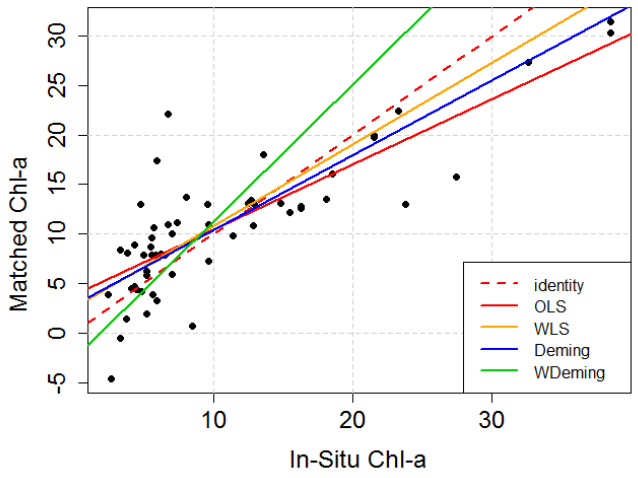
Dealing with data uncertainty: Regression tools for in-situ and satellite data comparisons

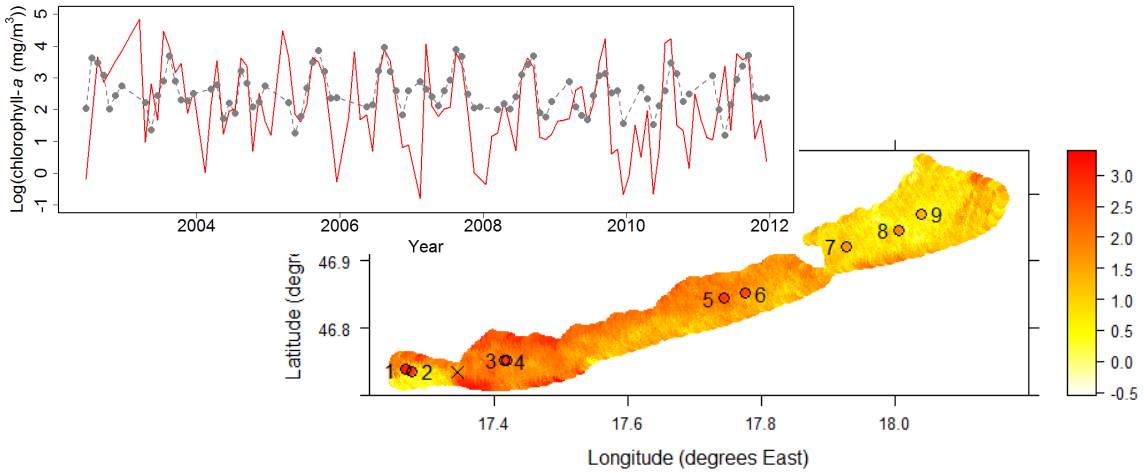
Ruth O'Donnell, Mengyi Gong, Craig Wilkie, **Claire Miller** & Marian Scott | University of Glasgow
Peter Hunter, Vagelis Spyrakos & Andrew Tyler | University of Stirling

Overview – tools investigated/developed

1. **Regression tools** accounting for error in both in-situ and satellite **matched data.**



2. **Regression tools** to combine in-situ and satellite data **at different temporal and spatial scales.**

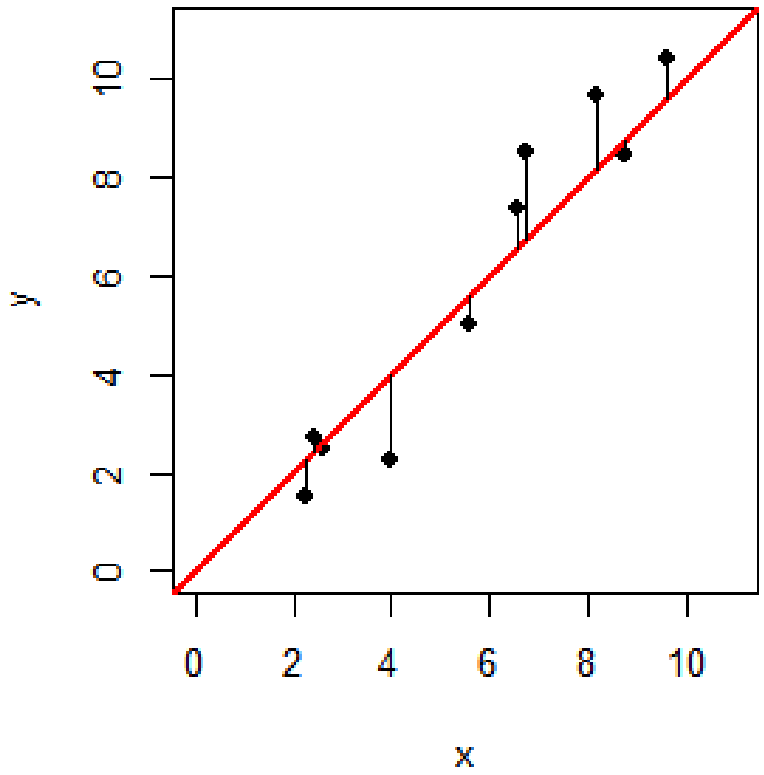


1. Errors in Variables Modelling

Standard regression:

Assumption - there **is no measurement error associated with x.**

If measurement error associated with both x and y is not accounted for, regression parameters will potentially be biased.



