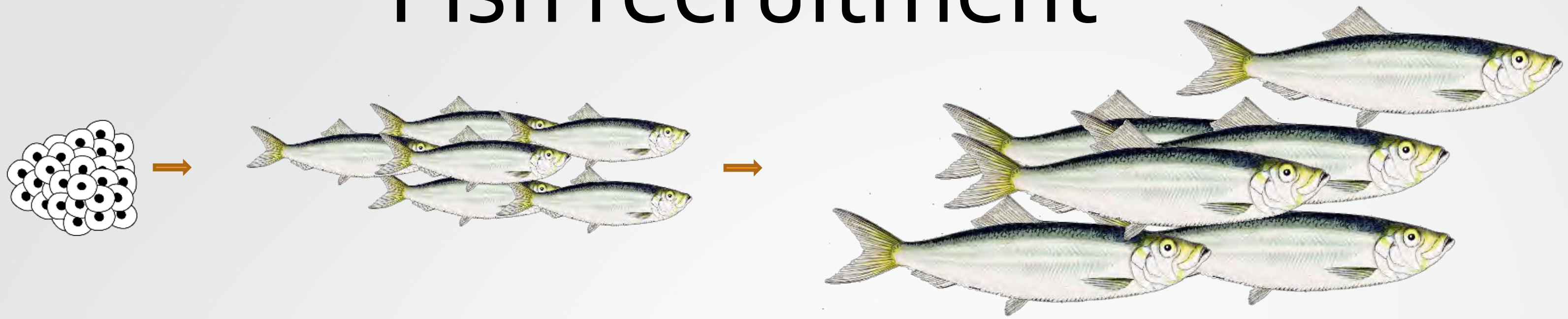
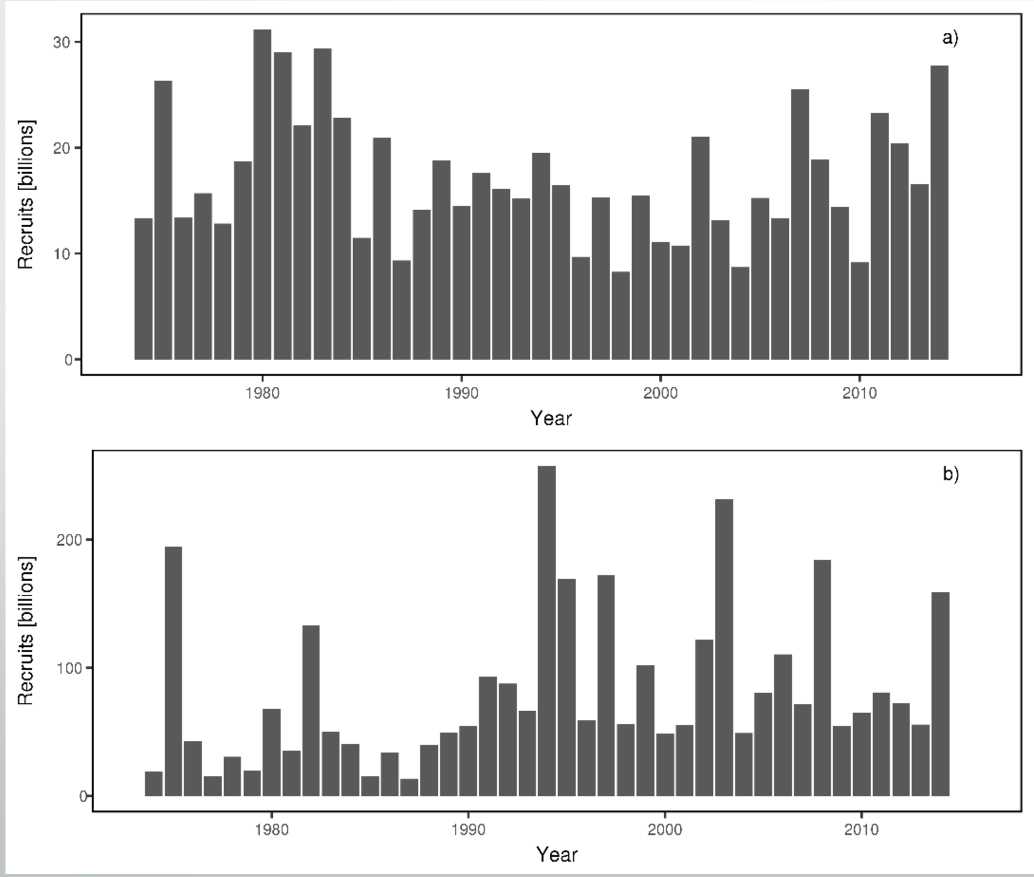
Incorporation of environmental drivers in the prediction of pelagic stocks recruitment in the Baltic Sea using random forest algorithms

Szymon Smoliński
ssmolinski@mir.gdynia.pl
National Marine Fisheries Research Institute
Department of Fisheries Resources

NATIONAL MARINE FISHERIES RESEARCH INSTITUTE

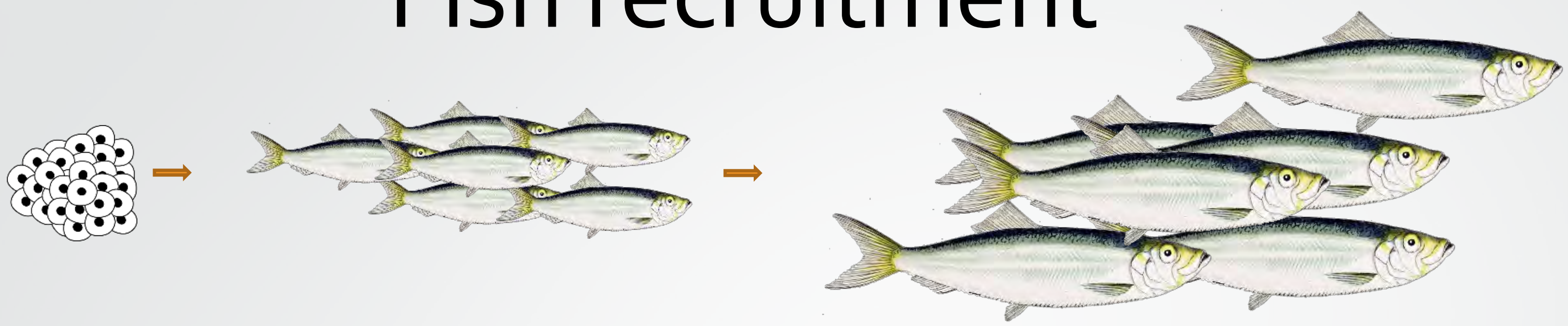
Fish recruitment





Recruitment of
Central Baltic herring (a)
and Baltic sprat (b).

