

RESEARCH

Open Access



# Correlating elements content in mosses collected in 2015 across Germany with spatially associated characteristics of sampling sites and their surroundings

Stefan Nickel\* and Winfried Schröder

## Abstract

**Background:** The aim of the study was a statistical evaluation of the statistical relevance of potentially explanatory variables (atmospheric deposition, meteorology, geology, soil, topography, sampling, vegetation structure, land-use density, population density, potential emission sources) correlated with the content of 12 heavy metals and nitrogen in mosses collected from 400 sites across Germany in 2015. Beyond correlation analysis, regression analysis was performed using two methods: random forest regression and multiple linear regression in connection with commonality analysis.

**Results:** The strongest predictor for the content of Cd, Cu, Ni, Pb, Zn and N in mosses was the sampled species. In 2015, the atmospheric deposition showed a lower predictive power compared to earlier campaigns. The mean precipitation (2013–2015) is a significant factor influencing the content of Cd, Pb and Zn in moss samples. Altitude (Cu, Hg and Ni) and slope (Cd) are the strongest topographical predictors. With regard to 14 vegetation structure measures studied, the distance to adjacent tree stands is the strongest predictor (Cd, Cu, Hg, Zn, N), followed by the tree layer height (Cd, Hg, Pb, N), the leaf area index (Cd, N, Zn), and finally the coverage of the tree layer (Ni, Cd, Hg). For forests, the spatial density in radii 100–300 km predominates as significant predictors for Cu, Hg, Ni and N. For the urban areas, there are element-specific different radii between 25 and 300 km (Cd, Cu, Ni, Pb, N) and for agricultural areas usually radii between 50 and 300 km, in which the respective land use is correlated with the element contents. The population density in the 50 and 100 km radius is a variable with high explanatory power for all elements except Hg and N.

**Conclusions:** For Europe-wide analyses, the population density and the proportion of different land-use classes up to 300 km around the moss sampling sites are recommended.

**Keywords:** Atmospheric deposition, Commonality analysis, European moss survey, Random forests, Heavy metals, Nitrogen

## Background

For As, Cd, Ni, Pb and N, there is good evidence that atmospheric deposition is one of the main factors for the accumulation of substances in mosses [12, 26, 40]. The level of these correlations varies from element to element depending on the reference areas studied, such as participating states or ecological spatial units [41–48]. For

the variance not explained by simple bivariate regression (deposition, accumulation of substances in mosses), further significant predictors could be identified and should be regarded: The spatial density of various land-use classes around the moss sampling sites (agricultural, forestry, urban-industrial, respective areal percentage), the distance and height of adjacent tree stands, the distance to the sea as an indicator of the sea spray effect, the population density, the altitude above sea level as well as the precipitation amount [11, 23, 24, 27, 31, 52]. In the German Moss Survey (GMS) which is part of the European

\*Correspondence: stefan.nickel@uni-vechta.de  
Chair of Landscape Ecology, University of Vechta, P.O.B. 1553,  
49364 Vechta, Germany

Moss Survey (EMS), so far potential influences of the resuspension of mineral dust particles from neighbouring areas [17] have not been investigated in detail. The same holds true for the potential influence of the precipitation amount during moss sampling, whereby it is expected that due to leaching processes from the atmosphere especially in highly polluted areas the element content increases with the precipitation quantity [1, 16]. Based on the data of the EMS2010, it was further shown that the spatial density of different land uses around the moss sampling sites would have to be investigated in larger radii than hitherto (> 100 km) in order to determine the spatial extent of influence in terms of correlations between element contents and the respective land use [27]. It was also found that the information on population density averaged for the moss sampling sites did not sufficiently reflect the real conditions there. Finally, results of deposition modelling [37] offer the opportunity to further substantiate the investigation of potential predictors of element accumulation in moss samples.

Against this background, the aim of the present study was therefore the statistical evaluation of the significance of a set of potentially explanatory variables for the estimation of element contents in moss samples that was updated and extended since the previous moss survey in Germany. It is hypothesized that the atmospheric deposition can be confirmed as most significant predictor of element contents in moss samples. To this end, the information on characteristics of the sampling sites and its surroundings and other data (Table 1) were analysed by means of multivariate statistics in order to identify and rank potential factors influencing the element contents in moss samples as well as to quantify and test the statistical relationships for significance. The investigation presented is the most comprehensive study aiming at correlating elements accumulated in moss samples due to atmospheric deposition. Germany is the only participant of the European Moss Survey that realized a large GIS-database encompassing not only element measurements from the surveys 1990, 1995, 2000, 2005, and 2015 but also data on characteristics of moss sampling sites and their surroundings which both can be integrated into statistical analysis. To this end, we introduced CA into environmental monitoring. The database can be made available by the German Environment Agency (Umweltbundesamt—UBA).

## Materials and methods

### Data

The empirical design of moss sampling and chemical analyses was detailed by Nickel & Schröder [26] and Schröder et al. [45]. The resulting measurement

values of 12 heavy metals (HM) and nitrogen (N) accumulated in moss samples collected on 400 sites across Germany in the summer of 2015 were used as target variables of the statistical analyses. The set of descriptors (= potential predictors) associated with the accumulated element deposition examined comprises a set of potentially explanatory variables, modified, updated or extended from a technical point of view compared to GMS2005, which were collected in the field according to the Moss Manual [13] or derived from it or from available data sources (Table 1).

Regarding the moss monitoring data, the limits of quantification for the elements determined, as well as the results of the quality control with the moss reference materials M2 and M3 [53], are published in Schröder & Nickel [43]. With the exception of Cr (recovery rate RR = 14%, M3), the element concentrations for M2 and M3 measured in this study did not differ significantly from the respective reference values (RR < 10%), meaning that the recovery rates were in most cases above 90%. Except for Zn (M2), the mean values of the measured heavy metal concentrations in the reference material (M2, M3) are within the confidence intervals reported by Steinnes et al. [53].

The collection of information about potential descriptors from available data sources was performed as follows: To determine the modelled heavy metal deposition (Cd, Hg, Pb) at the moss sampling sites, the site coordinates were spatially linked with the 50 km × 50 km deposition fields of the chemical transport model EMEP/MSC-E [55]. The total deposition of the year 2015 was used as well as the 3-year average of the atmospheric total deposition, corresponding to the 3-year shoots taken according to the sampling guideline (here: 2013–2015 for the GMS2015, since on the one hand a better comparability with the LOTOS-EUROS modelling was given and on the other hand EMEP modelling was not yet available at the time of the evaluation for the year 2016).<sup>1</sup> Accordingly, the modelled nitrogen deposition for the year 2015 and the period 2013–2015 was determined<sup>2</sup> by spatial linkage with the 50 km × 50 km grid of the EMEP/MSC-W model [51], specified for substance groups ( $N_{\text{Total}}$ ,  $\text{NO}_x$ ,  $\text{NH}_y$ ) and deposition processes (dry, wet, total).

<sup>1</sup> EMEP/MSC East: <http://www.msceast.org/index.php/pollution-assessment/emep-domain-menu/data-hm-pop-menu> (download 11.01.2018).

<sup>2</sup> EMEP/MSC West: Data sets '2013 v2016', '2014', '2015 emis2014' (for the year 2015 only the models '2015 emis2014' based on 2014 emission data were available at the time of the study); [http://emep.int/mscw/index\\_mscw.html](http://emep.int/mscw/index_mscw.html) (download 06.12.2017).

**Table 1 Site- and regional-specific descriptors of the element content in the EMS2015**

| Descriptor   | Abbreviation <sup>a</sup>  | Unit                                   |
|--|--|--|
| <i>Atmospheric deposition</i>  |  |  |
| Average Cd, Hg, Pb deposition for the years 2013–2015 and 2015 (EMEP) <sup>b</sup>   | cd_dep1315, cd_dep15<br>hg_dep1315, hg_dep15<br>pb_dep1315, pb_dep15   | [ $\mu\text{g m}^{-2} \text{a}^{-1}$ ] |
| Mean N-deposition of the years 2013–2015 and 2015 (EMEP) <sup>b</sup>  | N_1315, N_15 (see also explanations in Table 2)  | [ $\text{mg m}^{-2} \text{a}^{-1}$ ]   |
| Mean N-deposition of the years 2013–2015 (LOTOS-EUROS) <sup>b</sup>  | LE_all_N (see also explanations in Table 2)  | [ $\text{mg m}^{-2} \text{a}^{-1}$ ]   |
| <i>Meteorology</i>   |  |  |
| Local main wind direction <sup>c,d</sup>   | MainWindDirection  | –                                      |
| Average annual rainfall <sup>b</sup>   | pre1315_gk   | [mm/a]                                 |
| Average rainfall of the last 3 or 90 days (before sampling)  | pre3 days, pre90 days  | [mm]                                   |
| <i>Weather conditions<sup>c</sup></i>  |  |  |
| Geology/Soil/Relief  | Weather  | –                                      |
| <i>Soil type, humus species, exposure<sup>c,d</sup></i>  |  |  |
| Soil type, humus species, exposure <sup>c,d</sup>  | SoilTexture<br>HumusSpecies<br>SlopeDirection  | –                                      |
| <i>Source rock, soil type<sup>d</sup></i>  |  |  |
| Thickness of the humus layer <sup>c</sup>  | HumusLayer   | [cm]                                   |
| Slope inclination <sup>c</sup>   | SlopeGradient  | [°]                                    |
| Orographic height <sup>b,c</sup>   | elev_eu_gk   | [m ü. NN]                              |
| Background values of HM content (As, Cd, Cr, Cu, Hg, Ni, Pb, Sb, V, Zn) in topsoils (50th percentile) <sup>b</sup>   | HM_Ob50P   | [mg/kg]                                |
| <i>Sample extraction</i>   |  |  |
| Moss species, vegetation, occurrence, lime particles <sup>c</sup>  | MossSpecies<br>SamplingFrom<br>Frequency<br>VisibleDustParticles   | –                                      |
| <i>Vegetation structure</i>  |  |  |
| Distance of the extraction site from the projection of the tree or shrub layer of surrounding forests and shrubs (minimum, maximum, mean value) <sup>c</sup>                                 | DistTreeCrownsAverage<br>DistTreeCrownsMin<br>DistTreeCrownsMax<br>DistShrubsAverage<br>DistShrubsMin<br>DistShrubsMax | [m]                                    |
| Height of the tree layer of surrounding forests and shrubs (minimum, maximum, mean) <sup>c</sup>   | TsLayerHeightAverage<br>TsLayerHeightMin<br>TsLayerHeightMax   | [m]                                    |
| Degrees of coverage of the tree and shrub layer <sup>e</sup>   | TreeCoverage<br>ShrubCoverage<br>TreeShrubCoverage   | [%]                                    |
| <i>Leaf area index<sup>e</sup></i>   |  |  |
| Leaf area index <sup>e</sup>   | LAI (= sLAI), LAI2 (= wLAI)  | –                                      |
| <i>Potential emission sources</i>  |  |  |
| Spatial density of surrounding land uses (agricultural, forestry, urban-industrial) in radii of 1, 5, 10, 25, 50, 75, 100, 150, 200, 250, 300 km around the moss sampling point <sup>b</sup> | agr1–300<br>for1–300,urb1<br>– 300   | [%]                                    |
| Population density in radii 5, 50, 100 km around the moss sampling point <sup>b</sup>  | popdens5–100   | Inhabitant                             |