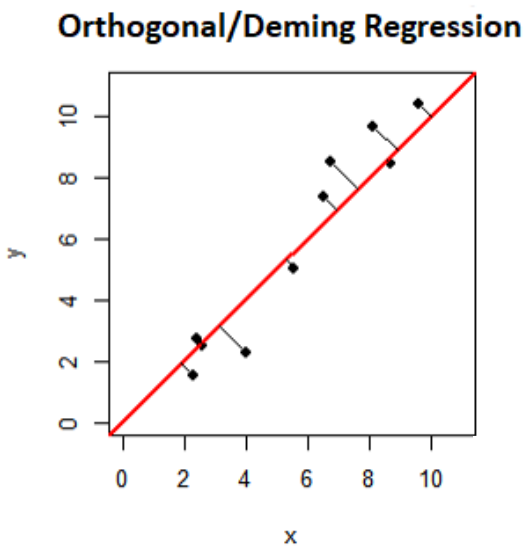
→ **Errors in variables models**

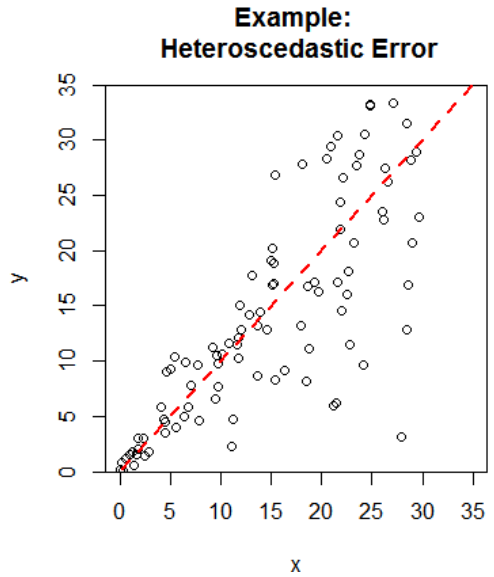
1. Errors in Variables Modelling

- **Orthogonal/Deming Regression –** the angle of minimising the differences from the observations to the line varies.

- **Deming regression** incorporates information on the **ratio of error variability** from the in-situ and the remote sensing data.

- **Weighted Deming regression** can be applied to additionally account for **heteroscedasticity** in the error (variability changes as mean changes).

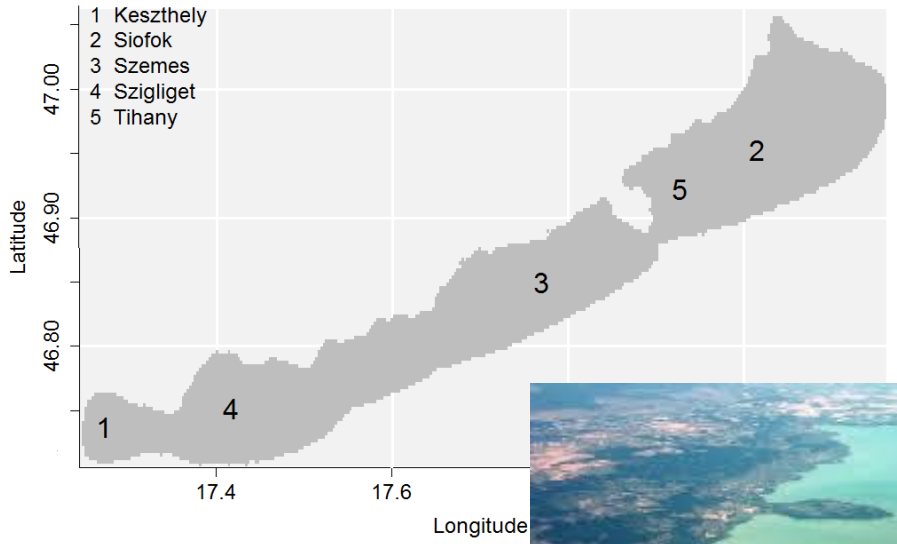




1. Errors in Variables Modelling – Lake Balaton

- **Lake Balaton, Hungary**.
- In-situ measurements taken at **5 sites**[1].
- Matched MERIS Chl-a retrievals[2]:
  - same day (within 3 hours) as;
  - mean of 9 (3 x 3) pixels (300m resolution)

  over the in-situ point.
- **63 matchups** were considered here.

