Intro to O/E modeling

John Van Sickle

U.S. Environmental Protection Agency
National Health and Environmental Effects Research Laboratory
Western Ecology Division
Corvallis, OR

O/E models
('Predictive' models, 'RIVPACS' models)

– For species assemblages (macroinvertebrates, periphyton, maybe fish).

-- Today's example: Macroinvertebrates sampled in EMAP-West stream survey.
(98 least-disturbed reference sites, 676 other sites).

Approach:

1) **Build a statistical model** to predict the assemblage that would be expected at any sampled site if that site were in reference condition.
-- Model built from assemblage data at reference sites.

2) **Apply the model** to any site.
-- Difference between expected and observed assemblages indicates site impairment.

A few details –

-- Models predict species presence/absence, but not abundances.

-- No need to compare observed vs expected assemblages species-by-species.

-- Instead, just compare observed (O) and expected (E) number of species.

-- If the ratio O/E differs significantly from 1.0, then site may be impaired.

-- We include only the common species (found at > 50% of reference sites) in calculation of O and E.

Predicting Expected Richness (E)

Model 1 --- A simple "null" model.

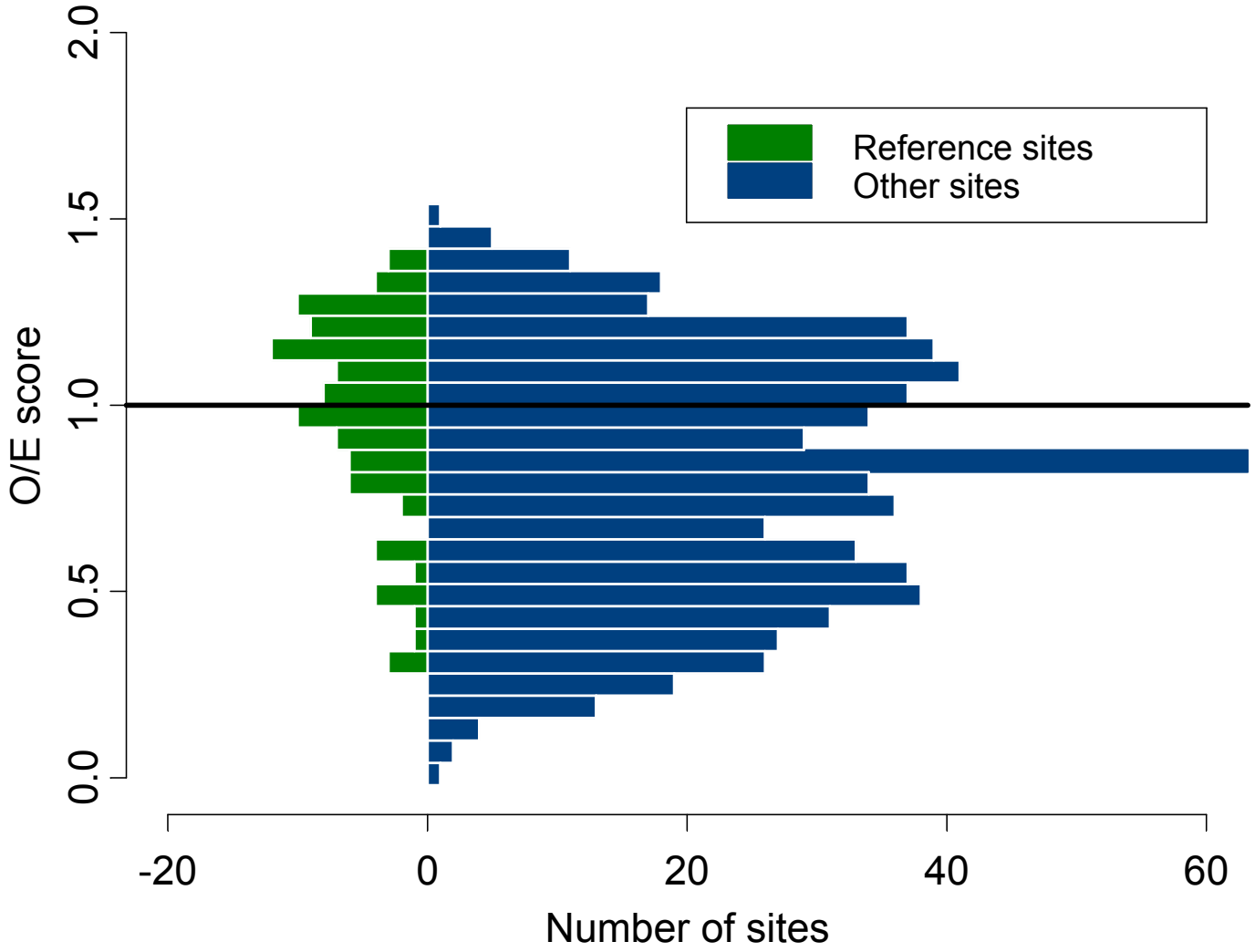
-- List the 27 common species found at the 98 reference sites.

-- Null-model E = Average number of these 27 that were observed at each reference site.
E = 17.1 species.