**Table 1 (continued)**

| Descriptor   | Abbreviation <sup>a</sup>  | Unit  |
|--|--|-------|
| Distance to local emission sources (veg. free area, agricultural land, etc.), Surface area, Arable land, Stable, Detached house, Rural settlement, Urban settlement, Unstable. Road, small country road, large country/federal road, railway, motorway, industry with high chimney, small industry, waste incineration plant, landfill, thermal power plant, building site, raw material extraction <sup>c,d</sup> | SDistNoneVegetationAreas<br>SDistAgriculturalAreas<br>SDistAnimalFarmingUnits<br>SDistPloughedAgriculturalFields<br>SDistSingleHouses, SDistVillage<br>SDistTown, SDistUnsealedRoads<br>SDistSmallPavedCountryRoads<br>SDistFederalRoads, SDistMotorways, SDistRailroadTracks<br>SDistIndustriesWithHighChimneys<br>SDistSmallIndustries<br>SDistWasteIncinerationFacilities<br>SDistDumpingGrounds<br>SDistCombustionEnergyPlants<br>SDistConstructionSites<br>SDistGravelPit | [m]   |
| Distance to the sea (North Sea and Baltic Sea) <sup>b</sup>  | CDistSea   | [km]  |
| Potential risk of erosion of arable soils by wind (1, 2, 5 km—radii) <sup>b</sup>  | pegwind_1–5  | [0–5] |

EMEP: Cooperative programme for monitoring and evaluation of the long-range transmission of air pollutants in Europe

LOTOS-EUROS: Open-source chemical transport model [22]

<sup>a</sup> See also explanation of the abbreviations below

<sup>b</sup> Potential predictors derived from available data sources

<sup>c</sup> Accompanying information collected in the field according to ICP [Vegetation [13]]

<sup>d</sup> Accompanying information derived from generally available data sources in accordance with the Sampling Directive (ICP [Vegetation [13]])

<sup>e</sup> Vegetation structure measures derived from c and d [42]

Further data on atmospheric N-deposition calculated with the LOTOS-EUROS model (LE, [38, 39]) were included: arithmetic averages specified for  $N_{\text{Total}}$ ,  $\text{NO}_x$ ,  $\text{NH}_3$ , and deposition processes (dry, occult, wet, total) for the period 2013–2015 (1 km × 1 km) were available. All values of the LE model in Eq/ha/year were converted to  $\text{mg m}^{-2} \text{a}^{-1}$  (1 Eq/ha/year =  $1.4 \text{ mg m}^{-2} \text{a}^{-1}$ ).

The precipitation was calculated for Germany from data of the German Weather Service (DWD)<sup>3</sup> as an arithmetic mean of the annual totals of the 1 km × 1 km grid of the monthly precipitation heights for the period 2013 to 2015 and thus updated compared to the previous campaign. In addition, the mean precipitation of the period three and 90 days prior to sampling was determined from regionalized (1 km × 1 km) data with daily resolution [35].

The orographic height of the moss sampling points was derived from the digital elevation model of the Shuttle Radar Topography Mission (SRTM, 90 m × 90 m) using its coordinates.<sup>4</sup>

In order to determine different spatial land-use densities around the sampling sites, the land-use classes of the Corine Land Cover 2012 [7] were grouped according

to urban-industrial land use (111—continuously built-up, 112—partially built-up, 121—industrial areas, 122—transport network, 123—port, 124—airport, 131—mining areas, 132—landfills, 133—building sites [14]); agricultural land use (211—arable land, 212—irrigated land, 213—rice crops, 221—vineyards, 222—fruit crops, 223—olive groves/plantations, 231—meadows and pastures, 241—mixed areas of permanent and annual crops, 242—complex management, 243—arable land and fallow land, 244—forest pastures) as well as forests and woodlands (311—deciduous forests, 312—coniferous forests, 313—mixed forests, 324—forest and bushes), and their percentage shares were classified for radii of 1, 5, 10, 25, 50, 75 km (CLC 100 m × 100 m) as well as for radii of 100, 150, 200, 250, 300 km (CLC 250 m × 250 m) around the 400 moss sampling sites of the GMS2015 by neighbourhood analysis (ESRI ArcGIS Focal Statistics) and subsequent spatial linkage with the moss survey network geometries.

The population density was derived from available data of the Gridded Population of the World (GPW, ~1 km × 1 km grid) of the year 2015 by averaging each within 5, 50 and 100 km radii around the 400 moss sampling sites.<sup>5</sup>

As an indicator for atmospheric transport from marine emission sources [53], the smallest distance to the

<sup>3</sup> DWD Climate Data Center (CDC), annual sum of the monthly rainfall grids for Germany, version v1.0: [ftp://ftp-cdc.dwd.de/pub/CDC/grids\\_germany/annual/](ftp://ftp-cdc.dwd.de/pub/CDC/grids_germany/annual/).

<sup>4</sup> <http://www2.jpl.nasa.gov/srtm/cbanddataproducts.html> (download 01.09.2015).

<sup>5</sup> <https://sedac.ciesin.columbia.edu/data/collection/gpw-v4>.

coastline of the North Sea and Baltic Sea was calculated for each of the 400 moss sites in Germany. The distances were then divided into ten distance classes: distances between 0 and 10 km to the coast were assigned the value 1, those between 10 and 20 km were assigned the value 2, and so on. In order to minimize the overlapping of the sea spray effect with other covariates such as elevation above sea level, distances beyond 100 km to the coast were assigned to a single class labelled with 11.

As an indicator of the influence of possible resuspension of mineral dust particles from neighbouring arable land, the potential erosion hazard of arable land by wind in Germany (1:1,000,000)<sup>6</sup> was selected. The arithmetic mean of the potentially possible soil erosion by wind on arable land (grid data set 250 m × 250 m) differentiated as hazard classes 0–5 in radii of 1, 2 and 5 km around the 400 moss sampling sites of GMS2015 was determined by neighbourhood analysis and assigned to the sampling areas.

As a further indicator for the heavy metal accumulation in moss samples, the 50th percentile of the background values of the content of As, Cd, Cr, Cu, Hg, Ni, Pb, Sb, V and Zn in topsoils was used. They were derived from information on bedrocks, soil horizons, main types of land use and settlement structures (data basis: Utilization and climate region differentiated map of the groups of soil source rocks in Germany 1:1,000,000) [3, 15].

All statistical analyses and modelling were carried out using statistical packages from the R programme system [28, 34]. The abbreviations used in the results figures and tables for the predictors are listed at the end of this text.

#### Bivariate correlation analysis

Nonparametric correlation analyses (Spearman) were performed for the element contents in moss samples (target variable) and the ordinal and metrically scaled variables contained in the descriptor set, and the results were checked for significance ( $p < 0.05$ ).

$$\begin{aligned} \text{Correlation coefficient (Spearman): } r_s \\ = 1 - \frac{6 \sum_{i=1}^n (R_{xi} - R_{yi})^2}{n(n^2 - 1)} \end{aligned} \quad (1)$$

The correlation coefficients of the bivariate analysis were interpreted according to Brosius [6] as very weak ( $< 0.2$ ), weak (0.2 to  $< 0.4$ ), medium (0.4 to  $< 0.6$ ), strong (0.6 to  $< 0.8$ ), very strong ( $\geq 0.8$ ).

#### Random forest regression

In a first step, random forest (RF, [4]) was used for multivariate statistical analysing the statistical association of the many potential predictors with the elements contents in moss samples (Table 1). RF preceded the multiple linear regression (MLR, [36]), since it allows more efficiently handling the numerous categorical variables contained in the descriptor set compared to MLR. Further, RF enables a limitation to significant categories for subsequent MLR. RF is based on the aggregation of a large number of decision trees. Depending on the scale dignity of the dependent variable, each decision tree is a classification or regression tree (CART, [5]), which determines the expressions of object characteristics that make the statistical distribution of objects in the stages of a target variable more homogeneous step by step through a sequence of divisions of the respective set of objects each into two subgroups [46–48]. Instead of generating a single prediction model from all available data, as with CART, RF typically generates ensembles of hundreds of trees [32], using randomized subsets of the data and the predictors (bags) to generate each tree. All RF trees are generated untrimmed in maximum size (no pruning), and the calculation results of all trees are included in the aggregation (mode for classifications; mean value for regressions).

Before the RF models were created, the contents of 12 HM and N in the moss samples (= target variables) were tested for normal distribution ( $\alpha = 0.05$ ) using the Shapiro–Wilk test [49], and, based on this, a log transformation of the target variables was carried out, if necessary, for a given left skew. For the RF modelling, all cases of the data set were used and missing values in the training data set were completed by imputation (= Missing Data Technique). The number of variables to consider at each node was set at one-third of the number of predictors [56]. The decision on the quantity of trees to be formed was based on diagrams in which continuously reduced numbers of trees were plotted against assigned error rates according to Eq. 2. Characteristic values of the model quality in RF are usually determined on the basis of the out-of-bag data (OOB data) removed from the training data set by randomization [20]. The mean squared error (MSE) is calculated from the sum of the quadratic deviations between measured values  $y_i$  and estimated values  $y_i^{\text{OOB}}$  divided by the number of samples. Ideally, the MSE is 0.

$$\text{Mean square error: } \text{MSE}_{\text{OOB}} = \frac{1}{n} \sum_{i=1}^n \{y_i - y_i^{\text{OOB}}\}^2 \quad (2)$$

A pseudo- $R^2$  (RSQ) is determined to indicate the percentage of the explained variance, whereby for the standard deviation SD the root of the uncorrected variance

<sup>6</sup> <https://download.bgr.de/bgr/Boden/pegwind1000/geotiff/pegwind1000>.

with the sample size is used as divisor. The proportion of explained variance is ideally 100%.

$$\text{Percentage of explained variance:RSQ[\%]} = 1 - \frac{\text{MSE}_{\text{OOB}}}{\text{SD}_y^2} \quad (3)$$

In a third step, all RF models for the four standard elements of the Convention on Long-Range Transboundary Air Pollution (Cd, Hg, Pb, N) and RF models with an explained variance >20% were optimized by backward selection until the predictors with the highest relevance for the estimation of the element contents in moss samples were identified. In addition to the model quality parameters mentioned above, two measurement values of the relative variable weight common for RF models were used for this purpose: The Increased Node Purity is a measure for the total increase in the homogeneity of the data set, which is caused by a split variable at all nodes of the random forest and is calculated for each split variable (= explanatory variable) as the mean increase in the Gini index over all individual decision trees according to Eq. 4 [21].

$$\begin{aligned} \text{Homogeneity growth:IncNodePurity}(X_m) \\ = \frac{1}{\text{NT}} \sum_T \sum_{t \in T: \nu(\text{st})=X_m} p(t) \Delta i(\text{st}, t) \end{aligned} \quad (4)$$

With:  $X_m$  = explanatory variable;  $T$  = decision tree;  $t$  = node,  $\text{NT}$  = number of decision trees;  $i(t)$  = homogeneity measure (here: Gini index);  $p(t) \Delta i(\text{st}, t)$  = homogeneity increase at the single node;  $\text{st}$  = cut (split);  $\nu(\text{st})$  = split criterion.

Second, the percentage increase in the mean square error (%IncMSE) was calculated by (1) randomly swapping (“falsifying”) the expressions of one predictor while retaining all other predictors, (2) repeating the procedure for each predictor, (3) for each RF model falsified by exactly one predictor, the mean square error (MSE) according to Eq. 2 and finally (4) the percentage increase in the mean square error (%IncMSE) is determined predictor-specifically by comparing the MSE of both models (unadulterated RF model; RF model falsified by the respective predictor). The higher the %IncMSE is, the higher the importance of the respective predictor in relation to the target variable is [9]. Explanatory variables of minor importance for the prediction of the target variables were iteratively removed from the predictor set by backward selection using the above parameters (IncNodePurity, %IncMSE) in order to form the optimized RF model with the highest proportion of the declared variance (RSQ according to Eq. 3) using the statistics program system R (R Core [34], Package “rattle”).

