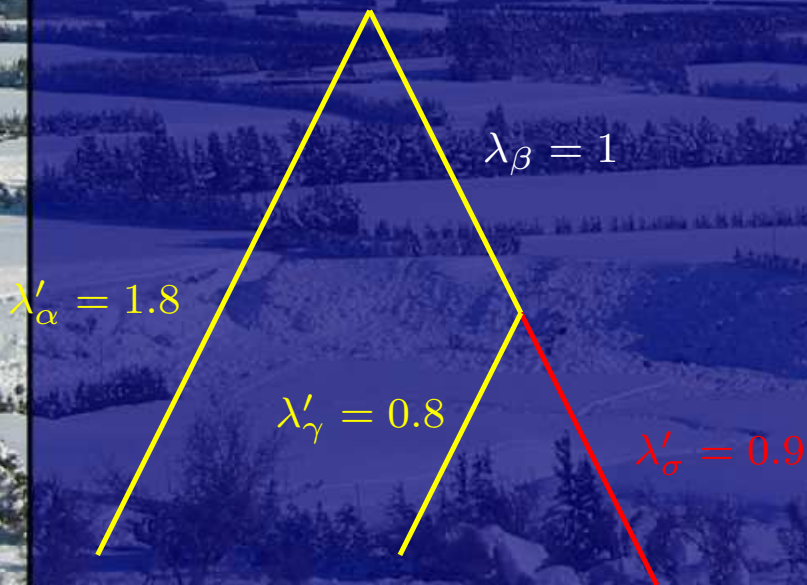
$$\lambda'_i := \lambda_i p(i) = \lambda_i \left(\prod_{j \in C_i} (1 - a_j) \right)$$

λ_i Length of i th edge

$p(i)$ Probability that i th edge will not survive without conservation

C_i Set of children of i th edge

a_j Survival probability of species j



$$p(\alpha) = 1 - 0.1 = 0.9$$

$$p(\gamma) = 1 - 0.2 = 0.8$$

$$p(\sigma) = 0.9$$

Rescaling Branch Lengths

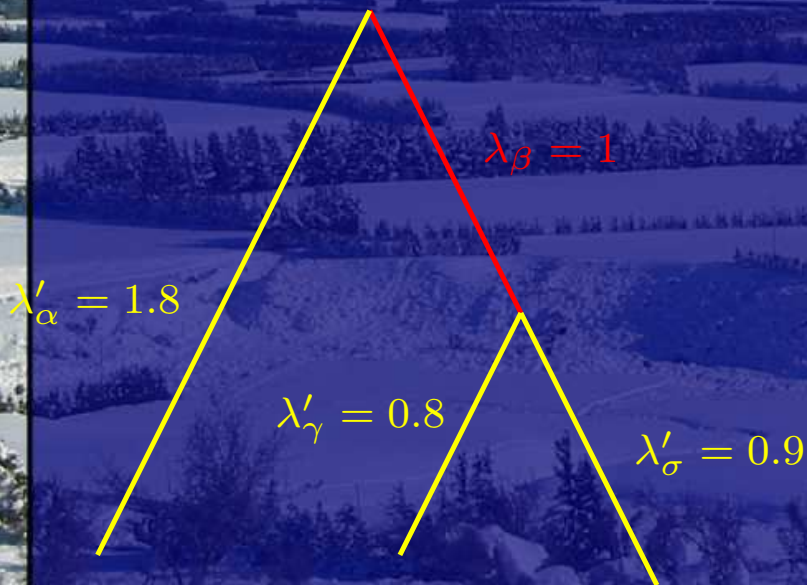
$$\lambda'_i := \lambda_i p(i) = \lambda_i \left(\prod_{j \in C_i} (1 - a_j) \right)$$

λ_i Length of i th edge

$p(i)$ Probability that i th edge will not survive without conservation

C_i Set of children of i th edge

a_j Survival probability of species j



$$p(\alpha) = 1 - 0.1 = 0.9$$

$$p(\gamma) = 1 - 0.2 = 0.8$$

$$p(\sigma) = 0.9$$

$$p(\beta) = (1 - 0.1)(1 - 0.2) = 0.72$$

Rescaling Branch Lengths

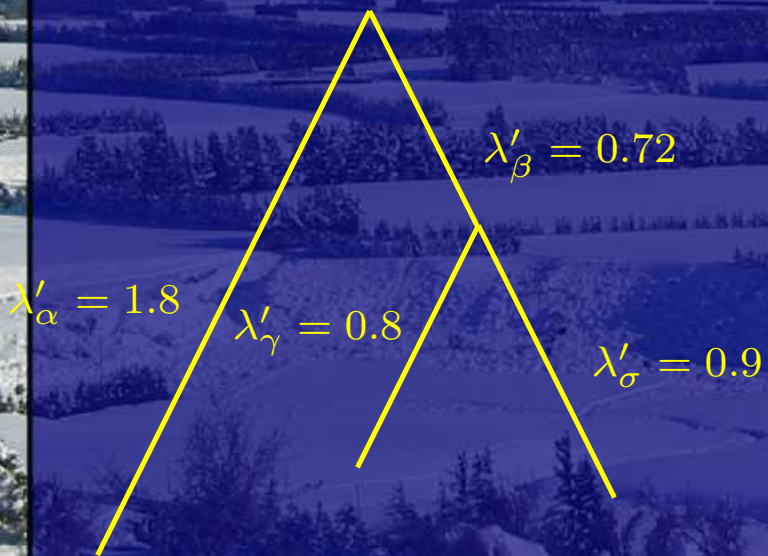
$$\lambda'_i := \lambda_i p(i) = \lambda_i \left(\prod_{j \in C_i} (1 - a_j) \right)$$

λ_i Length of i th edge

$p(i)$ Probability that i th edge will not survive without conservation

C_i Set of children of i th edge

a_j Survival probability of species j



$$p(\alpha) = 1 - 0.1 = 0.9$$

$$p(\gamma) = 1 - 0.2 = 0.8$$

$$p(\sigma) = 0.9$$

$$p(\beta) = (1 - 0.1)(1 - 0.2) = 0.72$$

Rescaling Branch Lengths

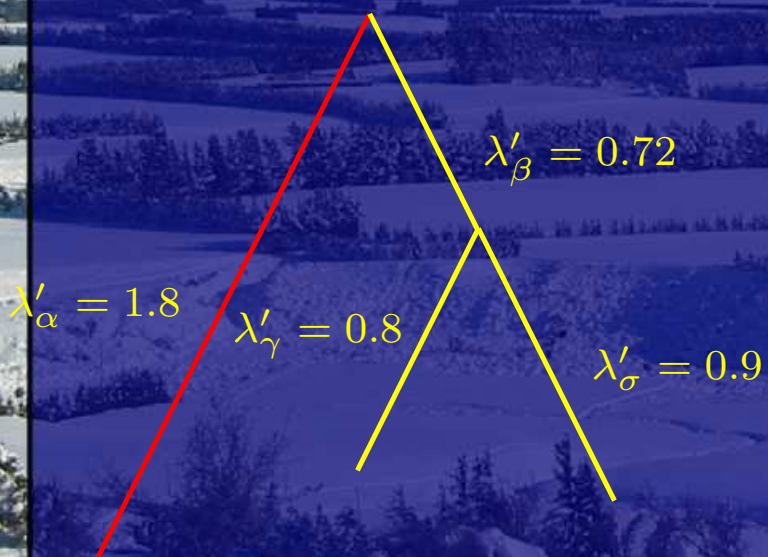
$$\lambda'_i := \lambda_i p(i) = \lambda_i \left(\prod_{j \in C_i} (1 - a_j) \right)$$

λ_i Length of i th edge

$p(i)$ Probability that i th edge will not survive without conservation

C_i Set of children of i th edge

a_j Survival probability of species j



$$p(\alpha) = 1 - 0.1 = 0.9$$

$$p(\gamma) = 1 - 0.2 = 0.8$$

$$p(\sigma) = 0.9$$

$$p(\beta) = (1 - 0.1)(1 - 0.2) = 0.72$$

Rescaling Branch Lengths

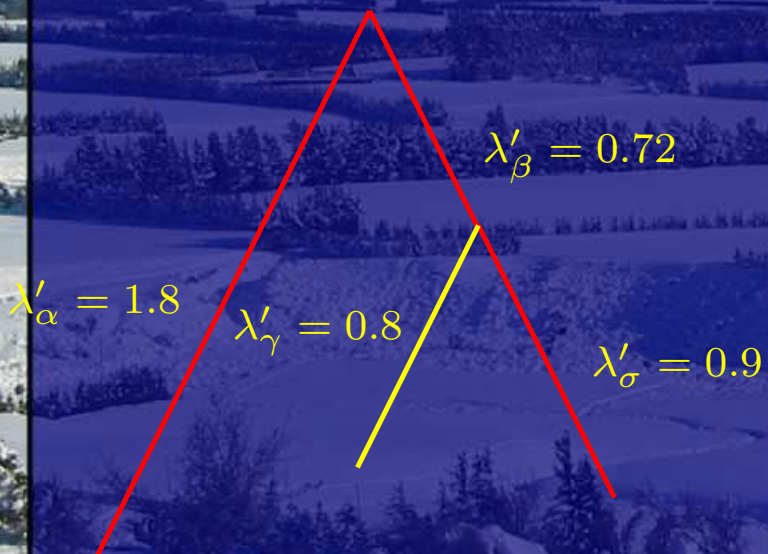
$$\lambda'_i := \lambda_i p(i) = \lambda_i \left(\prod_{j \in C_i} (1 - a_j) \right)$$

