Linking freshwater data

FISH.Link

JISC MRD Workshop, March 2011

Questions

1. What's the additional value of publishing linked data (as opposed to triplifying and/or providing open data)?
2. How do we support the production/annotation of datasets from small providers?
3. How can we support non-SPARQL experts in writing SPARQL queries/getting information out of triple stores?

FISH.Link

- Investigating Linked Data approaches to support data sharing in freshwater biology community

> *Will high altitude water bodies provide a refuge for species under pressure from invasive species taking advantage of changes in climate?*

- Integration with the FISH.Net repository
  - FISH.Net "Traffic Light" System characterising metadata states
  - FISH.Link at Green+: readiness for LD publication

http://www.flickr.com/photos/amulligan/103460333

FISH.Link

- This is not primarily about *scale*
- Data sets aren't huge.
- Small, hand-crafted idiosyncratic data sets.
- FISHNet carrots/sticks
  - Motivations: why *publish* LD (as opposed to consuming it)?

http://www.flickr.com/photos/brapps/3368998420

Initial Experiments

- Paper: Aquatic Plant Diversity
  Jones, Li, Maberly
- Extracted initial set of queries supporting research question
- Hand mapping existing datasets to support initial queries
- D2R, Jena, some scripting
- SPARQL queries generating "wide tables" of data for analysis
- Science/analysis still left in the hands of the experts

ECOGRAPHY 26: 411–420, 2003

Area, altitude and aquatic plant diversity

J. Iwan Jones, Wei Li and Stephen C. Maberly

Jones, J. I., Li, W. and Maberly, S. C. 2003. Area, altitude and aquatic plant diversity. – Ecography 26: 411–420.

Several explanations have been given for the decline in species richness with altitude. However, separating the influences of altitude, area, and isolation is difficult because of the conical shape of mountains. We used species lists of aquatic plants from > 300 lakes in a small geographical area to investigate the influence of altitude on species richness. Altitude and/or surface area were better predictors of species richness than any measure of water chemistry. The surface area and depth of individual lakes were not related to altitude, neither was the degree of isolation from other waterbodies. Although species range size increased with altitude, range sizes of all but the rarer species (in the data set) encompassed the lowest altitudes, indicating species loss rather than turnover and no influence of the Rapoport rescue effect. Nevertheless we found a decline in species richness with altitude, additive to the effect of area. Species were ascribed to attribute groups according to an a priori classification based on morphological and life-history traits. The number of attribute groups and number of species within each group increased with area, suggesting both increased diversity and coexistence within niches. With altitude, the number of attribute groups declined, but the number of species per group increased, consistent with a loss of richness and reduced competition. The species remaining at high altitudes were characterised by stress tolerant traits, associated with sites of low productivity.

Our results suggest an absolute effect of altitude on species richness, irrespective of other influences and consistent with a decline in productivity.

J. I. Jones (j.i.jones@qmul.ac.uk), School of Biological Sciences, Queen Mary, Univ. London, London, U.K. E1 4NS. – W. Li, Lab. of Aquatic Plant Biology, Wuhan Inst. of Botany, The Chinese Academy of Sciences, Wuhan, 430074, People's Republic of China. – S. C. Maberly, CEH Windermere, The Ferry House, Ambleside, Cumbria, U.K. LA22 0LP.

The relationship between area and species richness is probably one of the few general laws of ecology (Lawton 1999), with a weight of evidence and theoretical background supporting it (MacArthur and Wilson 1967, Rosenzweig 1995). In contrast, the decline in species richness with latitude, although widely applicable (Lawton 1999) and long recognised (Wallace 1878), still does not have a universally accepted mechanism at its root (Stevens 1989, Begon et al. 1990, Rohde 1992, Rosenzweig 1995, Willig and Lyons 1998, Chown and Gaston 2000, Lambers et al. 2002).

Although studies of the influence of latitude on species richness abound (Rosenzweig 1995, Chown and Gaston 2000), fewer have contemplated the influence of altitude (Rahbek 1995, Körner 2000). Nevertheless, a natural comparison exists between latitude and altitude, both representing a gradient of worsening climate. As with latitude, there are, however, several confounding factors.