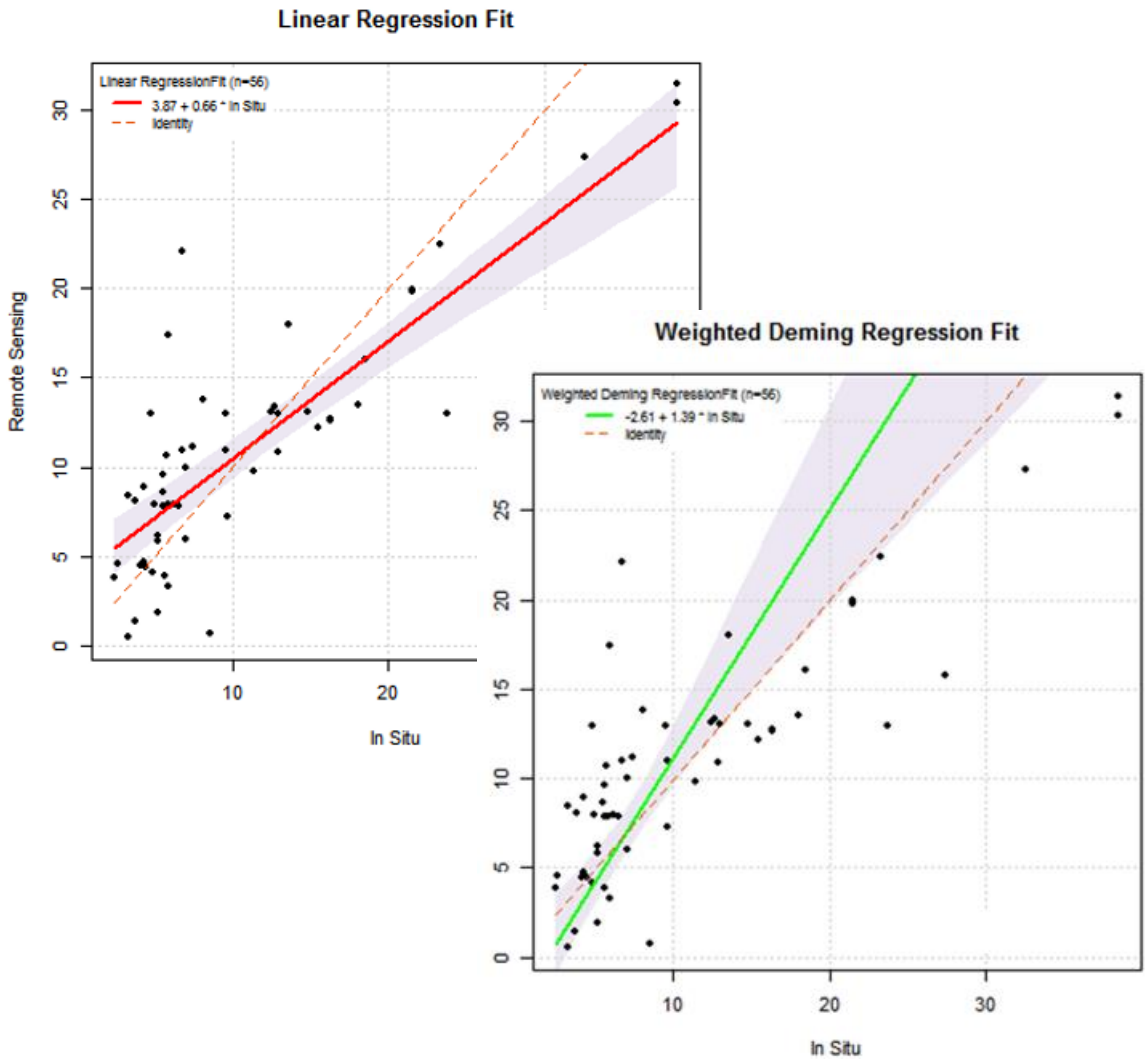


1. In-situ data were provided by Balaton Limnological Institute Centre for Ecological Research and the Central Transdanubian (Regional) Inspectorate for Environmental Protection, Nature Conservation and Water Management.
2. The retrievals discussed here were obtained by University of Stirling using work from (Gower et al., 2004) and (Palmer et al., 2015).

1. Errors in Variables Modelling

Linear regression and weighted Deming regression fit to Chl-a remote sensing data and in-situ data:

- estimated regression lines are solid lines;
- y=x identity line are dashed lines;
- shaded regions represent uncertainty associated with regressions.