λ_i Length of i th edge

$p(i)$ Probability that i th edge will not survive without conservation

C_i Set of children of i th edge

a_j Survival probability of species j



$$p(\alpha) = 1 - 0.1 = 0.9$$

$$p(\gamma) = 1 - 0.2 = 0.8$$

$$p(\sigma) = 0.9$$

$$p(\beta) = (1 - 0.1)(1 - 0.2) = 0.72$$

Uncertain Survival Probabilities

We have an algorithm for the scenario where:

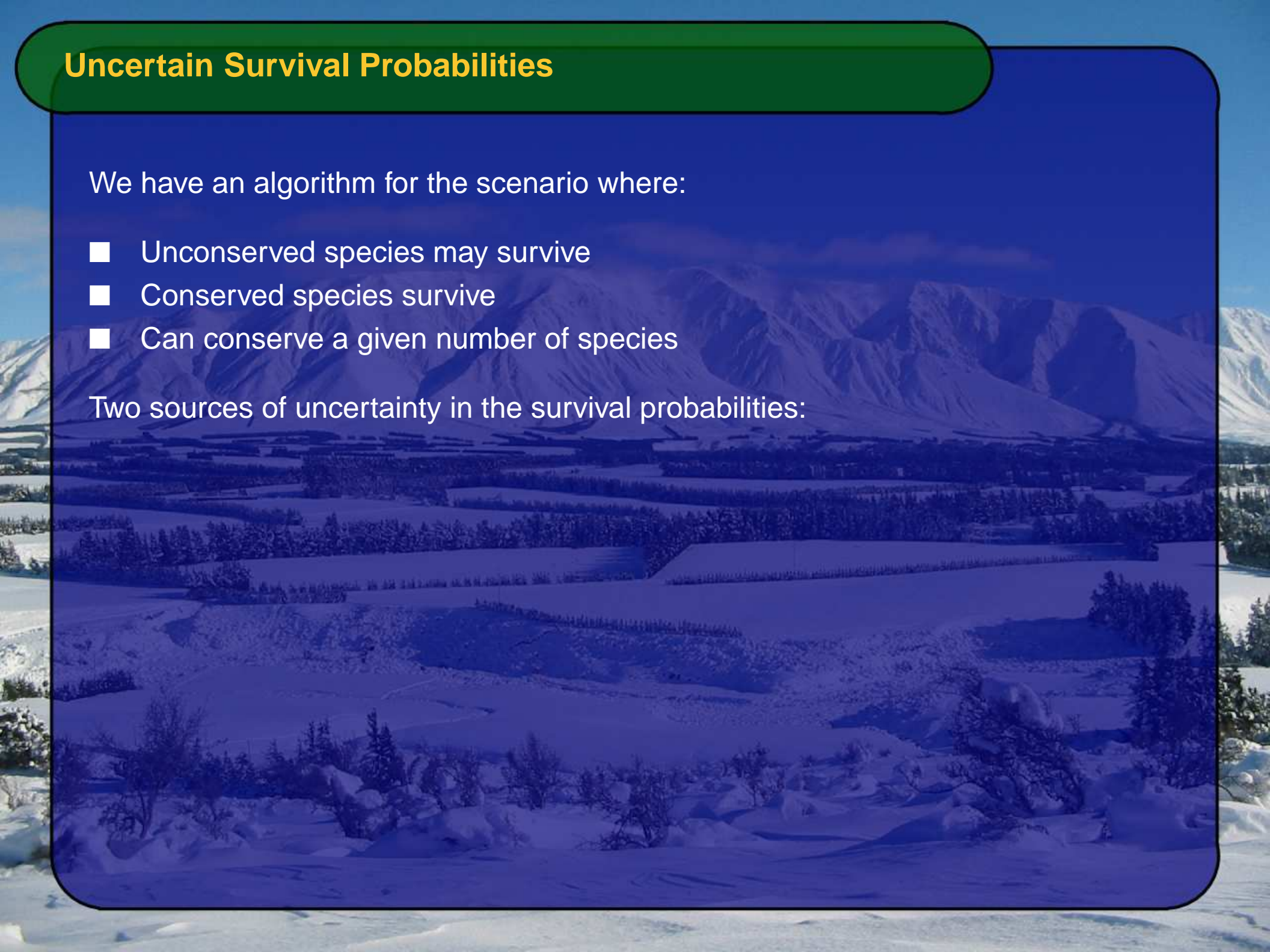
- Unconserved species may survive
- Conserved species survive
- Can conserve a given number of species

Uncertain Survival Probabilities

We have an algorithm for the scenario where:

- Unconserved species may survive
- Conserved species survive
- Can conserve a given number of species

Two sources of uncertainty in the survival probabilities:



Uncertain Survival Probabilities

We have an algorithm for the scenario where:

- Unconserved species may survive
- Conserved species survive
- Can conserve a given number of species

Two sources of uncertainty in the survival probabilities:

- Magnitude of probabilities

Uncertain Survival Probabilities

We have an algorithm for the scenario where:

- Unconserved species may survive
- Conserved species survive
- Can conserve a given number of species

Two sources of uncertainty in the survival probabilities:

- Magnitude of probabilities
- Relative probabilities

Uncertain Survival Probabilities

We have an algorithm for the scenario where:

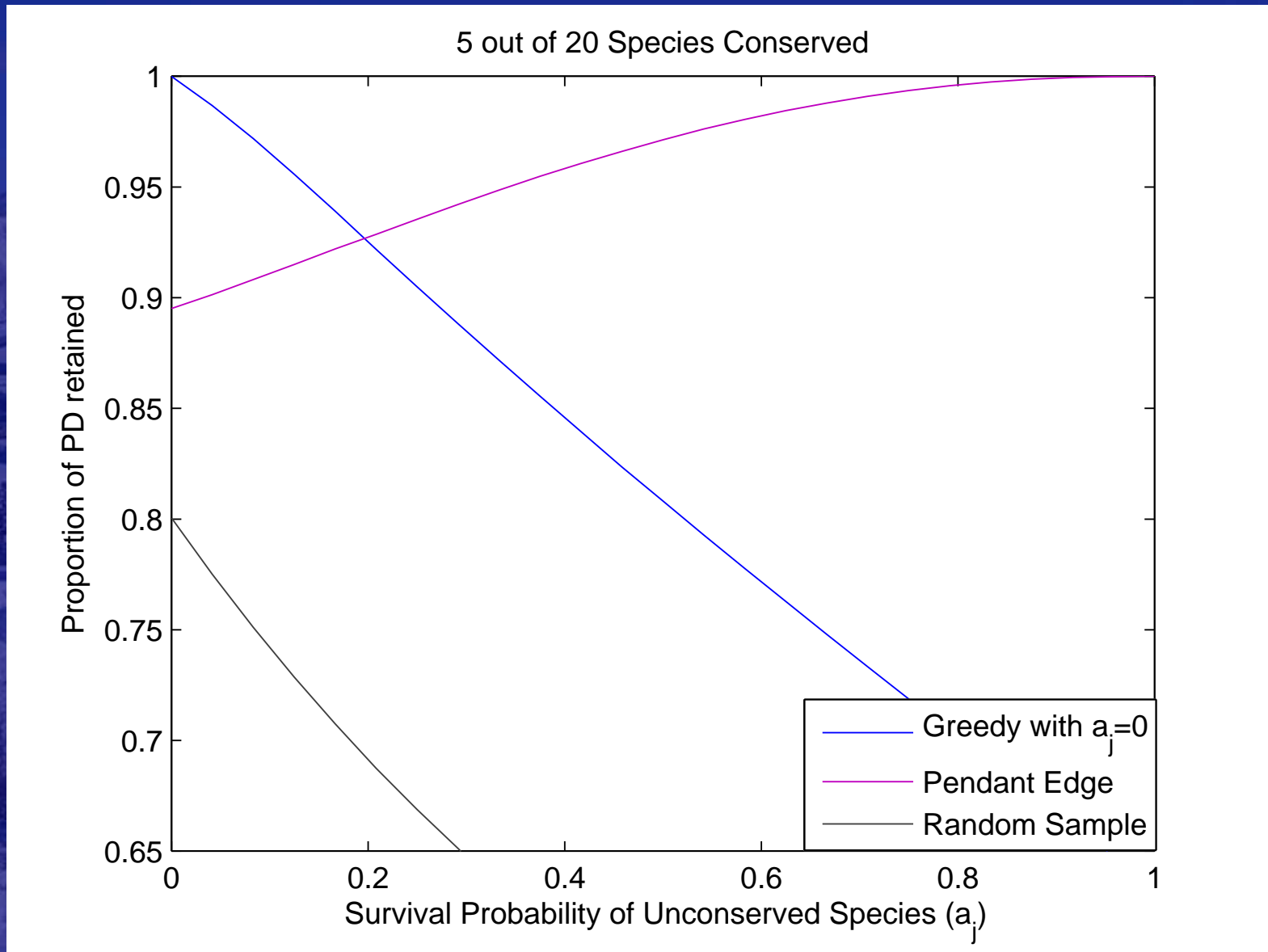
- Unconserved species may survive
- Conserved species survive
- Can conserve a given number of species

Two sources of uncertainty in the survival probabilities:

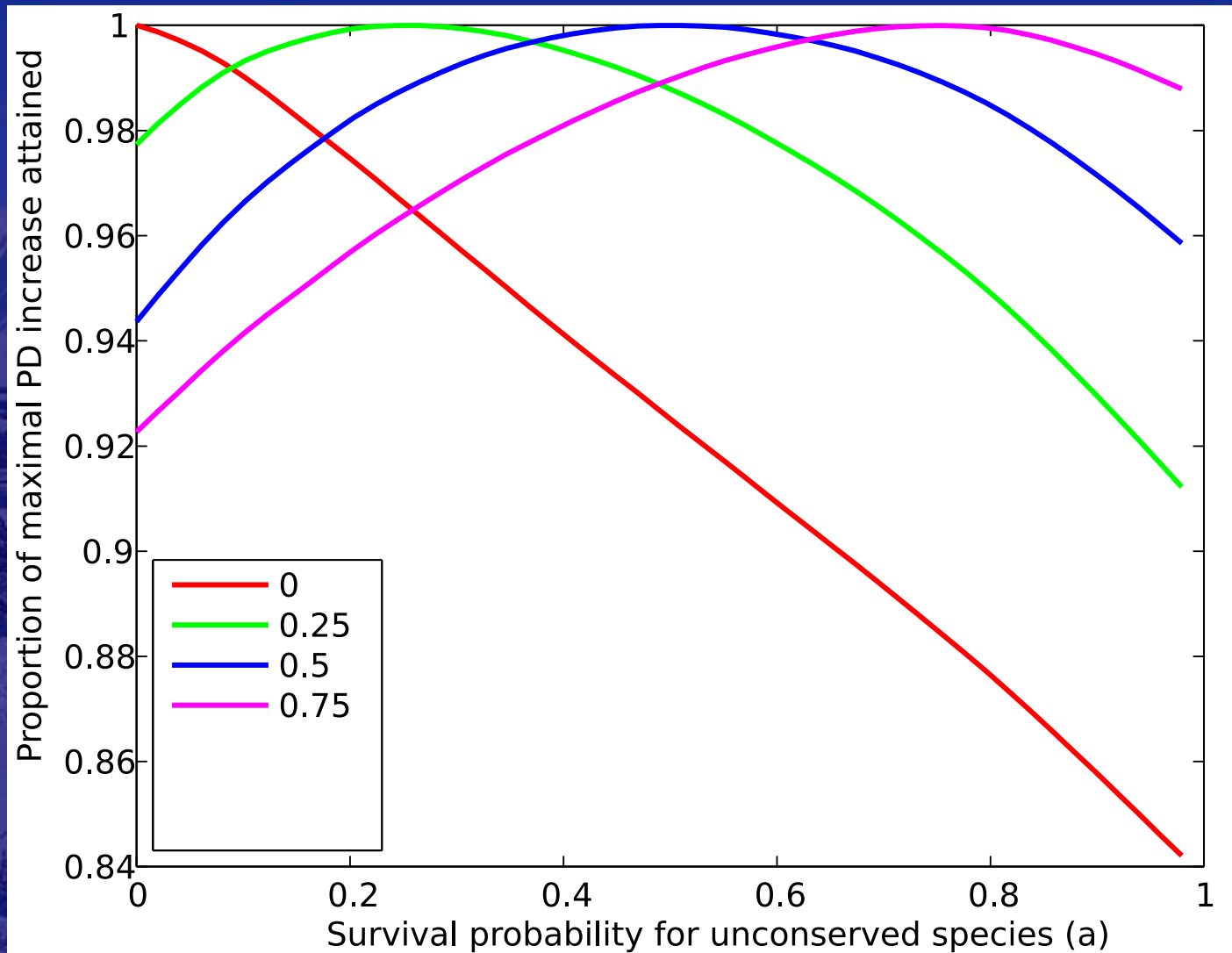
- Magnitude of probabilities
- Relative probabilities

The implicit assumption of zero survival is often made.

The Expected Effect of Incorrect Survival Probabilities



The Expected Effect of Incorrect Survival Probabilities



Dealing With Uncertain Survival Probabilities

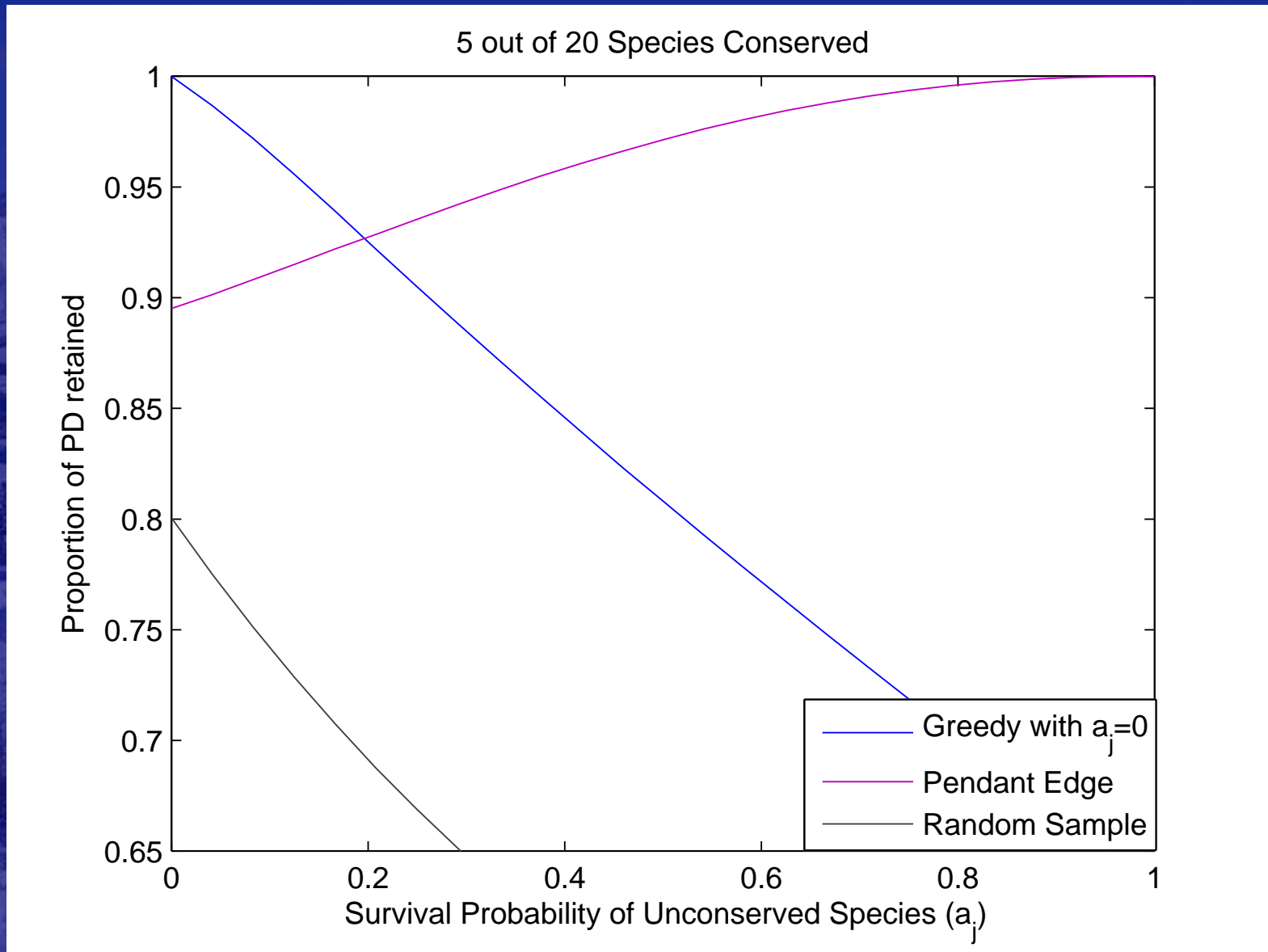
Alternative branch length rescaling:

$$\lambda'_i = \lambda_i \int_{\vec{a}} \phi(\vec{a}) \prod_{j \in C_i} (1 - a_j) d\vec{a}$$

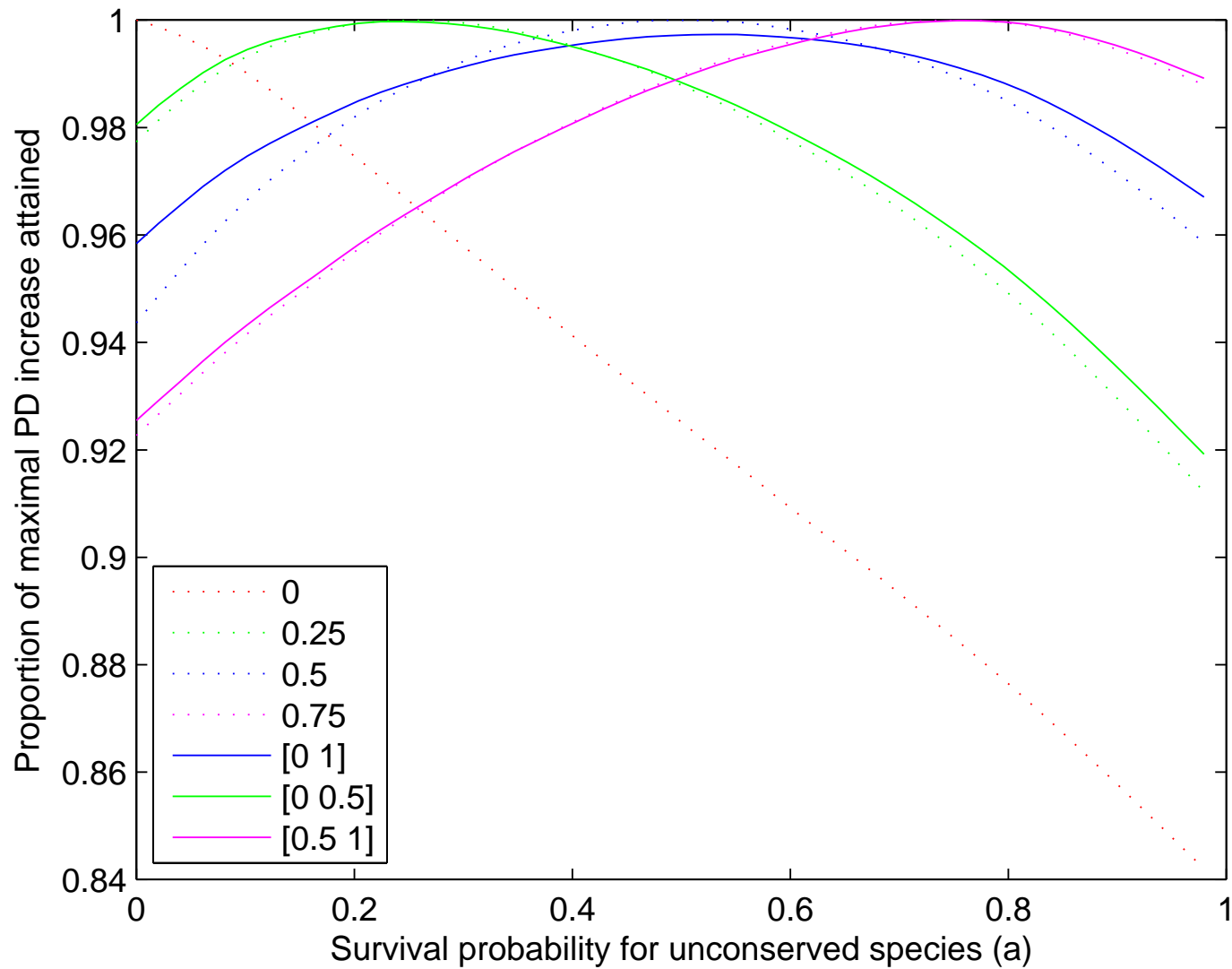
\vec{a} Survival probabilities of all species

