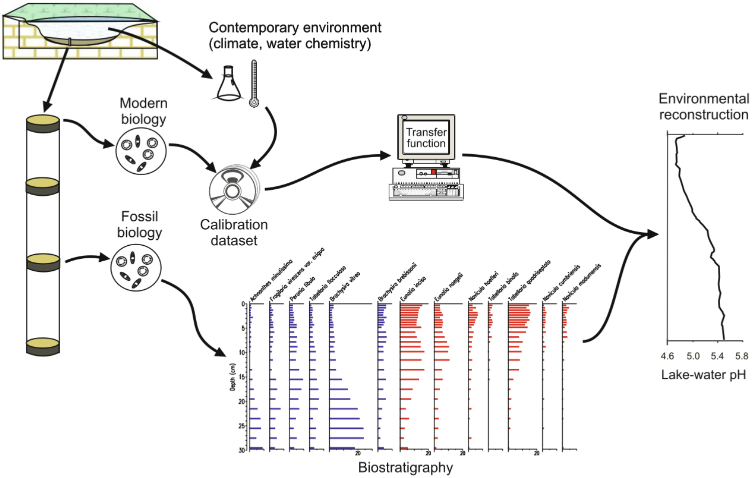
This presentation contains material by Steve Juggins, Guillaume Blanchet and Richard Telford

Jack Williams did not contribute to this presentation

Transfer functions

Transfer functions



Juggins and Birks (2012)

Transfer functions

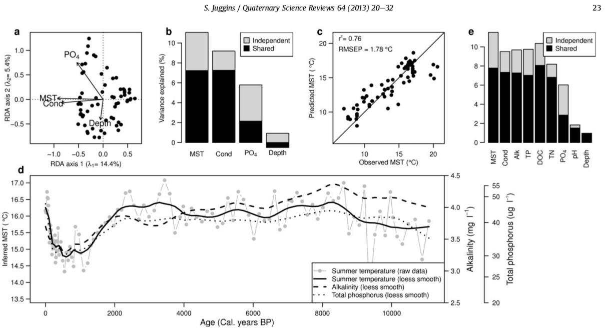
Two types of transfer functions:

Transfer functions

Two types of transfer functions:

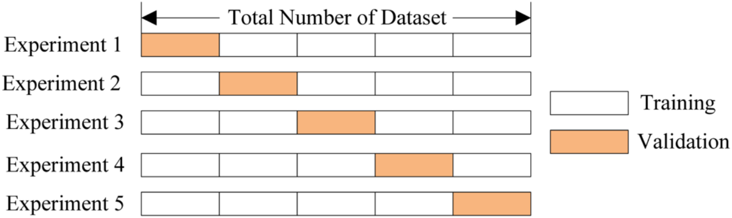
- Modern analogue technique (k-nearest neighbors)
- Weighted averaging

Transfer function: Performance




```
$object
            RMSE        R2     Avg.Bias Max.Bias    Skill
WA.inv 2.633236 0.7849437 8.237631e-16 6.904129 78.49437
WA.cla 2.972149 0.7849437 5.299770e-16 4.521565 72.60232

$crossval
            RMSE        R2     Avg.Bias Max.Bias    Skill
WA.inv 2.641335 0.7836189 0.0001948045 6.960012 78.36188
WA.cla 2.977577 0.7836767 0.0003111544 4.591617 72.50216

>
```

Transfer function: cross-validation



Special case: structure in data set

- strong similarity between samples/sites
- samples from the same site
  - build cross-validation groups following similarity of samples/sites
    - all similar (nested) samples/sites in one group

Transfer function: Weighted Averaging Partial Least Squares

Partial Least Squares Regression:

- independent variable (env. variable)
- dependent variables (species)
- seek linear combination of dependent data (species) that maximizes covariance with independent variable
- take residuals and seek linear combination of residuals that maximizes covariance with independent variable (further components usually orthogonal (independent) to first component)

WAPLS

Combine **Weighted Averaging** and **PLS**

Step 1: Estimate species optima

Step 2: Explore residuals for structure that improves fit with environmental variable

Step 3: Update optima now called **coefficients**

Overfitting

Calibration models: establishing relation between environment and species

relation env. species:

- aspects that are possible to generalize (signal)
- aspects that are specific to the calibration data set (noise)

Production of an analysis that corresponds too closely to a particular set of data