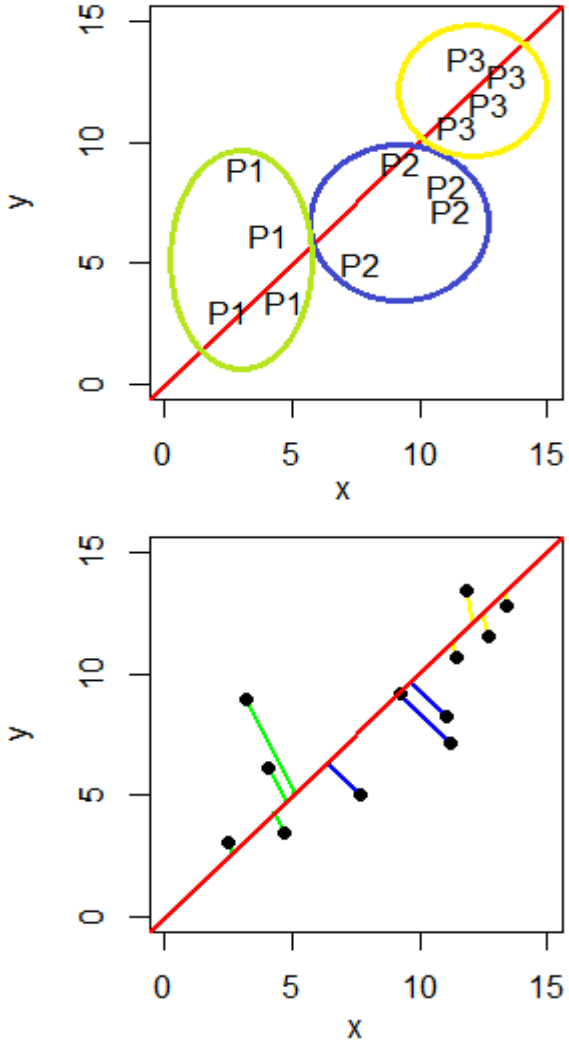
1. Errors in Variables Modelling - MDR

- **Modified Deming Regression (MDR)**

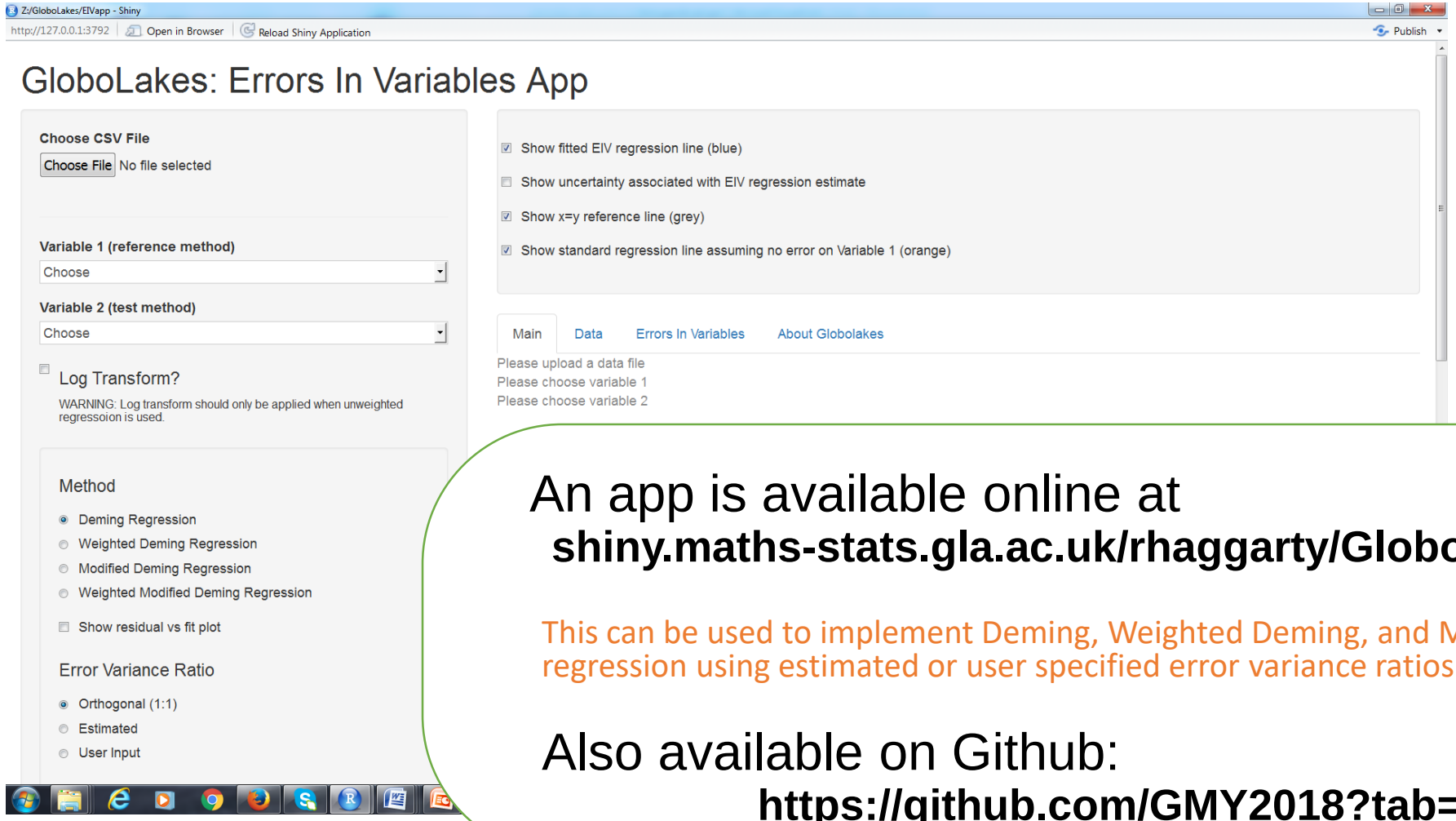
- Accounting for errors that have arisen from multiple populations (e.g. data from different lakes, or different basins within a lake)