$\phi(\vec{a})$ Probability density of survival probabilities

Weighted Greedy Solution



Weighted Greedy Solution



The Noah's Ark Problem

Given an edge-weighted phylogenetic tree, and values (a_j, b_j, c_j) for each taxon j , maximize $\mathbb{E}(PD|S)$ over all subsets S of taxa, subject to the constraint:

$$\sum_{j \in S} c_j \leq B.$$

- a_j Survival prob. of species j if it is not conserved
- b_j Survival prob. of species j if it is conserved
- c_j Cost of conserving species j

The Noah's Ark Problem

Given an edge-weighted phylogenetic tree, and values (a_j, b_j, c_j) for each taxon j , maximize $\mathbb{E}(PD|S)$ over all subsets S of taxa, subject to the constraint:

$$\sum_{j \in S} c_j \leq B.$$

- a_j Survival prob. of species j if it is not conserved
- b_j Survival prob. of species j if it is conserved
- c_j Cost of conserving species j

COMPUTATIONALLY EXPENSIVE: may need to examine all 2^n subsets of taxa

Algorithms for The Noah's Ark Problem

Algorithms exist for solving certain instances of the NAP:

- M. Steel Phylogenetic diversity and the greedy algorithm.2005
- Hartmann, K. and M. Steel. 2006. Maximising phylogenetic diversity in biodiversity conservation.
- Hartmann, K. and M. Steel. 2006. Phylogenetic diversity from combinatorics to ecology.
- Pardi, F. and N. Goldman. 2006. Resource aware taxon selection for maximising phylogenetic diversity.
- B.Q. Minh, S. Klaere, and A. von Haeseler. 2006. Phylogenetic Diversity within Seconds

and various extensions of it.

Algorithms for The Noah's Ark Problem

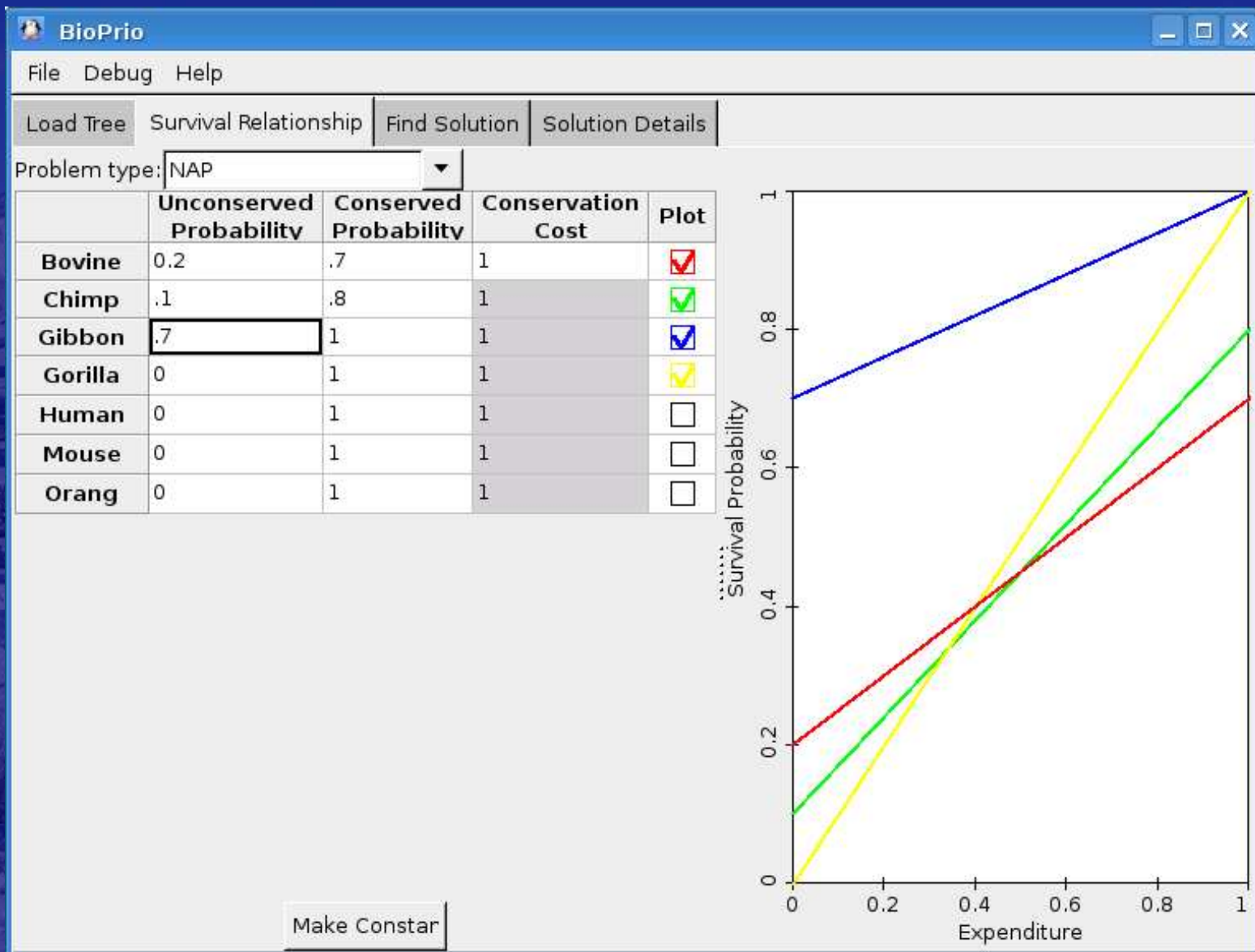
Some conservation managers have yet to appreciate the beauty of equations like this:

$$\Delta p(i) = \left(\prod_{m \in C_i - Z} (1 - a_m) \prod_{l \in C_i \cap Z - \{z\}} (1 - b_l) - \prod_{m \in C_i - Y - \{z\}} (1 - a_m) \prod_{l \in C_i \cap Y} (1 - b_l) \right) (a_z - b_z)$$

Hartmann, K. and M. Steel. 2006. Maximising phylogenetic diversity in biodiversity conservation.

Hence the literature is inaccessible for many conservation managers/biologists.

GUI for the NAP



GUI for the NAP

BioPrio

File Debug Help

Load Tree Survival Relationship Find Solution Solution Details

Budget: [Slider]

	Air portion	Pendant edge	Shapley value	31	32
Bovine	1.21	1.21	1.36	1	1
Chimp	0.37	0.37	0.511	1	0
Gibbon	0.665	0.665	0.74	1	1
Gorilla	0.453	0.453	0.556	1	1
Human	0.37	0.37	0.511	0	1
Mouse	1.21	1.21	1.36	1	1
Orang	0.515	0.515	0.602	1	1
Expected PD	5.26	5.26	5.26	5.38	5.38
Cost	5.76	5.76	5.76	6	6

Greedy traversal: Human Chimp

Show:
 Species specific indices
 Summary statistics
 Solutions

+ Add - Remove

0 0.2 0.4 0.6 0.8 1 1.2 1.4

Problems ...

