



## Partitioning the impact of environmental drivers and species interactions in dynamic aquatic communities

C. J. M. MUSTERS <sup>1,†</sup>, OLEKSANDRA IEROMINA,<sup>2</sup> S. HENRIK BARMENTLO,<sup>1</sup> ELLARD R. HUNTING,<sup>3,4</sup> MAARTEN SCHRAMA,<sup>1,5</sup> ELLEN CIERAAD,<sup>1</sup> MARTINA G. VIJVER,<sup>1</sup> AND PETER M. VAN BODEGOM<sup>1</sup>

<sup>1</sup>*Institute of Environmental Sciences, Leiden University, Leiden, The Netherlands*

<sup>2</sup>*Dutch Board for the Authorisation of Plant Protection Products and Biocides (Ctgb), Ede, The Netherlands*

<sup>3</sup>*School of Biological Sciences, University of Bristol, Bristol, UK*

<sup>4</sup>*Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, USA*

<sup>5</sup>*Naturalis Biodiversity Center, Leiden, The Netherlands*

**Citation:** Musters, C. J. M., O. Ieromina, S. H. Barmentlo, E. R. Hunting, M. Schrama, E. Cieraad, M. G. Vijver, and P. M. van Bodegom. 2019. Partitioning the impact of environmental drivers and species interactions in dynamic aquatic communities. *Ecosphere* 10(11):e02910. 10.1002/ecs2.2910

**Abstract.** Temperate aquatic communities are highly diverse and seasonally variable, due to internal biotic processes and environmental drivers, including human-induced stressors. The impact of drivers on species abundance is supposed to differ fundamentally depending on whether populations are experiencing limitations, which may shift over the season. However, an integrated understanding of how drivers structure communities seasonally is currently lacking. In order to partition the effect of drivers, we used random forests to quantify interactions between all taxa and environmental factors using macrofaunal data from 18 agricultural ditches sampled over two years. We found that, over the agricultural season, taxon abundance became increasingly better predicted by the abundances of co-occurring taxa and nutrients compared to other abiotic factors, including pesticides. Our approach provides fundamental insights in community dynamics and highlights the need to consider changes in species interactions to understand the effects of anthropogenic stressors.

**Key words:** abiotic; anthropogenic stressors; biotic; bottom-up; random forest; seasonal change; trophic level.

**Received** 23 July 2019; **accepted** 4 September 2019. Corresponding Editor: Fei Fang.

**Copyright:** © 2019 The Authors. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

† **E-mail:** musters@cml.leidenuniv.nl

### INTRODUCTION

Gaining knowledge from field data on the fundamental processes that shape biological communities is a huge challenge, but, if we are to effectively manage communities under threat, such knowledge is urgently needed. Many biological communities exhibit a strong seasonal variation in species composition and abundance. For instance, in aquatic communities, most invertebrates are relatively inactive in winter due to reduced water temperatures. Some species are present in the form of eggs or pupae that remain dormant until water temperatures increase in spring, while adult life stages of other aquatic

taxa seek refuge in organic matter layers or in adjacent terrestrial habitats (Chadd 2010). Microbial and algae production is present, but low, in winter (Sommer et al. 1986, Wetzel 2001). In spring, fauna populations grow in response to increasing primary production, which is the kick-off of a seasonal succession in species composition. Toward summer, population growth slows down because of intensified intra- and interspecific interactions (Sommer et al. 1986).

This phenomenon suggests a shift in processes that regulate the populations of most species, from processes of growth to processes of species interactions. Intra- and interspecific interactions may be dominated by competition for limiting

resources, leading to bottom-up controlled communities, or by predation and the like (parasitism, pathogeny, mutualism, etc.), leading to top-down-regulated communities (Shurin et al. 2006). In general, competition reduces species richness, while predation increases species richness by reducing interspecific competition (Terborgh 2015). While it seems likely that a seasonal shift occurs in many communities, this fundamental phenomenon has to our knowledge not been addressed at the community level, let alone quantified.

In addition to these autecological dynamics, communities are affected by environmental drivers, many of which are related to human activities such as land use, transport, and industry (Schwarzenbach et al. 2006, Ormerod et al. 2010, Halstead et al. 2014). Prevailing ecological concepts indicate that the effects of environmental drivers on a community will be different depending on the processes that regulate the populations of the majority of species (Carpenter et al. 1987, Chase and Leibold 2003, Vellend 2016). For example, in case of consumer species, human activities that introduce toxins, such as pesticides, into the environment may affect the vitality of species (Barmantlo et al. 2019) and therewith has relevance for population growth rate (and hence likely affect community composition most when populations are not yet resource-limited). On the other hand, human activities that affect resource availability and limitations, such as fertilization, may directly impact intra- and interspecific interactions (Scheffer et al. 1993). Environmental drivers not only may change the timing of the seasonal shift; both reducing growth rates by toxins and reducing resource limitations by nutrients may postpone the shift. They may also change the character of communities, for example, by changing them from top-down to bottom-up communities and thereby reducing their species richness (Scheffer et al. 1993, Terborgh 2015). While the fundamental question of how environmental drivers affect these different communities has long been recognized (Hunter and Price 1992), it has rarely been explored (but see Lancaster and Ledger 2015).

Understanding how environmental drivers affect individual species abundances, species interactions, and their dynamics is critically needed given the profound role of seasonal

dynamics in community composition and therefore also in ecosystem functioning. This requires quantification of the effects of environmental drivers on the community during the period when population growth dominates, as well as during the period when intra- and interspecific interaction dominates. In addition, knowledge about the timing of the seasonal switch between these two is essential in managing human influences on aquatic systems to maintain or restore species richness. It could, for instance, reveal time windows in which human disturbance is most damaging or when indicator species for specific human impacts should be monitored. Efforts to manage human impacts are to date still primarily governed by trial and error (Terborgh 2015, Hill et al. 2016).

Thus far, the high diversity of ecological communities and the associated number of potential interactions (Kéfi et al. 2015) have hampered quantitative analysis. As a consequence, field studies of seasonality in aquatic systems have determined changes in productivity, biomass, species richness, and interactions in a limited number of taxa or functional groups (Odum 1969, Sommer et al. 1986, Carpenter et al. 1987, Hill et al. 2016, Leslie and Lamp 2017, Ovaskainen et al. 2017, Little and Altermatt 2018), but rarely in all species interactions, including interspecific competition, of a complete community. An exception is the study of Lima-Mendez et al. (2015) who used a relatively new classifying technique, random forest, developed for machine learning (Strobl et al. 2009), to show that, in addition to environmental conditions, species interactions are important drivers of in marine plankton communities. Here, we expand on the random forest method and partition the relative importance of different drivers to gain fundamental ecological insights in community processes, articulating the role of interspecific species interactions therein.

Our method allows further specification of species interactions by categorizing the species of the community. Species can be categorized based on taxonomy or functional traits to show the interactions within and between higher level taxa, or between functional groups such as based on feeding mode. Grouping according to trophic level will allow assessment of whether a community is bottom-up or top-down. This can also be

used to make a distinction between interactions within the same trophic level (horizontal interactions), of which competition is the most important example, and interactions between species of different trophic level (vertical interactions), such as most forms of predation and the like.

We illustrate the applicability of this novel approach by analyzing a comprehensive dataset of aquatic ditch macrofaunal samples (Verdonshot et al. 2011). This dataset describes a set of ditches that are exposed to different types of land uses that create a mosaic of different abiotic pressures affecting the adjacent aquatic system (Ieromina et al. 2015, Hunting et al. 2016, Musters et al. 2019). Over the seasons, we expect to find in the macrofaunal community a shift in the relative importance of community processes, from those related to population growth toward an increased importance of interspecific interactions. Further, we expect that the impact of pesticides is greatest in the early seasons, when population growth dominates, so that the community might be sensitive to the vitality disturbing effect of these substances. In contrast, the impact of nutrients, limiting the biomass of food items, will be greatest in the late seasons when populations are at their peak and competition between species might be high.