  - Multiple error variance ratio values can be used.

  - Regression line is obtained by minimising distance lines at different angles simultaneously.



1. Errors in Variables Modelling - App



An app is available online at
shiny.maths-stats.gla.ac.uk/rhaggarty/GlobolakesEIV/

This can be used to implement Deming, Weighted Deming, and Modified Deming regression using estimated or user specified error variance ratios.

Also available on Github:
https://github.com/GMY2018?tab=repositories

2. Data fusion at different temporal/spatial scales

Lake Balaton

- Previous example - matched MERIS Chl-a retrievals:
  - same day (within 3 hours) as;
  - mean of 9 (3 x 3) pixels (300m resolution)

  over the in-situ point.

- **Nonparametric statistical downscaling**
  - a regression for fusion of data at different spatial and temporal resolutions.

2. Data fusion at different temporal/spatial scales

Lake Balaton (Jan 2002-March 2012)

- 9 in-situ[1] locations sampled monthly
- 7616 pixels with monthly MERIS Chl-a retrievals[2]

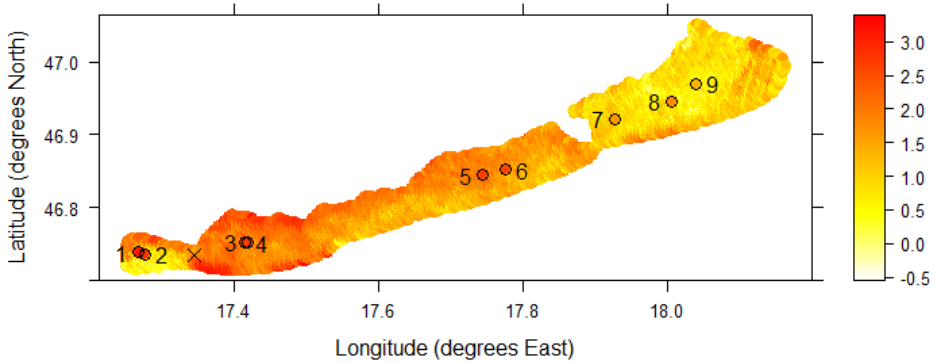


Figure: Remotely-sensed log chl-a data with in-situ data overlaid for March 2011.

1. Insitu data were provided by Balaton Limnological Institute Centre for Ecological Research and the Central Transdanubian (Regional) Inspectorate for Environmental Protection, Nature Conservation and Water Management.

2. Retrievals were provided by the ESA DUE DIVERSITY II project (http://www.diversity2.info/products/inlandwaters/). We acknowledge the ESA DUE DIVERSITY II project, Carsten Brockmann and Daniel Odermatt for providing ENVISAT data and derived indicator products.

2. Data fusion at different temporal/spatial scales

Regression

- to combine in-situ and remote sensing data using curves for matched point/pixel;
- Bayesian framework incorporates error on both in-situ and remotely sensed (RS) data;

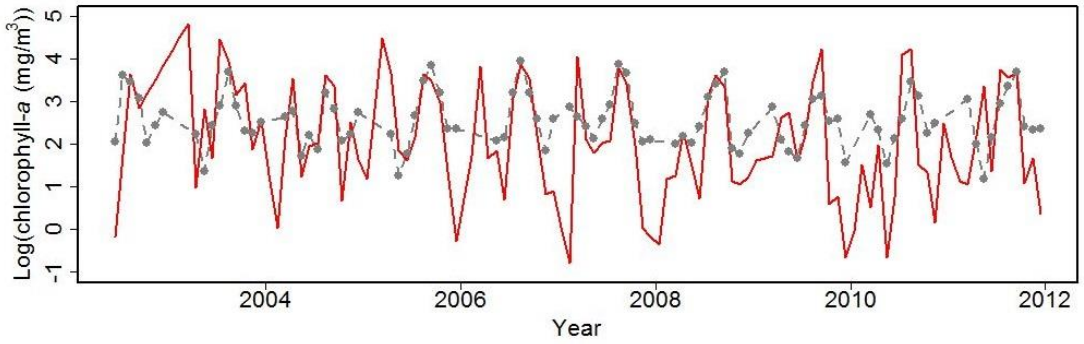


Figure: in-situ (points) and RS data (solid line) at location 1.