The NAP is critiqued for being too simple:

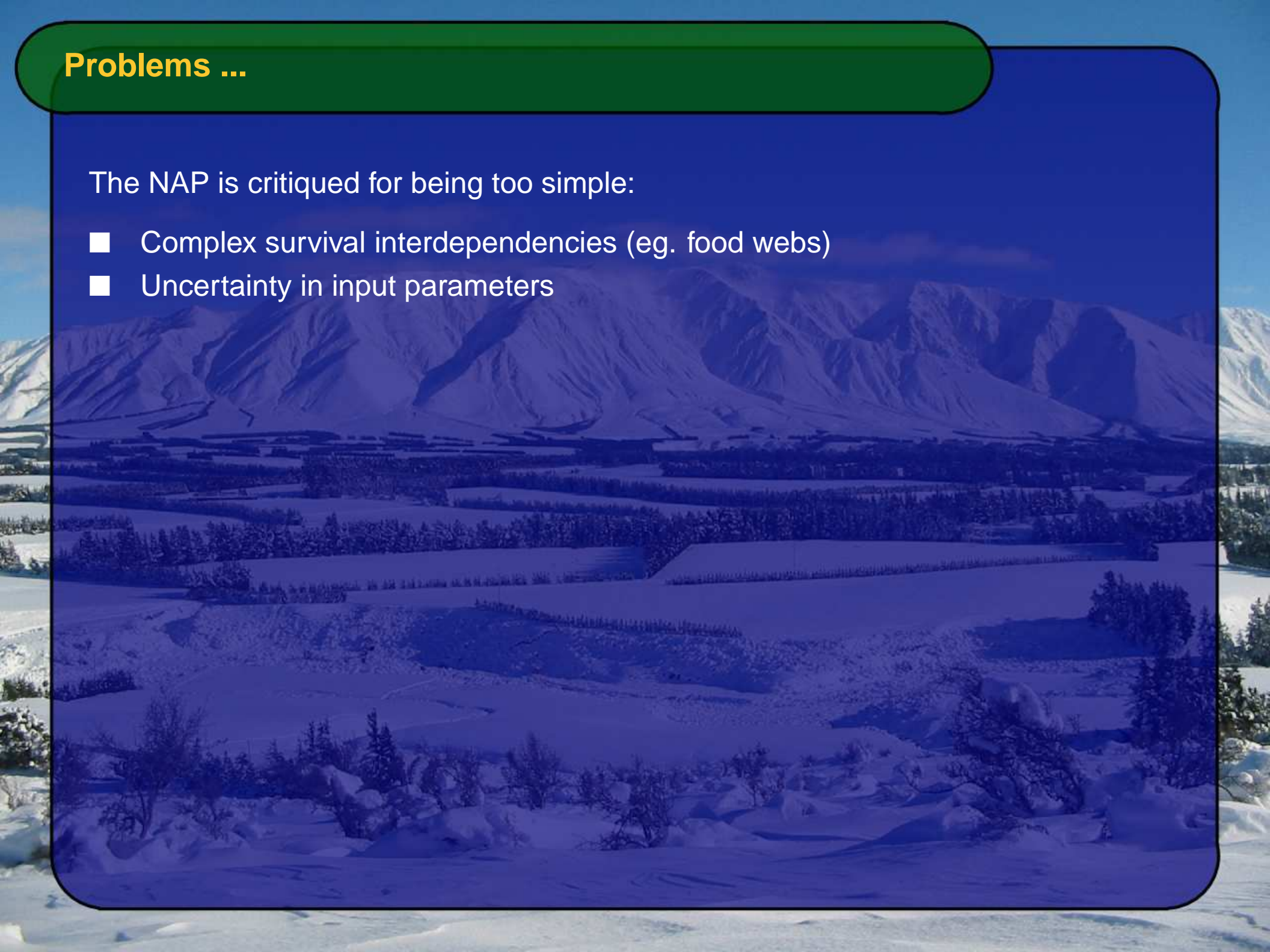
- Complex survival interdependencies (eg. food webs)



Problems ...

The NAP is critiqued for being too simple:

- Complex survival interdependencies (eg. food webs)
- Uncertainty in input parameters



Problems ...

The NAP is critiqued for being too simple:

- Complex survival interdependencies (eg. food webs)
- Uncertainty in input parameters
- Interdependencies among conservation mechanisms

Problems ...

The NAP is critiqued for being too simple:

- Complex survival interdependencies (eg. food webs)
- Uncertainty in input parameters
- Interdependencies among conservation mechanisms

AND too complex!!

Problems ...

The NAP is critiqued for being too simple:

- Complex survival interdependencies (eg. food webs)
- Uncertainty in input parameters
- Interdependencies among conservation mechanisms

AND too complex!!

Some conservation managers I have talked to want a method that:

- is easy to explain
- ranks species in a unique order

Problems ...

The NAP is critiqued for being too simple:

- Complex survival interdependencies (eg. food webs)
- Uncertainty in input parameters
- Interdependencies among conservation mechanisms

AND too complex!!

Some conservation managers I have talked to want a method that:

- is easy to explain
- ranks species in a unique order

Species specific indices

- Aim to divide the total length of a tree among the leaves

Problems ...

The NAP is critiqued for being too simple:

- Complex survival interdependencies (eg. food webs)
- Uncertainty in input parameters
- Interdependencies among conservation mechanisms

AND too complex!!

Some conservation managers I have talked to want a method that:

- is easy to explain
- ranks species in a unique order

Species specific indices

- Aim to divide the total length of a tree among the leaves
- (Some) are simple to calculate and explain

Problems ...

The NAP is critiqued for being too simple:

- Complex survival interdependencies (eg. food webs)
- Uncertainty in input parameters
- Interdependencies among conservation mechanisms

AND too complex!!

Some conservation managers I have talked to want a method that:

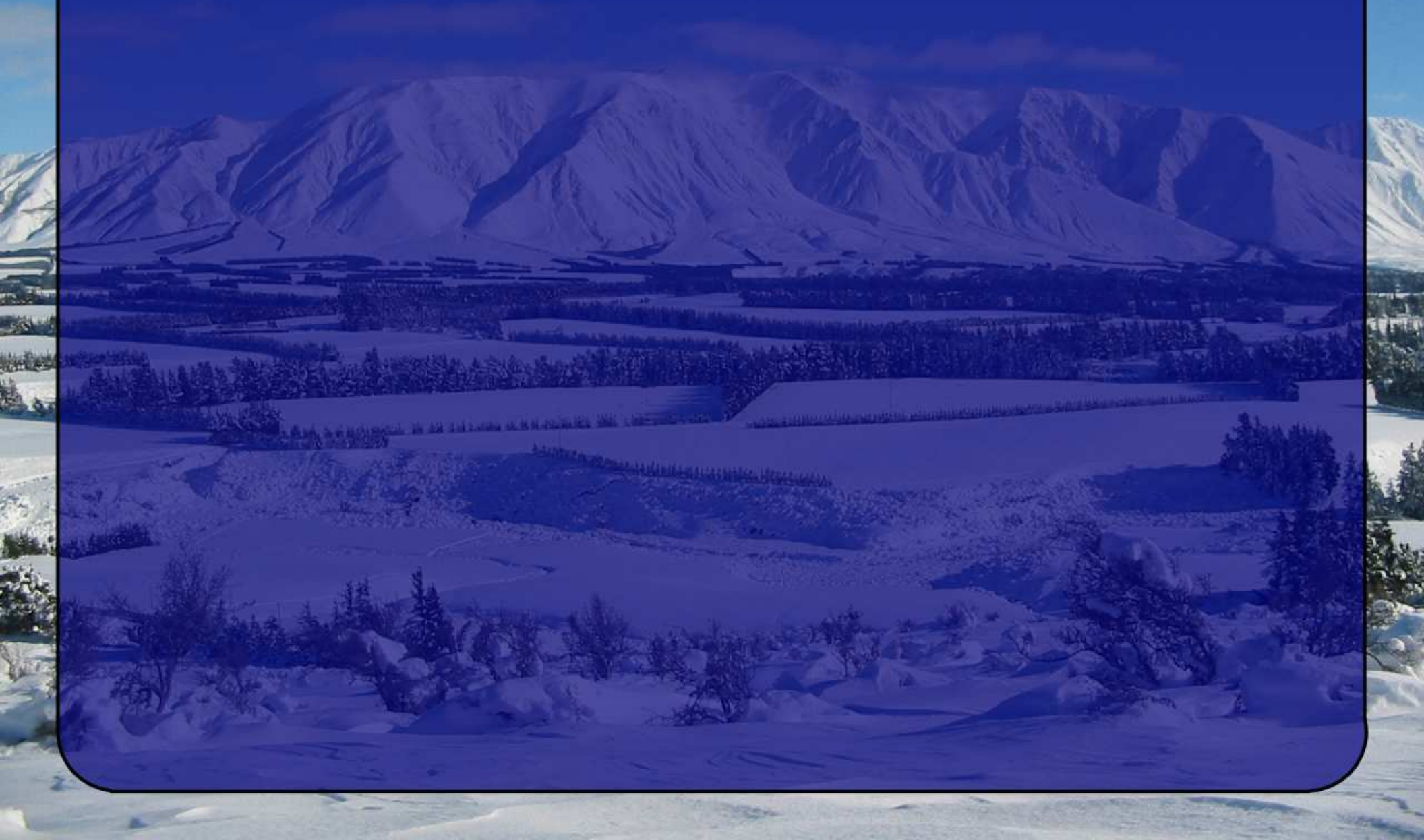
- is easy to explain
- ranks species in a unique order

Species specific indices

- Aim to divide the total length of a tree among the leaves
- (Some) are simple to calculate and explain
- But how good are they?