Over fitting in calibration results in poor performance in validation

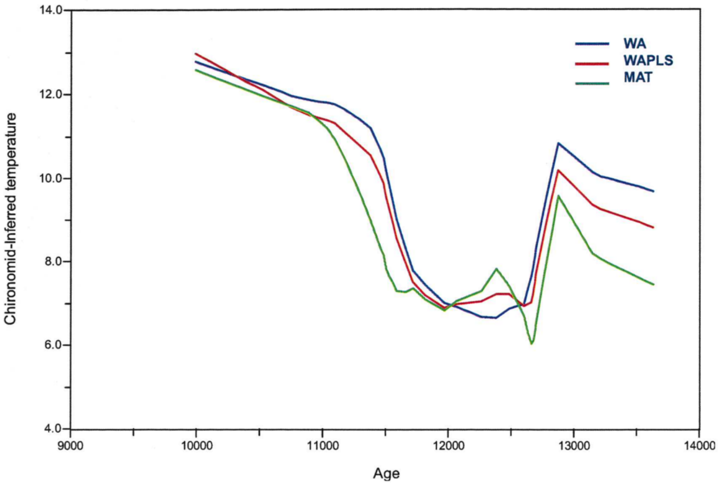
WAPLS: Diatoms and pH

```
$RMSE0
[1] 0.7694898

$object
            RMSE        R2     Avg.Bias   Max.Bias    Skill
Comp01 0.2991488 0.8490812  0.011292788 0.28498761 84.88637
Comp02 0.2298540 0.9108167  0.002854622 0.16684198 91.07726
Comp03 0.1864512 0.9413999  0.003542005 0.13921542 94.12883
Comp04 0.1482520 0.9629110  0.004191945 0.09276883 96.28811
Comp05 0.1255443 0.9733891 -0.001198717 0.09085347 97.33812

$crossval
            RMSE        R2   Avg.Bias  Max.Bias    Skill
Comp01 0.3235140 0.8247528 0.02009091 0.3177840 82.32415
Comp02 0.2898939 0.8586900 0.01672522 0.2308886 85.80705
Comp03 0.2954047 0.8536574 0.02329957 0.2711282 85.26232
Comp04 0.3233216 0.8277902 0.02388736 0.2783020 82.34516
Comp05 0.3483198 0.8028176 0.02422611 0.3079546 79.50959
```

Rule of thumb: to add a new component RMSEP needs to decrease by 10%

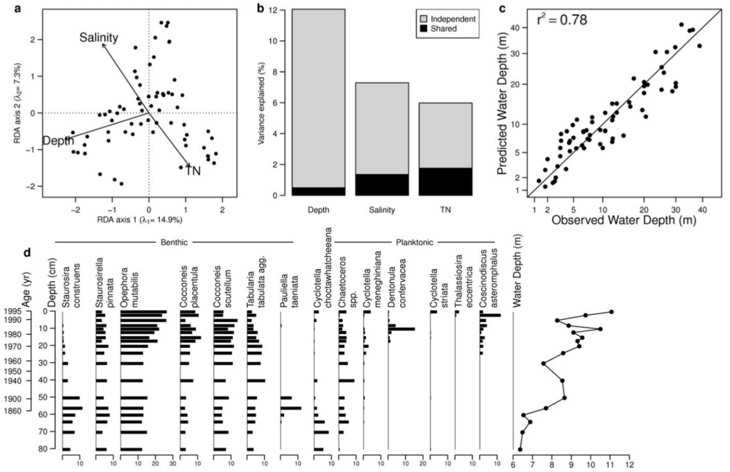
- RMSE: apparent
- RMSEP: cross-validated

Comparison of methods



Juggins (2012)

Problematic transfer functions



Juggins(2013)

Assumptions of transfer functions

1. The taxa in the modern training-set are systematically related to the environment in which they live.
2. The environmental variable(s) to be reconstructed is, or is linearly related to, an ecologically important determinant in the system of interest.
3. The taxa in the training-set are the same biological entities as in the fossil data and their ecological responses to the environmental variable(s) of interest have not changed over the time represented by the fossil assemblage.
4. The mathematical methods adequately model the biological responses to the environmental variable(s) of interest and yield numerical models that allow accurate and unbiased reconstructions.
5. Environmental variables other than the one of interest have negligible influence, or their joint distribution with the environmental variable does not change with time.

Assumptions of transfer functions

1. The taxa in the modern training-set are systematically related to the environment in which they live.
2. The environmental variable(s) to be reconstructed is, or is linearly related to, an ecologically important determinant in the system of interest.

Ecological knowledge

Test using constrained ordination (CCA)

Transfer function performance possibly improved by spatial autocorrelation

Assumptions of transfer functions

3. The taxa in the training-set are the same biological entities as in the fossil data and their ecological responses to the environmental variable(s) of interest have not changed over the time represented by the fossil assemblage.

Analogue quality

4. The mathematical methods adequately model the biological responses to the environmental variable(s) of interest and yield numerical models that allow accurate and unbiased reconstructions.