The conical shape of mountains means that the influence of altitude is compounded by one of area. The area contained within altitudinal zones is progressively reduced towards the summit and the species characteristic of higher altitudes influenced by a restricted area of habitat available to them. Studies of variation in species richness with altitude which have acknowledged this problem have sought to overcome it by including the area of arbitrary altitudinal bands (i.e. not necessar-

Accepted 20 January 2003

ISSN 0906-7590

http://www.flickr.com/photos/fabelfroh/

Initial Experiments

1. Range of altitude
2. Area of water body
3. Range of alkalinity
4. Covariation of ion concentration with altitude
5. Correlation of ion strength with alkalinity
6. Correlation of area and altitude
7. Correlation of depth and area
8. Correlation of area with presence of inflows
9. Correlation of area with number of inflows
10. Correlation of area with presence of out-flows
11. Correlation of area with concentration of nitrate
12. Correlation of altitude with presence of inflows
13. Correlation of altitude with number of inflows
13. Correlation of altitude with presence of out-flows
14. Correlation of altitude with distance to nearest body of standing water
15. Any measured variable as predictor of species richness
16. Altitude at which species occur, its range correlated with altitude
17. Frequency of occurrence of species and altitudinal range size
18. Mid-point of altitudinal range size
19. Number of observations
20. Attribute group per water body correlated with area
21. Attribute group correlated with altitude
22. Mean number of species per attribute group correlated with area and then altitude
23. Correlation of each variable with number of attribute groups per water body
24. Correlation of each variable with number of species per water body

ECOGRAPHY 26: 411–420, 2003

Area, altitude and aquatic plant diversity

J. Iwan Jones, Wei Li and Stephen C. Maberly

Jones, J. I., Li, W. and Maberly, S. C. 2003. Area, altitude and aquatic plant diversity. – Ecography 26: 411–420.

Several explanations have been given for the decline in species richness with altitude. However, separating the influences of altitude, area, and isolation is difficult because of the conical shape of mountains. We used species lists of aquatic plants from > 300 lakes in a small geographical area to investigate the influence of altitude on species richness. Altitude and/or surface area were better predictors of species richness than any measure of water chemistry. The surface area and depth of individual lakes were not related to altitude, neither was the degree of isolation from other waterbodies. Although species range size increased with altitude, range sizes of all but the rarer species (in the data set) encompassed the lowest altitudes, indicating species loss rather than turnover and no influence of the Rapoport rescue effect. Nevertheless we found a decline in species richness with altitude, additive to the effect of area. Species were ascribed to attribute groups according to an a priori classification based on morphological and life-history traits. The number of attribute groups and number of species within each group increased with area, suggesting both increased diversity and coexistence within niches. With altitude, the number of attribute groups declined, but the number of species per group increased, consistent with a loss of richness and reduced competition. The species remaining at high altitudes were characterised by stress tolerant traits, associated with sites of low productivity.

Our results suggest an absolute effect of altitude on species richness, irrespective of other influences and consistent with a decline in productivity.

J. I. Jones (j.i.jones@qmul.ac.uk), School of Biological Sciences, Queen Mary, Univ. London, London, U.K. E1 4NS. – W. Li, Lab. of Aquatic Plant Biology, Wuhan Inst. of Botany, The Chinese Academy of Sciences, Wuhan, 430074, People's Republic of China. – S. C. Maberly, CEH Windermere, The Ferry House, Ambleside, Cumbria, U.K. LA22 0LP.

The relationship between area and species richness is probably one of the few general laws of ecology (Lawton 1999), with a weight of evidence and theoretical background supporting it (MacArthur and Wilson 1967, Rosenzweig 1995). In contrast, the decline in species richness with latitude, although widely applicable (Lawton 1999) and long recognised (Wallace 1878), still does not have a universally accepted mechanism at its root (Stevens 1989, Begon et al. 1990, Rohde 1992, Rosenzweig 1995, Willig and Lyons 1998, Chown and Gaston 2000, Lambers et al. 2002).