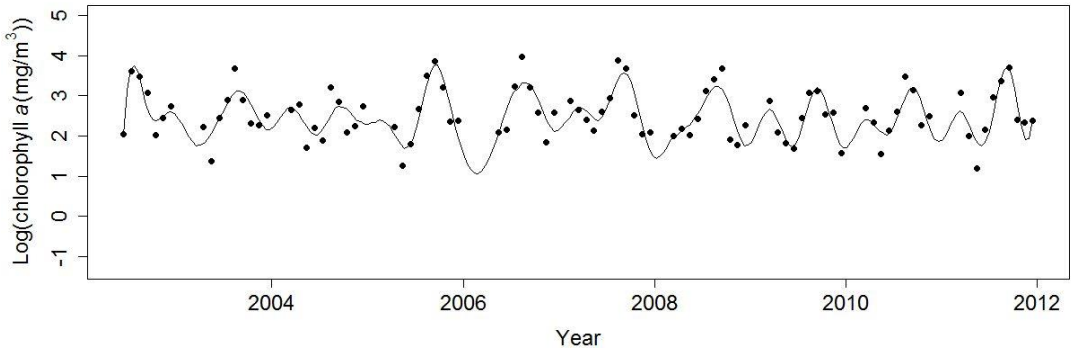


Figure: in-situ (points) and example fitted curve at location 1.

2. Data fusion at different temporal/spatial scales

Regression

- with spatially varying coefficients;
- spatial correlation incorporated;
- prediction possible at any specified points in grid range over time.

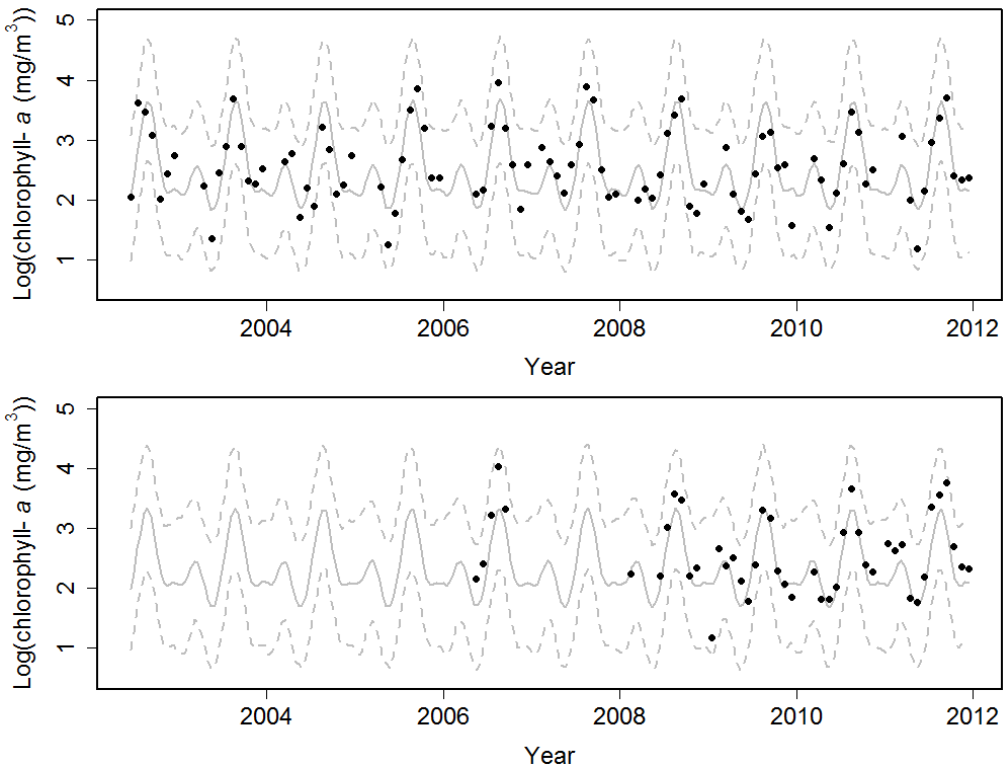
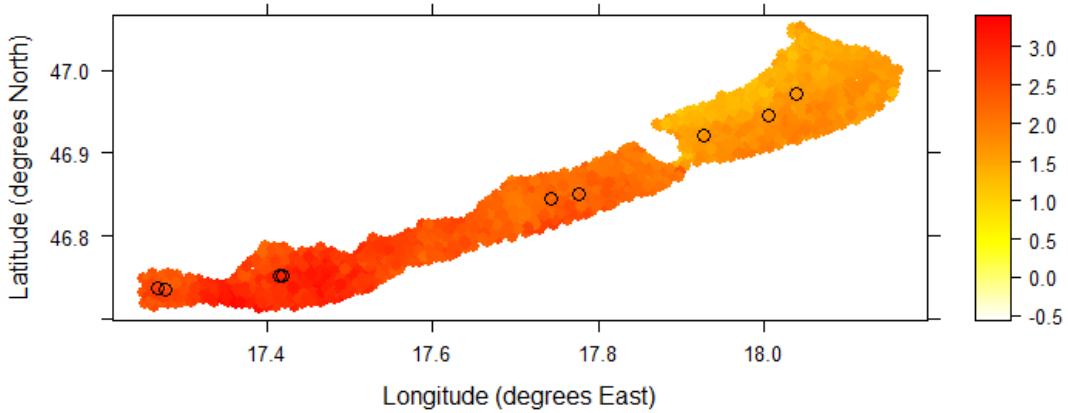