R2, RMSE, Significance tests, Spatial autocorrelation

Assumptions of transfer functions

5. Environmental variables other than the one of interest have negligible influence, or their joint distribution with the environmental variable does not change with time.

Part 1: almost always violated

Part 2: often violated (careful site selection)

Part 2: Space for time substitution might be problematic

Space for time substitution

- environmental variables important in space are also important in time

Growing season temperature: important in space and time

pH:

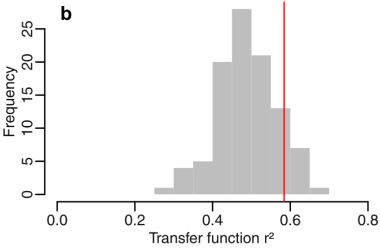
- importance in space caused by different bedrock types
- in time:
  - human influence
  - soil formation
  - otherwise pH reconstruction probably spurious

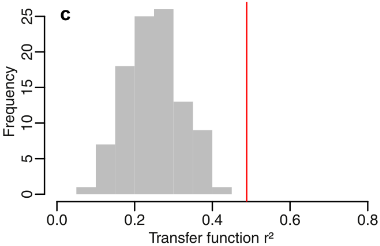
Assessing transfer functions and reconstructions:

- Mainly transfer function
- TF and reconstruction
- mainly reconstruction (next week)

Transfer function: Significance

How do transfer function methods perform when trained on random data?



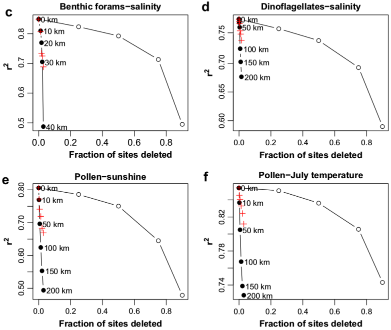


Transfer function: Significance

Simulate random environmental variable and train modern species:

- estimate cross-validated r^2
- compare to effective cross-validated r^2

Transfer function: Spatial autocorrelation

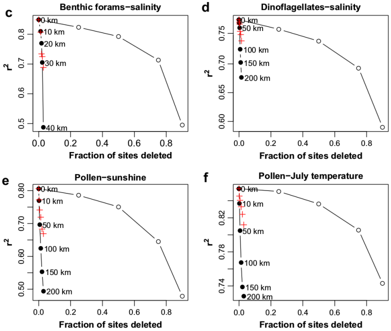


Telford and Birks (2009):

- Performance when removing sites:
  - random
  - spatially close
  - environmentally close

All decrease performance

Transfer function: Spatial autocorrelation



Spatially close > environmentally close: data set affected by spatial autocorrelation

rne in *palaeoSig*

Spatial autocorrelation:

Simulate random environmental variable with same spatial structure as observed environmental variable

Transfer function: WA

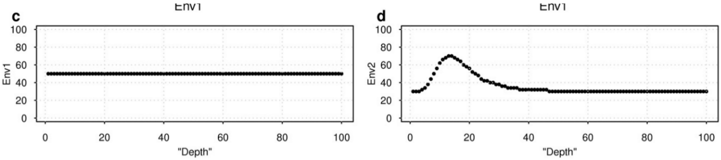
Juggins(2013):

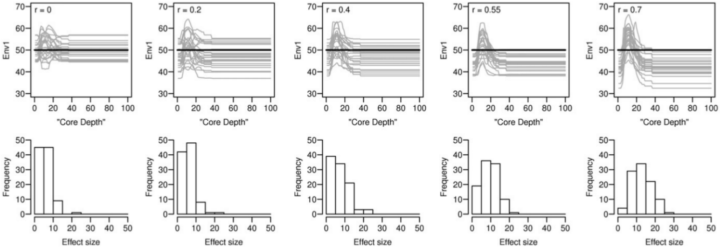
Correlation in modern data transmitted into fossil data

Problem if spatial and temporal correlation differ (space for time substitution)

Own experience MAT: correlation of used analogues can get transmitted into reconstruction

Transfer function: WA





Juggins(2013)

Transfer function and fossil data

Blog by Richard Telford:

https://quantpalaeo.wordpress.com/2014/05/03/transfer-function-and-palaeoenvironmenal-reconstruction-diagnostics/

Analogue quality:

Passively add fossil samples to ordination of training set samples

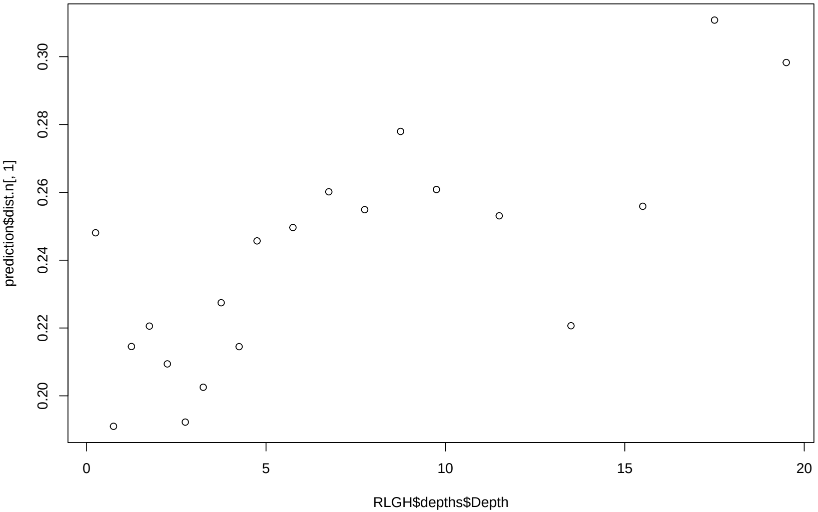
How well constrained are optima

Coverage of fossil data in transfer function

Analogue quality:

Dissimilarity between a fossil sample and closest modern analogue

More confidence in samples with good analogues in modern training set



Analogue quality:

What is a good analogue:

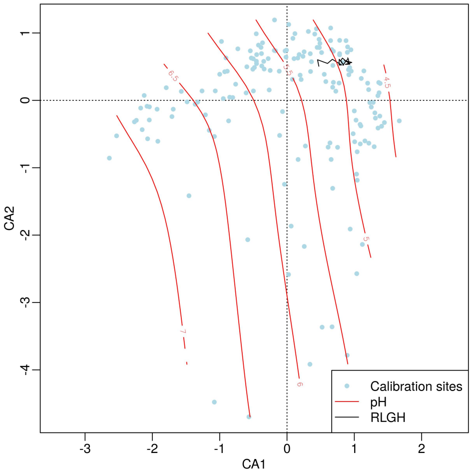
Gavin et al. 2003:

Pollen data assigned to biomes -within biome dissimilarities (good analogue) -among biome dissimilarity (poor analogue)

*Rule of thumb**

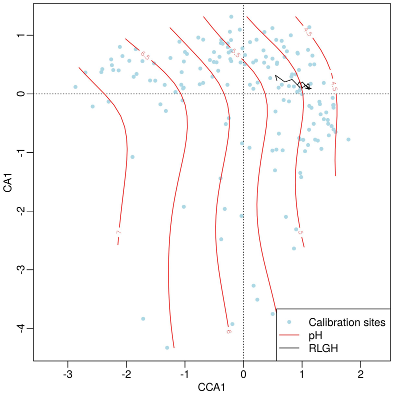
Estimate all dissimilarities within a training set - Dissimilarity $<$ 5% quantile: good analogue - Dissimilarity 5% - 10% quantile: fair analogue - Dissimilarity $>$ 10% quantile: no analogue

Transfer function and fossil data: Ordination



First pass assessment of analogue quality

Transfer function and fossil data: Ordination

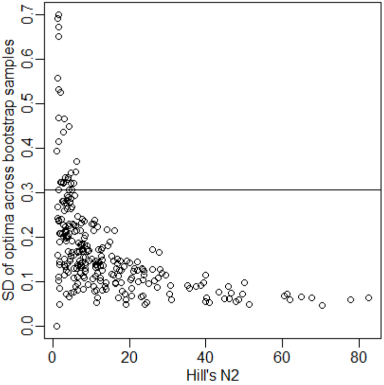


Also use function *residLen* in *analogue* package

How well is sample fitted in higher dimensions?

Transfer function and fossil data: Optima

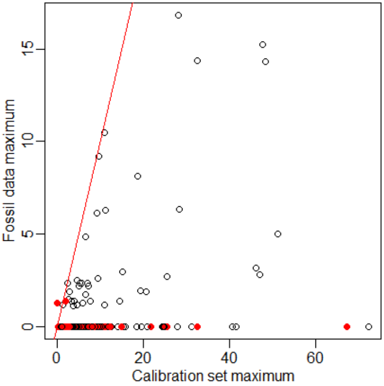
How much does re sampling the calibration data set affect optima



Transfer function and fossil data: Abundance of fossil taxa in training set

Similarity to analogue quality:

- are taxa important in fossil data set also abundant in calibration data set



Conclusions:

- Training sets now exist for a range of organisms and environmental variables for directly and indirectly inferring past environment and climate from biological remains
- A range of numerical methods exists for developing transfer functions
- Methods have advantages and disadvantages
- Producing a reconstruction is easy
- Identifying confounding effects and what can and can't be reconstructed is extremely difficult