## METHODS

### Research area

A detailed description of the research area, macrofaunal sampling strategy, and taxonomic identification level for each group is given in Ieromina et al. (2015, Musters et al. 2019). Briefly, the research area of ca. 1600 ha contains a network of ditches and is located in the bulb growing region of the Netherlands (center: Lat: 52°15'55.66", Lon: 4°28'27.94"). There is a small elevational gradient in the area: Elevation decreases gradually from a dune nature reserve (highest site is located at 4.26–4.50 m above sea level) toward polders consisting of bulb fields and pastures (lowest site is located at 0.49–0.25 m below sea level). The water flows mainly in a southwest direction. The nature reserve is situated in the northwestern part of the polder and is not contaminated from the north and northwest side. Previous research has shown clear differences in

human-affected influences in the area: from lower contamination loadings of agrochemicals at the higher elevation sites neighboring the nature reserve to increased contamination at lower sites located near the agricultural parcels (Ieromina et al. 2015).

### Data collection

A total of 18 sites in the freshwater ditch system were sampled repeatedly in the period April–November 2011–2012 with a time interval of 1–2 months: Ten sites were located in ditches next to flower bulb fields, four ditches next to grasslands, and four sites located in watersheds of the nature reserve close to the flower bulb area. The depth of the ditches was at least 0.7–1 m, selected ditches did not run dry during the year, and water flow was generally low.

Biotic samples were collected using a dipping net dragged over a total length of 5 m using a multi-habitat sampling strategy (Stowa, 2014). All animal specimens collected were taken to the laboratory and identified to the lowest taxonomic level feasible (OTU, hereafter called taxon). These included aquatic macroinvertebrates and small fishes. For summarizing our results, we grouped the taxa in some cases into fishes, insects, crustaceans, mollusks, and other invertebrates.

Floating macrophyte cover (Macroph) was estimated as a proxy for habitat structure. The following abiotic parameters were measured: pH, temperature (Temp, °C), dissolved oxygen (DO, mg/L), and dissolved organic carbon (DOC, mg/L). Measurements of dissolved nutrient concentrations (phosphate ( $\text{PO}_4^{3-}$ ), nitrite ( $\text{NO}_2^-$ ), nitrate ( $\text{NO}_3^-$ ), together indicated as Nutr) and pesticides commonly applied in bulb fields (chlorpropham, pirimiphos-methyl, tolclofos-methyl, carbendazim, ethiofencarb, imidacloprid, isoproturon, imazalil, methiocarb, and prochloraz; together indicated as Pest) were determined in the OMEGAM laboratory (Amsterdam, the Netherlands) using standard protocols. At most locations, pesticide concentrations above the detection threshold were found. Most insecticides have environmental quality standards below detection thresholds and were therefore high enough to expect an induction of effects. Ethiofencarb and methiocarb were excluded from further analyses because the concentrations were equal in all samples. The

temporal changes in all abiotic parameters and the results of tests of differences between months are given in the Supporting Information (Appendix S1: Fig. S1, Table S1).

To differentiate seasonal effects among functional groups, we defined functional group based on feeding mode as retrieved from the online database <http://www.freshwaterecology.info> (accessed in the years 2012–2014) supplemented by literature available through the Web of Science (<http://apps.webofknowledge.com/>). Feeding mode included seven modalities: predating, grazing, shredding, filter feeding, gathering, deposit feeding, and parasite type of feeding. The latter two modalities were only observed in low quantities and therefore not shown visually in our figures for clarity. If a taxon was characterized by one trait modality, this modality was assigned a coefficient of 1, and the other modalities of this trait were assigned 0. If a taxon was characterized by more than one trait modality, each of these modalities was assigned a coefficient ranging from 0 to 1, expressing the relative occurrence of the given modality. For each sample, the relative contribution of each feeding mode was derived by weighting feeding mode estimates for each taxon by their individual biomass (also obtained from <http://www.freshwaterecology.info>) and the abundance of the given taxon within the sample. This weighting avoids unduly impacts of small and rare taxa on community trait expressions and concurs to the biomass ratio hypothesis (Grime 1998). The changes in biomass, abundances, number of taxa, and feeding modalities over time and the results of tests of differences between months are given in Appendix S1: Figs. S2–S5, Tables S2–S4.

#### *Analyses of seasonal data*

Our methodological framework for analyzing differences between seasons includes four basic steps (Fig. 1). Each sample, collected at a given site on a given date, was considered a single observation in our analyses (step 1 in Fig. 1). The number of samples per month over two years is May, 30 biotic and 20 abiotic samples; June, 13 biotic and 9 abiotic samples; July, 28 biotic and 11 abiotic samples; September, 30 biotic and 19 abiotic samples; October, 14 biotic and 8 abiotic samples; and November, 30 biotic and 18 abiotic samples. To identify the most

important drivers that explain the abundance of each taxon in the community, we used recursive partitioning, more specifically random forests (Breiman 2001, step 2 in Fig. 1). Random forests consist of a large number of decision trees. In our case, these were regression trees. Each tree uses a random subset of samples and predictor variables as a learning set. Using random forests instead of a single regression tree prevents overfitting (Breiman 2001, Strobl et al. 2009). Random forests are known to be a superior classification technique (Fernández-Delgado et al. 2014), because they include non-linear relationships between the predictor variables and the response variable (here the abundance of a taxon). In addition, statistical interactions between predictor variables, where the relationship between the predictor variable and the response variable depends on the values of another predictor variable, are included, since on every node it is decided which predictor variable and which value of that predictor should be used for classifying the remaining set of cases (Strobl et al. 2009). To ensure that our analyses were not biased by extremely large abundances for some taxa, and zeroes in others, we applied a Hellinger transformation on all abundances (Borcard et al. 2011).

Our analysis deviates in important aspects from the approach recently outlined by Ovaskainen et al. (2017). First of all, our data do not need to meet the assumptions of regression analyses, such as linearity and normal distribution of residuals. Secondly, Ovaskainen et al. (2017) use a small number of community-level drivers only. Lastly, Ovaskainen et al. (2017) use time series. A benefit of this approach is that the causality of the interactions can be convincing and intraspecific interactions are included. We do not explicitly consider time series in our analysis. Instead, our approach allows for estimating interspecific interactions based on the co-occurrence of the taxa at different locations at one moment in time, thus enabling the study of changes in interactions over time.

The efficiency of a random forest to predict the abundance of a taxon may depend on the ability of the method to distinct random effects from real effects and on the number of samples it has as a learning set. To study the relationship between the number of samples and the

