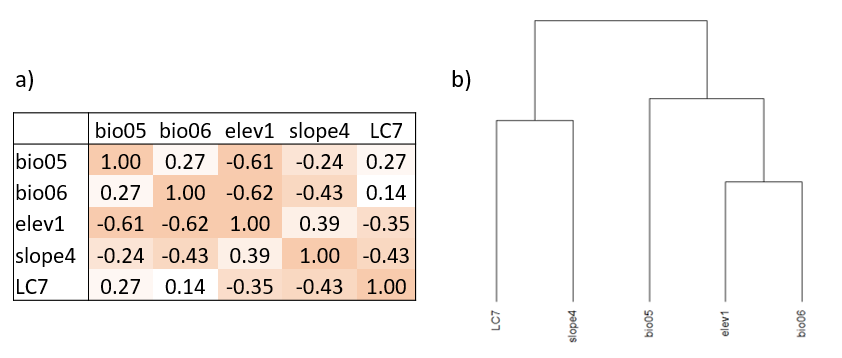


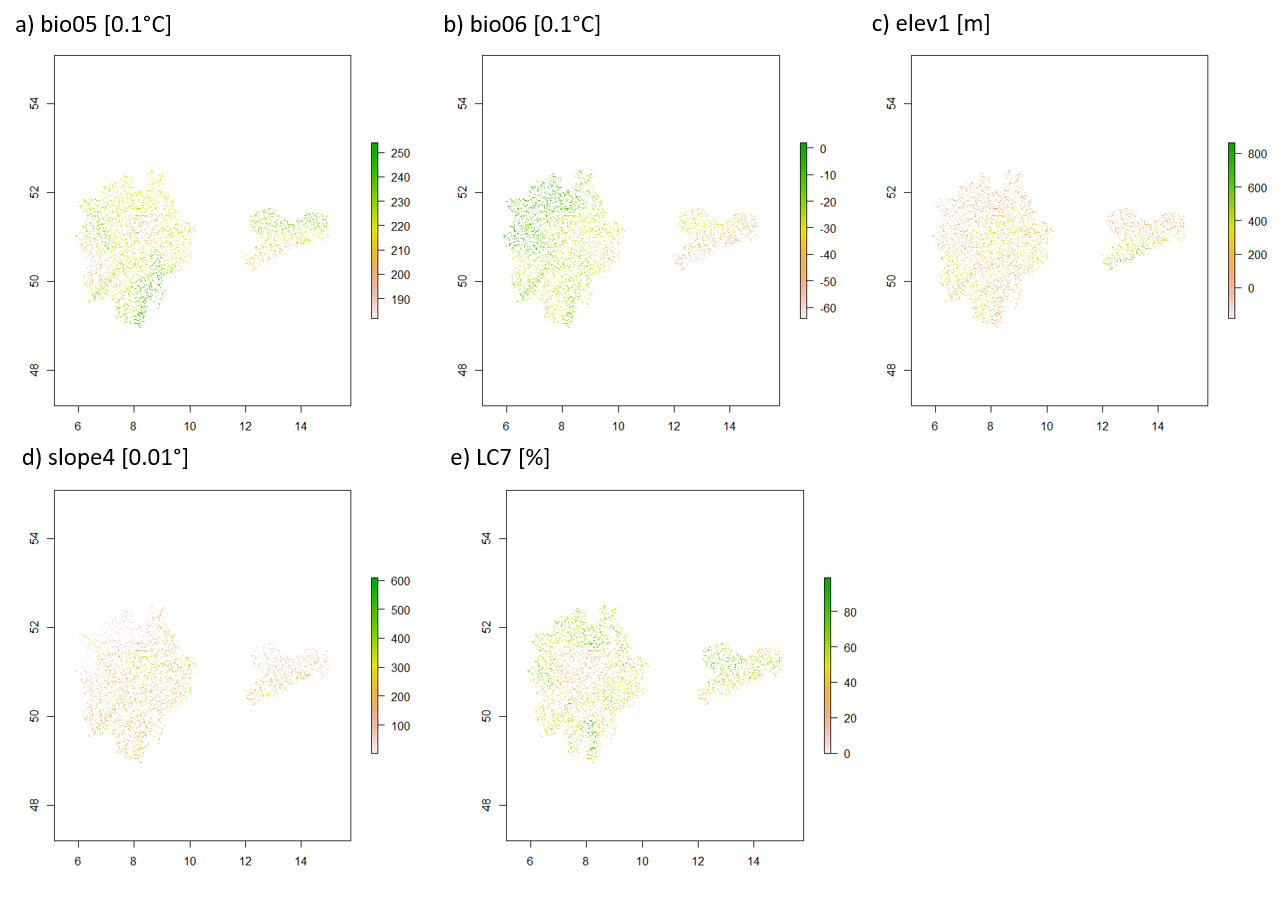
**Figure S1:** GBIF occurrences of the genus *Simulium* (Latreille, 1802) from Germany. Data presented according to the uncertainty of the given coordinates in meters (0m – N = 2243; 10m – N = 208; 250m – N = 3; 2500m – N = 636). We decided to consider occurrences with low coordinate uncertainty only. This is the case in the federal states of North Rhine-Westphalia, Rhineland-Palatinate, Hesse and Saxony (see Figure S2). Therefore, we have decided to restrict the study area (area for model training) to these four federal states. Map created in ESRI ArcGIS, version 10.8.1, projection: ETRS 1989 UTM.