Second, multiple linear regression (MLR) ([36], p 654) was chosen for the quantitative analysis of the relationships between the potential predictors and the element contents in moss samples. For reasons of efficiency, the categorical descriptors contained in Table 1 were limited to those that had already been identified as significant predictors using RF. First, 13 regression models (12 HM, N) were created in a preliminary analysis taking into account all metrically scaled variables using the total sample ( $n=400$ ), the multiple coefficient of determination ( $R^2$ ) was determined and, in addition to the models for the four standard elements of the Convention on Long-Range Transboundary Air Pollution (Cd, Hg, Pb, N), those models or elements were selected which exhibited significant  $R^2 > 0.45$ . The first step was to determine the multiple coefficient of determination ( $R^2$ ). The models were optimized by backward selection so that they only included significant predictors ( $p < 0.05$ ) in the  $t$  test. Finally, the influence of the above-mentioned categorical predictors on the gradients and axis sections of the thus-optimized linear models was investigated.

#### Multiple linear regression

For the interpretation of potential inter-correlations between the predictors (multicollinearities), the regression models were further investigated using commonality analysis (CA) (R Core [34], Package “yhat”). CA makes it possible to quantify the contribution of a predictor in the MLR model in such a way that it can be compared with the contributions of all other predictors [30, 54]. CA determines the statistical relevance of each predictor for explaining the variance by the model, expressed in multiple coefficient of determination  $R^2$ , both for each predictor individually and in all combinations with the other predictors [29]. The commonality coefficient (CC) serves as a measure for the contribution of each individual predictor or each predictor in all predictor combinations to the estimation of the target variables. The sum of the CCs of all predictors and predictor combinations yields the  $R^2$  of the regression model. In the case of uncorrelated predictors, their importance for predicting the target size can thus be assessed solely on the basis of the ranking of the CC of predictors (CCUnique). In the case of multicollinearities, the CCs of the predictor combinations (CCCommon) are essential indicators for assessing the extent and structure of the common contribution to the prediction of the target variables [25, 29, 33]. CCCommon can also have negative values, especially if some of the correlations between the predictors have negative signs [30]. Negative CCCommon finally have a negative effect on the level of CCTotal (= CCUnique + CCCommon) as an expression of the proportion of the predictor in multiple  $R^2$ . For performance reasons, the CA had to

be limited to those predictors that were already identified by the MLR as significant predictors, even if a CA would have provided another valuable source of information for variable selection for all conceivable combinations of variables.

## Results and discussion

### Bivariate correlation analysis

The bivariate statistical correlations between the descriptors and the HM content in moss samples yields significant weak to very weak correlations (Pb:  $r_s = 0.23$ ; Cd:  $r_s = 0.12$ ) between the modelled atmospheric heavy metal deposition (EMEP, mean of the years 2013–2015) and the corresponding levels in the moss samples of the 2015 campaign, and no significant correlations for Hg (Table 2). For the modelled deposition in 2015, the correlations are lower (Pb) or equal (Cd) compared to the average for the years 2013–2015.

Maximum statistical correlations between precipitation and HM content in moss samples are found for Ni ( $r_s = 0.32$ ; mean of the last 90 days before sampling) and for Pb ( $r_s = 0.23$ ; mean of the last 3 years before sampling). The 3-day average before sampling yields only very weak significant correlations (Al, As, Cr, Fe, V) for a few elements.

With respect to topographic parameters, orographic elevation shows significantly weak negative correlations with Cu and Zn contents ( $r_s = -0.28$  and  $-0.27$ , respectively) and significantly weak positive correlations ( $r_s = 0.35$ ) with Ni content in moss samples.

The vegetation structure as potential predictor for the small-scale variability of the HM content in moss samples shows the highest significant correlations for the mean distance to the projection of the tree crowns (Cu:  $r_s = -0.35$ , Hg:  $-0.28$ , Zn:  $-0.22$ ), but also for the mean height of the tree layer (Cd, Cu, Hg, each weakly positive). With the exception of Zn (simple leaf area index, weakly positive), the significance of the leaf area index clearly recedes.

The statistical relationship between the HM content in moss samples and the percentage of agricultural land around the sampling sites depends on the size of the selected radius: The lower the agricultural density is in radii between 1 and 150 km around the collection point, the higher the Al, Cd, Cr, Hg, Ni, Pb and Sb contents are in moss samples. This is most evident for Ni ( $r_s = -0.24$ ) and, with the exception of Pb, for the investigated radii of 25–75 km. In contrast, the higher the agricultural density in radii between 200 and 300 km, the higher the Cd, Cu, Hg, Pb, Sb and Zn contents in moss samples. The strongest positive correlation shows Pb within the 250 km radius ( $r_s = 0.21$ ).

The statistical correlations of the forest area percentages prove to be element specific: Positive correlations are found for Al, As, Cd, Cr, Ni and Pb, negative correlations for Cu, Hg and Zn, in each case in relation to the significant correlations. The strongest correlations for Cd and Pb are found with respect to small radii (1–5 km), for Cr and Ni with respect to medium radii (100 km) and for Al, As, Cu, Hg, Ni, Zn with respect to large radii (250–300 km) around the moss sampling sites.

The spatial densities of urban-industrial uses show positive correlations in nine of the 12 HM investigated (Cd, Cr, Cu, Fe, Ni, Pb, Sb, V and Zn). With the exception of Cd for the radii of 25–75 km, the strongest statistical relationships can be detected. The highest correlation coefficient for Pb for the radii 50 km and 75 km is  $r_s = 0.3$ . For Cd, the larger radii (200–250 km) are relevant. Al and Hg show no significant correlations, As negative very weak correlations (300 km radius). The investigation of the population density in different radii around the sampling sites showed the comparatively strongest significant positive correlations mostly with the population density in the radius 50 km (Fe, Cr, Cu, Ni, Pb, V, Zn), at Cd in the radius 5 km and at Sb in the radius 100 km. The highest correlation coefficient is obtained for Ni ( $r_s = 0.3$ ). There are no significant dependencies between Al, As and Hg contents in moss samples and population density.

The distance of the sampling areas to potential local emission sources showed remarkable significant correlations only in a few cases: railway lines (Cu:  $r_s = -0.23$ ,  $n = 179$ ; Zn:  $r_s = -0.21$ ,  $n = 179$ ), industrial installations with high chimneys (Hg:  $r_s = -0.32$ ,  $n = 109$ ) and landfills (Zn:  $r_s = -0.57$ ,  $n = 13$ ).

The 50th percentiles of the HM background values in topsoils only provide evidence for positive statistical relationship ( $r_s = 0.19$ ) for Ni. All other HMs show no significant or negative correlations supporting the thesis that poikilohydry mosses do not absorb metals from soils.

The potential erosion risk of arable soils by wind (1, 2, 5 km radii) is proved to correlate at low level with the content of Cu ( $r_s = 0.22$ ; radius 5 km), Hg ( $r_s = 0.12$ ; radius 2 km) and Zn ( $r_s = 0.22$ ; radius 5 km), while Al, Ni and Pb show negative correlations, strongest within a radius of 5 km.

The descriptor “Distance to the sea (North Sea and Baltic Sea)” shows exclusively significant relations with the HM contents except for Fe, Hg, Sb and V. The direction of the correlation is different depending on the element: negative correlations exist for Cu and Zn, positive correlations in all other cases. The negative correlations—at least for Cu and Zn—are consistent with earlier investigations by Berg et al. [2], according to which sea salt cations (Ca, Mg) introduced into ecosystems can

**Table 2 Correlation coefficients (Spearman) for the relationship between contents of 12 heavy metals in moss samples and the selected descriptor set**

| Element descriptors   | Al      | As     | Cd      | Cr      | Cu      | Fe      | Hg      | Ni      | Pb      | Sb      | V       | Zn      |
|-----------------------|---------|--------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| HM_dep1315            |         |        | 0.12*   |         |         |         |         |         | 0.23**  |         |         |         |
| HM_dep15              |         |        | 0.12*   |         |         |         |         |         | 0.15**  |         |         |         |
| pre1315_gk            |         |        |         |         |         |         |         | 0.18**  | 0.23**  | 0.14**  |         |         |
| pre90 days            |         |        | 0.11*   |         | -0.15** |         | -0.18** | 0.32**  | 0.14**  | 0.10*   | 0.14**  | -0.19** |
| pre3 days             | 0.18**  | 0.13** |         | 0.19**  |         | 0.11*   |         |         |         |         |         |         |
| HumusLayer            |         |        |         | -0.11*  | 0.11*   |         |         |         |         |         |         |         |
| SlopeGradient         |         |        | 0.12*   |         |         |         |         | 0.15**  |         |         |         | -0.11*  |
| elev_eu_gk            | 0.18**  |        |         |         | -0.28** |         | -0.15** | 0.35**  |         |         |         | -0.27** |
| DistTreeCrownsAverage | 0.11*   |        | -0.15** | 0.11*   | -0.35** |         | -0.28** | 0.19**  |         |         |         | -0.22** |
| DistTreeCrownsMin     |         |        | -0.13*  | 0.10*   | -0.29** |         | -0.26** | 0.13*   |         |         |         | -0.15** |
| DistTreeCrownsMax     | 0.17**  |        | -0.14** | 0.17**  | -0.32** |         | -0.18** | 0.21**  |         |         | 0.11*   | -0.22** |
| DistShrubsAverage     | 0.13*   |        | -0.24** |         | -0.19** |         |         |         |         |         |         | -0.11*  |
| DistShrubsMin         |         |        | -0.11*  |         |         |         |         |         |         |         |         |         |
| DistShrubsMax         | 0.22**  |        | -0.22** |         | -0.24** |         |         |         |         |         |         | -0.17** |
| TsLayerHeightAverage  | -0.12*  |        | 0.23**  | -0.11*  | 0.25**  |         | 0.24**  |         | 0.15**  | 0.14**  | -0.12*  | 0.19**  |
| TsLayerHeightMin      |         |        | 0.22**  |         | 0.16**  |         | 0.22**  |         | 0.15**  | 0.11*   |         | 0.13*   |
| TsLayerHeightMax      |         |        | 0.21**  |         | 0.13*   |         |         | 0.15**  | 0.16**  | 0.13**  |         |         |
| TreeCoverage          | 0.17**  | 0.11*  |         |         |         |         |         |         | -0.13*  | -0.12*  |         | -0.11*  |
| ShrubCoverage         | 0.15**  | 0.12*  |         |         |         |         |         |         | -0.13** | -0.14** |         | -0.10*  |
| TreeShrubCoverage     | -0.19** |        |         | -0.20** | 0.13**  | -0.16** |         |         |         |         | -0.20** | 0.21**  |
| LAI                   |         |        | 0.15**  | -0.20** |         |         |         |         | -0.14** |         |         | -0.11*  |
| LAI2                  |         |        |         | -0.16** |         |         |         |         |         |         |         |         |
| agr1                  |         |        |         |         |         |         |         |         |         |         |         |         |
| agr5                  |         |        | -0.11*  |         |         |         |         | -0.10*  | -0.15** | -0.11*  |         |         |
| agr10                 | -0.10*  |        | -0.12*  |         |         |         |         | -0.12*  | -0.14** |         |         |         |
| agr25                 | -0.11*  |        | -0.12*  |         |         |         |         | -0.16** | -0.14** | -0.11*  |         |         |
| agr50                 | -0.14** |        | -0.13** | -0.11*  |         |         | -0.12*  | -0.22** | -0.10*  |         |         |         |
| agr75                 | -0.16** |        | -0.14** | -0.14** |         |         | -0.11*  | -0.24** |         |         |         |         |
| agr100                | -0.15** |        |         |         |         |         |         | -0.23** |         |         |         |         |
| agr150                | -0.12*  |        |         |         |         |         |         | -0.17** |         |         |         | 0.11*   |
| agr200                |         |        |         |         | 0.12*   |         | 0.12*   |         | 0.19**  | 0.11*   |         | 0.12*   |
| agr250                |         |        | 0.15**  |         | 0.10*   |         | 0.14**  |         | 0.21**  |         |         |         |
| agr300                |         |        | 0.18**  |         |         |         | 0.12*   |         | 0.20**  |         |         |         |
| for1                  |         |        | 0.12*   |         |         |         |         | 0.11*   | 0.10*   |         |         |         |
| for5                  |         |        |         |         |         |         |         | 0.13*   | 0.10*   |         |         |         |





**Table 2 (continued)**

| Element descriptors         | Al      | As     | Cd    | Cr     | Cu      | Fe | Hg     | Ni      | Pb     | Sb | V | Zn      |
|-----------------------------|---------|--------|-------|--------|---------|----|--------|---------|--------|----|---|---------|
| SDistDumpingGrounds         |         |        |       |        |         |    |        |         | 0.47*  |    |   | -0.57*  |
| SDistCombustionEnergyPlants |         |        |       |        |         |    |        | 0.31**  | 0.12*  |    |   | -0.17** |
| CDistSea                    | 0.25**  | 0.13** | 0.11* | 0.15** | -0.18** |    | 0.11*  | -0.22** |        |    |   | 0.19**  |
| pegwind_1                   |         |        |       |        | 0.20**  |    | 0.12*  | -0.24** |        |    |   | 0.21**  |
| pegwind_2                   |         |        |       |        | 0.20**  |    |        | -0.29** | -0.12* |    |   | 0.22**  |
| pegwind_5                   | -0.13** |        |       |        | 0.22**  |    | -0.12* | 0.19**  |        |    |   | -0.19** |
| HM_Ob50P                    |         |        |       |        |         |    |        |         |        |    |   |         |

Significance codes: \* = 0.05; \*\* = 0.01

influence the bioaccumulation of As, Cd, Cu, Pb, V and Zn in moss samples by competing with the heavy metals at the exchange sites in the moss tissue and thus competition from sea salt ions can lead to underestimation of atmospheric deposition [57]. The statistical relationships between the descriptors examined and the nitrogen content in the moss samples are summarized in Table 3.

