Maximum Entropy as a Foundation for Theory Building in Ecology

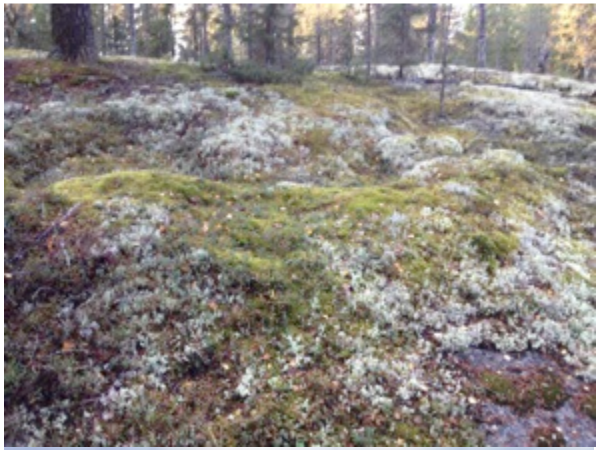
John Harte

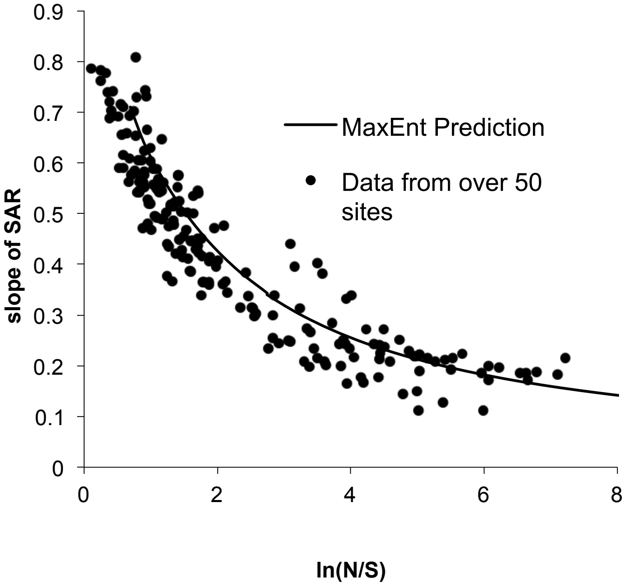
UC Berkeley

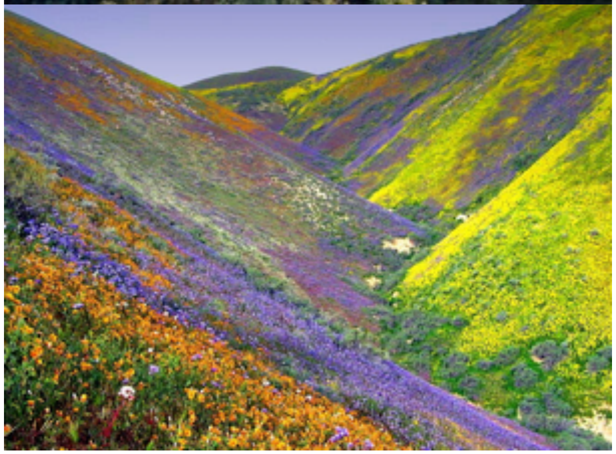
NIMBios

Knoxville TN

April 8-10, 2015







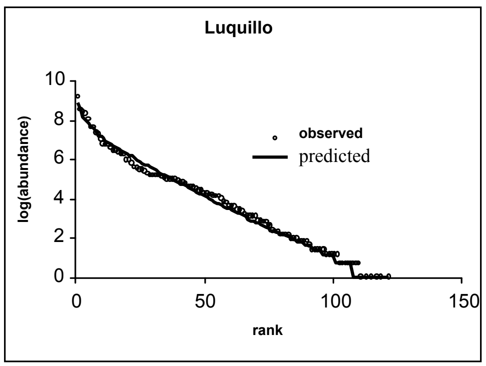
Prevalent Patterns in Ecology & Useful Metrics to quantify them

Prevalent Patterns in Ecology	Useful Metrics to quantify them
1. # species increases, with diminishing returns, on area censused.	Species-Area Relationship. (SAR)
2. Most species are rare, some abundant.	Species-Abundance Distribution (SAD)
3. Some individuals are big, most small.	Individuals Size Distribution
4. Common species have small individuals.	Size-abundance distribution
5. Individuals within species tend to spatially aggregate.	Spatial-Abundance Distribution
6. More trophic specialists than generalists.	Foodweb node-linkage distribution

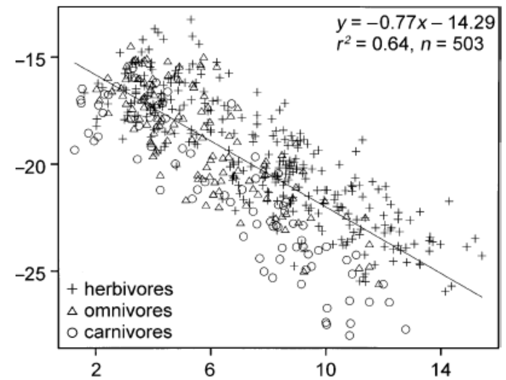
Macroecology: Patterns in the abundance, spatial distribution, & energetics of species…

Of special interest: patterns that are widely observed across habitats, taxa, & spatial scales