Fish recruitment



Which factors have impact on pelagic fish recruitment?

Köster *et al.*, 2003

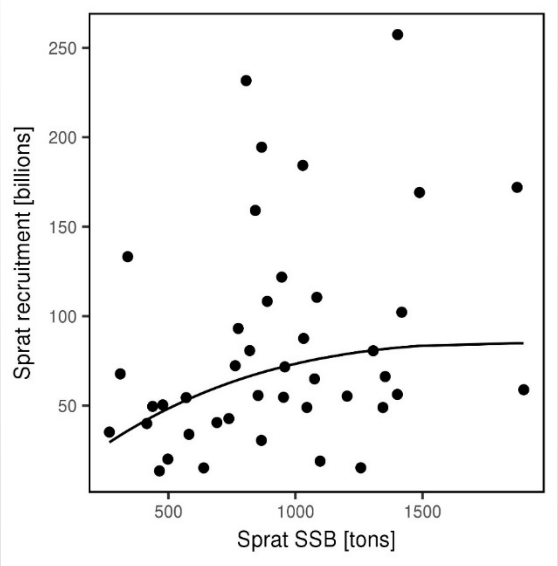
Cardinale *et al.*, 2009

Margonski *et al.*, 2010

Bartolino *et al.*, 2014

Gröger *et al.*, 2014

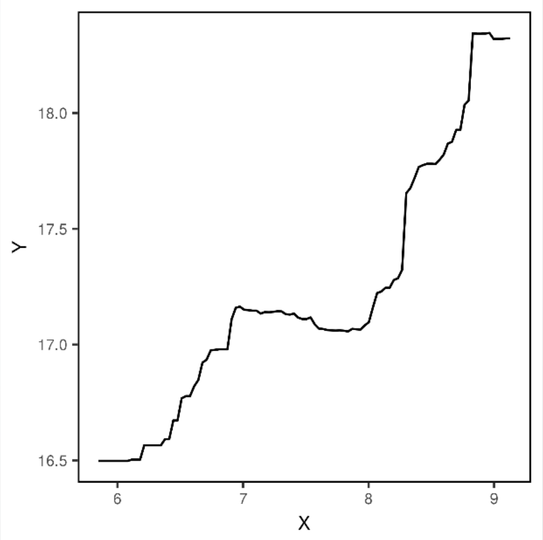
- Spawning stock biomass
- Spawners characteristic
- Interactions with other species
- <u>Environmental factors</u> (hydrological conditions, climate)

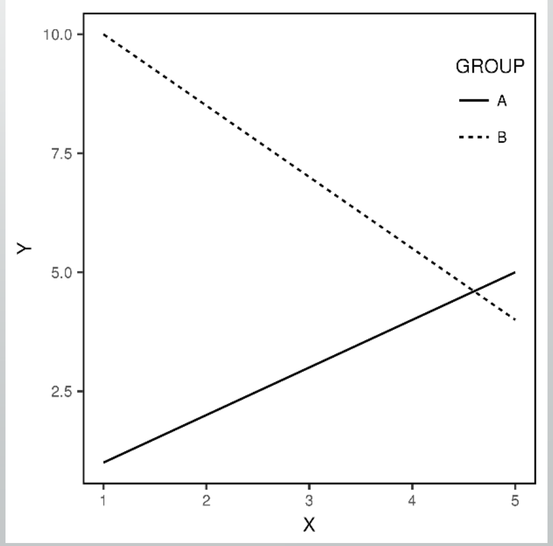


Ricker model of Baltic sprat recruitment.