**Figure S2:** Study area. The four federal states (North Rhine-Westphalia, Rhineland-Palatinate, Hesse and Saxony) with highly precise GBIF occurrence data (see Figure S1) are highlighted. Map created in ESRI ArcGIS, version 10.8.1, projection: ETRS 1989 UTM.



**Figure S3:** Intercorrelation of environmental variables used as predictors. a) Correlation matrix with Pearson correlation coefficients (r) and b) Dendrogramm of the dissimilarity (1-|r|) between environmental variables used as predictors: bio05 - maximum temperature of warmest month; bio06 - minimum temperature of coldest month (bio06); elev1 - minimum elevation across sub-catchment, i.e. altitude of the record points; slope4 - average slope across sub-catchment and LC7 - percentage of cultivated vegetation (weighted average) across the sub-catchment.



**Figure S4:** Spatial patterns of the environmental variables used as predictors: bio05 - maximum temperature of warmest month; bio06 - minimum temperature of coldest month (bio06); elev1 - minimum elevation across sub-catchment, i.e. altitude of the record points; slope4 - average slope across sub-catchment and LC7 - percentage of cultivated vegetation (weighted average) across the sub-catchment.

**Table S1:** Numbers of occurence records used for model training, mean Area under the ROC curves (AUC) ± standard deviation over 20 replicate runs for each of the 12 black fly species.

|  |  |  |
| --- | --- | --- |
| **Species** | **N** | **AUC (mean over 20 replicates ± SD)** |
| *S. argryreatum* | 148 | 0.8285 ± 0.0594 |
| *S. aureum* | 43 | 0.7902 ± 0.0582 |
| *S. cryophilum* | 101 | 0.7999 ± 0.0692 |
| *S. equinum* | 128 | 0.7522 ± 0.0812 |
| *S. erytrocephalum* | 60 | 0.8267 ± 0.0754 |
| *S. lineatum* | 40 | 0.7926 ± 0.079 |
| *S. monticola* | 41 | 0.9107 ± 0.0533 |
| *S. ornatum* | 467 | 0.6315 ± 0.0553 |
| *S. reptans* | 117 | 0.7998 ± 0.078 |
| *S. trifasciatum* | 41 | 0.6224 ± 0.1155 |
| *S. varieatum* | 89 | 0.8879 ± 0.0423 |
| *S. vernum* | 175 | 0.7155 ± 0.0718 |

**Table S2**: Cosine similarity of modelled fresh water habitat suitability with wards method as cluster algorithm. S. equ – *Simulium equinum*, S. lin – *S. lineatum*, S. aur – *S. aureum*, S. ery – *S.* *erythrocephalum* (Group C); S. cry – *Simulium cryophilum*, S. ver – *S. vernum*, S. orn – *S. ornatum*, S. tri – *S.* trifasciatum (Group B); S. var – *Simulium variegatum*, S. rep – *S. reptans*, S. arg – *S. argyreatum*, S. mon – *S. monticola* (Group A)