There are significant correlations only with the EMEP total depositions calculated for the year 2015 ( $r_s=0.11$ ). The highest correlations result with respect to the wet deposition of oxidized nitrogen ( $r_s=0.27$ ). For the N-deposition calculated with LOTOS-EUROS (LE, average of the years 2013–2015), the correlations are higher compared to EMEP. The strongest dependence on the N content in moss samples is in the proportion of wet N-deposition ( $r_s=0.31$ ). For dry deposition this is also significantly positive ( $r_s=0.24$ ), and for occult deposition it is significantly negative ( $r_s=-0.22$ ). There is a weak significant relationship ( $r_s=0.28$ ) to the total deposition weighted according to the proportion of use in each LE grid.

The mean rainfall (3 years or 90 days before sampling) indicates very weak negative correlations. With increasing orographic height, the N contents in the moss samples decrease significantly ( $r_s=-0.34$ ).

With regard to the influence of the canopy drip effect on the N content in moss samples, significantly weak relationships to the mean distance from neighbouring trees ( $r_s=-0.4$ ) and to the mean height of the tree layer ( $r_s=0.3$ ) are formed. The correlations with the leaf area index (LAI), on the other hand, are clearly receding, as is the case with heavy metals. In comparison with the results of the investigations of the canopy drip effect [42, 45], this is surprising, since the significance of the leaf area index for the prediction of contents in the moss samples was higher than for the height and distance of adjacent tree stands. It is important to emphasize that this only applies to the relationships between the quotients of the pairwise combined location categories (eaves, semi-eaves and open land locations) in the small-scale studies on the canopy drip effect. Probably the quotients more adequately reflect the small-scale varying canopy drip effect and are not superimposed by other large-scale varying influencing variables (e.g. land use).

The spatial densities of different land-use classes around the moss sampling sites show significant positive relationships for the proportion of agricultural land in the radii 1–150 km with comparatively highest correlation coefficients in the 5 and 10 km radius ( $r_s$  0.26 each). For all radii 1–300 km, there are significant negative correlations with a maximum for the radius 200 km ( $r_s=0.41$ ). There are no significant correlations with the density of urban-industrial uses or population density.

The significant positive correlation between the N content in the moss samples and the characteristic value for the potential erosion hazard of the arable soils by wind is remarkable (1 km:  $r_s=0.34$ ; 2 km:  $r_s=0.34$ ; 5 km:  $r_s=0.35$ ). Evidence for technically plausible significant dependencies of the N content in moss samples on the distances to potential emission sources was found for: arable land ( $r_s=-0.13$ ,  $n=247$ ) and building sites ( $r_s=-0.94$ ,  $n=6$ ). All other descriptors do not provide significant or implausible results. For the distance to the sea, the N contents in the moss samples, as for Cu and Zn, indicate a significant negative correlation ( $r_s=-0.28$ ).

### Regression and random forest models

RF models with  $RSQ \geq 20\%$  (=proportion of the explained variance) result for the six heavy metals Cd, Cu, Hg, Ni, Pb, Zn and for nitrogen (Table 4). The MLR models for Cd, Cu, Hg, Ni, Pb, Zn, and N show  $R^2 \geq 0.2$  (Table 5). The RF and MLR models with the highest explanatory power were reached by using the log-transformed HM contents in moss samples. For Al, As, Cr, Fe, Sb and V, the preliminary analysis with RF had  $RSQ < 20\%$ , so that these were not used for statistical modelling (RF, MLR). The comparative element-specific consideration of both regressions results in the following findings.

### Cd—Cadmium

The proportion of variance explained by the RF model is  $RSQ=31\%$  (Additional file 1: Figure S1); the coefficient of determination of the MLR model  $R^2=0.36$  (Table 5), i.e. approx. 36% of the variance of the Cd content in moss samples, is explained by multiple linear regression. The strongest predictor in both models is the moss species (Additional file 1: Figures S1 and S2). This is indicated by the characteristic values of the relative importance of the predictors of the RF model (Increased Node Purity, Increased MSE) and in the commonality coefficient (CC) of the MLR model. After logging off the three moss species-specific axis sections of the MLR model shown in Table 5, it can be seen that the element contents in *Plesch* are estimated by the Cd model to be  $-33\%$  lower than in *Hypcup* and *Psepur*, i.e. the entire linear model for *Plesch* always shifts the values of the target variables significantly by  $-33\%$ . A significant predictor (RF, MLR) is the population density in 100 km around the moss sampling sites, not the population density in the radius of 5 km or 50 km. The higher the population density is within a radius of 100 km, the higher the Cd content is in moss samples, which indicates the influence of urban emissions. This finding initially seems to correspond with the relatively strong significance of

**Table 3 Correlation coefficients (Spearman) for the relationship between nitrogen content in moss samples and the selected descriptor set**

| Descriptors           | N       |
|-----------------------|---------|
| N_15                  | 0.11*   |
| N_dry_ox15            | -0.19** |
| N_wet_ox15            | 0.27**  |
| N_tot_ox15            | 0.13*   |
| N_wet_red15           | 0.17**  |
| N_tot_red15           | 0.11*   |
| LE_all_NH4            | 0.21**  |
| LE_all_N              | 0.28**  |
| LE_all_NH4            | 0.30**  |
| LE_dryall_N           | 0.24**  |
| LE_dryall_NH4         | 0.26**  |
| LE_occultall_N        | -0.22** |
| LE_occultall_NH4      | -0.21** |
| LE_occultall_NO3      | -0.24** |
| LE_wet_N              | 0.31**  |
| LE_wet_NH4            | 0.32**  |
| LE_wet_NO3            | 0.20**  |
| pre1315_gk            | -0.13*  |
| pre90 days            | -0.17** |
| SlopeGradient         | -0.19** |
| elev_eu_gk            | -0.34** |
| DistTreeCrownsAverage | -0.40** |
| DistTreeCrownsMin     | -0.38** |
| DistTreeCrownsMax     | -0.39** |
| DistShrubsAverage     | -0.18** |
| DistShrubsMax         | -0.24** |
| TsLayerHeightAverage  | 0.30**  |
| TsLayerHeightMin      | 0.20**  |
| TsLayerHeightMax      | 0.17**  |
| ShrubCoverage         | -0.12*  |
| LAI                   | 0.16**  |
| agr1                  | 0.22**  |
| agr5                  | 0.26**  |
| agr10                 | 0.26**  |
| agr25                 | 0.24**  |
| agr50                 | 0.24**  |
| agr75                 | 0.20**  |
| agr100                | 0.18**  |
| agr150                | 0.15**  |
| for1                  | -0.17** |
| for5                  | -0.26** |
| for10                 | -0.29** |
| for25                 | -0.31** |
| for50                 | -0.35** |
| for75                 | -0.34** |
| for100                | -0.34** |
| for150                | -0.37** |
| for200                | -0.41** |

**Table 3 (continued)**

| Descriptors                     | N       |
|---------------------------------|---------|
| For250                          | -0.40** |
| for300                          | -0.39** |
| SDistPloughedAgriculturalFields | -0.13*  |
| SDistUnsealedRoads              | 0.18**  |
| SDistFederalRoads               | 0.16*   |
| SDistRailroadTracks             | -0.17*  |
| SDistConstructionSites          | -0.94*  |
| CDistSea                        | -0.28** |
| pegwind_1                       | 0.34**  |
| pegwind_2                       | 0.34**  |
| pegwind_5                       | 0.35**  |

Significance codes: \* = 0.05; \*\* = 0.01

the spatial density of urban-industrial uses in the radius 75–300 km with the strongest signal at 150 km around the sampling point (RE, MLR). In the MLR model, however, the relationship proves to be less significant compared to population density and above all negative, i.e. the higher the urban area share, the lower the Cd content in moss samples. In both models, the agricultural area within a radius of 300 km around the sampling sites also has a high explanatory power with a significantly positive relationship to the Cd contents in moss samples. In the RF model, the precipitation quantity 2013–2015 has a high variable weight, in the MLR model, higher Cd contents are observed with higher precipitation. Two other variables that can be associated with the filtering effect of the trees can also be assigned medium importance as explanatory variables: the simple leaf area index (LAI, derived from surrounding land use) and the stand height. It is noteworthy that the predictive force of the modelled Cd deposition recedes—as was already the case with correlation analysis. In the RF model, the total deposition (EMEP; mean of the years 2013–2015) still shows an average relative variable weighting; in the MLR model, it is no longer included as a predictor. Further predictors with a comparatively low but significant influence on the Cd contents in the moss samples form: the mean rainfall of the last 90 days before sampling, the agricultural and forested area in 5–10 km and the forested area in 250–300 km around the moss sampling sites, the mean distance to the projection of the canopy of adjacent tree stands, the degree of tree cover and the slope of the sampling sites as well as the substrate from which the moss samples were collected (soil, tree stump).

**Table 4** Characteristics of the random forest models for the relationship between the log-transformed (ln) HM contents and N contents in moss samples and the predictors given in the following figures

| Predictors model parameters | Target variable (element content in moss samples) |        |        |        |        |       |        |
|-----------------------------|---|--------|--------|--------|--------|-------|--------|
|                             | ln Cd   | ln Cu  | ln Hg  | ln Ni  | ln Pb  | ln Zn | N      |
| N                           | 400   | 400    | 400    | 400    | 400    | 400   | 400    |
| n <sub>tree</sub>           | 500   | 500    | 500    | 500    | 500    | 500   | 500    |
| m <sub>try</sub>            | 3   | 4      | 3      | 3      | 3      | 2     | 3      |
| Model quality indices       |   |        |        |        |        |       |        |
| MSE                         | 0.1882  | 0.0774 | 0.1878 | 0.2239 | 0.2617 | 0.077 | 0.0992 |
| RSQ                         | 30.70   | 41.62  | 19.83  | 34.50  | 21.99  | 38.71 | 50.18  |

ln log-transformed values, n sample size, n<sub>tree</sub> number of regression trees in the random forest model, m<sub>try</sub> number of variables used for each split, MSE mean squared residuals, RSQ share of explained variance

### Cu—copper

With an RSQ of 42%, the RF model for copper shows the second highest explanatory power after nitrogen among the seven RF models presented in this article (Table 4). The same applies to the MLR model with an  $R^2$  of 0.38 (Table 5). The variables with the comparatively highest predictive force in both models are the maximum distance to adjacent tree populations (negative), the density of forest use in a radius of 300 km (negative) and the population density in a radius of 50 km around the moss collection point (positive) (Additional file 1: Figures S3 and S4). This indicates, above all, the influence of the canopy drip effect and the influence of urban emissions. There is also a significant dependence with the moss species. For the same characteristic values, the estimation model of multiple linear regression always yields 14% higher Cu contents in *Plesch* and 18% higher Cu contents in *Psepur* than in *Hypcup* (Table 5). Further predictors with rather subordinate relative importance in the RF model are the orographic height, the spatial density of urban-industrial uses (75–300 km), the humus species and the mean precipitation (90-day average).