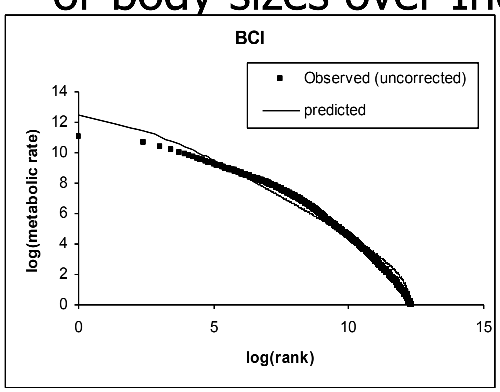
Species-Abundance Distribution



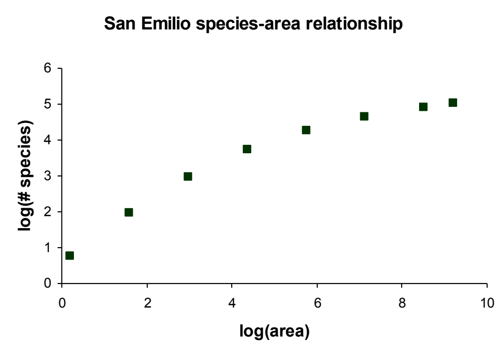
Body Size-Abundance Relationship



Distribution of Metabolic Rates or body sizes over Individuals



Species-Area Relationship



The Dilemma faced by Ecosystem Modelers:

- **Many mechanisms and processes:**

  predation, mutualism, competition, dispersal, speciation, birth, death, pollination, cannibalism, migration, …

- **Many traits and behaviors:**

  specific leaf area, body size, speed, phenology, food preferences, rooting depth, mating strategies, coloration, temperature/drought tolerance, nutrient acquisition strategies, temporal allocation strategies …

- **Stochastic environments, historical contingency**

All influence Patterns in Macroecology

Hence basing models on explicit mechanisms, traits & behaviors generally results in

Arbitrary choices of dominant mechanisms

Many adjustable parameters

Models that are not readily falsifiable.

The Maximum Entropy Theory of Ecology (METE)

Goal: To predict prevalent patterns in "macroecology"

- **Across taxa:** plants, bugs, birds,…
- **Across spatial scale:** small patches to large biomes
- **Across habitats:** forests, meadows, deserts, tundra,…

- **without adjustable parameters**
- **without pre-judging what specific mechanisms drive the system**

And thereby

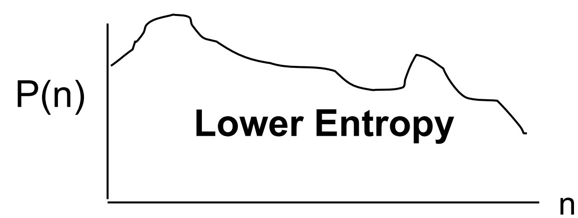
- **gain insight into the forces that shape ecosystems**
- **make reliable predictions that can aid in conservation**

The Maximum Entropy Concept:
Just what is being maximized?

Here "**entropy**" refers to **information entropy**, $I = - \Sigma_n P(n) \log(P(n))$,
not thermodynamic entropy

Information entropy is a measure of the lack of structure or detail in the probability distribution describing your knowledge.

Maximizing entropy $\leftrightarrow$ finding the smoothest possible probability distribution that is compatible with the constraints that arise from prior knowledge.



If both of these distributions are consistent your prior knowledge, you should prefer the one with higher entropy. It makes fewer implicit unwarranted assumptions

Jaynes, E. T., 1982, On the Rationale of Maximum Entropy Methods, *Proc. IEEE.*, 70, 939;

MaxEnt and the State Variable Concept

In **Thermodynamics**, these state variables characterize the system:

P: pressure

V: volume

T: temperature

n: number of moles

PV=nRT, Boltzmann distribution of energy levels, entropy law, equipartition, binomial distribution of molecules in space … can all be derived from MaxEnt, with constraints provided by these state variables. (Jaynes 1957a, b)

**As in statistical physics,
we will predict microstate structure
from macrostate constraints.**

In **Ecology** we start with:

A_0 *: area of ecosystem or census plot*

S_0 *: total number of species in* A_0

N_0 *: total number of individuals amongst all those species*

E_0 *: total metabolic rate of all those individuals*

A ROAD MAP of the ASNE MODEL of METE

INPUT DATA

***State Variables*:**

- **Area: A**
- **total # Species: S**
- **total # Individuals: N**
- **total Metabolic rate E**

THEORY

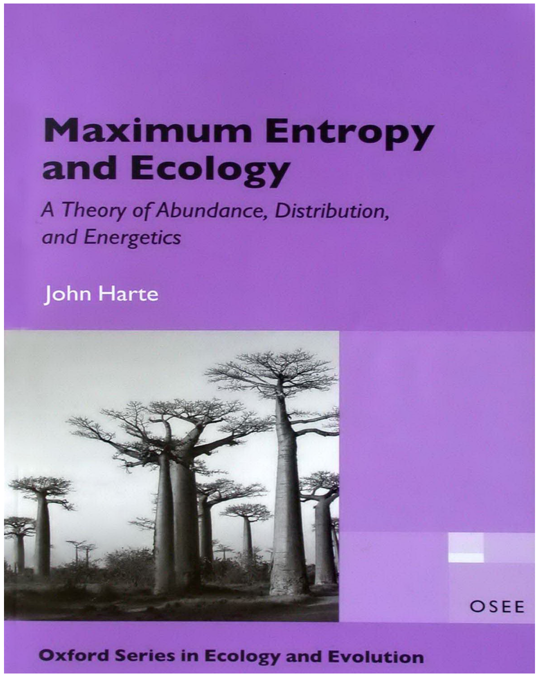
***MaxEnt*:**

An inference procedure based on information theory

PREDICTIONS

(*Metrics of Ecology*)

- Species-Area Relationships
- Endemics-Area Relationships
- Abundance & Body Size Distributions
- Spatial Aggregation Patterns
- Web Structure & Dynamics
- Species Distribution across Genera, Families, etc.