Species specific indices

- Pendant edge



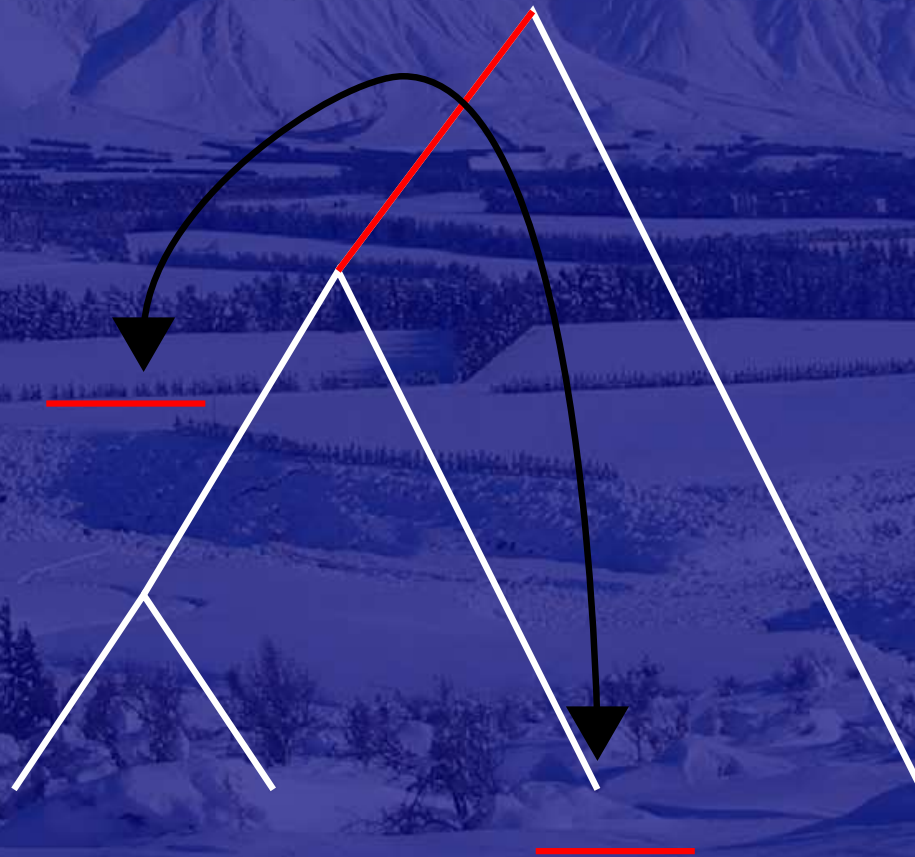
Species specific indices

- Pendant edge
- Fair proportion



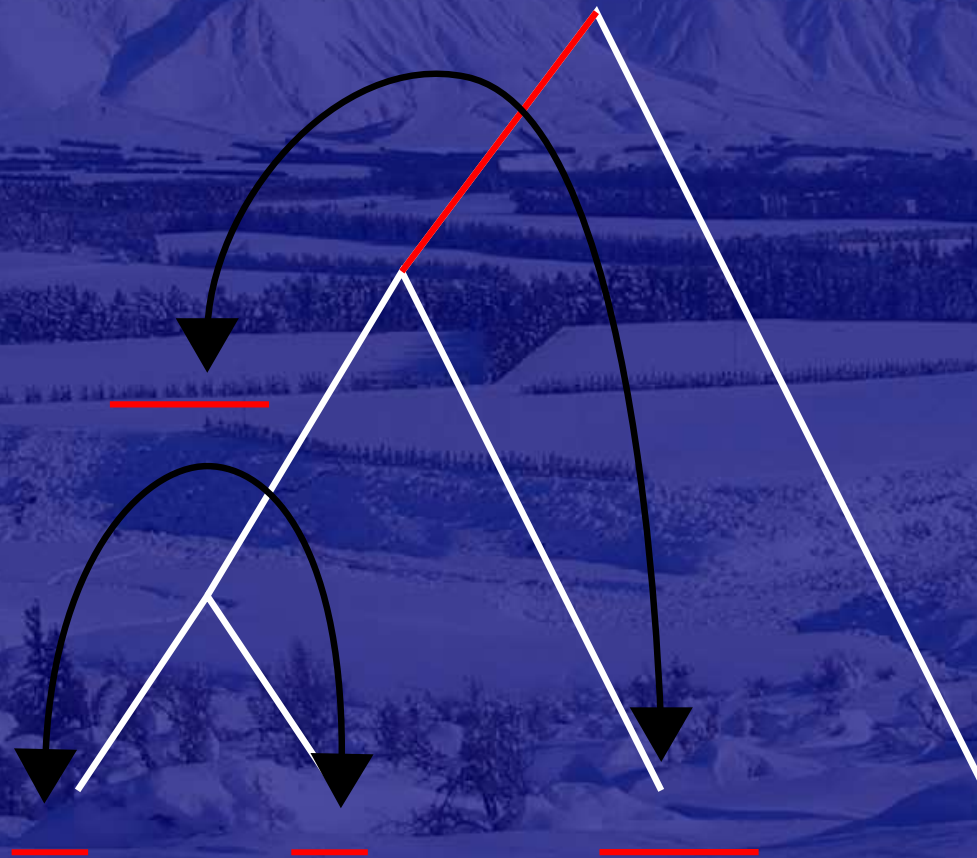
Species specific indices

- Pendant edge
- Fair proportion
- Equal Splits

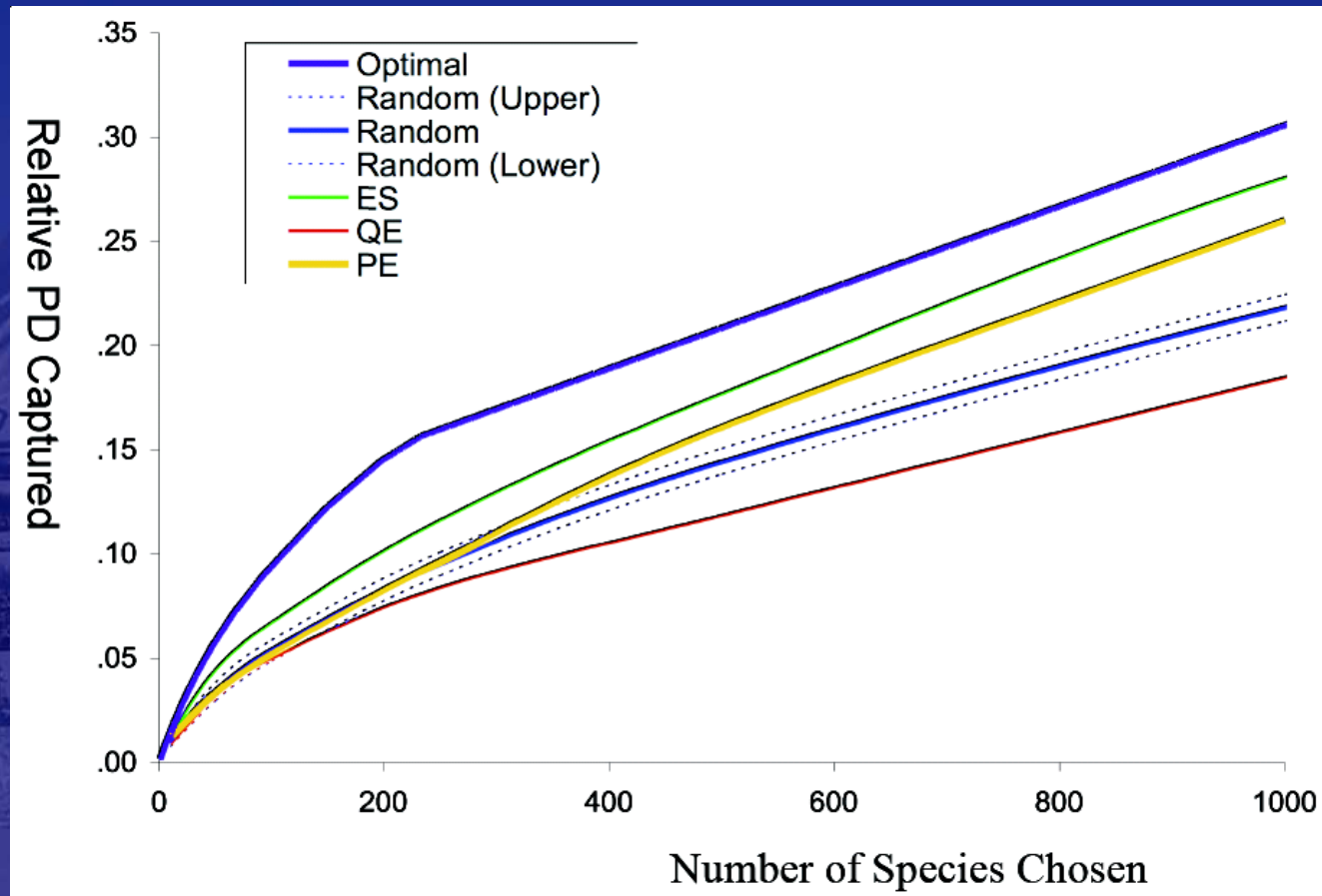


Species specific indices

- Pendant edge
- Fair proportion
- Equal Splits



Indices VS NAP

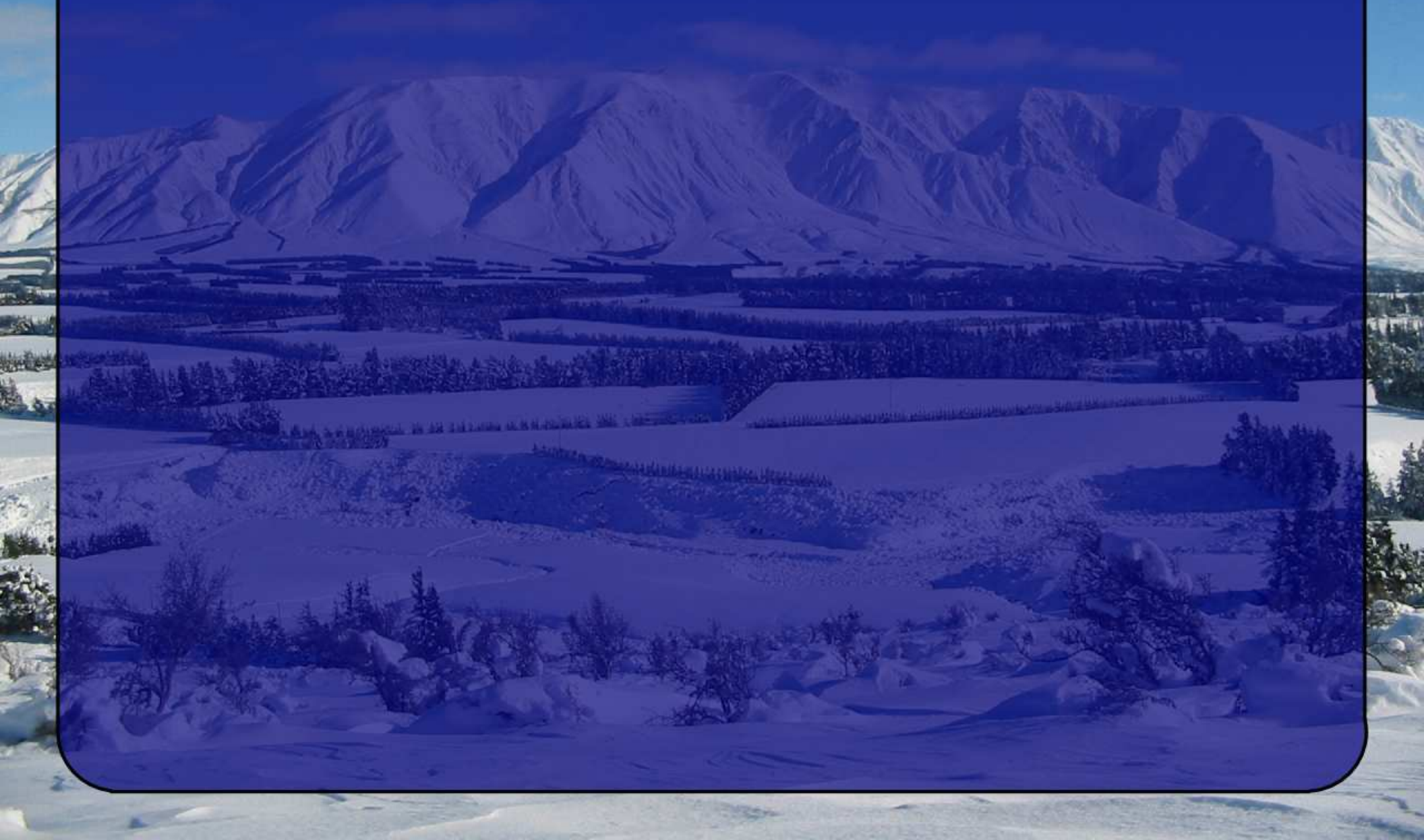


Evolutionarily distinct species often capture more phylogenetic diversity than expected (submitted)