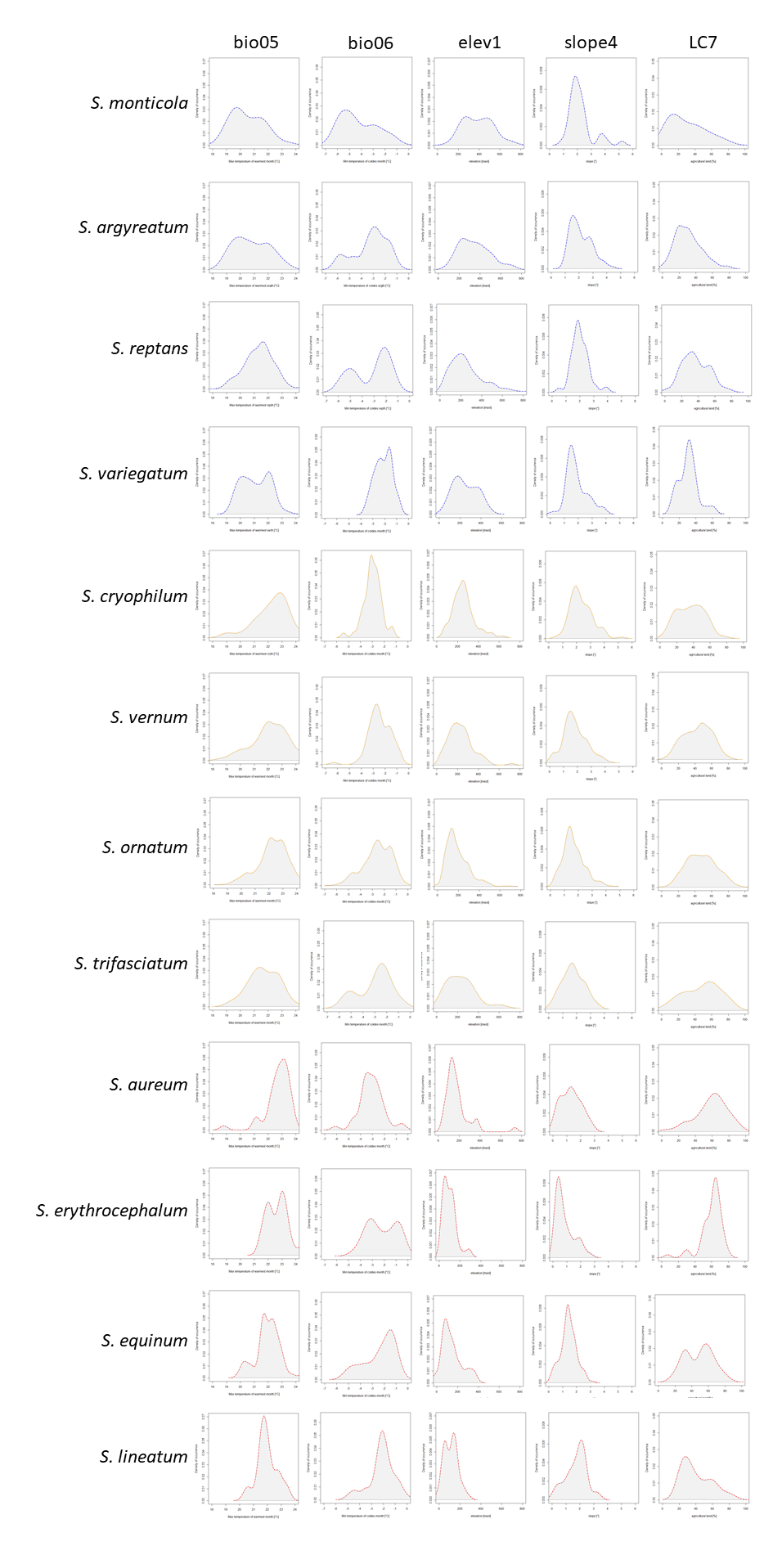


**Figure S5:** Cluster dendrogram based on the cosine similarity of modelled fresh water habitat suitability with wards method as cluster algorithm. S. equ – *Simulium equinum*, S. lin – *S. lineatum*, S. aur – *S. aureum*, S. ery – *S.* *erythrocephalum* (Group C); S. cry – *Simulium cryophilum*, S. ver – *S. vernum*, S. orn – *S. ornatum*, S. tri – *S.* trifasciatum (Group B); S. var – *Simulium variegatum*, S. rep – *S. reptans*, S. arg – *S. argyreatum*, S. mon – *S. monticola* (Group A). See Table S2 for the cosine similarity values.

a)

b)

**Figure S6:** Variable contribution. bio05 - maximum temperature of warmest month; bio06 - minimum temperature of coldest month (bio06); elev1 - minimum elevation across sub-catchment, i.e. altitude of the record points; slope4 - average slope across sub-catchment and LC7 - percentage of cultivated vegetation (weighted average) across the sub-catchment. a) relative contributions of the environmental variables to the Maxent model accumulated over the 12 considered species b) species specific percent contribution. On average, summer temperature (bio05) contributes the most to the models, followed by percentage of agricultural land in the sub-catchment (LC7). Overall, however, the relative contribution of variables varies among species. The distribution of *S. erythrocephalum*, for example, is strongly affected (>80% contribution) by geomorphological variables (elevation and slope), while *S. trifasciatum* and *S. monticola* are largely driven by climatic factors (bio05, bio06).



**Figure S8**: Single density functions of occurrence records of the 12 black fly species along the five environmental gradients considered as predictor variables in the niche modelling. Species are grouped according the cosine similarity if modelled freshwater habitat suitability (see Figure 1). bio05: maximum temperature of warmest month; bio06: minimum temperature of coldest month; elev1: minimum elevation across sub-catchment, i.e. altitude of the record points; slope4: average slope across sub-catchment; lc07: percentage of cultivated vegetation across the sub-catchment.