Environmental data in recruitment modeling

www.marine.copernicus.eu

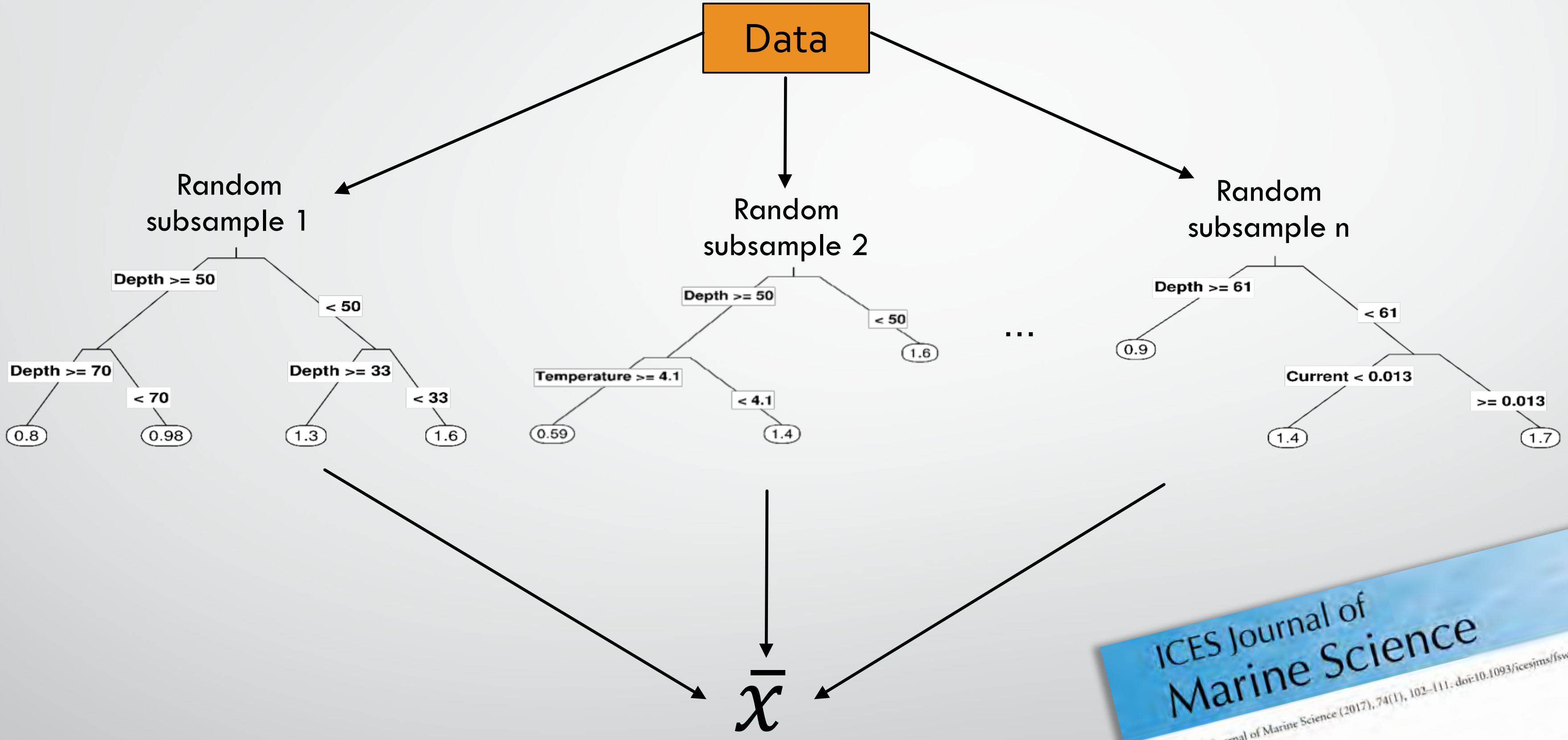




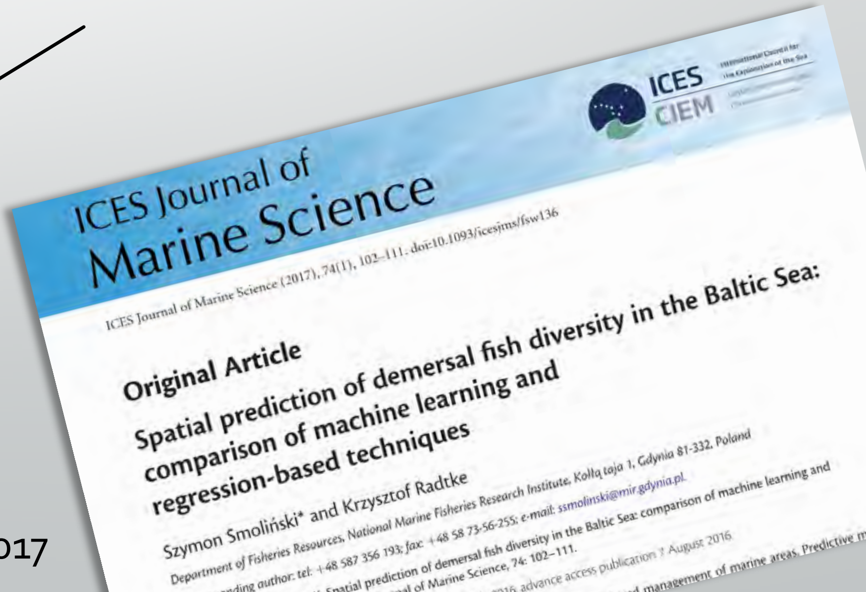
Examples of nonlinear relationships and interactions between variables

Random forests

Breiman L (2001) Random forests. *Machine learning*, **45**, 5–32.



Smoliński & Radtke, 2017


ICES Journal of Marine Science

ICES Journal of Marine Science (2017), 74(1), 102–111. doi:10.1093/icesjms/fsw136

Original Article

Spatial prediction of demersal fish diversity in the Baltic Sea: comparison of machine learning and regression-based techniques

Szymon Smoliński* and Krzysztof Radtke

Department of Fisheries Resources, National Marine Fisheries Research Institute, Kołłątaja 1, Gdynia 81-332, Poland

Boruta algorithm

- an extension of the random forest method which utilizes the importance measure generated by the original algorithm
- compares, in the iterative fashion, the importance of original attributes with importance of their randomized copies (shadow attributes)
- used for feature selection in the systems with unclear mechanisms of investigated processes

Kursa & Rudnicki, 2010

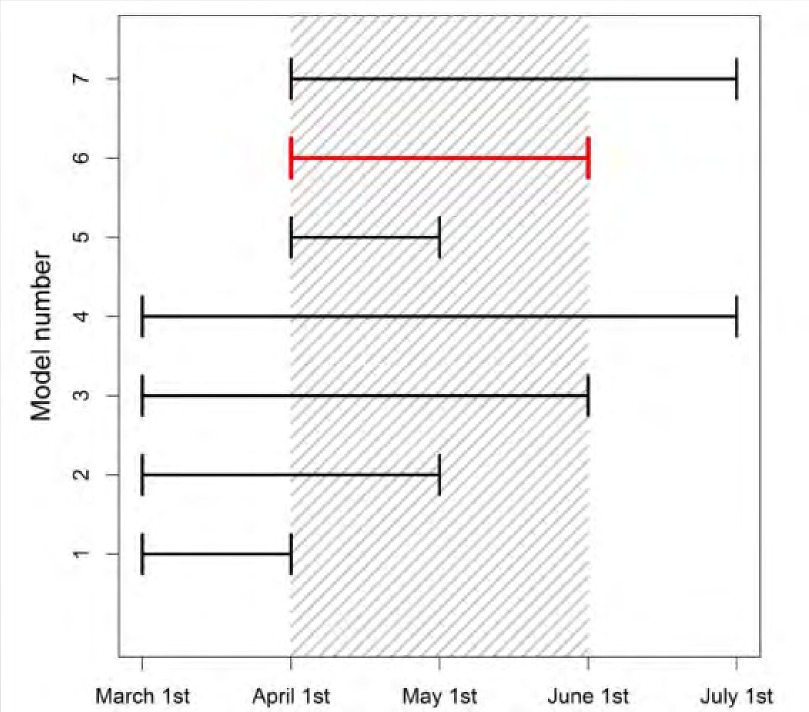
Journal of Statistical Software
September 2010, Volume 36, Issue 11.
http://www.jstatsoft.org/