### Hg—Mercury

The share of the total variance of Hg accumulation in the moss samples explained by the RF model is RSQ=20% (Table 4). It is similar to the MLR model with  $R^2=0.21$  (Table 5). The 90-day average precipitation height proves to be a strong predictor in both models (Additional file 1: Figures S5 and S6), whereby the Hg content in the moss samples decreases significantly with increasing precipitation, which could be associated with leaching effects. In addition, in the RF model the proportion of agricultural land within the radius of 50 km has a high variable weight, and within the MLR model in the radius of 75 km at least an average relative variable weight. A low-to-medium predictive force is the forest density in the radius 75–300 km (RF) and 75–200 km (MLR). With regard to

all other predictors, the results of the regression analysis differ significantly: While the random forest regression reveals the mean distance to adjacent tree populations, the orographic height and the modelled total deposition (EMEP; mean of the years 2013–2015) as influencing factors. The multiple linear regression reveals the height of the tree layer, the wood cover of the extraction area and the moss species. This means that in both models different indicators of the vegetation structure play a role, which reflects the relevance of the canopy drip effects on the Hg contents in moss samples. The multiple linear regression results in a moss species-dependent shift of the logarithmic intercept of *Plesch* by –13% and of *Psepur* by –18% compared to *Hypcup* (Table 5).

### Ni—Nickel

According to the indicators of goodness of fit of both models (RF: RSQ=35%; MLR:  $R^2=0.32$ ), approx. 32–35% of the total variance is explained by the selected predictors (Tables 4 and 5). The strongest dependence of the Ni contents in the moss samples exists with the moss species. In the MLR model, the moss species as a categorical variable reduces the axis section by –31% compared to *Hypcup* at *Plesch* and by –39% at *Psepur*, calculated on the basis of the logarithmic values from Table 5. Another strong predictor is the density of forest use in 100 km around the moss sampling sites (RF, MLR) (Additional file 1: Figures S7 and S8), where the higher the proportion of forest area, the higher the Ni contents. Similarly, in both models population density has a strong influence on Ni contents in moss samples, with the radius 50 km in the RF model and the radius 100 km in the MLR model having the relatively largest variable weight. The positive correlation illustrates the influence of urban emissions on the Ni contents in moss samples. RF shows a subordinate variable significance for the orographic height and the density of urban-industrial uses in

**Table 5 Characteristics of the MLR models for the relationship between the log-transformed (ln) HM contents and N concentrations in mosses and the predictors**

| Variable                   | Target variable (element content in moss samples) |            |            |            |            |            |            |
|----------------------------|---|------------|------------|------------|------------|------------|------------|
|                            | ln Cd   | ln Cu      | ln Hg      | ln Ni      | ln Pb      | ln Zn      | N          |
| Axis intercept (intercept) |   |            |            |            |            |            |            |
| MossSpecies:Hypcup         | −4.1259***  | 2.1396***  | −2.8503*** | −0.6498*** | −1.0461*** | 3.2543***  | 1.4662***  |
| MossSpecies:Plesch         | −4.5237***  | 2.2703***  | −2.9865*   | −1.0266*** | −1.2981*** | 3.2504     | 1.3841     |
| MossSpecies:Psepur         | −4.1519   | 2.3067***  | −3.0545*** | −1.1412*** | −1.398***  | 3.569***   | 1.68***    |
| Slope (slope)              |   |            |            |            |            |            |            |
| agr5                       | 1.0823*   |            |            |            |            |            |            |
| agr10                      | −1.4968**   |            |            |            |            |            |            |
| agr75                      |   | −0.8641*** | −5.2626*** |            |            |            |            |
| agr100                     |   |            | 4.7779***  |            |            |            |            |
| agr150                     |   |            |            |            | −2.8215**  |            | −0.7759**  |
| agr200                     |   |            |            |            | 5.0957***  |            |            |
| agr300                     | 3.8929***   |            |            |            |            |            |            |
| CDistSea                   |   |            |            | −0.0402**  |            |            |            |
| DistTreeCrownsAverage      |   |            |            |            |            | −0.0042**  | −0.0078*** |
| DistTreeCrownsMax          |   | −0.0078*** |            |            |            |            |            |
| for5                       | 1.379**   |            |            |            |            | 0.3196***  |            |
| for10                      | −1.6679***  |            |            |            |            | 0.9194**   | −0.5501*** |
| for25                      |   |            |            |            |            | −1.6428*** |            |
| for50                      |   |            |            | −1.56***   |            |            |            |
| for75                      |   |            | −4.8525*** |            |            |            |            |
| for100                     |   |            | 6.0917***  | 3.7169***  |            |            |            |
| for200                     |   |            | −2.3588*** |            |            |            |            |
| For250                     | −5.6749*  |            |            |            |            | −1.135***  | −5.5923*** |
| for300                     | 4.8276*   | −1.4352*** |            |            |            |            | 5.0524**   |
| LAI                        | 0.0332***   |            |            |            |            |            |            |
| LAI2                       |   | 0.0315**   |            |            |            | 0.0346***  |            |
| LE_all_N                   |   |            |            |            |            |            | 0.0016***  |
| LE_all_NH4                 |   |            |            |            |            |            | −0.0017*** |
| pegwind_1                  |   |            |            |            |            |            | 0.1112**   |
| popdens100                 | 0.0017***   |            |            | 0.0014***  |            |            |            |
| popdens5                   |   |            |            |            | −0.0003*   | −0.0003**  |            |
| popdens50                  |   | 0.0007***  |            |            |            | 0.0004***  |            |
| pre1315_gk                 | 0.0005***   |            |            |            | 0.0009***  |            |            |
| pre90 days                 |   |            | −0.0068*** |            |            |            |            |
| ShrubCoverage              |   |            | −0.0046**  |            |            |            |            |
| SlopeGradient              | 0.0078**  |            |            |            |            |            |            |
| TreeCoverage               | −0.004*   |            |            |            |            |            |            |
| TreeShrubCoverage          | 0.0067**  |            | 0.0028**   |            |            |            |            |
| TsLayerHeightAverage       |   |            | 0.0082**   |            |            |            | 0.0114***  |
| TsLayerHeightMin           | 0.0085**  |            |            |            | 0.0117***  |            |            |
| urb1                       |   |            |            |            | 0.9271**   |            |            |
| urb10                      |   |            |            |            |            | 1.5446**   |            |
| urb25                      |   |            |            |            | 2.575***   |            | −1.2321**  |
| urb150                     | −4.7086*  |            |            |            |            |            |            |
| Model quality indices      |   |            |            |            |            |            |            |
| DF                         | 369   | 381        | 385        | 393        | 386        | 379        | 386        |
| RSE                        | 0.4285  | 0.289      | 0.4383     | 0.4849     | 0.4797     | 0.2863     | 0.3451     |

**Table 5 (continued)**

| Variable          | Target variable (element content in moss samples) |       |       |       |       |       |      |
|-------------------|---|-------|-------|-------|-------|-------|------|
|                   | In Cd   | In Cu | In Hg | In Ni | In Pb | In Zn | N    |
| R <sup>2</sup>    | 0.36  | 0.38  | 0.21  | 0.32  | 0.32  | 0.36  | 0.42 |
| adjR <sup>2</sup> | 0.33  | 0.37  | 0.19  | 0.31  | 0.31  | 0.35  | 0.40 |

In log-transformed values, DF degree of freedom, RSE root of residual standard error, R<sup>2</sup> coefficient of determination, adjR<sup>2</sup> corrected coefficient of determination

Significance codes: \* = 0.05; \*\* = 0.01; \*\*\* = 0.001

the radii 75 and 150 km and the MLR for the distance to the sea.

#### Pb—Lead

The quality indices of both models differ significantly with RSQ = 22% (RF) and R<sup>2</sup> = 0.32 (MLR) (Tables 4 and 5). Nevertheless, there are similarities between the two models with regard to the strongest prediction for the Pb content in moss samples: precipitation (3-year average), urban-industrial area within a radius of 25 km around the moss sampling point and the type of moss (Additional file 1: Figures S9 and S10). On average, higher Pb values can be seen here with higher precipitation and higher density of urban-industrial uses, which indicates the influence of urban emissions. The moss-specifically determined axis sections of the MLR model show that the Pb content is always estimated to be –22% lower for *Plesch* and –30% lower for *Psepur* than for *Hypcup*. This is followed by the percentage of agricultural land within radii of 200–300 km (RF) and 150–200 km (MLR), the percentage of urban land within radii of 50–75 km (RF) and 1 km (MLR) and the population density within a radius of 100 km (RF) and 5 km (MLR). The proportion of agricultural land in the 150 km radius shows a negative relationship in the MLR model and a positive relationship in the 200 km radius, which makes it difficult to interpret the opposite gradients. Further significant, but rather subordinate, influences on the target variable result from the 90-day average of the precipitation sum (RF) and the minimum stand height of surrounding groves (MLR). The modelled atmospheric deposition plays no role in either model.