D. Redding, K. Hartmann, A. Mimoto, D. Bokal, M. DeVos and A. Mooers

Findings

- Most indices allocate pendant edges to corresponding leaves



Findings

- Most indices allocate pendant edges to corresponding leaves
- Pendant edges contain roughly half the tree's PD



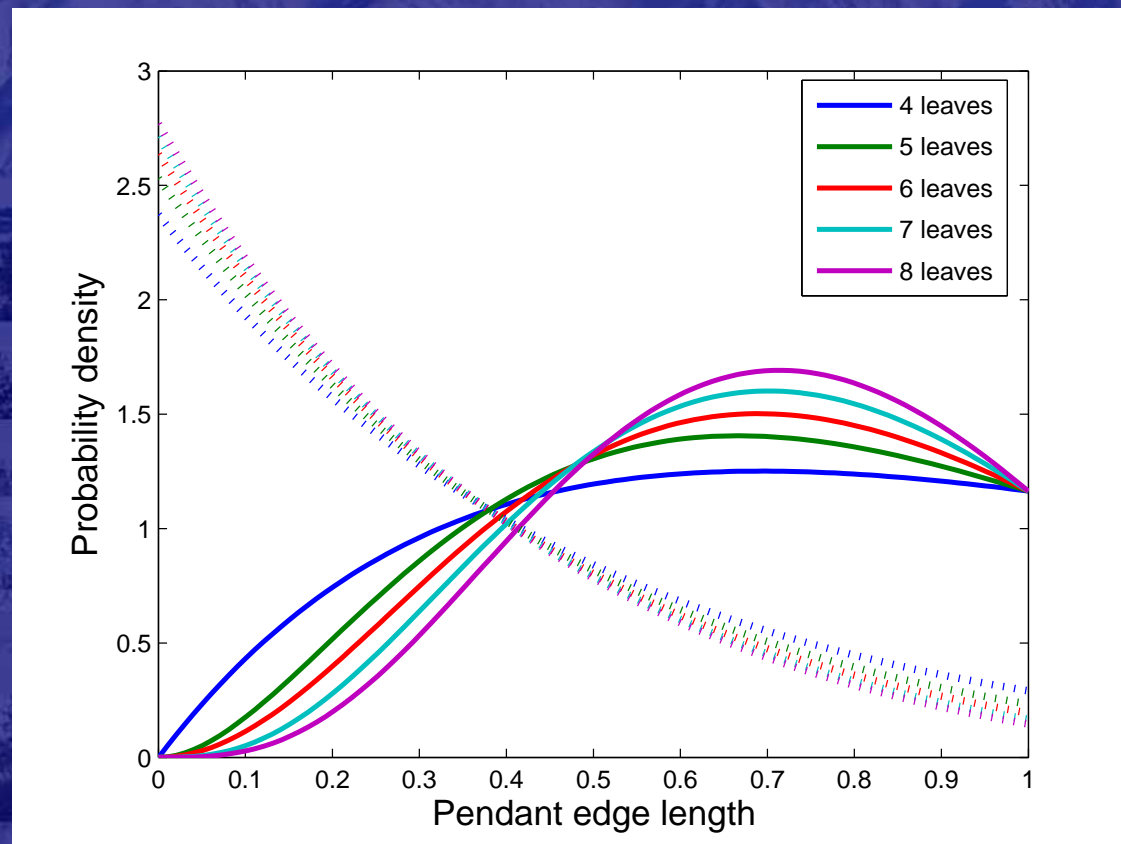
Findings

- Most indices allocate pendant edges to corresponding leaves
- Pendant edges contain roughly half the tree's PD

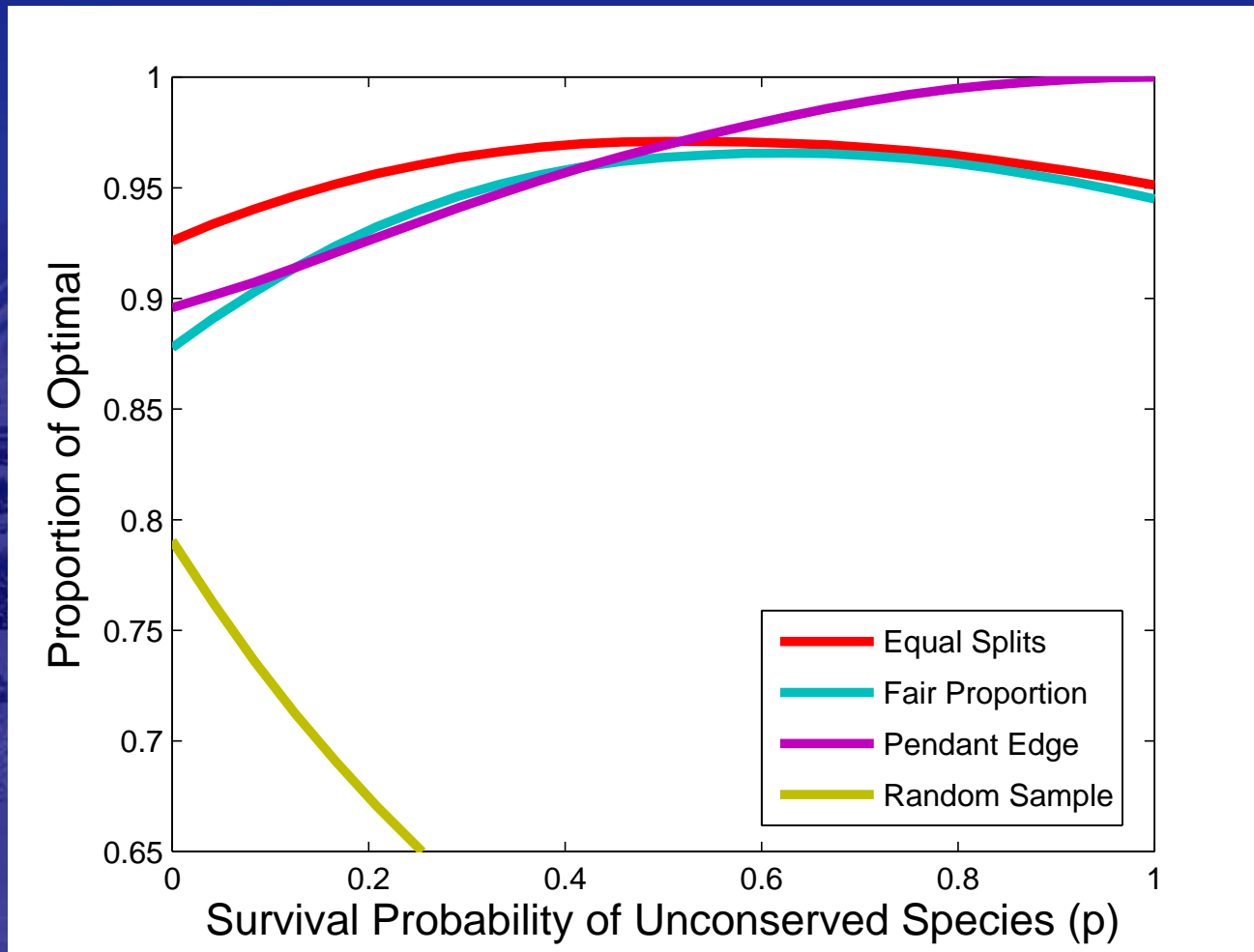


Findings

- Most indices allocate pendant edges to corresponding leaves
- Pendant edges contain roughly half the tree's PD