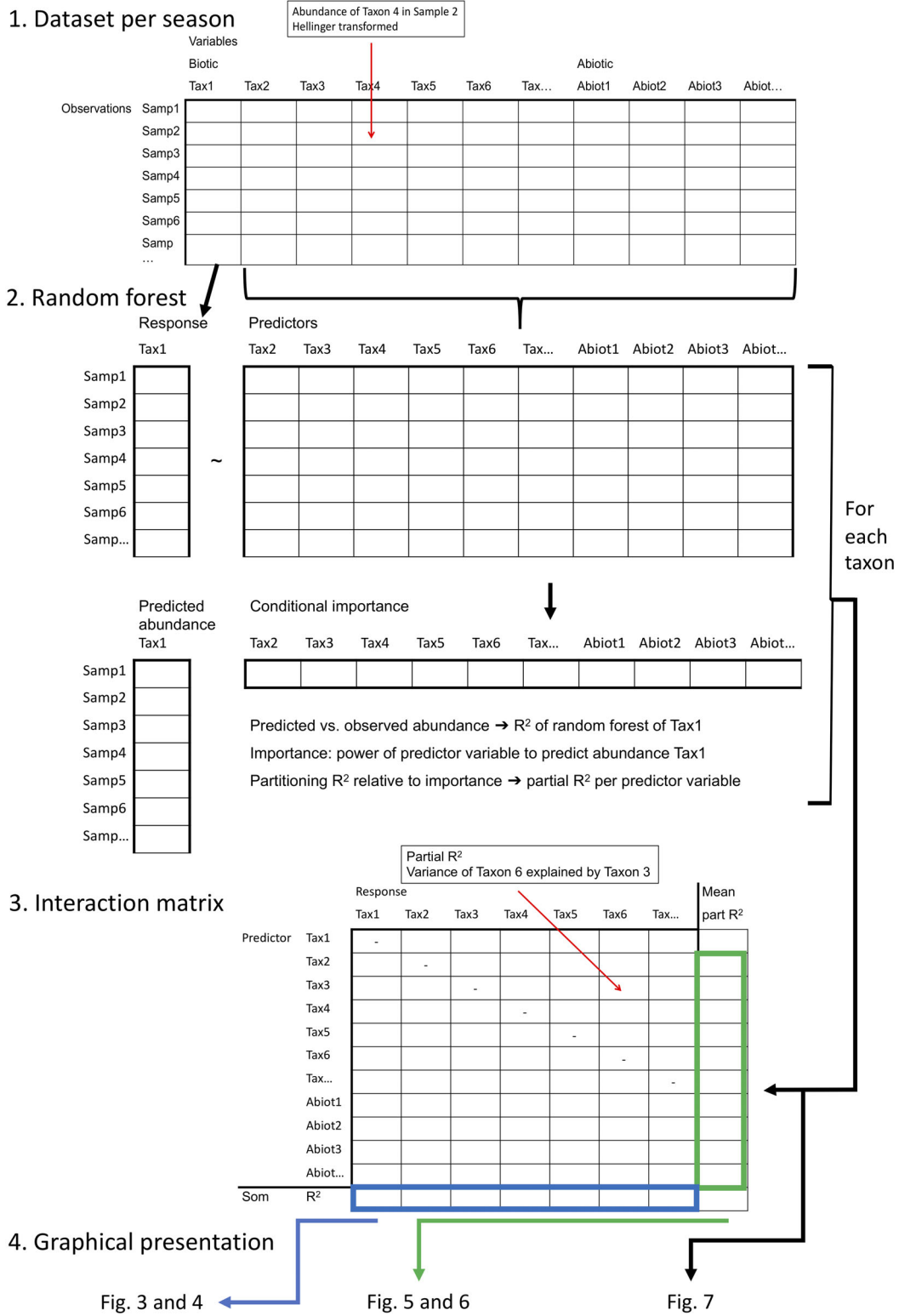


Fig. 1. Flowchart of analysis per season. See text for extensive explanation.

outcomes of our data analysis, we used a present-absent dataset of biotic samples from different locations in our research area collected in two other years (viz. 2013–2014), in a limited period of time within the year (viz. April–May). Each location was sampled once a year. This analysis (described in Appendix S2: Fig. S1) shows that our random forest analysis indeed depends on sample size and that the sample size should be higher than 25 to attain predictor variables that have predictive power higher than zero (Musters and van Bodegom 2018). It also shows that our data analysis does not predict any taxon to be present with a chance of deviating from zero in case of a randomized dataset of 206 samples. The constraint on the number of samples limited the temporal resolution of our analysis. As an alternative for an analysis per month, we created a moving window, that is, we combined the samples of three sequential months, pooled over two years, which yields four points in time, four seasons, for our analysis: spring (May, June, July; 40 samples), spring–summer (June, July, September; 39 samples), summer–autumn (July, September, October; 38 samples), and autumn (September, October, November; 45 samples). We consider pooling over two years justified, because we have no reasons to believe that the shift in the relative importance of community processes will be different between the years. However, we added year to our predictor variables to correct for differences between years. Our moving average approach shows the changes between seasons, but cannot be used to statistically test those differences, because the observations per season are overlapping, and therefore, the results per season are not statistically independent. We ignored the slight difference in sample size, after checking that this did not affect the interpretation of our results when focusing on changes over time (Appendix S3: Fig. S1).

The predictor variables per season were the abundances of all taxa minus the taxon of which the abundance was to be predicted, all chemistry variables, taxon richness per sample, land use of sample site (flower bulb growing, grassland, or nature reserve), and month and year of sampling (2011 or 2012), resulting in 134 predictor variables in spring, 132 in spring–summer, 126 in summer–autumn and 123 in autumn. Missing values were imputed with the function `rfImpute()` of the

`randomForest` package of R (Cutler et al. 2007) using 500 trees and 5 iterations. This function replaces the missing value with the value of the, according to the random forest, closest other sample.

The difference between the predicted abundance of each taxon in each sample based on all predictor variables as determined by the random forests and the actual abundance can be regarded as a residual. Hence, an analogue of  $R^2$ , that is, the proportion of the variance explained by the random forest, hereafter called  $R^2$ , can be calculated (Ellis et al. 2012). Furthermore, the power of a predictor variable to predict a taxon's abundance can be estimated by calculating its importance. This is done by comparing the prediction accuracy of random forests with the actual vs. randomly permuted values of that predictor variable. The decrease in prediction accuracy is a measure of the importance of that predictor (Strobl et al. 2009). However, it has been shown that this measure is biased toward correlated predictors (Strobl et al. 2007). Therefore, the alternative measure of importance called the conditional importance and developed to solve this problem by conditional permutation, was used, that is, the importance under the condition that all other predictor variables have a constant value (Strobl et al. 2008). We regard each conditional importance of a taxon for predicting another taxon as an estimate of the unidirectional interaction between the two taxa. So, the conditional importance of taxon A found in the random forest of taxon B can be regarded as the predictive power of taxon A for predicting the abundance of taxon B. The opposite direction, the predictive power of taxon B for predicting the abundance of taxon A, is the conditional importance of taxon B found in the random forest of taxon A. The contributions of the predictor variables per taxon can be averaged over all taxa, that is, over the complete community (Ellis et al. 2012). This procedure assumes that the sampling sites differ in their values of predictor variables and abundance of the taxa, but that each taxon can be present at any of the sampling sites, and that the effects of dispersal and order of arrival can be ignored (Little and Altermatt 2018). Indeed, our research area is so small and the sampling sites are so well connected, that the distribution of our taxa is not spatially delimited

(Viana et al. 2015, Musters et al. 2019). The conditional random forests and importance were calculated with the party package of R (Hothorn et al. 2006, Strobl et al. 2007, 2008). All default settings were kept, except for the number of trees, which was set to 500 and the number of predictor variables tested per node, which was set to the square root of the total number of predictors as recommended by Strobl et al. (2009). Each random forest was run 10 times in order to assess the variability in the outcomes due to the random procedures.

Next, for each taxon of the community, the conditional importance of each predictor variable can be expressed as the part of the  $R^2$  of that taxon that is explained by that predictor variable (Ellis et al. 2012), which we will call the partial  $R^2$  (step 3 in Fig. 1). We used the formulas (1) to (3) of Ellis et al. (2012) for calculating  $R^2$  per predicted taxon and the partial  $R^2$  per predictor variable. Negative values of both can occur due to the random procedures of random forests (Strobl et al. 2009, Musters and van Bodegom 2018). The minimum values of the negative  $R^2$  are indicative for the range within which the random forest has no predictive power. So, we replaced all  $R^2$  and partial  $R^2$  lower than the absolute minimum value by zero.

Finally, after analysis, to illustrate and summarize our results (step 4 in Fig. 1), we summed the partial  $R^2$  of the predictor variables within the following groups of variables: Biotics (all taxa, taxonomic richness, macrophytes), Nutrients (phosphorus, nitrite, nitrate), Pesticides (all pesticides), Other chem (dissolved oxygen [DO], Dissolved organic carbonates [DOC], pH), and Other abiotics (land use, temperature, month, and year). Since these groups differ in the number of predictor variables, the summed partial  $R^2$  per group will also differ, but this is no problem because we are interested in the effect of the main drivers, that is, the total effect of the group of predictor variables, on the community. All analyses were conducted in R 3.3.2 (R Development Core Team 2017).

## RESULTS

Taxonomic alpha diversity hardly changed over the seasons, while gamma diversity decreased slightly (Fig. 2a). A decrease in gamma diversity over the months was found in

both sampling years (Appendix S1: Fig. S5b). Taxonomic composition changed in that the number of fishes and insects decreased and mollusks increased over the months (Appendix S1: Fig. S5c). For 54–72 taxa (around 60% of the taxa), the random forest had a predictive power ( $R^2$ ) higher than zero (Fig 2b). We further refer to these as the predicted taxa. These are taxa of which the abundance is predicted by at least one of the predictor variables, being both the other taxa and the abiotic factors. The number of predicted taxa was lowest in summer–autumn (Fig. 2b). About 25% of the taxa had a partial  $R^2$  for predicting at least one other taxon higher than zero. We refer to these as the predictor taxa. These are taxa of which the abundance predicts at least one other taxon. The highest  $R^2$  found among the predicted taxa increased over the seasons (Fig 2c).