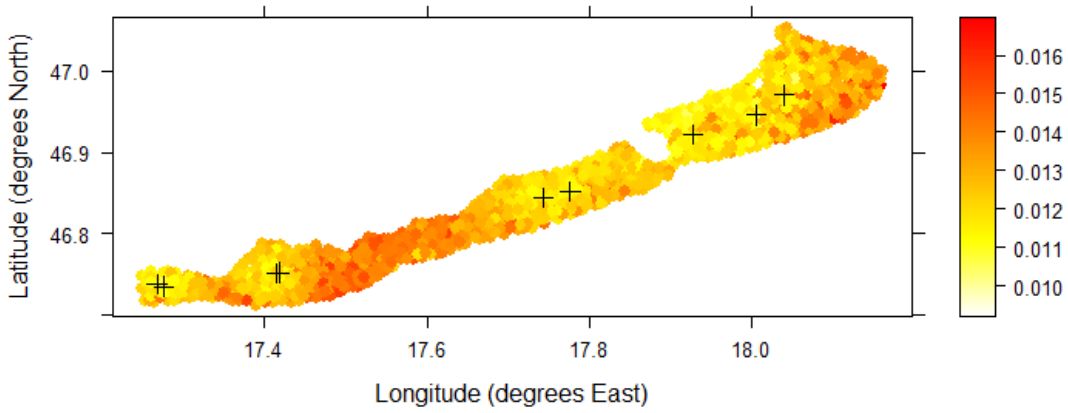
Figure: predictions at locations 1 (top) and 2 (bottom) with in-situ data as points.

2. Data fusion at different temporal/spatial scales

Lake Balaton: log chl-a, March 2011

- log chl-a predictions (top);
- standard errors (bottom);
- symbols indicate in-situ monitoring locations.