Feature Selection with the Boruta Package

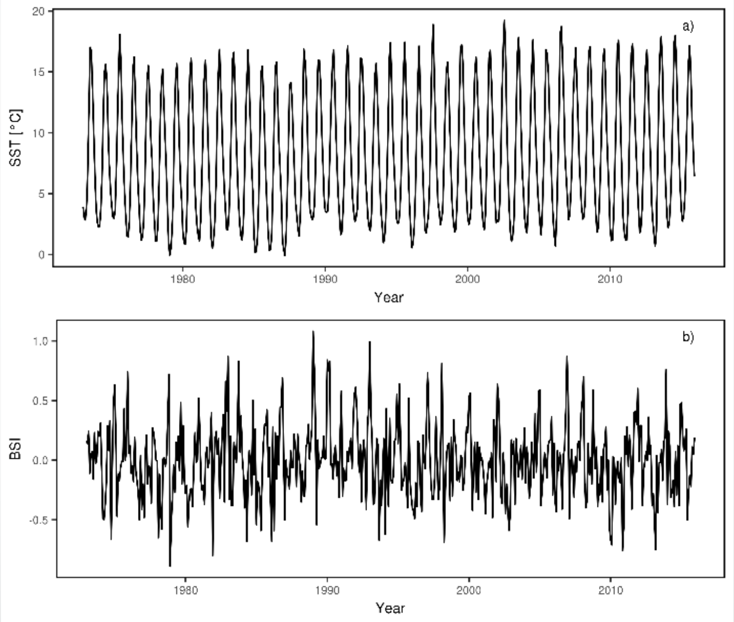
Miron B. Kursa
University of Warsaw

Witold R. Rudnicki
University of Warsaw

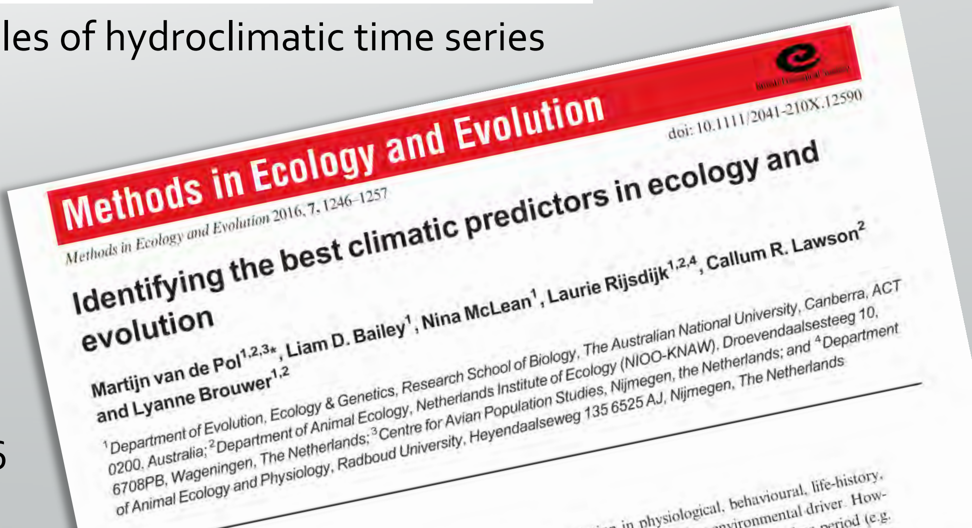
Variables in the temporal scale



Bailey & van de Pol, 2016



Examples of hydroclimatic time series


Methods in Ecology and Evolution

Methods in Ecology and Evolution 2016, 7, 1246–1257

doi: 10.1111/2041-210X.12590

Identifying the best climatic predictors in ecology and evolution

Martijn van de Pol[1,2,3]*, Liam D. Bailey[1], Nina McLean[1], Laurie Rijsdijk[1,2,4], Callum R. Lawson[2] and Lyanne Brouwer[1,2]

[1]Department of Evolution, Ecology & Genetics, Research School of Biology, The Australian National University, Canberra, ACT 0200, Australia; [2]Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Droevendaalsesteeg 10, 6708PB, Wageningen, The Netherlands; [3]Centre for Avian Population Studies, Nijmegen, the Netherlands; and [4]Department of Animal Ecology and Physiology, Radboud University, Heyendaalseweg 135 6525 AJ, Nijmegen, The Netherlands

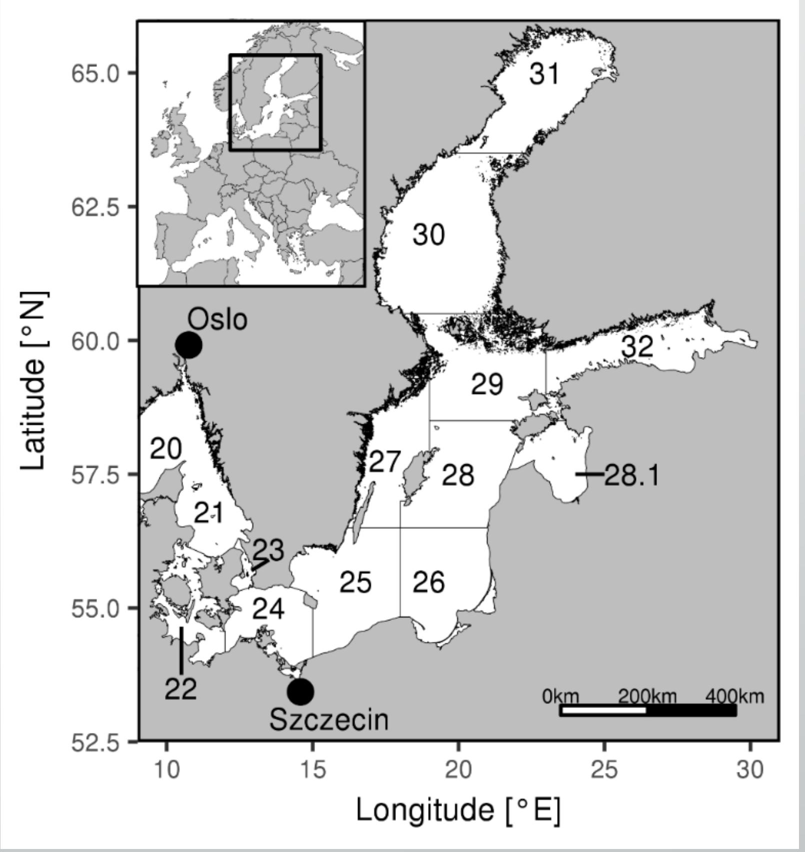
van de Pol *et al.*, 2016

Aims of the work

- Investigation of hydroclimatic factors, that may have influence on recruitment of:
  - herring (*Clupea harengus*) in ICES Subdivisions 25–29, 32, excluding Gulf of Riga
  - sprat (*Sprattus sprattus*) in ICES Subdivisions 22–32

- Hypotheses:
- Hydroclimatic conditions have impact on the recruitment success
- Signals from different time windows may have different effects

- Incorporation of random forests in the „sliding window” method



Map of the Baltic Sea with indicated ICES subdivisions.

Methods

List of variables used in the modelling of Baltic herring and sprat recruitment.