Indices for time scales

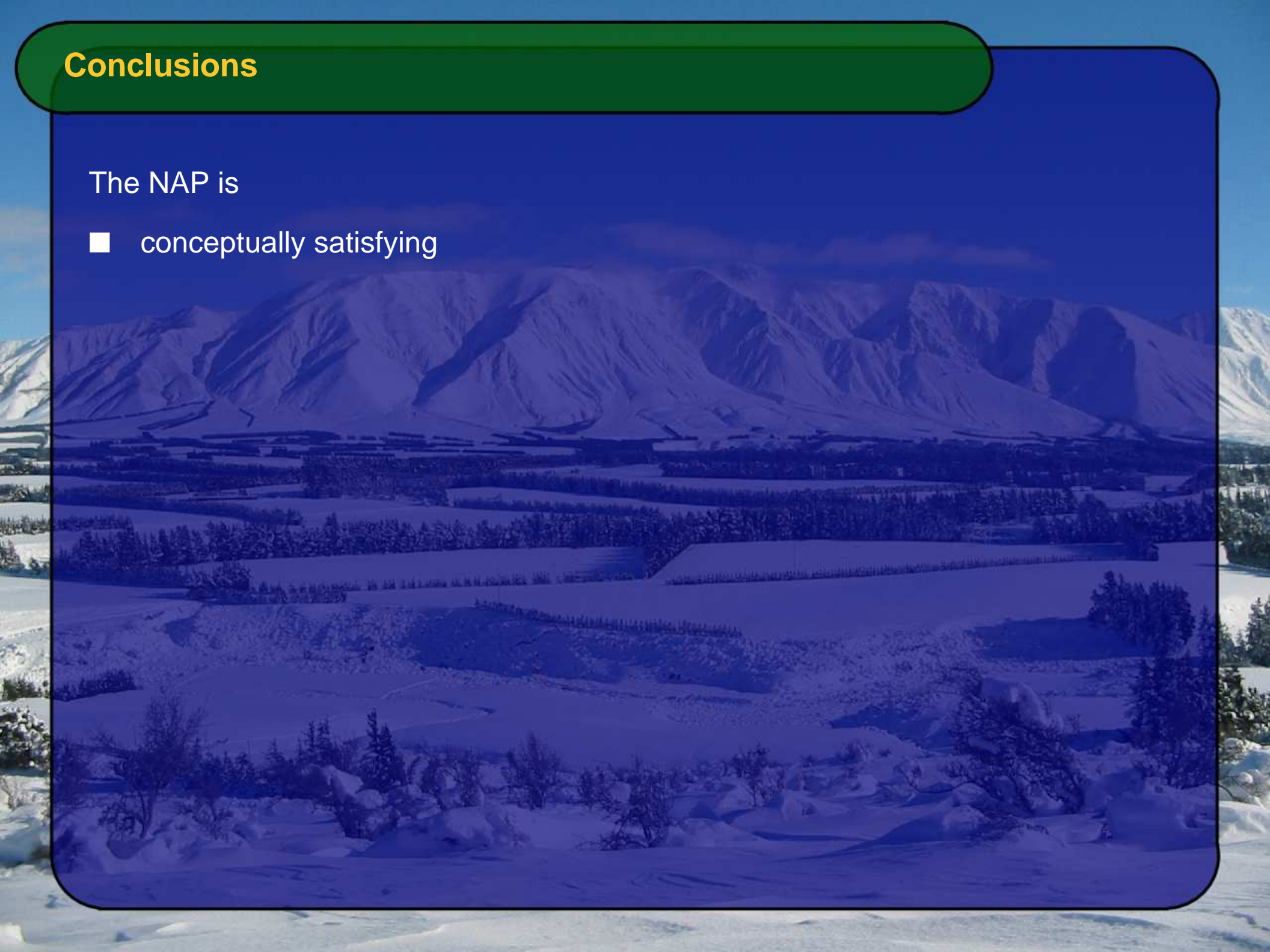


Prioritising species for conservation with uncertain survival probabilities (in prep)
K. Hartmann, S. Paquette, A. Mooers

Conclusions

The NAP is

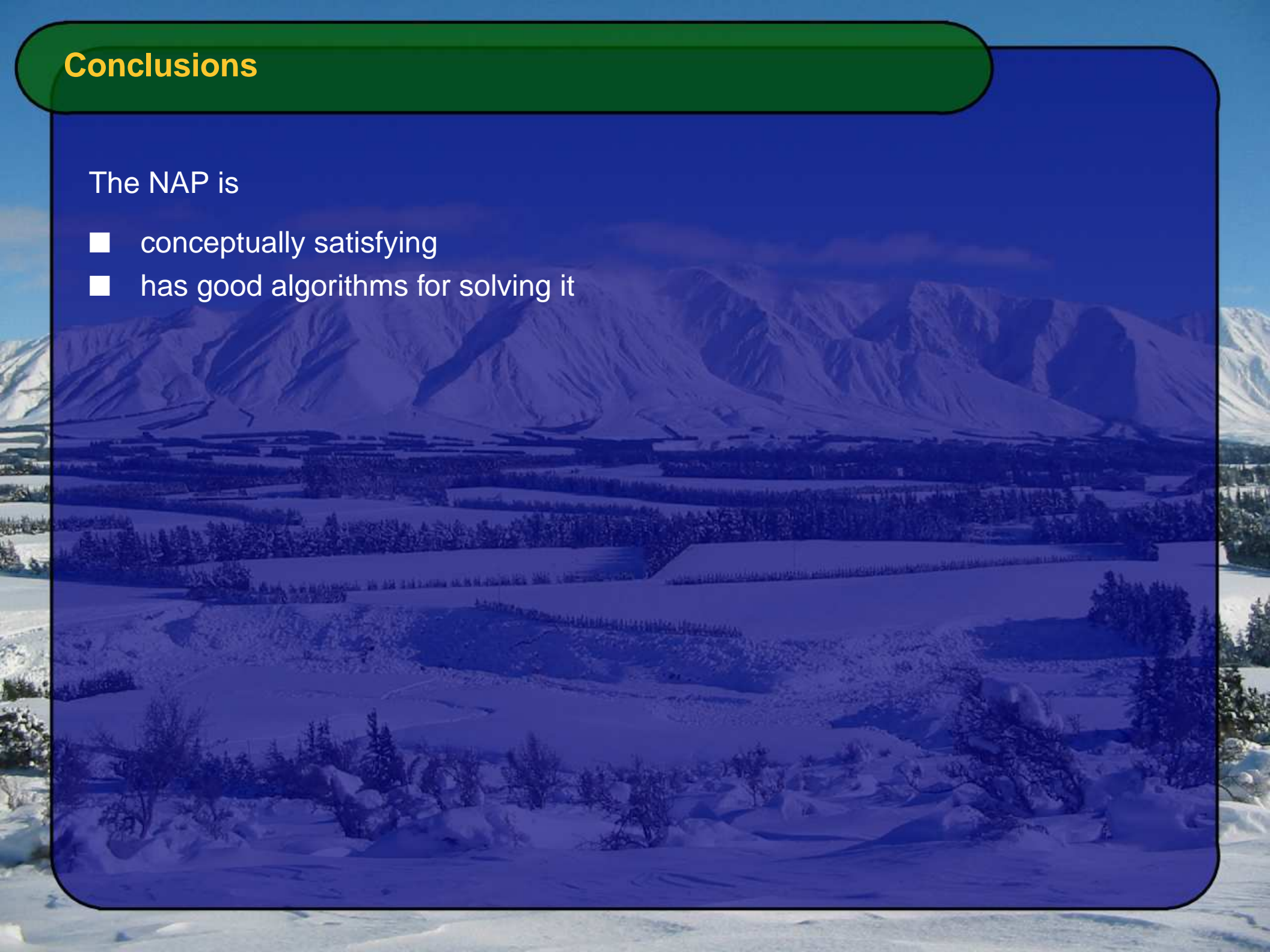
- conceptually satisfying



Conclusions

The NAP is

- conceptually satisfying
- has good algorithms for solving it



Conclusions

The NAP is

- conceptually satisfying
- has good algorithms for solving it
- has a GUI to allow simple application

Conclusions

The NAP is

- conceptually satisfying
- has good algorithms for solving it
- has a GUI to allow simple application
- is robust to uncertainty in some input parameters

Conclusions

The NAP is

- conceptually satisfying
- has good algorithms for solving it
- has a GUI to allow simple application
- is robust to uncertainty in some input parameters
- much work is being done on extensions to it

Conclusions

The NAP is

- conceptually satisfying
- has good algorithms for solving it
- has a GUI to allow simple application
- is robust to uncertainty in some input parameters
- much work is being done on extensions to it

However there is some resistance to its application

Conclusions

The NAP is

- conceptually satisfying
- has good algorithms for solving it
- has a GUI to allow simple application
- is robust to uncertainty in some input parameters
- much work is being done on extensions to it

However there is some resistance to its application

If indices are going to be used anyway then:

- We can inform the choice of index

Conclusions

The NAP is

- conceptually satisfying
- has good algorithms for solving it
- has a GUI to allow simple application
- is robust to uncertainty in some input parameters
- much work is being done on extensions to it

However there is some resistance to its application

If indices are going to be used anyway then:

- We can inform the choice of index
- Whilst continuing to suggest more sophisticated methods

Conclusions

The NAP is

- conceptually satisfying
- has good algorithms for solving it
- has a GUI to allow simple application
- is robust to uncertainty in some input parameters
- much work is being done on extensions to it

However there is some resistance to its application

If indices are going to be used anyway then:

- We can inform the choice of index
- Whilst continuing to suggest more sophisticated methods
- Enjoy solving some interesting mathematical problems in the meantime

Conclusions

The NAP is

- conceptually satisfying
- has good algorithms for solving it
- has a GUI to allow simple application
- is robust to uncertainty in some input parameters
- much work is being done on extensions to it

However there is some resistance to its application

If indices are going to be used anyway then:

- We can inform the choice of index
- Whilst continuing to suggest more sophisticated methods
- Enjoy solving some interesting mathematical problems in the meantime