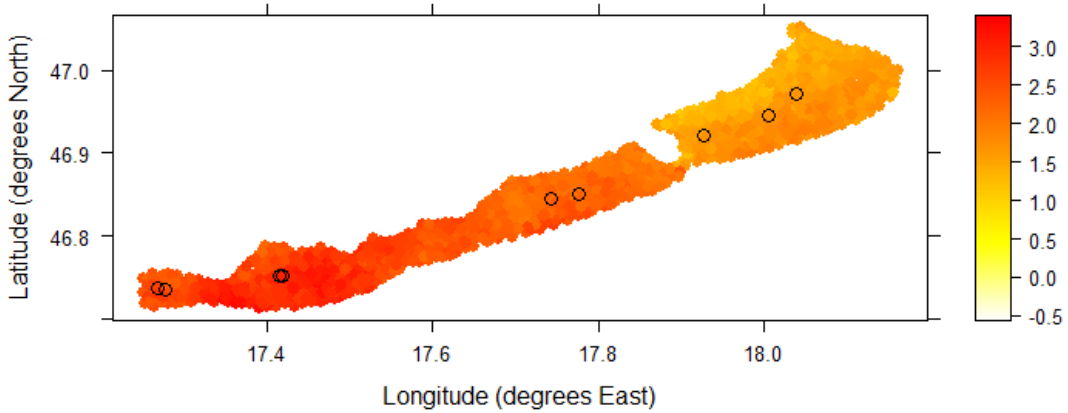


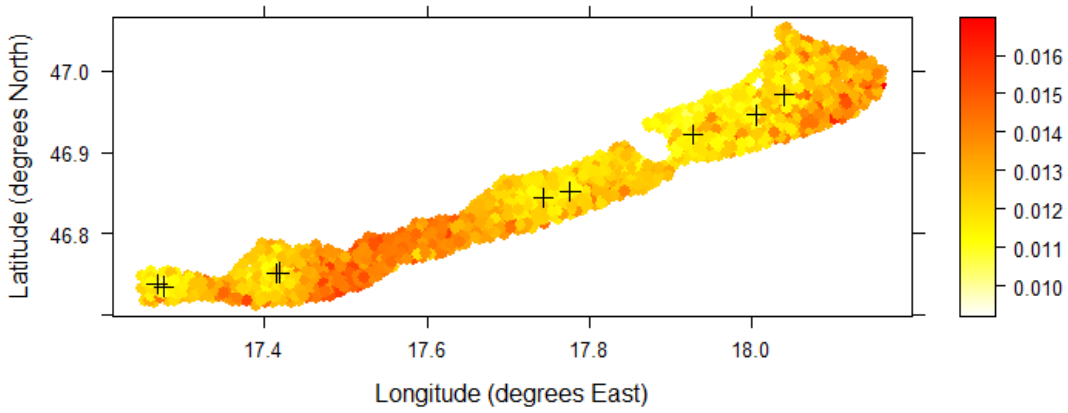


2. Data fusion at different temporal/spatial scales

- **Nonparametric statistical downscaling for the fusion of data of different spatiotemporal support**, Craig Wilkie, Claire Miller, Marian Scott, Ruth O'Donnell, Peter Hunter, Evangelos Spyrakos, Andrew Tyler– submitted to **Environmetrics**

- **Draft R package available at** http://dx.doi.org/10.5525/gla.researchdata.651

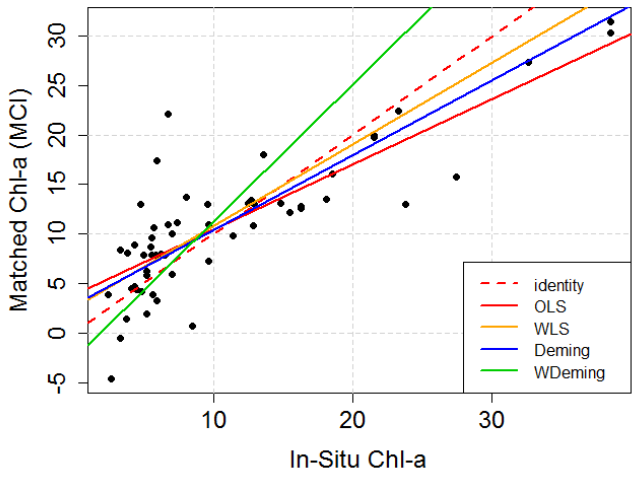




Summary – tools investigated/developed

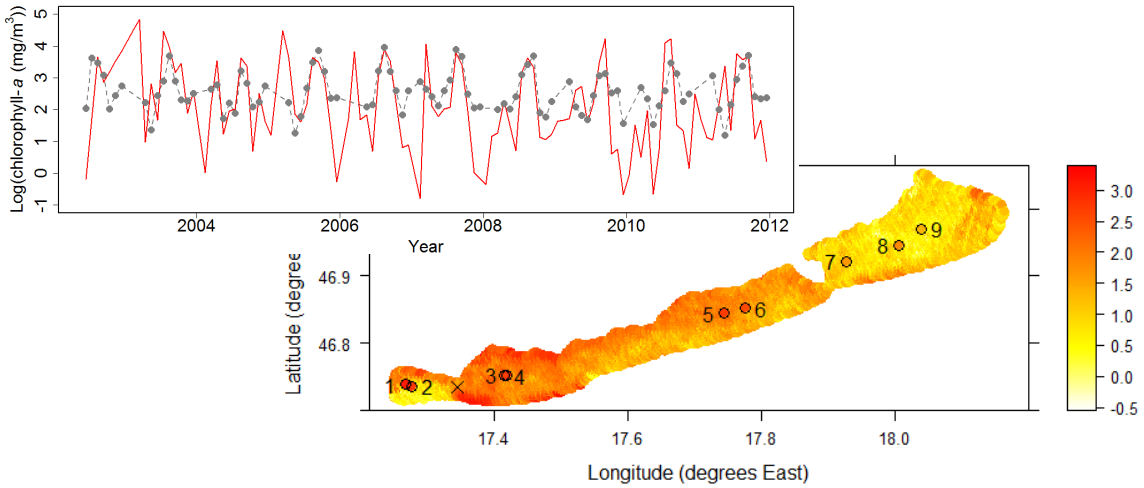
1. **Regression tools** accounting for error in both in-situ and satellite **matched data**

http: //shiny.maths-stats.gla.ac.uk/rhaggarty/GlobolakesEIV/



2. **Regression tools** to combine in-situ and satellite data at **different temporal and spatial scales**

http://dx.doi.org/10.5525/gla.researchdata.651



Thank you

Claire Miller
School of Mathematics & Statistics
University of Glasgow

e claire.miller@glasgow.ac.uk
w www.stats.gla.ac.uk/~claire

www.globolakes.ac.uk

@globolakes

This research was funded by the Natural Environment Research Council (NE/E009328, NE/J022810/1)

NERC
SCIENCE OF THE ENVIRONMENT