Variables abbreviation	Description	Source
Biological data		
her/sprR	Herring or sprat recruitment at age 1	(ICES, 2016)
her/sprSSB	Herring or sprat spawning stock biomass	(ICES, 2016)
her/sprWAAx	Herring or sprat weight at age x	(ICES, 2016)
codTSB	Total stock biomass of cod in subdivisions 25-32	(ICES, 2013)
Hydroclimatic data		
SST	Mean sea surface temperature (Monthly resolution)	(Huang et al., 2015)
BSI	Mean Baltic Sea Index (Monthly resolution)	(Lehmann et al., 2002)

Data analysis

Separate runs for each stock and hydroclimatic variable (SST and BSI)

For each time window (range of 36 months):

- Calculate the value of investigated variable for each observation using aggregate statistic (e.g. mean)
- Add new variable to the data matrix
- Select variables using Boruta algorithm (test of variable relevance)
- Cross-validate random forest grown on selected variables:
  - Randomly split the data into the 5 subsets
  - For each subset: train model on 4 subsets and test model on remaining subset

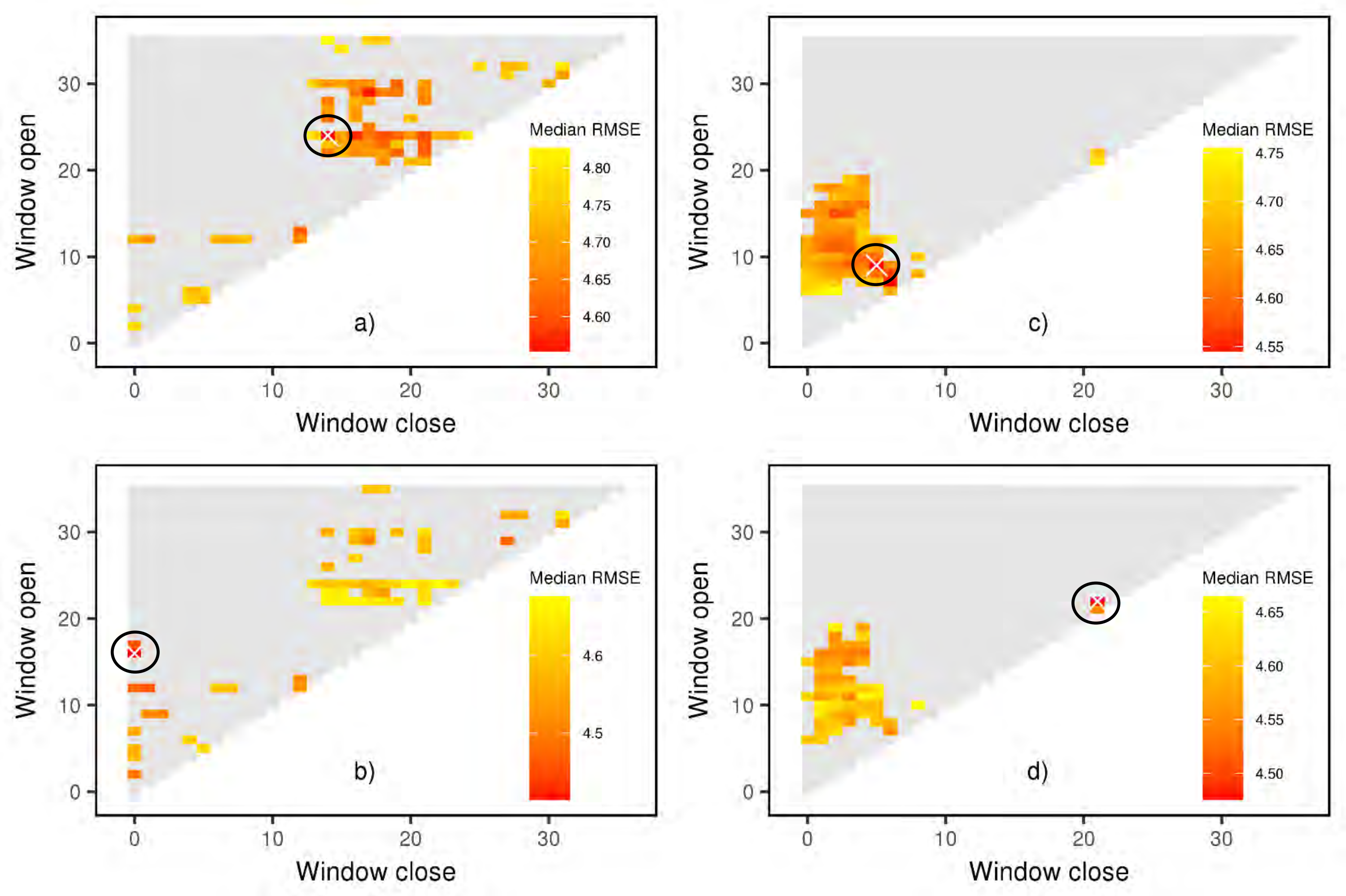
100x

- Variable relevance
- RMSE (root mean square error)