Applications

Estimating extinction rates under habitat loss; extinction debt

Estimating species richness at spatial scales too large to census

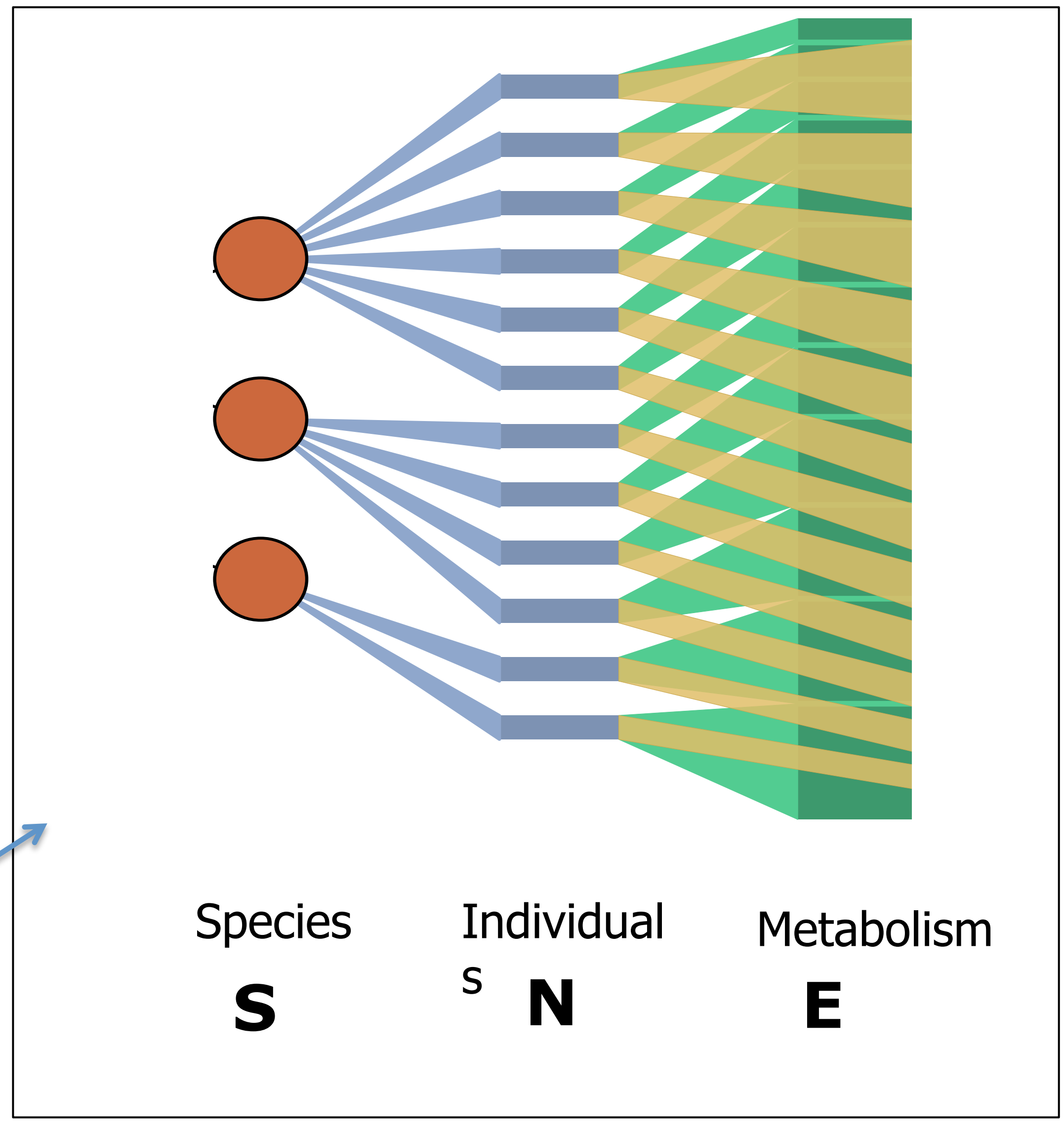
Optimizing use of sparse data

Census Data for Testing Theory

The Architecture of ASNE Version of METE

“mēt” (v):
to allocate

Area
A



Two probability distributions comprise the Model:

1. $$R(n,\, \varepsilon | A_0, S_0, N_0, E_0)$$

abundance

metabolic energy rate

State variables

R is defined over the species and individuals in an area A_0.

$R \cdot d\varepsilon$ = probability that if a species is picked from the species pool, then it has abundance n, and if an individual is picked at random from that species then its metabolic energy requirement is in the interval $(\varepsilon,\, \varepsilon + d\varepsilon)$

.

Harte et al. (2008) *Ecology* 89:2700-2711*;*

(2009) *Ecology Letters* 12: 789-797

Harte: Oxford U. Press: June 2011

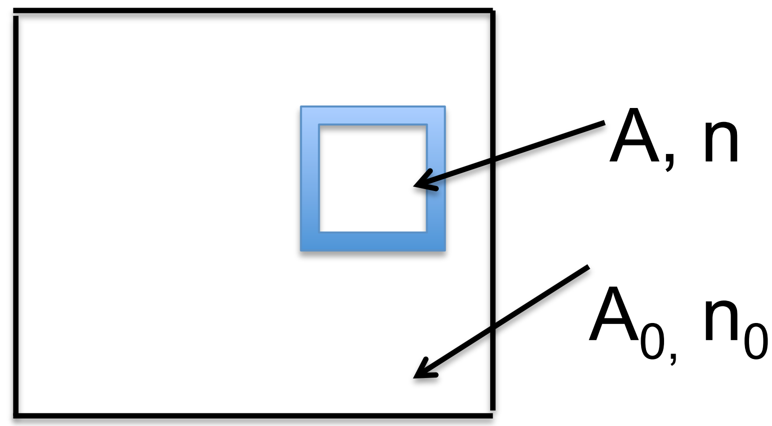
"Maximum Entropy and Ecology"

2. … and a species-level spatial distribution,

$$\Pi(n|A,n_0,A_0)$$

describing aggregation of individuals within species:

Π = probability that n individuals in A if n_0 in A_0



From R and Π, most of the metrics of macroecology can be derived.