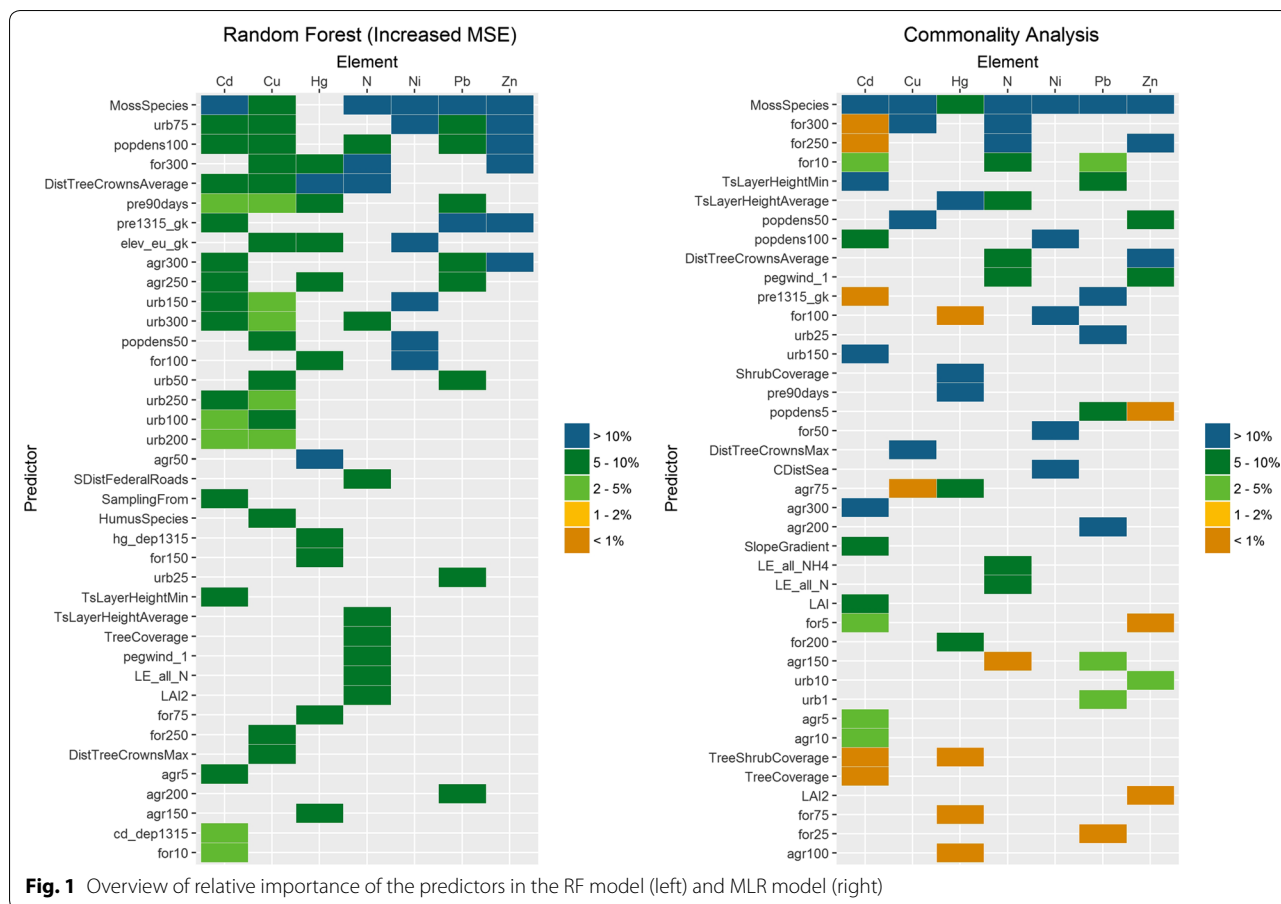
#### Zn—Zinc

The variance of the Zn content in moss samples is explained by the predictors remaining after backward selection to 39% for RF (RSQ = 39%) and to 36% for MLR (R<sup>2</sup> = 0.36) (Tables 4 and 5). The predictors with the strongest explanatory power in both models are the density of forest use within radii 250–300 km (RF, MLR), the population density within 100 km (RF) or 50 km (MLR) and the moss species. The MLR shows that a high proportion of forest and a low population density within the

relevant radii mean lower Zn bioaccumulation. In addition, the MLR model estimates element contents in *Psepur* to be 37% higher than *Hypcup* and *Plesch*. In the RF regression, the precipitation amount (3-year average) and the agricultural area portion (300 km) are characterized by a mean variable weight (Additional file 1: Figure S11). In the ML regression, this is true for the mean distance to adjacent treetops and the wind erosion sensitivity of the soils within a 1-km radius around the moss sampling sites (Additional file 1: Figure S12). Further significant predictors, however, with comparatively low commonality coefficients, are the urban-industrial use within a radius of 10 km and the weighted leaf area index derived from the surrounding land use and the wood cover of the sample sites.

#### N—nitrogen

The regression models for nitrogen show the highest quality of fit with RSQ = 50% (RF) and R<sup>2</sup> = 0.42 (MLR) compared to the metals (Tables 4 and 5). The explained variance is higher, since N is involved as a macronutrient element in numerous metabolic processes in mosses [10]. In both models, the density of forest uses (250–300 km) and the type of moss (Additional file 1: Figures S13 and S14) have the highest variable significance. For *Psepur*, the MLR always results in 15% higher N content in the moss samples than for *Hypcup* and *Plesch*. In the RF regression, the population density in 100 km around the sampling point is of highest importance for the estimation of the N content in moss samples, whereas in the MLR model it is of no importance. The distance to treetops has a high variable significance in the RF model and a medium variable significance in the MLR model, whereby the N content in the moss samples decreases with increasing distance. This illustrates the influence of the canopy drip effect on the N content in moss samples. Atmospheric deposition (LE, mean total deposition of the years 2013–2015) is of only medium importance in both models. In the RF model, the weighted leaf area index (LAI2), the degree of coverage of the trees on the extraction area and the mean stand height are also of minor to medium importance, which can be linked to the canopy drip effect. Equally important are the potential risk of erosion of arable soils by wind (radius 1 km) and



the distance to federal highways, indicating the influence of local emission sources (dust, traffic emissions).

**Predictors**

The predictor-specific consideration of both regressions (Fig. 1) produces the following picture, whereby it should be noted that all categorical variables—with the exception of the moss species—were only examined by using RF.

The classification reflects the percentage of the relative significance of the predictor as the sum of the relative significance of all element-specific relevant predictors. Their relative importance was quantified by the percentage increase in the mean square error (IncreasedMSE) for the RF model and the commonality coefficient (CCTotal) for the MLR model.

**Atmospheric deposition**

With regard to the modelled atmospheric deposition (LE and EMEP; 2015, mean of the years 2013–2015), it can be seen that this has a lower predictive force for the respective element contents in the moss samples than in previous campaigns. For N (LE; 2013–2015) and Hg (EMEP;

2013–2015) the calculated deposition in moss monitoring has a mean variable significance, for Cd (EMEP; 2013–2015) a minor significance and for Pb no significance. One reason for the declining explanatory power of deposition at Cd and Pb can be seen in the decrease in atmospheric deposition between 1990 and 2015 [43] and the associated increase in the relative importance of other predictors (surrounding land use, population density). The median of the Cd deposition calculated with the EMEP/MSC-E model decreased from 41.29 to 26.06  $\mu\text{g}/\text{m}^2$  a and for Pb from 1660.33 to 1052.52  $\mu\text{g}/\text{m}^2$  a in the period 2005–2015. Comparing the two chemical transport models, the predictive force of the N-deposition calculated with the EMEP/MSC-W model is significantly lower than that derived by LOTOS-EUROS. One explanation for this could be the comparatively lower resolution and poor data quality of the EMEP models used for the year 2015, which are still based on the emissions of the year 2014.

**Meteorology**

The mean precipitation sum 2013–2015 can be determined as an important influencing factor for Cd, Pb and



Zn, the 90-day average for Hg and Pb and subordinated also for Cd and Cu. However, the weather conditions (3-day average) cannot be determined as predictor for any element content; the same is true for the local main wind direction. The hypothesis that higher precipitation at the time of sampling induces higher element contents, especially in polluted areas [1], could not be confirmed by the data of GMS2015. First, in the MLR model the relatively high variable weight of the 90-day mean for Hg proves to be a negative relation, and second, in the RF model the 3-year mean for Pb and Cd each shows higher variable relevance than the 90-day mean.

#### **Geology/soil/relief**

Among the topographic parameters, the orographic height (Cu, Hg and Ni) and the slope inclination (Cd) are the strongest predictors and among the soil science parameters the humus species as, for example, mull or moder (Cu). In contrast, cardinal points of slopes, bedrock, soil type, thickness of the humus layer, and HM content in the topsoil are not involved in any of the regression models.

#### **Moss species**

The moss species in all models has a very high (Cd, Cu, Ni, Pb, Zn and N) or high variable meaning (Hg). This confirms the results of earlier multivariate static analyses [12, 18, 19, 23, 40]. It is important to emphasize, however, that the boundary conditions uncovered in these investigations, under which the moss species have different substance contents, can themselves vary spatially. This applies both to the large-scale consideration and to the small-scale consideration, where the micro-local variability of the deposition also does not permit a clear conclusion as to whether it is the moss species or the micro-local variance of the deposition that causes the possibly different element contents. Siewers et al. [50] do not recommend a mathematical conversion of moss species-specific element content data, since they are also element specific and range within the site-specific variability. Consequently, reliable determinations of the influence of the moss species on the accumulated deposition are only possible in systematic, statistically verified laboratory experiments in which all factors such as light, soil and temperature are constant and only the moss species and the elements and their content in the artificially generated wet and dry deposition are systematically varied. Such laboratory studies should be carried out in Moss Monitoring 2020. While the MLR models for the nutrient elements Cu, Zn and N for *Psepur* and *Plesch* estimate significantly higher to equally high element contents compared to *Hypcup*, the MLR models estimate significantly lower to equally high metal content for Cd, Hg, Ni

and Pb (Table 5). The influence of the moss species on element accumulation in combination with other predictors could thus be quantified element-specifically by the MLR model. However, moss species-specific transformation of the element contents should not be concluded from these results. In addition to the moss species, only the vegetation (tree stump, soil) has an influence on the Cd contents in moss samples. The frequency of moss occurrence plays no role as a predictor for the elements investigated here. The significance of the presence of lime particles could not be verified as any lime particles were observed in the field.

#### **Vegetation structure**

Of the 14 vegetation structure measures studied, the mean distance to adjacent tree populations forms the comparatively strongest predictor (Cd, Cu, Hg, Zn and N), followed by the height of the vegetation cover (Cd, Hg, Pb and N), the leaf area index (Cd and N; subordinated for Zn) and finally the degree of coverage of the tree layer (Ni; subordinated for Cd and Hg). This confirms the results of the multivariate statistical analyses of GMS2005 [31] regarding distance and stand height for N. (Leaf area index and degree of coverage were not investigated in GMS2005.) In contrast to the small-scale investigations of the crown drift effect [42, 45], the distance to trees for all six HM and N proves to be a stronger predictor than the leaf area index. One reason for this could be that the LAI derived from the surrounding land use (e.g. deciduous forest and coniferous forest) also has a region-specific distribution. For example, deciduous forests are more widespread in lowland areas and coniferous forests are more widespread in mountain forest areas. Therefore, the LAI, unlike the distance to trees, is an indicator not only for the small-scale filter effect of local tree populations, but also for large-scale varying influences (e.g. deposition). The latter has no influence if, as in the case of Schröder & Nickel [42], relations of characteristic values between canopy drip and open land moss sampling sites are used instead of absolute values as here. This leads to the hypothesis that correlation and regression within ecological land classes (ELCE, [41, 48]) and ecosystem types [44] should show stronger statistical relationships for the LAI, as these represent more homogeneous units with regard to predominant tree species and environmental characteristics.

#### **Land-use density**

In many cases, the spatial density of various land-use classes around the moss sampling sites has a high explanatory power for the element contents in moss samples. In the case of forests and woods, it is mainly the respective areal percentages within radii of 100–300 km that are

significant predictors for Cu, Hg, Ni and N. In the case of urban areas, there are element-specific radii between 25 and 300 km (Cd, Cu, Ni, Pb and N), and in the case of agricultural areas, there are mostly radii between 50 and 300 km which are statistically relevant. Land-use density within radii < 25 km is rarely of importance (Cd, Pb and N). It follows from this that spatial land-use density indicates less the influence of individual local emission sources in the immediate vicinity of moss sampling sites than the varying atmospheric deposition over large areas. This is also supported by the results of the geostatistical variogram analysis, according to which the ranges of the spatial autocorrelations of the substance content in the moss samples between 67 km (Hg) and 223 km (Cd) were determined for seven elements [45]: Annex A5.14). The findings that the larger radii are usually more important than the smaller ones are initially surprising if one assumes that the filter effect plays a role locally or for vertical rather than horizontal mass transport directions. This indicates that the density of the surrounding land use is less indicative of the filtering effect of the forests than of the presence or absence of other influencing factors. The spatial density of land use indicates above all the large-scale variability of emission sources and their influence on the element contents in moss samples, i.e. this overlays the small-scale variability of the emission sources (e.g. dusts blown away from agricultural and settlement areas with their substance content). This interpretation is also supported by the fact that the distances between local emission sources and the moss sampling sites play a rather subordinate role as a predictor (see below).

#### **Population density**

The population density within the 50 and 100 km radius around the moss sampling sites is a variable with a high explanatory power for all elements except Hg and N. Population density within radius of 5 km thereby is of minor importance. The higher the population density is within radii  $\geq 50$  km, the higher the average content of Cd, Cu, Ni, Pb and Z is in the moss samples, as the corresponding MLR models show (Table 5). The direction of the relationship cannot be determined unambiguously by the RF model alone, since the possibilities of interpretation are fundamentally impeded by the hundreds of individual trees involved in the RF regression. If one considers at least the first tree of the RF ensemble, which is equivalent to an analysis using CART, a positive relationship results for N as for HM. This means that the higher the values of the two classes of population density generated by the split variable (= population density) are at the respective node, the higher the values for the target variable (= N content in moss samples) are estimated at the following

two nodes. As with the density of surrounding land use, population density is less likely to indicate the influence of local emission sources than the large-scale influences responsible for element contents in the moss samples. In order to uncover possible optima of the relevant radii, population density within radii > 100 to 300 km should therefore be included in subsequent investigations.