Identification of optimal signal

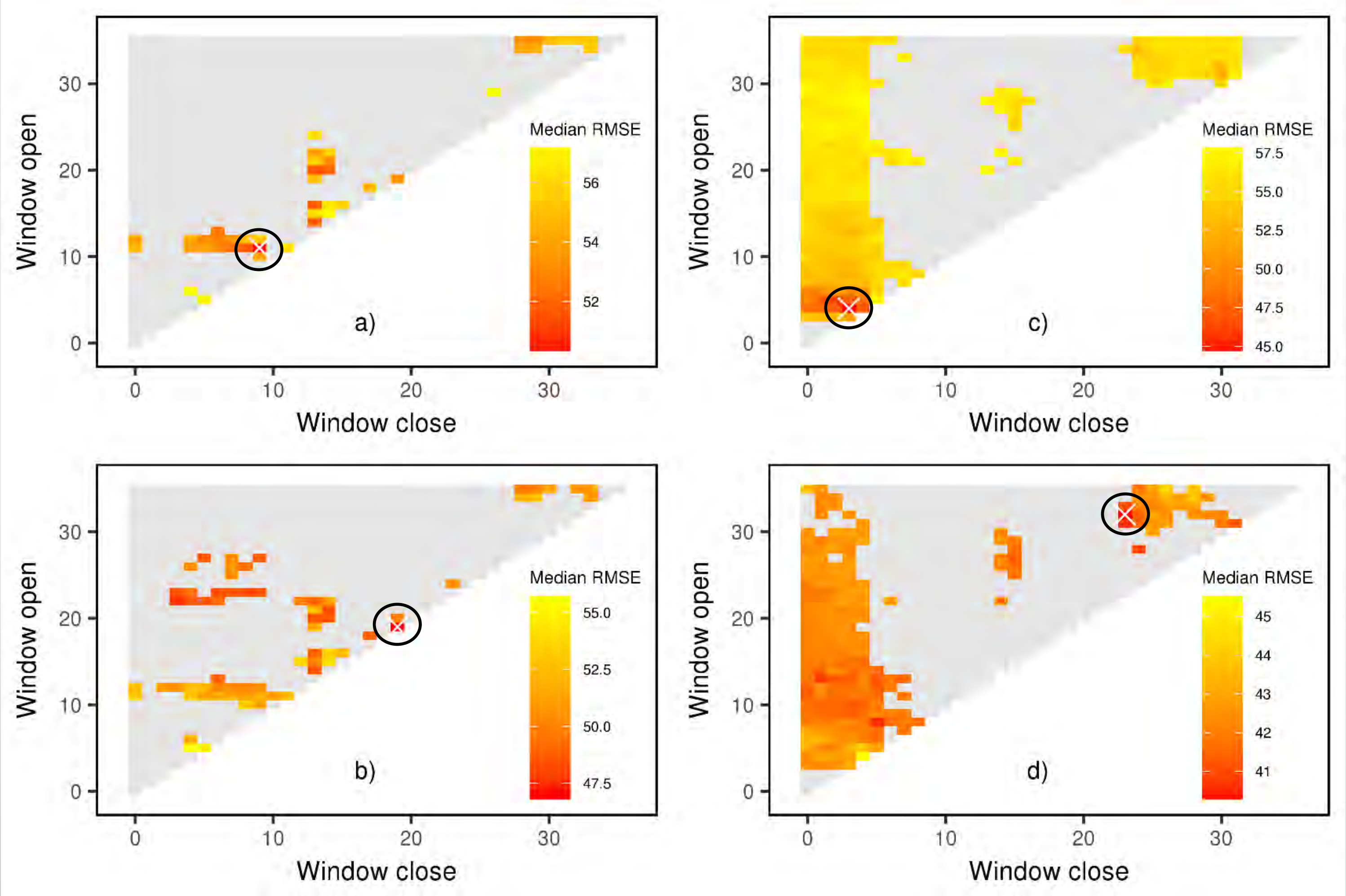
Further analysis

Results



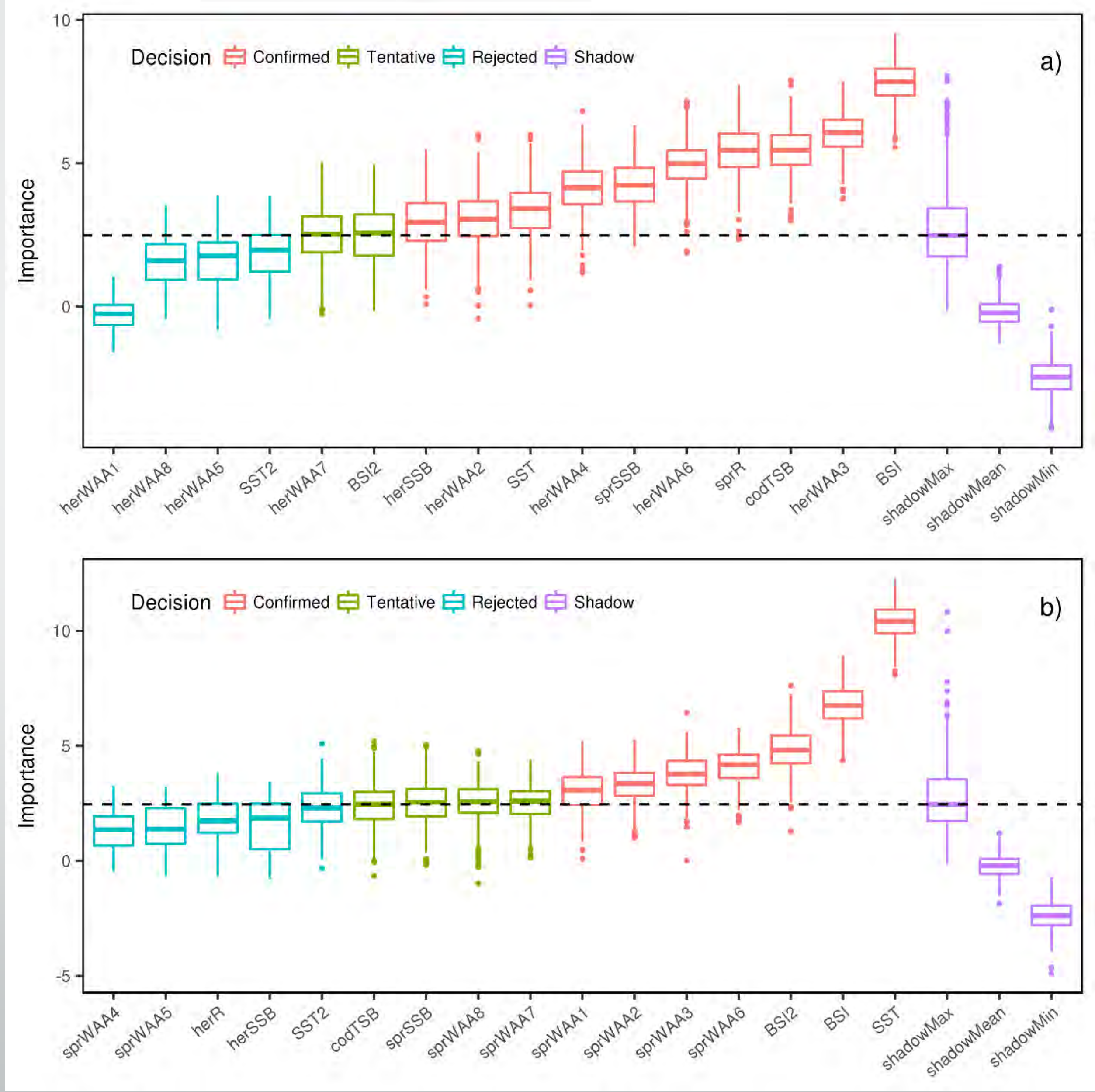
Results of random forest “sliding window” analysis for the environmental effects on herring recruitment. Outcome of first and second step of optimal signals identification for BSI (a, b) and SST (c, d) were shown on the plots.

Results



Results of random forest "sliding window" analysis for the environmental effects on sprat recruitment. Outcome of first and second step of optimal signals identification for BSI (a, b) and SST (c, d) were shown on the plots.