MaxEnt gives

$$R(n,\varepsilon \mid S_0, N_0, E_0) = \frac{e^{-\lambda_1 n} e^{-\lambda_2 n \varepsilon}}{Z(\lambda_1, \lambda_2)}$$

$$\Pi(n \mid n_0, A, A_0) = \frac{e^{-\lambda_\Pi n}}{Z_\Pi}$$

Species-level Spatial distribution

(Summing over n)

$$\Psi(\varepsilon \mid S_0, N_0, E_0) \approx \lambda_2 \cdot \beta \cdot \frac{e^{-\gamma}}{(1-e^{-\gamma})^2}$$

Distribution of metabolic rates over individuals
$\gamma(\varepsilon) = \lambda_1 + \lambda_2(\varepsilon)$

(Integrating over ε)

$$\Phi(n) = \frac{1}{\lambda_2 Z n} e^{-\beta n}$$

Species Abundance distribution

fraction of occupied cells

$= n_0/(n_0 + A_0/A)$

Abundance-occupancy relation

$$\Theta(\varepsilon \mid n) = \frac{R}{\Phi} = \lambda_2 n e^{-\lambda_2 n(\varepsilon - 1)}$$

Intraspecific metabolic rate distribution

(Taking the mean)

$$< \varepsilon(n) >= 1 + \frac{1}{n\lambda_2}$$

Abundance-metabolism relation for species

$$S(A) = S_0 \sum_{n_0=1}^{N_0} \Phi(n_0) * [1 - \Pi(0 \mid n_0, A, A_0)]$$

The Species-Area relationship

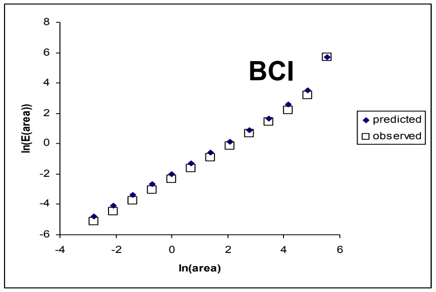
$$E(A) = S_0 \sum_{n_0=1}^{N_0} \Phi(n_0) * \Pi\ (n_0 \mid n = n_0, A, A_0)$$

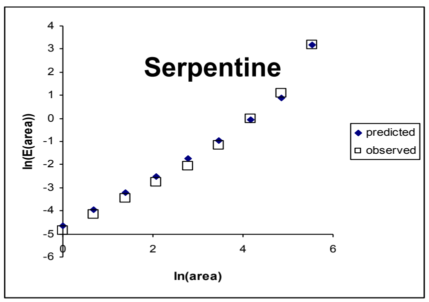
The Endemics-Area relationship

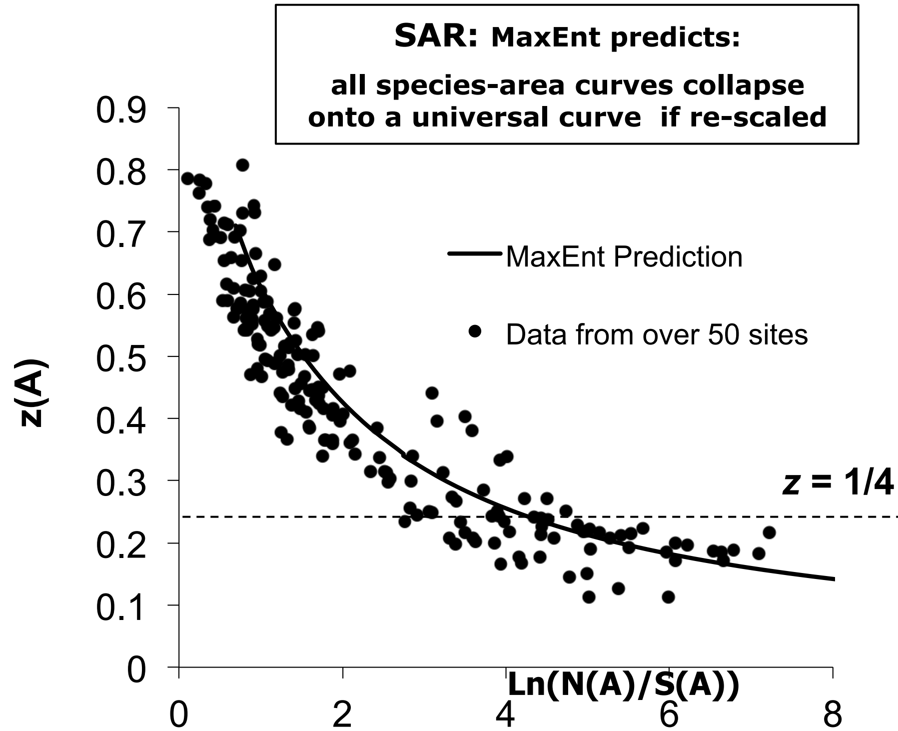
Derivations in: Harte, J., Oxford U. Press:2011
"Maximum Entropy and Ecology"