Although studies of the influence of latitude on species richness abound (Rosenzweig 1995, Chown and Gaston 2000), fewer have contemplated the influence of altitude (Rahbek 1995, Körner 2000). Nevertheless, a natural comparison exists between latitude and altitude, both representing a gradient of worsening climate. As with latitude, there are, however, several confounding factors.

The conical shape of mountains means that the influence of altitude is compounded by one of area. The area contained within altitudinal zones is progressively reduced towards the summit and the species characteristic of higher altitudes influenced by a restricted area of habitat available to them. Studies of variation in species richness with altitude which have acknowledged this problem have sought to overcome it by including the area of arbitrary altitudinal bands (i.e. not necessar-

Accepted 20 January 2003

ISSN 0906-7590

r analysis

http://www.flickr.com/photos/fabelfroh/

Initial Experiments

- Paper: Aquatic Plant Diversity
  Jones, Li, Maberly
- Extracted initial set of queries supporting research question
- Hand mapping existing datasets to support initial queries
- D2R, Jena, some scripting
- SPARQL queries generating "wide tables" of data for analysis
- Science/analysis still left in the hands of the experts

ECOGRAPHY 26: 411–420, 2003

Area, altitude and aquatic plant diversity

J. Iwan Jones, Wei Li and Stephen C. Maberly

Jones, J. I., Li, W. and Maberly, S. C. 2003. Area, altitude and aquatic plant diversity. – Ecography 26: 411–420.

Several explanations have been given for the decline in species richness with altitude. However, separating the influences of altitude, area, and isolation is difficult because of the conical shape of mountains. We used species lists of aquatic plants from > 300 lakes in a small geographical area to investigate the influence of altitude on species richness. Altitude and/or surface area were better predictors of species richness than any measure of water chemistry. The surface area and depth of individual lakes were not related to altitude, neither was the degree of isolation from other waterbodies. Although species range size increased with altitude, range sizes of all but the rarer species (in the data set) encompassed the lowest altitudes, indicating species loss rather than turnover and no influence of the Rapoport rescue effect. Nevertheless we found a decline in species richness with altitude, additive to the effect of area. Species were ascribed to attribute groups according to an a priori classification based on morphological and life-history traits. The number of attribute groups and number of species within each group increased with area, suggesting both increased diversity and coexistence within niches. With altitude, the number of attribute groups declined, but the number of species per group increased, consistent with a loss of richness and reduced competition. The species remaining at high altitudes were characterised by stress tolerant traits, associated with sites of low productivity.

Our results suggest an absolute effect of altitude on species richness, irrespective of other influences and consistent with a decline in productivity.

J. I. Jones (j.i.jones@qmul.ac.uk), School of Biological Sciences, Queen Mary, Univ. London, London, U.K. E1 4NS. – W. Li, Lab. of Aquatic Plant Biology, Wuhan Inst. of Botany, The Chinese Academy of Sciences, Wuhan, 430074, People's Republic of China. – S. C. Maberly, CEH Windermere, The Ferry House, Ambleside, Cumbria, U.K. LA22 0LP.

The relationship between area and species richness is probably one of the few general laws of ecology (Lawton 1999), with a weight of evidence and theoretical background supporting it (MacArthur and Wilson 1967, Rosenzweig 1995). In contrast, the decline in species richness with latitude, although widely applicable (Lawton 1999) and long recognised (Wallace 1878), still does not have a universally accepted mechanism at its root (Stevens 1989, Begon et al. 1990, Rohde 1992, Rosenzweig 1995, Willig and Lyons 1998, Chown and Gaston 2000, Lambers et al. 2002).

Although studies of the influence of latitude on species richness abound (Rosenzweig 1995, Chown and Gaston 2000), fewer have contemplated the influence of altitude (Rahbek 1995, Körner 2000). Nevertheless, a natural comparison exists between latitude and altitude, both representing a gradient of worsening climate. As with latitude, there are, however, several confounding factors.