-- Model application:

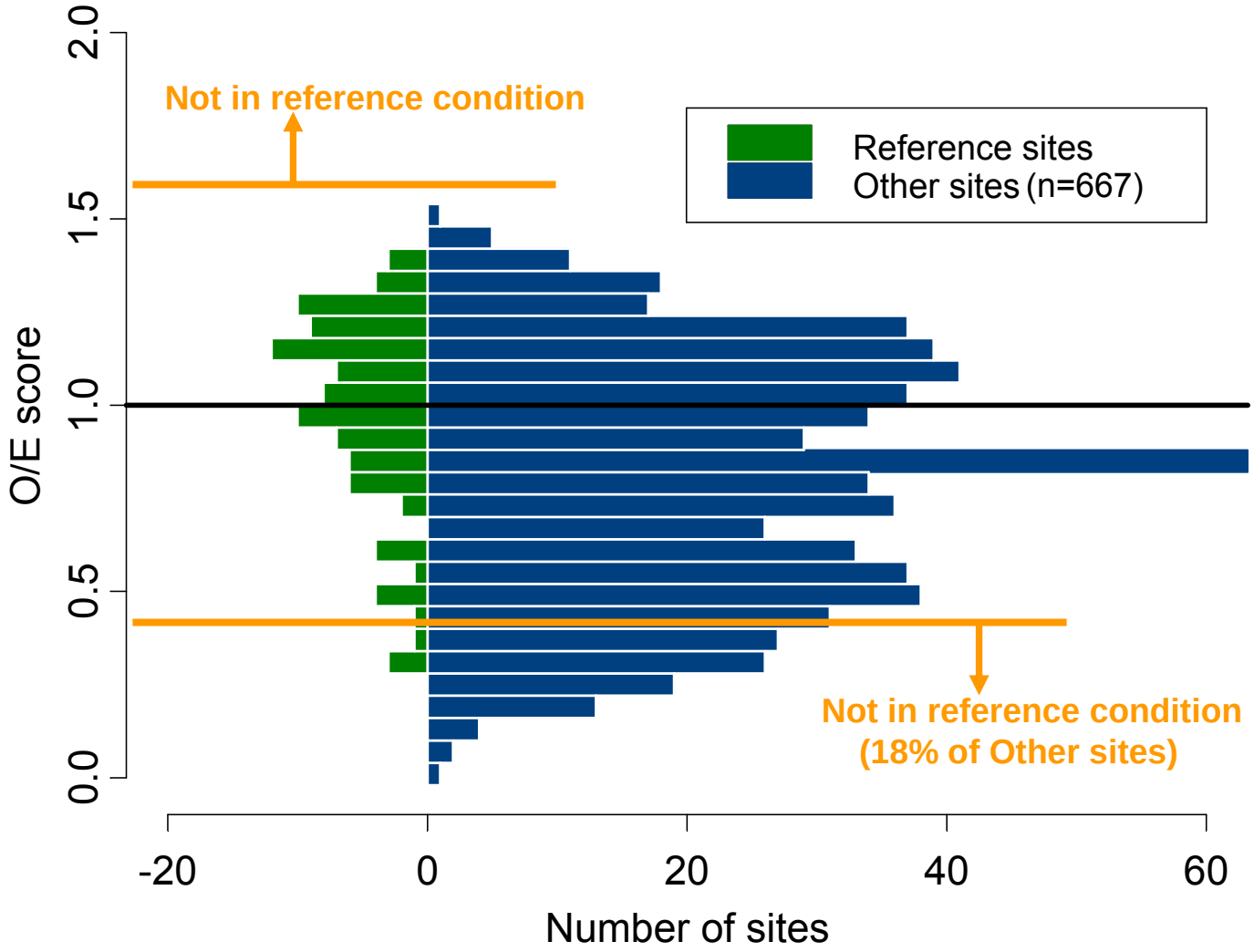
At a particular site, suppose you capture **O = 14** of the 27 species.

Then **O/E** = 14/17.1 **= 0.82**, for that site.


Distributions of null-model O/E scores at EMAP-West sites
Reference sites
Other sites
O/E score
0.0
0.5
1.0
1.5
2.0
-20
0
20
40
60
Number of sites

Assessment based on null model:

Sites outside [1.0 ± 2SD] of the reference distribution are declared to be “not in reference condition”.



Problem:

-- Null model does not explain variation in reference-site assemblages due to “natural” factors.

→ Null model has high variability in O/E at reference sites
→ Little power for detecting impairment.

Possible solution:

-- Model 2: RIVPACS-type predictive model

Adjusts the expected assemblage for natural-factor effects.

Structure of RIVPACS-type predictive model

- Cluster Reference sites into *K* groups, based on their assemblages

Stage 1 – Development (Reference sites)

- Build discriminant function model to predict group membership probabilities g_{jk} for sites *j*, using 'natural gradient' environmental predictor variables.

Stage 2 – Development (Reference sites)

- Predict P_{ij} = Occurrence probability of species *i* at site *j*. $P_{ij} = \Sigma_k g_{kj} f_{ki}$, where f_{ki} = observed relative occurrence frequency of species *i* in site group *k*.