Some of the Validated Predictions

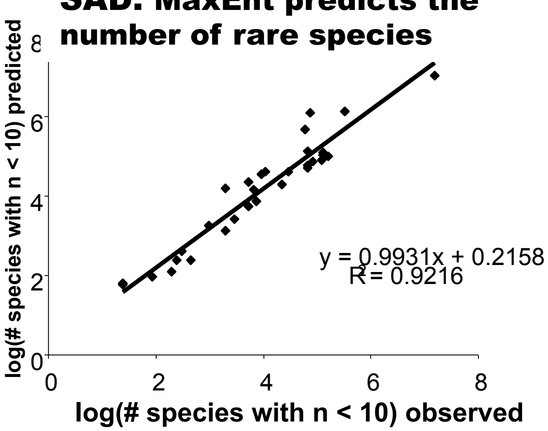
Endemics Area Relationship



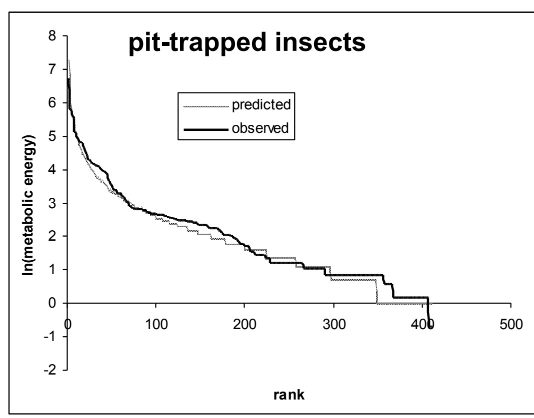


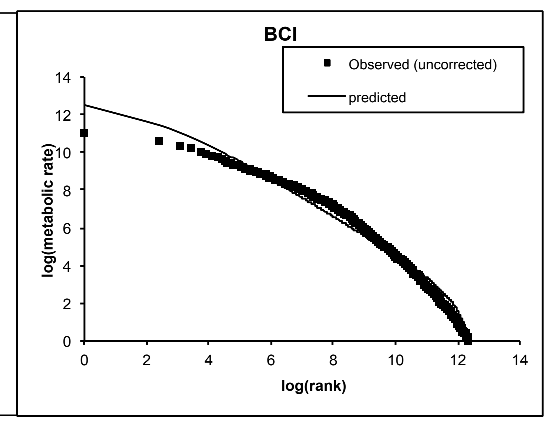


SAD: MaxEnt predicts the number of rare species



Tests of the energy distribution





Original Version of METE: Static ASNE Model

Successes:

- Species abundance distributions
- Spatial distributions
- Species Area relationships
- Metabolic rate distribution across individuals

Gaps and Failures

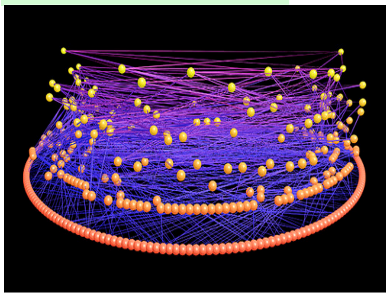
- Systems undergoing rapid change
- The energy equivalence rule
- Population dynamics
- Networks
- Multiple resources