The best explained taxa varied by season (Fig. 3): Crustaceans are well explained in autumn and fish in spring and summer–spring, while mollusks are best explained in all seasons (Fig. 4a). Divided into functional groups, the grazers (dominated by mollusks) were best explained, and increasingly so over the seasons. The filter feeders and the shredders were not as well explained, but the differences between functional groups were small (Fig. 4b).

The explanatory power of different sets of predictor variables for the community as a whole, expressed as partial  $R^2$ , changed over the seasons (Fig. 5). In all cases, land use was a dominant predictor, that is, within the top 5 of predictors, as was one of the time variables (year or month). Further, of the nutrients, phosphate was the most important predictor. Only in spring one pesticide appeared as a dominant predictor, namely the fungicide carbendazim. DOC, DO, and pH were important in spring and remained relatively important over the seasons, though less so than in spring.

The importance of different groups of predictors changed over the seasons (Fig. 6a; Appendix S3: Fig. S2 presents the same information, but without grouping the partial  $R^2$  of most abiotic drivers). The set of biotic predictors, that is, species interactions, were relatively important for explaining the taxa abundances, and this importance increased over time. Nutrients were less important as a predictor, but their

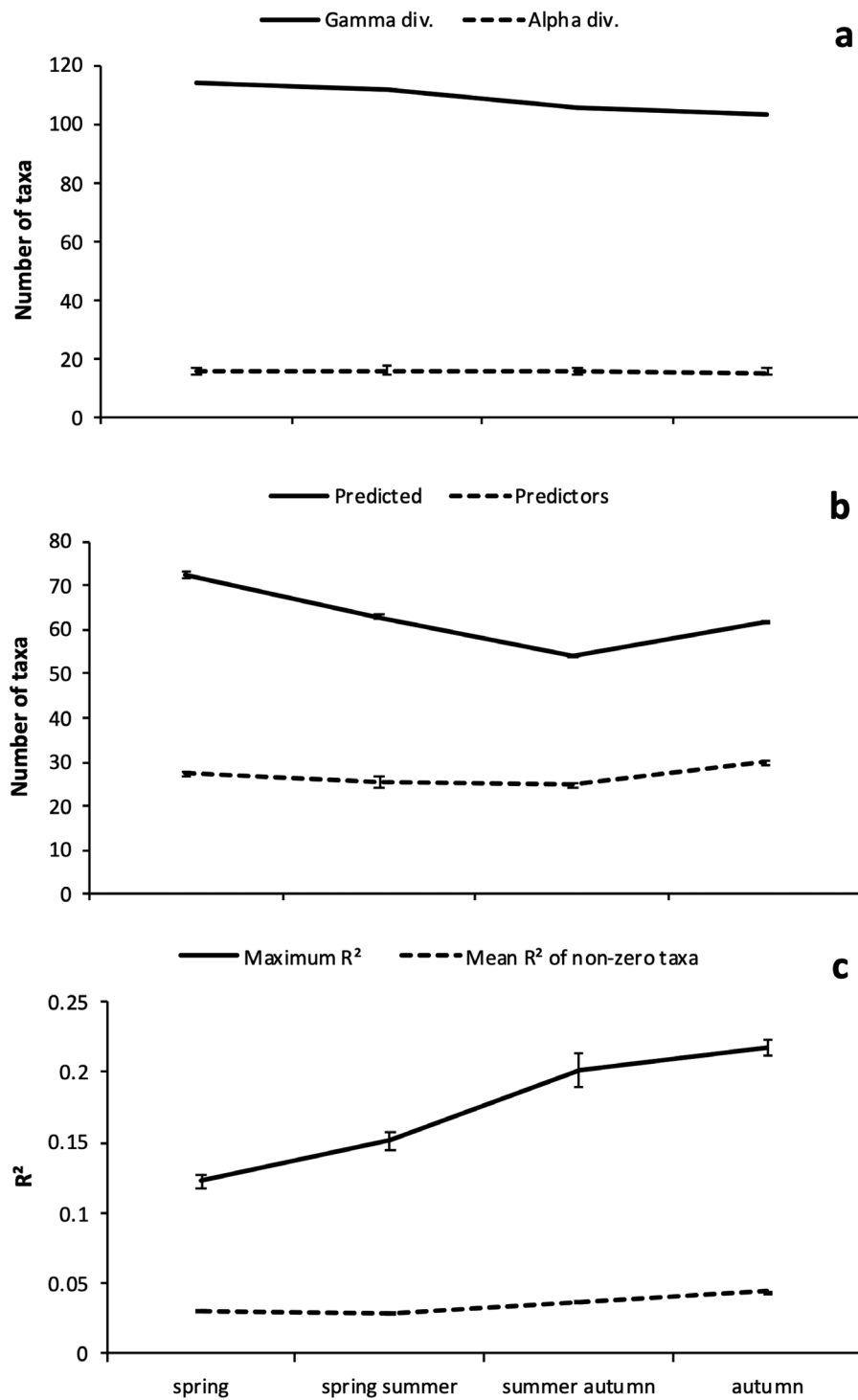


Fig. 2. (a) Taxonomic diversity; (b) predicted taxa (taxa that had an explained  $R^2 > 0$ ) and predictors (taxa that had a partial  $R^2 > 0$  for predicting the abundance of at least one other taxon); (c) the mean  $R^2$  and the highest  $R^2$  of predicted taxa. The error bars represent the 95% confidence interval based on 10 sets of random forests.



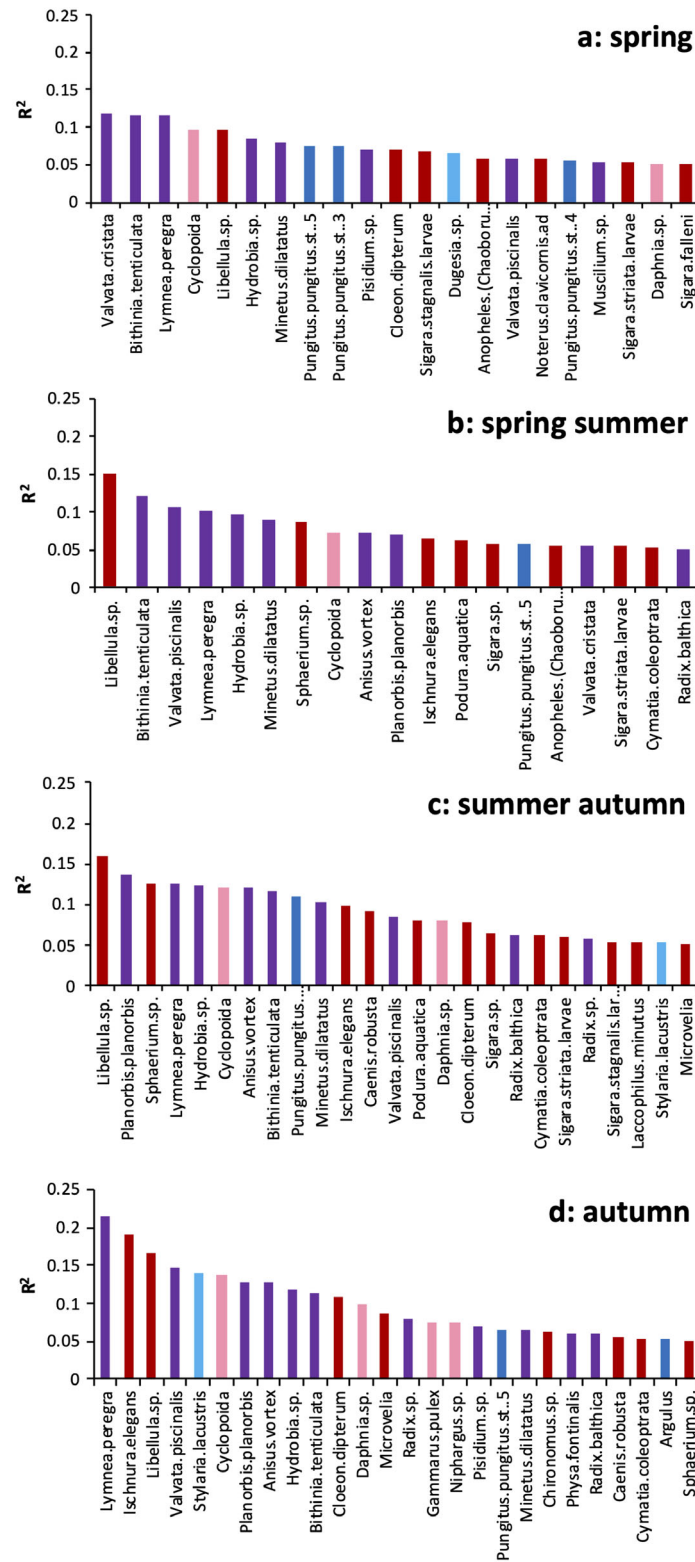


Fig. 3. The mean explained variance of the abundance of the taxa with an  $R^2 > 0.05$  in the four seasons. The  $R^2$