-Expected number of species at site *j* = $E_j = \Sigma_i P_{ij}$

-Compare E_j to O_j = Observed number of species at site *j*.

-If O_j/E_j differs significantly from1.0, then site *j* is "impaired" (i.e., not in Reference condition).

Stage 3 – Prediction and Assessment (All sites)

Example of predictive model for EMAP- West

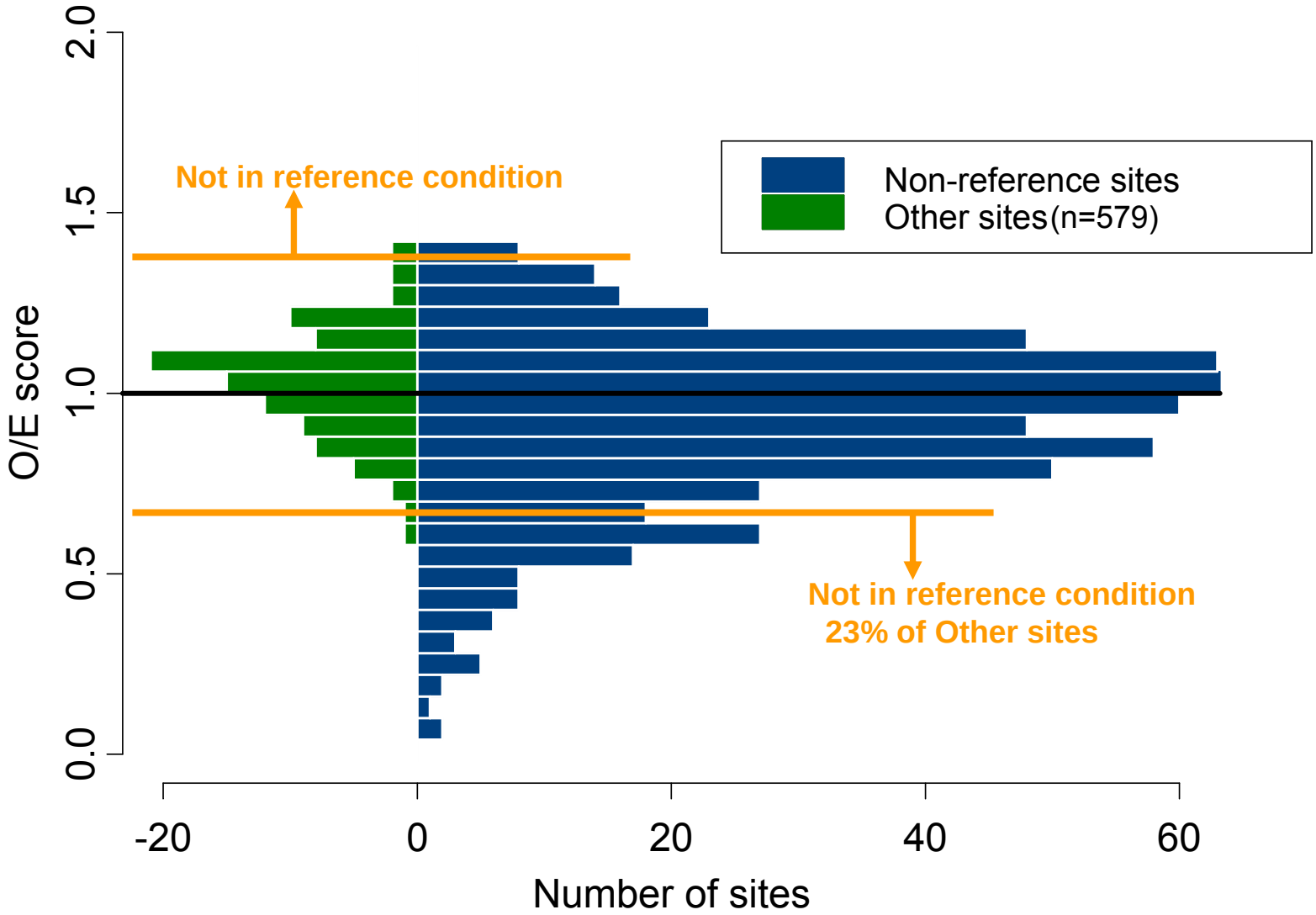
-- 3 clusters of sites.

-- 5 natural-factor predictor variables:

- -- Channel width
- -- Elevation
- -- Mean watershed slope
- -- Latitude
- -- Longitude

-- 88 of 676 'Other' sites require model extrapolation
→ Predicted Expected richness may not be valid.

EMAP-West predictive model: O/E scores



Requirements for building/applying O/E models

For null model:

-- Assemblage data (presence/absence).

-- Large (N ≥ 25), representative set of reference sites.
('Reference' status determined independently of assemblages.)

For RIVPACS-type model:

-- Assemblage data (presence/absence or abundance).

-- Larger set of reference sites (N ≥ 70).

-- Data at all sites, for natural (nonanthropogenic) factors that:
a) Vary noticeably across the sampled region/rivers.
b) May affect assemblage composition.