The conical shape of mountains means that the influence of altitude is compounded by one of area. The area contained within altitudinal zones is progressively reduced towards the summit and the species characteristic of higher altitudes influenced by a restricted area of habitat available to them. Studies of variation in species richness with altitude which have acknowledged this problem have sought to overcome it by including the area of arbitrary altitudinal bands (i.e. not necessar-

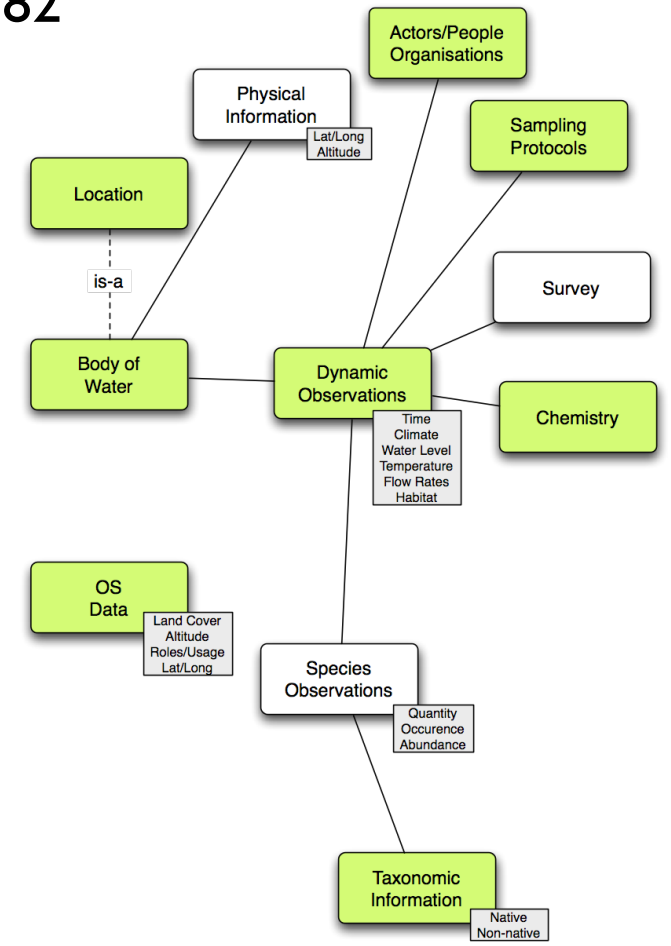
Accepted 20 January 2003

ISSN 0906-7590

http://www.flickr.com/photos/fabelfroh/

Initial Data Sets

- Major ions, Lakes & Tarns, 1982
- Major ions, Duddon & Windermere, 1982
- Cumbria Tarns/Stokoe Tarns
- Ferry House/Windermere Level & Temp Data
- Willby Species Groupings
- Meteorological Data
- RIVPACS
- BIOSYS



Difficult Aspects

- Column Names
- Multiple names for the same things
  - Alkalinity/ALK_1/ Average_Alk
- Similar names for different things
  - ALK_1/ALK_2
- Multiple values/dirty data
  - Altitude
  - Area
- Miscellaneous
  - "Nearest Water"
  - Tarns vs Tarn Groups

Tools to assist transition from FISHNet Orange to Green

http://www.flickr.com/photos/guitarfish/

Observations

- Datasets are largely about *observations*
  - Provenance: who, what, where, when, how
  - Associated Actors/Organisations
  - Associated Locations
  - Methodologies/Protocols
- Derived observations
  - 5 measurements, average recorded
- Models for investigations
  - ISA: Investigation/Study/Assay for managing experimental metadata
  - Top-down vs bottom-up

Where are we?

- Datasets in triple store
- SPARQL queries pulling out data tables allowing recreation of analysis
  - Plus analysis using different source data (e.g. 2010 rather than 2003)
  - How to produce/write these queries?

- Benefits of *integration* via RDF
- Benefits of data *publication*
- But,... *Linked* Data?

Questions

1. What's the additional value of publishing linked data (as opposed to triplifying and/or providing open data)?
2. How do we support the production/annotation of datasets from small providers?
3. How can we support non-SPARQL experts in writing SPARQL queries/getting information out of triple stores?

Commercial Break...