(Fig. 3. *Continued*)

was averaged over 10 conditional random forests. Colors indicate the taxonomic group: blue: fish; red: insects; pink: crustacea; purple: mollusks; and light blue: other taxa.

importance also increased over time. Pesticides were relatively unimportant and decreased in importance over time. The other chemical and abiotic factors remained relatively important over time. In general, this is also true when looking at the separate taxonomic groups and feeding mode (Appendix S3: Figs. S3, S4).

To show the relative importance of groups of taxa in the species interactions, we summarized

the partial  $R^2$  of the taxa, grouped by taxonomy and by feeding mode (Fig. 6b, c, respectively). Insects and mollusks as groups were important predictors of the taxa abundances and the importance of the mollusks as predictors increased over time. Grazers, which are mainly mollusks (more specifically gastropods), were clearly the most important functional group of the community and its importance increased over the season

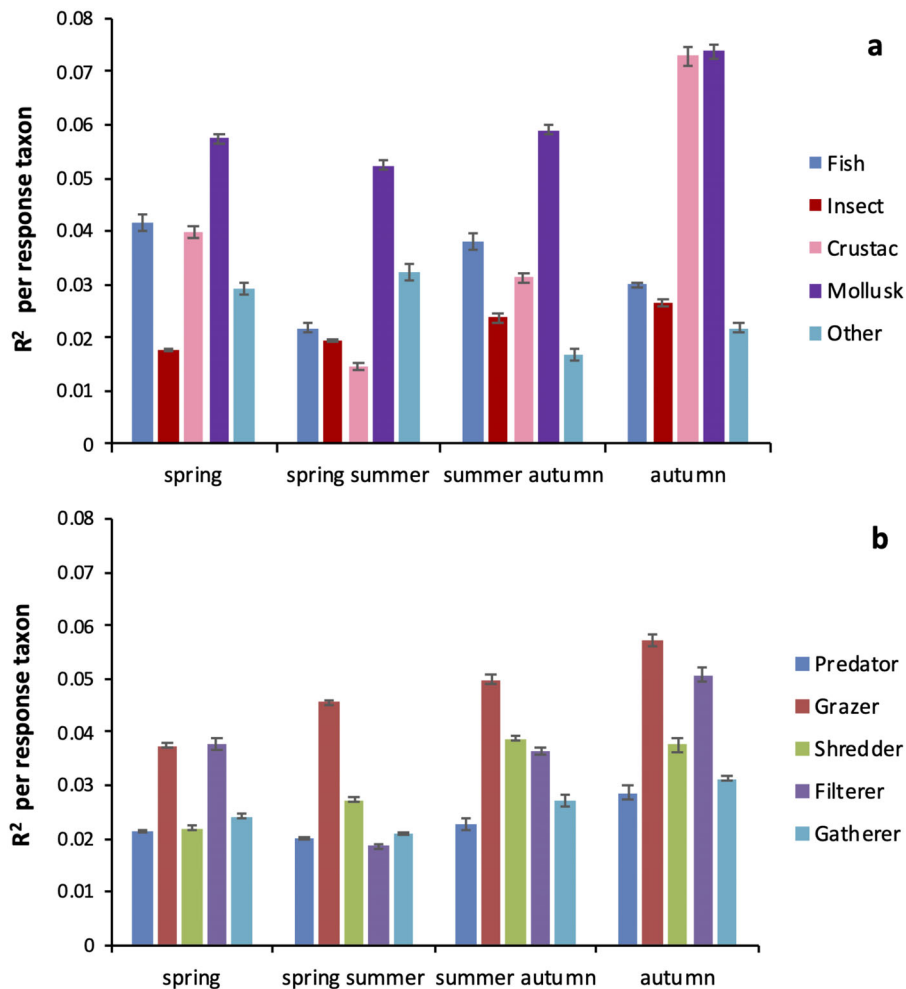


Fig. 4. Seasonal change in average variance explained of the abundance of taxa according to their taxonomic group (a) and feeding mode (b). The error bars are the 95% confidence interval based on 10 sets of conditional random forests.

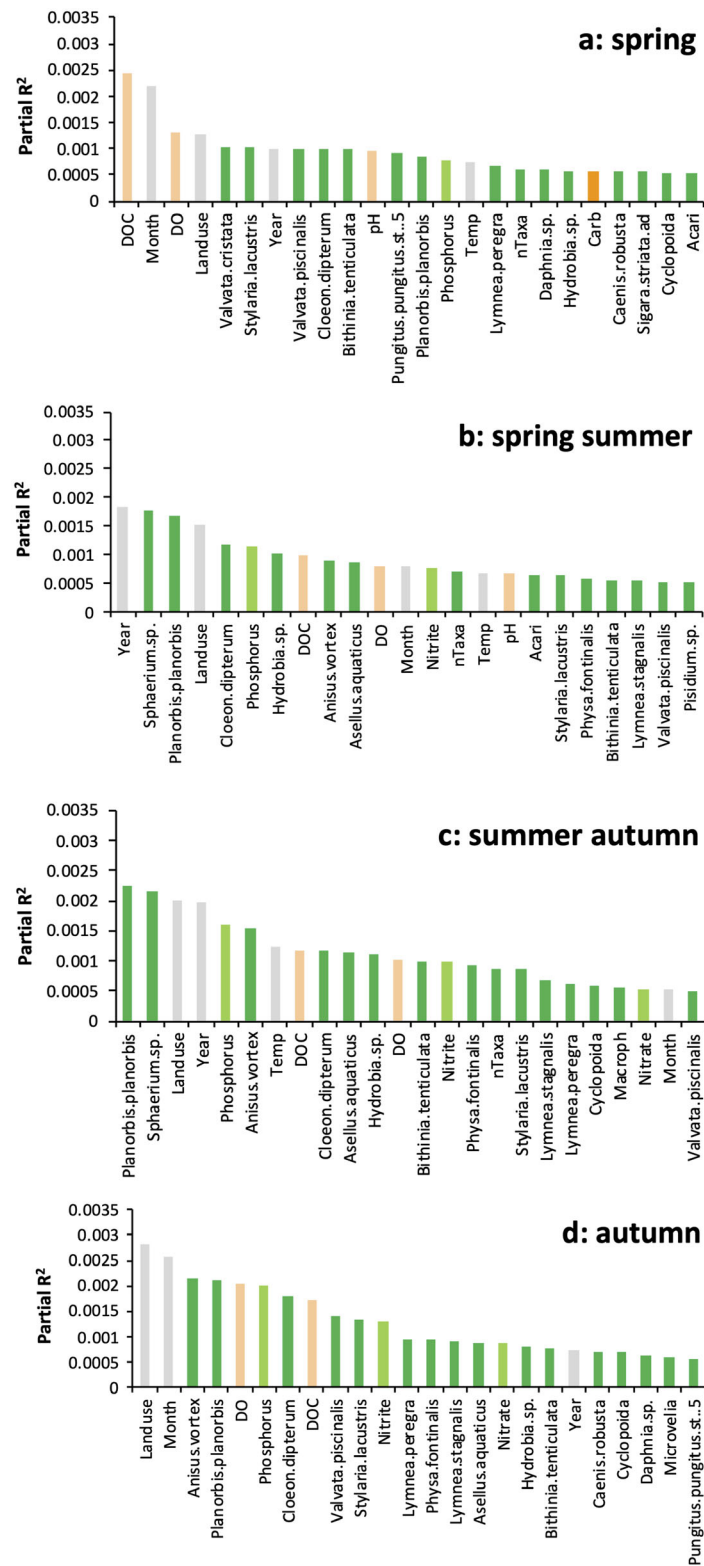


Fig. 5. The mean partial R<sup>2</sup> of the predictor variables with a partial R<sup>2</sup> > 0.0005. The partial R<sup>2</sup> was averaged

(Fig. 5. *Continued*)

over 10 conditional random forests. Colors indicate the predictor group: green, biotics; light green, nutrients; orange, pesticides; light orange, other chemicals; and gray, other abiotics.