At the Frontier of METE

core model: A, S, N, E

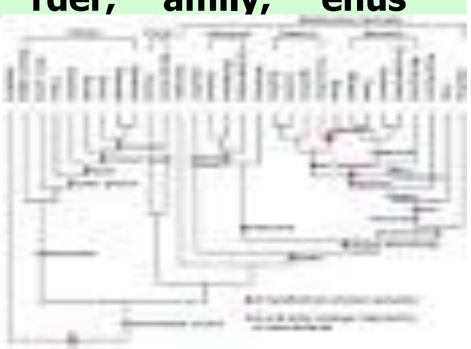
Network structure

Linkages



Evolutionary constraints: taxonomy/ phylogeny

Order, Family, Genus



Coexistence & demographics

population dynamics, stabilizing mechanisms

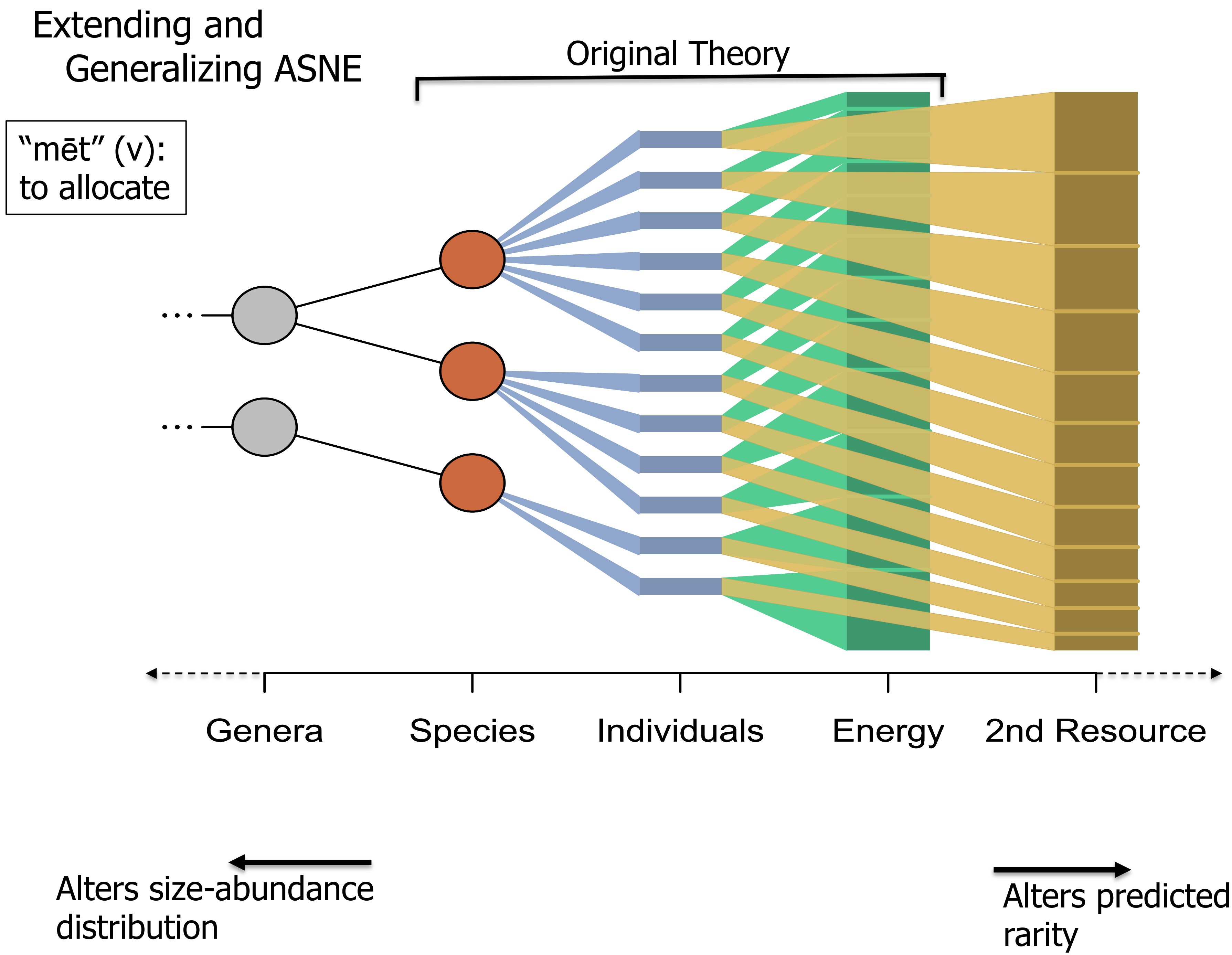
S= k log(W)

Resource constraints: niche partitioning

Water, Phosphorus,..

Dynamic theory: disturbance ecology

S(t), N(t), E(t)


Extending and Generalizing ASNE
"mēt" (v): to allocate
Original Theory
...
...
Genera
Species
Individuals
Energy
2nd Resource
Alters size-abundance distribution
Alters predicted rarity

Including additional resource constraints (in addition to energy, *E*)

The log-series SAD becomes:

$$\Phi(n) \sim \frac{e^{-\lambda n}}{n^r}$$

r **- 1** = # additional resources

The inclusion of additional resource constraints predicts increased rarity

Extension of METE to higher taxonomic levels

Example: inclusion of family as a category

(ASNE → AFSNE)

State Variables:

- A_0 = Area
- F_0 = # families
- S_0 = # species
- N_0 = # individuals
- E_0 = total metabolic rate

Motivation:
patterns in macroecology could depend on species richness of genera, families,…

The probability function *Q* replaces *R*

$Q(m,n,\varepsilon|G_0, S_0, N_0, E_0)$**, defined as follows:**

New variable:
m = # species in higher taxonomic category

Pick a family;
***Q* is the probability it has *m* species, and**
if you pick one of those species from that family,
that it has *n* individuals, and
if you pick one of those individuals from that species,
that it has metabolic rate ε.

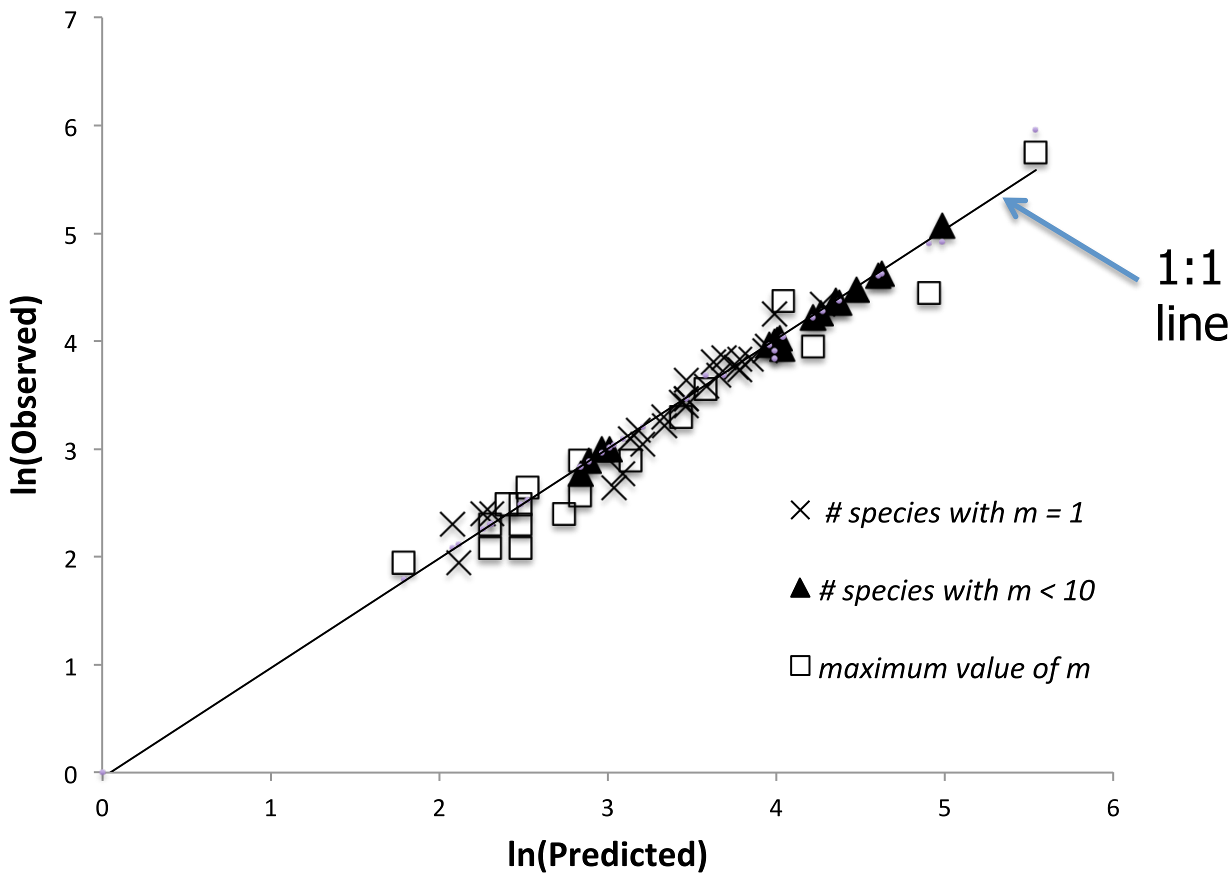
The constraints:

$$< m >= \frac{S_0}{F_0} = \sum_{m,n,\varepsilon} mQ$$

$$< n_F >= \frac{N_0}{F_0} = \sum_{m,n,\varepsilon} mnQ$$

$$< \varepsilon_F >= \frac{E_0}{F_0} = \sum_{m,n,\varepsilon} mn\varepsilon Q$$

Test of predicted distributions of species across families for arthropods, plants, birds, and microorganisms



- Arthropod data from Basset et al. (2011), and Gruner (2007)
- Bird data consist of ten transects chosen randomly from the Breeding Bird Census (Sauer et al. 2014)
- Plant data from: census plots at Cape Point Preserve (Slingsby, pers. comm.); the Smithsonian Tropical Forest Research Institute plots at BCI (Condit 1998, Condit et al. 2004; Hubbell et al. 2005), Luquillo (Thompson et al. 2002); Sherman and Cocoli (Pyke et al. 2001; Condit et al. 2004); Yasuni (Valencia et al. 2003; 2004);
- Microbiome data (Wu et al. 2013; Larry Smarr pers. comm.).

The taxonomically extended theory predicts observed patterns in macroecology that depend on species richness of higher taxonomic levels:

1. The most abundant species should belong to families or genera that contain relatively few species.
(*Consistent with Amazon tree data: ter Steege et al., 2013*)

2. Rare species should be over-represented in species-rich families or genera.
(*Consistent with vascular plant data: Schwartz & Simberloff, 2001; Lozano & Schwartz, 2005*)

3. Species with the largest body sizes, and therefore largest metabolic rates of individuals, belong to families or genera with the fewest species. Moreover, the variance of body size across species should be greatest in families or genera with the fewest species
(*Both predictions consistent with mammal data: Smith et al., 2004*)

Extending METE from Static to Dynamic

Static systems :

MaxEnt **adequately predicts the form of many of the metrics of macroecology**

Lacking is theory describing Rates of Change in these Metrics during the Processes of:

- **Speciation and Extinction**
- **Succession**
- **Adaptive Responses (e.g., to "global change")**

Using static theory during these processes is like using PV=nRT in a tornado

Possible approaches:

1. **Maximum Entropy Production**
2. **Non-extensive entropy**
3. **Dynamic, stochastic theory of state variables: use "master equation" incorporating dominant mechanisms as transition probabilities**
4. **Maximum resource allocation entropy**

Master Equation Approach

Five Transition Probabilities

Growth, $\boldsymbol{g}$: $(S, N, E) \rightarrow (S, N, E+1)$

Birth, $\boldsymbol{b}$: $(S, N, E) \rightarrow (S, N+1, E)$;

Loss of a singleton results in an extinction

Death/emigration, $\boldsymbol{d}$: $1\text{-}\mu$: $(S, N, E) \rightarrow (S, N\text{-}1, E\text{-}w)$; μ: $(S\text{-}1, N\text{-}1, E\text{-}w)$

Immigration, $\boldsymbol{m}$: $1\text{-}\sigma$: $(S, N, E) \rightarrow (S, N+1, E+w)$; σ: $(S+1, N+1, E+w)$

Speciation, λ: $(S, N, E) \rightarrow (S+1, N, E)$

The immigrant is a new species

drive the state variable distributions:

$$P(S,N,E,t) = P(S,t|N,E) * P(N,t) * P(E|N,t)$$

Slow	Intermediate	Fast
$\boldsymbol{\lambda}$, $\boldsymbol{m\sigma}$, $\boldsymbol{d\mu}$	***m, b,d***	$\boldsymbol{g}$

For example, Master Equation for $P(N)$:

$$P(N,t) = [1 - b(N)-d(N)-m]*P(N,t-1) + [m+b(N-1)]*P(N-1,t-1) + d(N+1)*P(N+1,t-1)$$