#### **Other potential sources of emissions**

Of the numerous potential local emission sources recorded in GMS2015 (Table 1), only the distance to federal roads (N; traffic emissions) shows a high variable significance in the RF regression. The distance to the sea is a relevant factor only for Ni: the greater the distance is, the lower the Ni content, measured in the moss samples, is. It is at least known for Ni that it tends to accumulate in the oceans and is emitted into the atmosphere as sea spray aerosol [8]. In both models (RF, MLR), the potential erosion hazard of arable soils by wind within 1 km around the sampling point is shown to be a comparatively strong predictor for N in both models (RF, MLR). As the risk of wind erosion increases, so does the N content in the moss samples, which indicates an actual re-emission of nitrogen-containing particles from arable land and deposition on the mosses. Thus, the wind erosion hazard of arable soils at N proves to be a more significant predictor than the proportion of agricultural land in the same radius.

Overall, it should be noted that the following descriptors can be regarded as meaningful additions to the extended set of variables compared to the GMS2005: land-use density in radii of 150 to 300 km, population density in extended radii 50 and 100 km, wind erosion hazard of arable land in radius 1 km, 3-year and 90-day mean of precipitation sum. As in GMS2005, in GMS2015 the distances to potential emission sources, the background values of the heavy metal content in topsoils, the distance to the sea, and the local main wind direction are of less importance as predictors.

#### **Conclusions**

Beyond correlation analysis, regression analysis was performed using random forest regression and multiple linear regression in connection with commonality analysis to quantify the statistical relations between atmospheric deposition of elements accumulated in moss samples and characteristics of sampling sites and their surroundings. This combined methodology revealed that strongest predictor for the content of Cd, Cu, Ni, Pb, Zn and N in moss samples was the sampled species. In 2015, the atmospheric deposition showed a lower predictive power compared to earlier campaigns. Uncertainties arise due to the specified inaccuracies of the analytical data, which can

significantly affect the validity of the conclusions, especially at low concentration levels. Regarding the statistical methods applied, it can be concluded that compared to individual decision trees such as CART, RF tends to be more robust to changes in training data set, outliers and over-adaptations due to random predictor and data selection. By allowing each tree to grow to maximum size, RF tries to maintain a certain predictive power. The associated problem of over-adjustment of each individual tree could be offset by randomizing the predictors. Analogous to the above advantages, the GMS2015 used RF instead of CART for the evaluation of substance accumulation in moss samples. The RF characteristic values of the variable weight compared to CART also offer the advantage of more direct comparability with those of multiple linear regression in conjunction with the commonality analysis used for the first time in environmental monitoring as exemplified by the GMS2015.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12302-019-0260-7>.

**Additional file 1.** Additional Figures S1–S14.

## Abbreviations

agr1–300: Spatial density of surrounding agricultural land use in radii of 1, 5, 10, 25, 50, 75, 100, 150, 200, 250, 300 km around the moss sampling point; Al: aluminium; As: arsenic; CA: commonality analysis; Ca: calcium; CARBIDE: heavy metal (heavy metal); CART: classification and regression trees; CC: commonality coefficient; Cd: cadmium; cd\_dep1315: modelled total cadmium deposition (EMEP; mean value of the years 2013 to 2015); CDC: Climate Data Center; CDistSea: distance to the sea (North Sea and Baltic Sea); CLC: Corine Land Cover; Cr: chromium; Cu: copper; DistShrubsAverage: distance of the sampling site to surrounding shrubs (mean value); DistShrubsMax: distance of the extraction site to surrounding shrubs (maximum); DistShrubsMin: distance of the extraction site to surrounding shrubs (minimum); DistTreeCrownsAverage: distance of the sampling site to the projection of the tree layer of surrounding forests and woody plants (mean value); DistTreeCrownsMax: distance of the sampling site to the projection of the tree layer of surrounding forests and shrubs (maximum); DistTreeCrownsMin: removal of the extraction site for projection of the tree layer of surrounding forests and shrubs (minimum); DWD: German Weather Service; ELCE: Ecological Land Classes of Europe; elev\_eu\_gk: orographic height; EMEP: European Monitoring and Evaluation Programme; EMS: European Moss Survey; Fe: iron; for1–300: spatial density of surrounding forest land use in radii of 1, 5, 10, 25, 50, 75, 100, 150, 200, 250, 300 km around the moss sampling point; frequency: frequency at place of growth; GIS: geographical information system; GMS: German Moss Survey; GPW: Gridded Population of the World; Hg: mercury; hg\_dep1315: modelled total mercury deposition (EMEP; mean value of the years 2013 to 2015); HM: heavy metals; HM\_dep1315: modelled total deposition of the respective heavy metal indicated (EMEP; mean value of the years 2013 to 2015); HM\_dep15: modelled total deposition of the respective heavy metal indicated (EMEP; 2015); HM\_Ob50P: background values of HM content in the topsoil (50th percentile); HumusLayer: thickness of the humus layer; HumusSpecies: humus species as for example mull or moder; Hypcup: *Hypnum cupressiforme*; ICP: International Cooperative Programme; LAI: leaf area index; LAI2: weighted leaf area index; LE: LOTOS-EUROS; LE\_all\_N: modelled total deposition of nitrogen (LOTOS-EUROS; average of the years 2013 to 2015); LE\_all\_NH4: modelled total deposition of reduced nitrogen (LOTOS-EUROS; average of the years 2013 to 2015); LE\_dryall\_N: modelled dry nitrogen deposition (LOTOS-EUROS; average of the years 2013 to 2015);

LE\_dryall\_NH4: modelled dry deposition of reduced nitrogen (LOTOS-EUROS; average of the years 2013 to 2015); LE\_occultall\_N: modelled occult deposition of nitrogen (LOTOS-EUROS; average of the years 2013 to 2015); LE\_occultall\_NH4: modelled occult deposition of reduced nitrogen (LOTOS-EUROS; average of the years 2013 to 2015); LE\_occultall\_NO3: modelled occult deposition of oxidized nitrogen (LOTOS-EUROS; average of the years 2013 to 2015); LE\_wet\_N: modelled wet deposition of nitrogen (LOTOS-EUROS; average of the years 2013 to 2015); LE\_wet\_NH4: modelled wet deposition of reduced nitrogen (LOTOS-EUROS; average of the years 2013 to 2015); LE\_wet\_NO3: modelled wet deposition of oxidized nitrogen (LOTOS-EUROS; average of the years 2013 to 2015); MainWindDirection: local main wind direction; Mg: magnesium; MLR: multiple linear regression; MossSpecies: moss species; MSC-E: Meteorological Synthesizing Center-East; MSC-W: Meteorological Synthesizing Center-West; MSE: mean squared error; n: sample size; N: nitrogen; N\_1315: total deposition of nitrogen, average from 2013 to 2015 (EMEP); N\_15: total deposition of nitrogen in 2015 (EMEP); N\_dry\_ox15: dry deposition of oxidized nitrogen in 2015 (EMEP); N\_tot\_ox15: total deposition of oxidized nitrogen in 2015 (EMEP); N\_tot\_red15: total deposition of reduced nitrogen in 2015 (EMEP); N\_wet\_ox15: wet deposition of oxidized nitrogen in 2015 (EMEP); N\_wet\_red15: wet deposition of reduced nitrogen in 2015 (EMEP); Ni: nickel; OOB: out-of-bag; p: level of significance; Pb: lead; pb\_dep1315: modelled lead deposition (EMEP) from 2013 to 2015 (EMEP); pegwind\_1–5: potential risk of erosion of arable soils by wind (1, 2, 5 km—radii); Plesch: *Pleurozium schreberi*; popdens5–100: population density within radius 5, 50, 100 km around the moss sampling point; pre1315\_gk: average rainfall for the years 2013 to 2015; pre3days: average rainfall in the last 3 days before the date of sampling; pre90days: average rainfall during the 90 days preceding the date of sampling; Psepur: *Pseudoscleropodium purum*; RF: random forest; r: correlation coefficient (Spearman); RSQ: pseudo-R<sup>2</sup>; SamplingFrom: vegetation (substrate: soil, tree stump); Sb: antimony; SD: standard deviation; SDistAgriculturalAreas: distance to agricultural land; SDistAnimalFarmingUnits: distance to agricultural estates; SDistCombustion-EnergyPlants: distance to thermal power plant; SDistConstructionSites: distance to construction sites; SDistDumpingGrounds: distance to landfills; SDistFederalRoads: distance to federal highways; SDistGravelPit: distance to gravel pit, sand and stone quarry; SDistIndustriesWithHighChimneys: distance to industrial plants with high chimneys; SDistMotorways: distance to motorways; SDistNoneVegetationAreas: distance to vegetation-free areas; SDistPloughedAgriculturalFields: distance to arable land; SDistRailroadTracks: distance to railway lines; SDistSingleHouses: distance to single houses; SDistSmallIndustries: distance to smaller industrial plants; SDistSmallPaved-CountryRoads: distance to national roads; SDistTown: distance to larger settlements; SDistUnsealedRoads: distance to district roads, etc.; SDistVillage: distance to smaller settlements; SDistWasteIncinerationFacilities: distance to waste incineration plants; ShrubCoverage: degrees of coverage of the shrub layer; SlopeDirection: exposure; SlopeGradient: slope inclination; SoilTexture: soil type; SRTM: Shuttle Radar Topography Mission; TreeCoverage: degrees of coverage of the tree layer; TSLayerHeightAverage: height of the tree layer of surrounding forests and woody plants (mean); TSLayerHeightMax: height of the tree layer of surrounding forests and woody plants (maximum); TSLayerHeightMin: height of the tree layer of surrounding forests and shrubs (minimum); UBA: umweltbundesamt; urb1–300: spatial density of surrounding urban-industrial land use in radii of 1, 5, 10, 25, 50, 75, 100, 150, 200, 250, 300 km around the moss collection point; VisibleDustParticles: visible particles; Weather: weather; Zn: zinc.

## Acknowledgements

We would like to thank the Federal Environment Agency (Dessau-Roßlau, Germany) for financial support and professional advice.

## Authors' contributions

WS headed the computations executed by SN. WS drafted the article. Both authors read and approved the final manuscript.

## Funding

Federal Environmental Agency, Dessau-Roßlau, Germany (FKZ 3715 63 212 0).

## Availability of data and materials

The datasets generated and/or analysed during the current study are not publicly available due to copyright but are available from the corresponding author on a reasonable request.