(Fig. 6c). The important role of the grazers in explaining abundances of other functional groups is illustrated in the interaction webs in Fig. 7. The interaction webs illustrate that grazers were the most important predictors for all functional groups, including the grazers themselves, which means that some grazing taxa predicted the abundance of other grazing taxa. Filter feeders predict themselves partly only in spring. No clear trends were visible in the other groups but the predators stood out in being unimportant predictors, especially in spring–summer and summer–autumn, and they hardly predict themselves (Fig. 7).

Because the importance of a group of predictor variables highly depends on the number of variables per group—the partial  $R^2$  are summed—we present the average partial  $R^2$  per groups of predictor variables in the Supplementary Information (Appendix S3: Fig. S5). These results show that an average biotic predictor was relatively unimportant as compared to a nutrient, a chemical other than a pesticide and other abiotic predictors (Appendix S3: Fig S5a).

## DISCUSSION

In this study, we developed and present an approach based on random forests to analyze the seasonal change in an aquatic community in Dutch ditches to partition the effects of a multitude of driving factors on the abundance of each of the taxa in the community. Our results show how the predictive power of drivers of aquatic ditch communities changes within the agricultural growing season: Early in the season, the community is already regulated by species interactions, and as the season progresses, it transitions toward being more strongly regulated by a combination of species interactions and nutrients later in the season. In other words, there is no strict switch between processes, from dominated by population growth toward dominated by species interactions, across seasons but rather an intensification of bottom-up control. This could

be a consequence of our approach of working with overlapping seasons.

Already in spring, we find that the community, in terms of taxon abundances, is mainly explained by the abundances of co-occurring taxa. We regard this as indicating direct and indirect interspecific interactions, although we should stress that co-occurrence may also be explained by common drivers among taxa. Our result is in accordance though with the results of a study into the relative importance of the environment vs. species interactions in marine plankton (Lima-Mendez et al. 2015). The importance of pesticides for explaining the abundances of taxa, that is, the composition of the community, is low, even in spring—at the peak of their importance. The low impact of pesticides can be regarded as surprising given the high and persistent pesticide loads in parts of the area due to the flower bulb industry (Hunting et al. 2016, Barmantlo et al. 2018; <http://www.pesticidesatlas.nl>), although it may also suggest that the community is well-adapted to pesticides due to long-term exposure, so that very sensitive taxa may be missing from the area. The decrease in importance over the seasons up to early autumn is also exhibited in the other chemical predictors, namely pH, DO, and DOC. When lumping pesticides and the other chemicals, the suggestion that these drivers become less important over the seasons is even stronger (Appendix S3: Fig. S6). The reverse is true for the importance of other taxa as predictors, which increases over the season, suggesting that the importance of interactions between taxa increases over time. This is also true for the importance of the nutrients (Fig. 6a.; Appendix S3: Fig. S2). The increase in average and highest  $R^2$  per taxon over time (Fig. 2) as well as the decreased dissimilarity between samples (Musters et al. 2019) coincides with this transition, reinforcing the idea that the community becomes more strongly regulated over time, either by life histories of taxa, nutrients, or both.

Concerning the predictive power of taxa, we find a marked difference in the importance of the

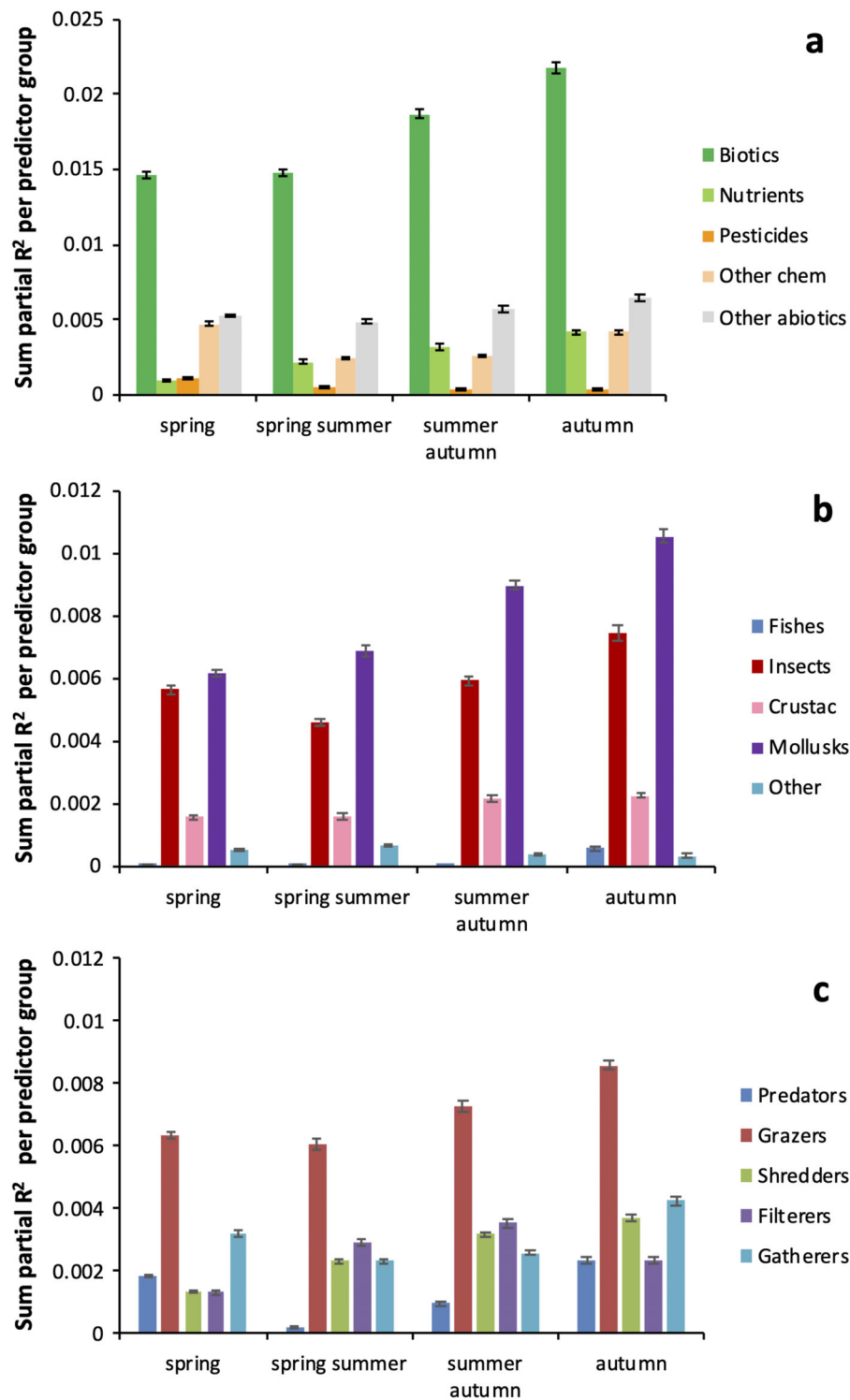


Fig. 6. Seasonal change in community importance of the groups of predictor variables (a), and taxa according to their taxonomic group (b) and feeding mode (c). The error bars are the 95% confidential interval based on 10 sets of conditional random forests.