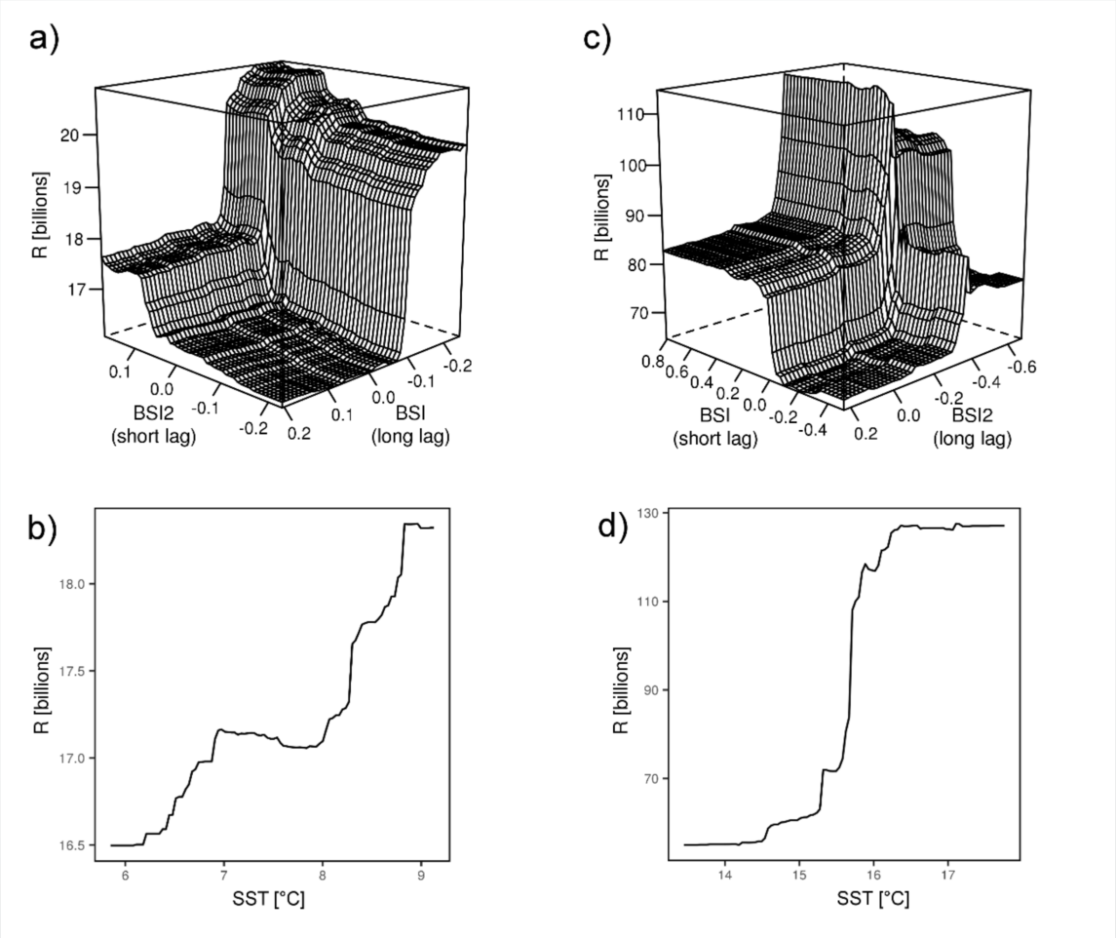
Results



Relevance of variables in the random forest model of herring (a) and sprat (b) recruitment according to the results of Boruta.

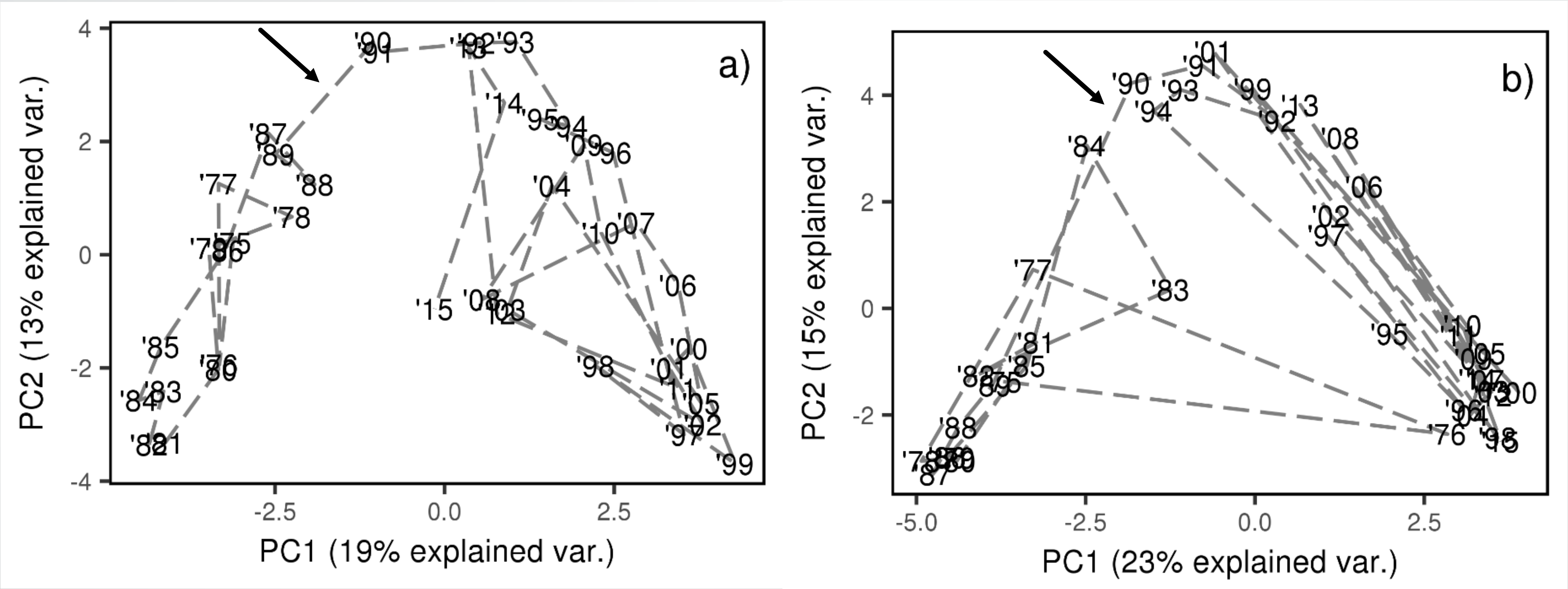
Results

	Window open	Window close	Effect
Herring			
BSI	24	14	-
BSI2	16	0	+
SST	9	5	+
Sprat			
BSI	11	9	+
BSI2	19	19	-
SST	4	3	+



Partial dependence plots for hydroclimatic variables for random forest predictions of herring (a, b) and sprat (c, d) recruitment. Effects of different BSI signals (according to results of 1st and 2nd step of "sliding window" analysis) were presented.