With plausible assumptions about the functional forms of

$g(E,N)$, $b(N,E)$, $d(N,E)$, $\lambda(S,N,E)$, $\mu(S,N,E)$, $\sigma(S, N, E)$

the steady-state equations are solvable analytically,
exactly with respect to to the fast rates
and in the van Kampen approximation
with respect to the intermediate and slow rates.
(The dynamics needs to be solved numerically)

In one solved example, the form of $P(S,N,E)$ in steady state is:

$$P(S,N,E) \sim \frac{(\alpha \log(N))^S) S^{-\beta} e^{-\delta N}}{E N^{\rho}}$$

where the constants α, β, δ, ρ characterize the transition rates

Superstatistics then connects the metrics of macroecology to the dynamics of the state variables

The time-dependent stochastic Structure Function, R^*, is:

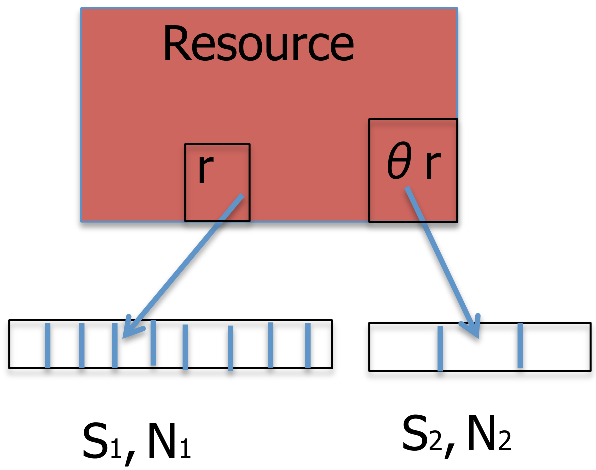
$$\boldsymbol{R^*(n,e,t) = \iiint dS\, dN\, dE\ \ R(n,e|S,N,E) * P(S,N,E,t)}$$

From ASNE

From the master eq.

Need suitable initial conditions for $P(S,N,E,t)$:

e.g., bare ground: $P = \delta_{E,0}\ \delta_{N,0}\ \delta_{S,0}$



Species Co-existence Revisited

a. Conventional approach: $b_i - d_i = f_i(N_i)$

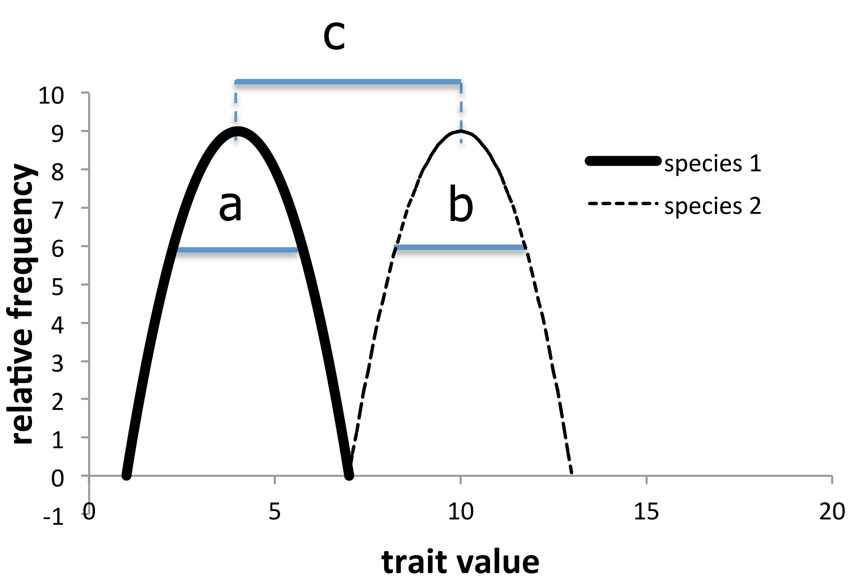
(in Lotka-Volterra eqs., coexistence if $\alpha_{ii} > \alpha_{ij}$)

b. Zhang and Harte: use $\boldsymbol{S = k \log(W)}$

W = # of possible allocations of resource units consistent with a macrostate (S species, each with n_i individuals)

$= W_{between} \times (W_{within})^{D_r}$

(W's from combinatorics)



$$D_r = \frac{\text{distinguishability of individuals within species}}{\text{distinguishability of individuals between species}}$$

A possible empirical evaluation:

$$D_r \sim (a+b)/2c$$

(genetics-based definition also possible)

RESULTS

(Yu Zhang and JH, submitted)

i. $b_i - d_i = f_i(N_i, \theta, D_r)$; f_i **predicted by maximizing log(*W*)**

Density dependence is an emergent property of
S = k log(W), not an imposed assumption.

ii. A pair of species can coexist iff G(θ, D_r) > 0; *G* predicted by maximizing *W*

Under predicted conditions, two species can coexist on one resource because there are more ways to allocate the resource if the species co-exist than if one drives the other to extinction.

Coexistence possible only if:
the intra-specific distinguishability of individuals is sufficiently less than
the inter-specific distinguishability of individuals

Shannon, Jaynes

Two complementary facets of the same thing!

Boltzmann

Top-down MaxEnt

Bottom-up MaxEnt

State Variables → Static Macroecology Metrics

Dynamics ← **#Resources**

What are the smoothest possible distributions compatible with the state variables?

Spatial distributions

SARs

Metabolic rate distributions

Generalizations to Higher Taxa

Demographic rates

Co-existence

Predator-prey dynamics, (prey resource allocated to predators)

What demographic rates and surviving species and abundances are associated, over time, with the maximum number of resource allocations?

Will the two approaches agree??

Species Abundance Distribution
Size-abundance relationship
Network Structure

We are approaching, from two entropic directions, the goal of a unified theory of ecology

& Thank you for listening!

Questions?