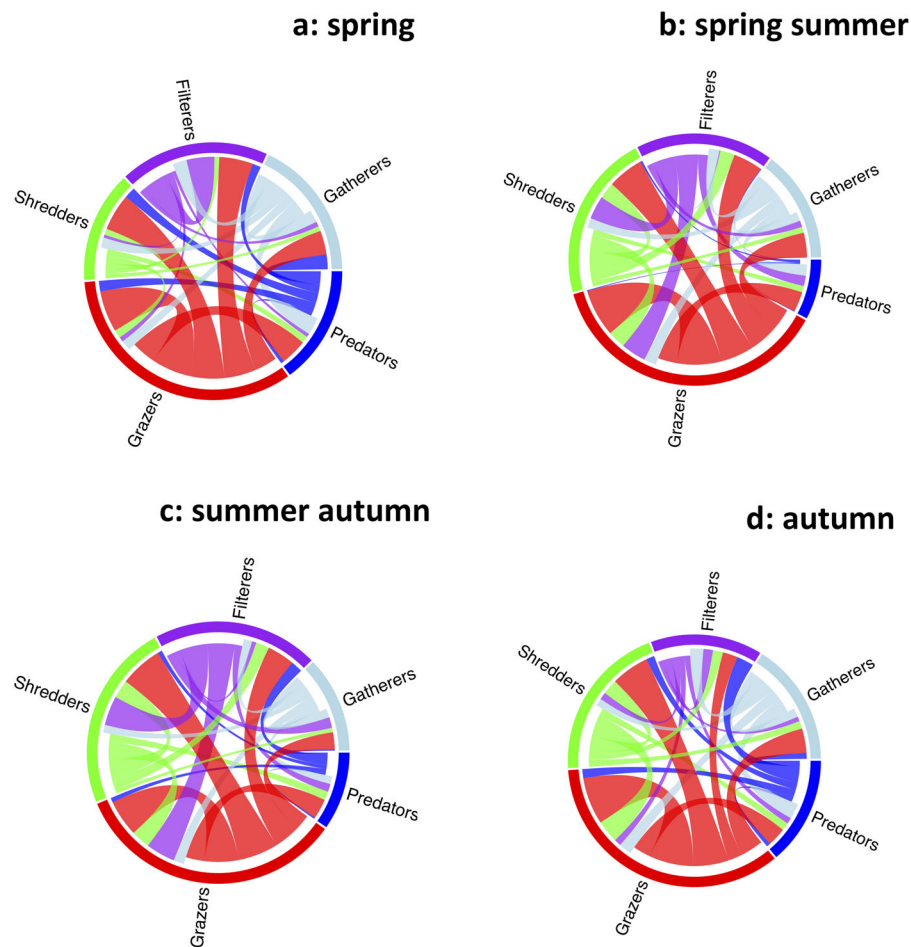


Fig. 7. Interactions between functional groups per season. Indented: predicting. Non-indented: predicted.

different functional groups. Predators, which constitute on average 6.7% of the total biomass, had the lowest explanatory power (Fig. 7). This observation, together with the fact that the relative biomass of predators consistently decreases over time (Appendix S1: Fig. S4), suggests that the change toward stronger regulation of the community over time is not due to an increasing importance of top-down regulation. Instead, the community seems to be bottom-up regulated, with grazed periphyton and epiphytes at its base. The grazers in this system (mainly mollusks, dominated by gastropods) become increasingly important over the season. The importance of shredders and filter feeders, which is less than that of the grazers, also seems to increase over time, at least up until early autumn (Fig. 6c). This

might coincide with an increase in importance of the crustaceans. Insects form an important group, in terms of the number of taxa in this group, but on average an insect taxon is a less important predictor than an average mollusk or crustacean (Appendix S3: Fig. S5b). Further, as the season progresses, nutrients become more important as predictors, which again indicates an increase over time of the importance of bottom-up processes for the community.

Although the abundance of ~60% of the taxa is predicted by the random forests, only ~25% of the taxa participate in these predictions (Fig. 2b). This suggests that only a limited number of taxa shape the community and that about 40% of the taxa are governed either by drivers that are not captured in our predictor variables or by neutral processes,

that is, by migration or reproduction that are independent of abiotic factors or interactions with other species (Hubbell 2001). However, as was stated by Pimm (2002), the better one looks at a community, the more interactions one sees. Or, in statistical terms, the number of predicted taxa and predictor taxa will depend on the sample size of the study (Appendix S3: Fig. S2.1b), so that these percentages will probably be study-specific. Digging deeper into the interspecific interactions, we find that horizontal interaction is a dominant form of species interaction. This is most obvious for the grazers and the shredders—both are strongly predicted by abundances of other taxa within the same functional group (Fig. 7)—but less so for the other functional groups. Competition is the obvious type of interaction to think of, and the importance of this type of interaction increases over time. Stronger competition is supposed to lead to lower species richness (Terborgh 2015), of which we observe a signal in the decrease in gamma diversity (Fig. 2). Our results also show that other types of interactions occur. Fish, for example, are explained quite well, but do not predict the abundances of other taxa (Fig. 4a vs. 6b). The same can be said about predators (Fig. 4b vs. 6c). The abundance of the taxa in these groups is thus explained by vertical interactions while competition does not seem to play a significant role for these groups. Interestingly, filter feeders show clear competition in spring only, which indicates that the food of filter feeders, algae and small organic particles, is relatively abundant in the other seasons.

This all being said, we must acknowledge that the  $R^2$  per taxon was usually low and never exceeded 0.21 (Fig. 3). Low percentages of explained variance are quite common in studies of macroinvertebrates in aquatic systems (Leslie and Lamp 2017, Little and Altermatt 2018). As a consequence, partial  $R^2$ s are also small, even though they are non-zero, so that the differences we found between seasons are actually small. The warning of Carpenter et al. (1985) against using statistics to find cause-and-effect relationships is still valid. For example, our method based on co-occurrence is unable to detect intraspecific interactions, but may reflect indirect interactions between species in communities (Chase and Leibold 2003, Vellend 2016), and may miss time lags between predictor and response

variables (Evans et al. 2018). However, random forests at least give us the opportunity to quantify relationships between taxa abundances and between taxa abundances and environmental factors in one analysis, which is relatively new in ecology. Hence, our results should not be regarded as more than indications for the interpretations that we present here.

Our reservations aside, the proposed approach is promising as it allows identifying seasonal dynamics in the role of drivers governing temperate freshwater communities. We observed an intensification of bottom-up control by increased effects of nutrients and species interactions on taxa abundances over the growing season. This observed shift is particularly relevant since it is not considered in studies assessing the hazards and risks of anthropogenic stressors. As such, the timing of shifts may create opportunities for management. For example, our results suggest that for the conservation of aquatic invertebrate communities, pesticides should at least not be applied in early spring. Our results highlight the need for efforts to monitor and better quantify and understand the diversity and health of the environment accounting for the dynamic nature of communities and their drivers.

## ACKNOWLEDGMENTS

The authors are grateful to B. Schaub of Water Board Rijnland for his help, E. Gertenaar for assistance in the fieldwork, M. Wouterse for DOC measurements, and B. Koese for help with taxonomic identification of macrofaunal samples. CM designed the study, did the statistical modeling and analyses, and wrote the draft paper; OI did field sampling and taxonomic identification and constructed the datasets; OI and HB structured the data; EH, MS, ES, MV, and PvB contributed to the study design and the conceptual improvement of the manuscript; all authors substantially revised the subsequent drafts.

## LITERATURE CITED

Barmantlo, S. H., M. Schrama, E. R. Hunting, R. Heutink, P. M. van Bodegom, G. R. de Snoo, and M. G. Vijver. 2018. Assessing combined impacts of agrochemicals: aquatic macroinvertebrate population responses in outdoor mesocosms. *Science of the Total Environment* 631–632:341–347.