Results



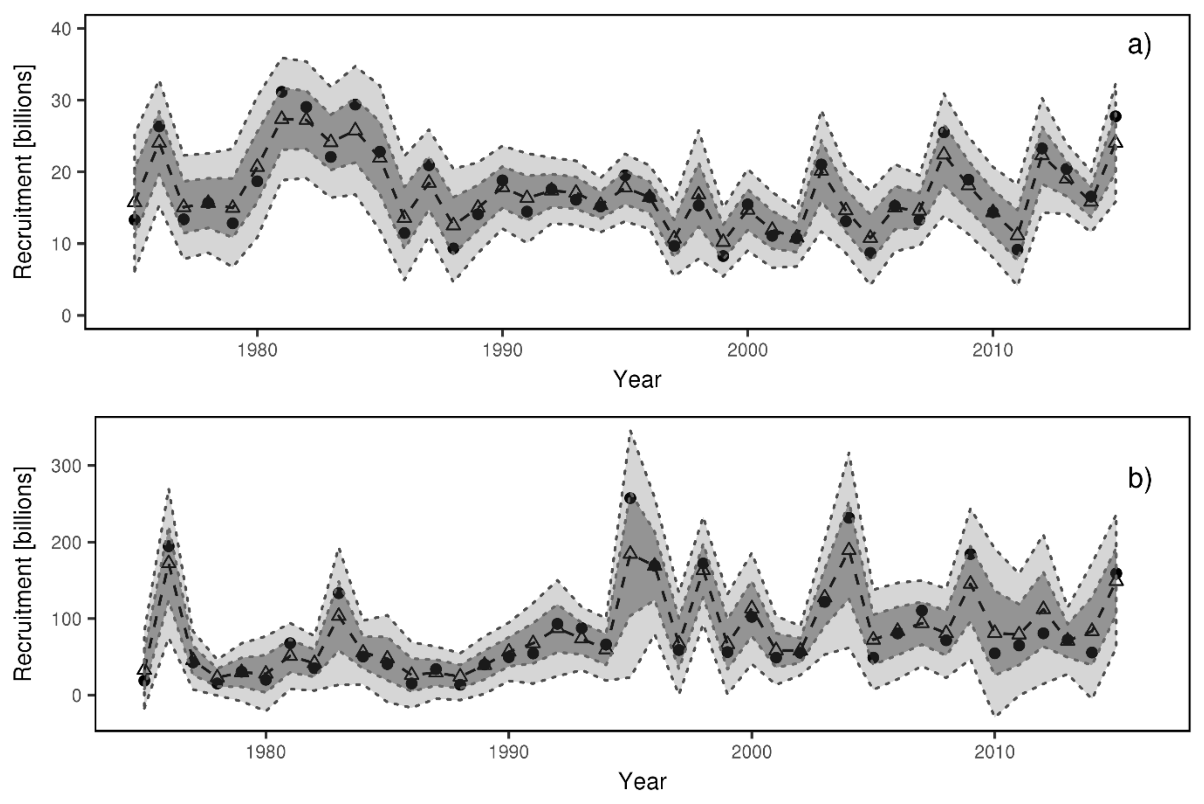
Plots of two first principal component scores derived by PCA based on proximity matrix of recruitment years of herring (a) and sprat (b).

Results

Prediction accuracy (R^2±sd) of models derived from cross-validation.

Variables included	Herring	Sprat
All relevant variables	**0.59**±0.211	**0.50**±0.229
Biological variables	0.44±0.237	0.17±0.191
SSB	0.11±0.127	0.13±0.159

Plots of observed (points) and predicted (open triangles) recruitment of herring (a) and sprat (b).



Conclusions

- Incorporation of environmental data improves accuracy of Baltic pelagic fish recruitment prediction
- SST and BSI have significant impact on recruitment processes
- Observed relationships of pelagic fish recruitment and hydroclimatic conditions were nonlinear
- The same environmental variable from different time windows may have counteracting effects (Kruuk *et al.* 2015, GCB)
- Data mining methods and random forests may be used to obtained new knowledge from ecological data

References

- Bailey LD, van de Pol M (2016) climwin: An R Toolbox for Climate Window Analysis. *PLoS ONE*, 11, 1–27.
- Bartolino V, Margonski P, Lindegren M et al. (2014) Forecasting fish stock dynamics under climate change : Baltic herring (Clupea harengus) as a case study. *Fisheries Oceanography*, 23, 258–269.
- Breiman L (2001) Random forests. *Machine learning*, 45, 5–32.
- Cardinale M, Möllmann C, Bartolino V et al. (2009) Effect of environmental variability and spawner characteristics on the recruitment of Baltic herring Clupea harengus populations. *Marine Ecology Progress Series*, 388, 221–234.
- Gröger JP, Hinrichsen HH, Polte P (2014) Broad-scale climate influences on spring-spawning herring (Clupea harengus, L.) recruitment in the Western Baltic Sea. *PLoS ONE*, 9, 1–14.
- Huang B, Banzon VF, Freeman E et al. (2015) Extended reconstructed sea surface temperature version 4 (ERSST.v4). Part I: Upgrades and intercomparisons. *Journal of Climate*, 28, 911–930.
- ICES (2013) Report of the Baltic Fisheries Assessment Working Group 2013 (WGBFAS), 10 - 17 April 2013, ICES Headquarters, Copenhagen. ICES CM 2013/ACOM:10. 747.
- ICES (2016) Report of the Baltic Fisheries Assessment Working Group 2016 (WGBFAS), 12 - 19 April 2016, ICES Headquarters, Copenhagen. ICES CM 2016/ACOM:11. 793.
- Köster FW, Hinrichsen H, Schnack D et al. (2003) Recruitment of Baltic cod and sprat stocks : identification of critical life stages and incorporation of environmental variability into stock-recruitment relationships. *Scientia Marina*, 67, 129–154.
- Kruuk LEB, Osmond HL, Cockburn A (2015) Contrasting effects of climate on juvenile body size in a Southern Hemisphere passerine bird. *Global Change Biology*, 21, 2929–2941.
- Kursa MB, Rudnicki WR (2010) Feature Selection with the Boruta Package. *Journal Of Statistical Software*, 36, 1–13.
- Lehmann A, Krauss W, Hinrichsen HH (2002) Effects of remote and local atmospheric forcing on circulation and upwelling in the Baltic Sea. *Tellus A*, 54, 299–316.
- Margonski P, Hansson S, Tomczak MT, Grzebielec R (2010) Climate influence on Baltic cod, sprat, and herring stock-recruitment relationships. *Progress in Oceanography*, 87, 277–288.
- van de Pol M, Bailey LD, McLean N, Rijsdijk L, Lawson CR, Brouwer L, Gimenez O (2016) Identifying the best climatic predictors in ecology and evolution. *Methods in Ecology and Evolution*, 7, 1246–1257.
- Smoliński S, Radtke K (2017) Spatial prediction of demersal fish diversity in the Baltic Sea: comparison of machine learning and regression-based techniques. *ICES Journal of Marine Science*, 74, 102–111.