- Barmantlo, S. H., L. M. Vriend, R. A. H. van Grunsven, and M. G. Vijver. 2019. Environmental levels of neonicotinoids reduce prey consumption, mobility and emergence of the damselfly *Ischnura elegans*. *Journal of Applied Ecology*, <https://doi.org/10.1111/1365-2664.13459>
- Borcard, D., F. Gillet, and P. Legendre. 2011. *Numerical ecology* with R. Springer, New York, New York, USA.
- Breiman, L. 2001. Random forests. *Machine Learning* 45:5–32.
- Carpenter, S. R., J. F. Kitchell, and J. R. Hodgson. 1985. Cascading trophic interactions and lake productivity. *BioScience* 35:634–639.
- Carpenter, S. R., J. F. Kitchell, J. R. Hodgson, P. A. Cochran, J. J. Elser, M. M. Elser, D. M. Lodge, D. Kretchmer, X. He, and C. N. von Ende. 1987. Regulation of lake primary productivity by food web structure. *Ecology* 68:1863–1876.
- Chadd, R. 2010. Assessment of aquatic invertebrates. Pages 63–72 in C. Hurford, M. Schneider and I. Cowx, editors. *Conservation monitoring in freshwater habitats (practical guide and case studies)*. Springer, New York, New York, USA.
- Chase, J. M., and M. A. Leibold. 2003. *Ecological niches. Linking classical and contemporary approaches*. The University of Chicago Press, Chicago, Illinois, USA.
- Cutler, D. R., T. C. Edwards Jr., K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007. Random forests for classification in ecology. *Ecology* 88:2783–2792.
- Ellis, N., S. T. Smith, and C. R. Pitcher. 2012. Gradient forests: calculating importance gradients on physical predictors. *Ecology* 93:156–168.
- Evans, T. R., D. Salvatore, M. van de Pol, and C. J. M. Musters. 2018. Adult firefly abundance is linked to weather during the larval stage in the previous year. *Ecological Entomology* 44:265–273.
- Fernández-Delgado, M., E. Cernadas, and S. Barro. 2014. Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning* 15:3133–3181.
- Grime, J. P. 1998. Benefits of plant diversity to ecosystems: immediate, filter and founder effects. *Journal of Ecology* 86:902–910.
- Halstead, N. T., T. A. McMahon, S. A. Johnson, T. R. Raffel, J. M. Romansic, P. W. Crumbrine, and J. R. Rohr. 2014. Community ecology theory predicts the effects of agrochemical mixtures on aquatic biodiversity and ecosystem properties. *Ecology Letters* 17:932–941.
- Hill, M. J., C. D. Sayer, and P. J. Wood. 2016. When is the best time to sample aquatic macroinvertebrates in ponds for biodiversity assessment? *Environmental Monitoring Assessment* 2016:188–194.
- Hothorn, T., P. Buehlmann, S. Dudoit, A. Molinaro, and M. van der Laan. 2006. Survival ensembles. *Biostatistics* 7:355–373.
- Hubbell, S. P. 2001. *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton, New Jersey, USA.
- Hunter, M. D., and P. W. Price. 1992. Playing chutes and ladders: heterogeneity and the relative role of bottom-up and top-down forces in natural communities. *Ecology* 73:724–732.
- Hunting, E. R., J. A. Vonk, C. J. M. Musters, M. H. S. Kraak, and M. G. Vijver. 2016. Effects of agricultural practice on organic matter degradation in ditches. *Scientific Reports* 6:21474.
- Ieromina, O., W. J. G. M. Peijnenburg, C. J. M. Musters, and M. G. Vijver. 2015. The contribution of pesticides to the variance in community composition of aquatic macrofauna in the field. *Basic and Applied Ecology* 17:125–133.
- Kéfi, S., E. L. Berlow, E. A. Wieters, L. N. Joppa, S. A. Wood, U. Brose, and S. A. Navarrette. 2015. Network structure beyond food webs: mapping nontrophic and trophic interactions on Chilean rocky shores. *Ecology* 96:291–303.
- Lancaster, J. J., and M. E. Ledger. 2015. Population-level responses of stream macroinvertebrates to drying can be density-independent or density-dependent. *Freshwater Biology* 60:2559–2570.
- Leslie, A. W., and W. O. Lamp. 2017. Taxonomic and functional group composition of macroinvertebrate assemblages in agricultural drainage ditches. *Hydrobiologia* 787:99–110.
- Lima-Mendez, G., et al. 2015. Determinants of community structure in the global plankton interactome. *Science* 348:1262073-1–1262073-9.
- Little, C. J. and F. Altermatt. 2018. Do priority effects outweigh environmental filtering in a guild of dominant freshwater macroinvertebrates? *Proceedings of the Royal Society B* 285:20180205.
- Musters, C. J. M., and P. M. van Bodegom. 2018. Analysis of species attributes to determine dominant environmental drivers, illustrated by species decline in the Netherlands since the 1950s. *Biological Conservation* 219:68–77.
- Musters, C., J. M., E. R. Hunting, M. Schrama, E. Cieraad, S. H. Barmantlo, O. Ieromina, M. G. Vijver, and P. M. van Bodegom. 2019. Spatial and temporal homogenization of freshwater macrofaunal communities in ditches. *Freshwater Biology*. <https://dx.doi.org/10.1111/fwb.13415>
- Odum, E. P. 1969. The strategy of ecosystem development. *Science* 164:262–270.



- Ormerod, S. J., M. Dobson, A. G. Hildrew, and C. R. Townsend. 2010. Multiple stressors in freshwater ecosystems. *Freshwater Biology* 55(Suppl. 1):1–4.
- Ovaskainen, O., G. Tikhonov, D. Dunson, V. Grøtan, S. Engen, B. Sæther, and N. Abrego. 2017. How are species interactions structured in species-rich communities? A new method for analyzing time-series data. *Proceedings of the Royal Society B: Biological Science* 284:20170768.
- Pimm, S. L. 2002. *Food webs*. With a new foreword. The University of Chicago Press, Chicago, Illinois, USA.
- R Development Core Team. 2017. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>
- Scheffer, M., S. H. Hosper, M.-L. Meijer, B. Moss, and E. Jeppesen. 1993. Alternative equilibria in shallow lakes. *TREE* 8:275–279.
- Schwarzenbach, R. P., B. I. Escher, K. Fenner, T. B. Hofstetter, C. A. Johnson, U. von Gunten and B. Wehrli. 2006. The challenge of micropollutants in aquatic systems. *Science* 313:1072–1077.
- Shurin, J. B., D. S. Grune, and H. Hillebrand. 2006. All wet or dried up? Real differences between aquatic and terrestrial food webs. *Proceedings of the Royal Society B: Biological Science* 273:1–9.
- Sommer, U., Z. M. Gliwicz, W. Lampert, and A. Duncan. 1986. The PED\*-model of seasonal succession of planktonic events in fresh water. *Archiv für Hydrobiologie* 106:433–471.
- Stowa. 2014. *Handboek Hydrobiologie*. [http://handboekhydrobiologie.stowa.nl/Het\\_Handboek/Het\\_Handboek.aspx](http://handboekhydrobiologie.stowa.nl/Het_Handboek/Het_Handboek.aspx)
- Strobl, C., A.-L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis. 2008. Conditional variable importance for random forests. *BMC Bioinformatics* 9:307.
- Strobl, C., A.-L. Boulesteix, A. Zeileis, and T. Hothorn. 2007. Bias in Random Forest Variable Importance Measures: illustrations. Sources and a Solution. *BMC Bioinformatics* 8:25.
- Strobl, C., J. Malley, and G. Tutz. 2009. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* 14:323–348.
- Terborgh, J. W. 2015. Toward a trophic theory of species diversity. *Proceedings of the National Academy of Sciences of USA* 112:11415–11422.
- Vellend, M. 2016. *The theory of ecological communities*. Monographs in population biology 57. Princeton University Press, Princeton, New Jersey, USA.
- Verdonschot, R. C. M., H. E. Keizer-Vlek, and P. F. M. Verdonschot. 2011. Biodiversity value of agricultural drainage ditches; a comparative analysis of the aquatic invertebrate fauna of ditches and small lakes. *Aquatic Conservation: Marine and Freshwater Ecosystems* 21:715–727.
- Viana, D. S., et al. 2015. Assembly mechanisms determining high species turnover in aquatic communities over regional and continental scales. *Ecography* 38:1–8.
- Wetzel, R. G. 2001. *Limnology*. Lake and river ecosystems. Third edition. Academic Press, San Diego, California, USA.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online at: <http://onlinelibrary.wiley.com/doi/10.1002/ecs2.2910/full>

Appendix S1: Analyses of monthly data.

Appendix S2: Effect of sample size and randomization.

Appendix S3: Analyses of seasonal data.