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

Received: 5 August 2019 Accepted: 21 September 2019

Published online: 12 October 2019

**References**

- Amodio M, Catino S, Dambruoso PR, de Gennaro G, Di Gilio A, Giungato P, Laiola E, Marzocca A, Mazzone A, Sardaro A, Tutino M (2014) Atmospheric deposition: sampling procedures, analytical methods, and main recent findings from the scientific literature. *Adv Meteorol*. <https://doi.org/10.1155/2014/161730>
- Berg T, Røyset O, Steinnes E (1995) Moss (*Hylocomium splendens*) used as biomonitor of atmospheric trace element deposition: estimation of uptake efficiencies. *Atmos Environ* 29:353–360
- Birke M, Rauch U, Raschka HUA (2007) Geochemischer Atlas Bundesrepublik Deutschland. Verteilung anorganischer und organischer Parameter in Oberflächenwässern und Bachsedimenten. 641 S
- Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
- Breiman L, Friedman J, Olshen R, Stone C (1984) Classification and regression trees. Wadsworth, Belmont
- Brosius F (2018) SPSS. Umfassendes Handbuch zu Statistik und Datenanalyse. mitp Verlags GmbH & Co, Frechen
- EAE (European Environment Agency) (2016) Corine land cover 2012 raster data <https://www.eea.europa.eu/data-and-maps/data/clc-2012-raster>. Accessed 13 Dec 2017
- Eisler R (2007) Eisler's encyclopedia of environmentally hazardous priority chemicals. Elsevier Science, 986 S
- Foulkes AS (2009) Applied statistical genetics with R: for population-based association studies. Auszug E-Book, Springer-Verlag, New York
- Harmens H, Norris DA, Cooper DM, Mills G, Steinnes E, Kubin E, Thöni L, Aboal JR, Alber R, Carballeira A, Coskun M, De Temmerman L, Frolova M, Frontasyeva M, González-Miqueo L, Jeran Z, Leblond S, Liiv S, Maňkóvá B, Pesch R, Poikolainen J, Rühling Å, Santamaria JM, Simoneie P, Schröder W, Suchara I, Yurukova L, Zechmeister HG (2011) Nitrogen concentrations in mosses indicate the spatial distribution of atmospheric nitrogen deposition in Europe. *Environ Pollut* 159:2852–2860
- Holy M, Leblond S, Pesch R, Schröder W (2009) Assessing spatial patterns of metal bioaccumulation in French mosses by means of an exposure index. *Environ Sci Pollut Res* 16(5):499–507
- Holy M, Pesch R, Schröder W, Harmens H, Ilyin I, Alber R, Aleksiyenak Y, Blum O, Coşkun M, Dam M, De Temmermann L, Fedorets N, Figueira R, Frolova M, Frontasyeva M, Goltsova N, González ML, Grodzinska K, Jeran Z, Korzekwa S, Krmár M, Kuni E, Kvietkus K, Larsen M, Leblond S, Liiv S, Magnússon S, Maňkóvá B, Mocanu R, Piispanen J, Rühling Å, Santamaria JM, Steinnes E, Suchara I, Thöni L, Turcsányi G, Urumov V, Wolterbeek HT, Yurukova L, Zechmeister HG (2010) First thorough identification of factors associated with Cd, Hg and Pb concentrations in mosses sampled in the European Surveys 1990, 1995, 2000 and 2005. *J Atmos Chem* 63:109–124
- ICP Vegetation (International Cooperative Programme on Effects of Air Pollution on Natural Vegetation and Crops) (2014) Monitoring of atmospheric deposition of heavy metals, nitrogen and POPs in Europe using bryophytes. Monitoring manual 2015 survey. United Nations Economic Commission for Europe Convention on Long-Range Transboundary Air Pollution. ICP Vegetation Moss Survey Coordination Centre, Dubna, Russian Federation, and Programme Coordination Centre. Bangor, Wales, UK
- Keil M, Kiefl R, Strunz G (2005) CORINE land cover 2000—Germany. Oberpfaffenhofen (Final Report. German Aerospace Center, German Remote Sensing Data Center)
- LABO (2017) Background values for inorganic and organic substances in soils. In: Rosenkranz et al. (ed) Bodenschutz Supplementares Handbuch der Maßnahmen und Empfehlungen für Schutz, Pflege und Sanierung von Böden, Landschaft und Grundwasser. Erich Schmidt Publishers, Berlin
- Lazo P, Qarri F, Allajbeu S, Bektashi L, Stafilov T (2018) Temporal and spatial distribution of multi-element atmospheric deposition in Albania (2010–2015 Moss Survey). In: Harmens H, Mills G (eds) 31th ICP vegetation task force meeting: 5th–8th March, Dessau-Roßlau, Germany. ICP Vegetation Coordination Centre. Centre for Ecology & Hydrology, Bangor
- Lazo P, Steinnes E, Qarri F, Allajbeu S, Kane S, Stafilov T, Frontasyeva MV, Harmens H (2018) Origin and spatial distribution of metals in moss samples in Albania: a hotspot of heavy metal contamination in Europe. *Chemosphere* 190:337–349
- Lequy E, Dubos N, Witte I, Pascaud A, Sauvage S, Leblond S (2017) Assessing temporal trends of trace metal concentrations in mosses over France between 1996 and 2011: a flexible and robust method to account for heterogeneous sampling strategies. *Environ Pollut* 220:828–836
- Lequy E, Saby NP, Ilyin I, Pascaud A, Sauvage S, Leblond S (2017) Spatial analysis of trace elements in a moss bio-monitoring data over France by accounting for source, protocol and environmental parameters. *Sci Total Environ* 590–591:602–610
- Liaw A, Wiener M (2002) Classification and regression by random forest. *R News* 2(3):18–22
- Loupe G, Wehenkel L, Sutera A, Geurts P (2013) Understanding variable importances in forests of randomized trees. <http://papers.nips.cc/paper/4928-understanding-variable-importances-in-forests-of-randomized-trees.pdf>. Accessed 24 Jan 2018
- Manders AMM, Bultjes PJH, Curier L, Denier van der Gon HAC, Hendriks C, Jonkers S, Kranenburg R, Kuenen JJP, Segers AJ, Timmermans RMA, Visschedijk AJH, Wichink Kruit RJ, van Pul WAJ, Sauter FJ, van der Swaluw E, Swart DPJ, Douros J, Eskes H, van Meijgaard E, van Ulfst B, van Velthoven P, Banzhaf S, Mues AC, Stern R, Fu G, Lu S, Heemink A, van Velzen N, Schaap M (2017) Curriculum vitae of the LOTOS–EUROS (v2.0) chemistry transport model. *Geosci Model Dev* 10:4145–4173. <https://doi.org/10.5194/gmd-10-4145-2017>
- Meyer M (2017) Site-specifically differentiated recording of atmospheric nitrogen and heavy metal inputs by means of mosses with consideration of the eaves effect and supplementary investigations on the relationship between nitrogen inputs and accompanying vegetation. *Diss Univ Vechta* 1–262 + 86 S. <http://dx.doi.org/10.23660/voado-16>. Accessed 24 June 2019
- Meyer M, Schröder W, Pesch R, Steinnes E, Uggerud HT (2014) Multivariate association of regional factors with heavy metal concentrations in moss and natural surface soil sampled across Norway between 1990 and 2010. *J Soils Sediments* 14(11):1–15
- Mood AM (1971) Partitioning variance in multiple regression analyses as a tool for developing learning models. *Am Educ Res J* 8:191–202
- Nickel S, Schröder W (2017) Reorganisation of a long-term monitoring network using moss as biomonitor for atmospheric deposition in Germany. *Ecol Ind* 76:194–206
- Nickel S, Schröder W, Wosniok W, Harmens H, Frontasyeva MV, Alber R, Aleksiyenak J, Barandovski L, Blum O, Danielsson H, de Temmermann L, Dunaev AM, Fagerli H, Godzik B, Ilyin I, Jonkers S, Jeran Z, Pihl Karlsson G, Lazo P, Leblond S, Liiv S, Mankovska B, Martínez-Abaigar J, Piispanen J, Poikolainen J, Popescu IV, Qarri F, Radnovic D, Santamaria JM, Schaap M, Skudnik M, Špiric Z, Stafilov T, Steinnes E, Stihl C, Suchara I, Thöni L, Uggerud HT, Zechmeister HG (2017) Modelling and mapping heavy metal and nitrogen concentrations in moss in 2010 throughout Europe by applying Random Forests models. *Atmos Environ* 156:146–159
- Nimon KF, Lewis M, Kane R, Haynes RM (2008) An R package to compute commonality coefficients in the multiple regression case: an introduction to the package and a practical example. *Behav Res Methods* 40:457–466
- Nimon KF, Oswald FL (2013) Understanding the results of multiple linear regression: beyond standardized regression coefficients. *Organ Res Methods*. <https://doi.org/10.1177/1094428113493929>
- Pedhazur EJ (1997) Multiple regression in behavioral research: explanation and prediction, 3rd edn. HarcourtBrace, Fort Worth
- Pesch R, Schröder W, Genssler L, Goeritz A, Holy M, Kleppin L, Matter Y (2007) Moss monitoring 2005/2006: heavy metals IV and total nitrogen. Berlin (Environmental Research Plan of the Federal Minister for the Environment, Nature Conservation and Nuclear Safety. R&D project 205 64 200, final report, on behalf of the Federal Environment Agency); 90 p.,

- 11 Table, 2 Figure (text section); 51 p. + 41 maps, 34 tables, 46 diagrams (appendix)
32. Prasad AM, Iverson LR, Liaw A (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* 9:181–199
  33. Ray-Mukherjee J, Nimon K, Mukherjee S, Morris DW, Slotow R, Hamer M (2014) Using commonality analysis in multiple regressions: a tool to decompose regression effects in the face of multicollinearity. *Methods Ecol Evol* 5:320–328
  34. R Core Team (2013). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>. Aufgerufen am. Accessed 17 Dec 2018
  35. Rauthe M, Steiner H, Riediger U, Mazurkiewicz A, Gratzki A (2013) A Central European precipitation climatology—part I: generation and validation of a high-resolution gridded daily data set (HYRAS). *Meteorol Z* 22(3):235–256. <https://doi.org/10.1127/0941-2948/2013/0436>
  36. Sachs L, Hedderich J (2009) Applied statistics. Method collection with R. Springer, Berlin
  37. Schaap M, Hendriks C, Kranenburg R, Kuenen J, Segers A, Schlutow A, Nail HD, Ritter A, Banzhaf S (2018) PINETI-III: modelling and mapping of atmospheric inputs from 2000 to 2015 for the assessment of the ecosystem-specific threat to biodiversity in Germany. Final Report FKZ 3714 64 2010149 Federal Environment Agency, Dessau-Roßlau, pp 149. [https://www.umweltbundesamt.de/sites/default/files/medien/1410/publikationen/2018-10-17\\_texte\\_79-2018\\_pineti3.pdf](https://www.umweltbundesamt.de/sites/default/files/medien/1410/publikationen/2018-10-17_texte_79-2018_pineti3.pdf). Accessed 24 June 2019
  38. Schaap M, Roemer M, Sauter F, Boersen G, Timmermans R, Builjes PJH, Vermeulen AT (2005) LOTOS-EUROS: Documentation, TNO report B&O-A R 2005/297
  39. Schaap M, Timmermans RMA, Roemer M, Boersen GAC, Bultjes PJH, Sauter FJ, Velders GJM, Beck JP (2008) The LOTOS-EUROS model: description, validation and latest developments. *Int J Environ Pollut* 32(2):270–290
  40. Schröder W, Holy M, Pesch R, Harmens H, Fagerli H, Alber R, Coşkun M, De Temmermann L, Frolova M, González-Miqueo L, Jeran Z, Kubin E, Leblond S, Liiv S, Mankovská B, Piispanen J, Santamaria JM, Simonè P, Suchara I, Yurukova L, Thöni L, Zechmeister HG (2010) First Europe-wide correlation analyses identifying factors best explaining the total nitrogen concentration in mosses. *Atmos Environ* 44:3485–3491
  41. Schröder W, Hornsman I, Pesch R, Schmidt G, Fränze S, Wünschmann S, Heidenreich H, Markert B (2008) Nitrogen and metal accumulation in mosses of two central European regions as a mirror of their land use? *Environ Sci Pollut Res* 20:62–74
  42. Schröder W, Nickel S (2018) Site-specific investigation and spatial modelling of canopy drip effect on element concentrations in moss. *Environ Sci Pollut Res* 25(27):27173–27178
  43. Schröder W, Nickel S (2019) Spatial structures of heavy metals and nitrogen accumulation in moss specimens sampled between 1990 and 2015 throughout Germany. *Environ Sci Eur* 31(33):1–15 + Suppl:1–8. <https://doi.org/10.1186/s12302-019-0216-y>
  44. Schröder W, Nickel S, Jenssen M, Riediger J (2015) Methodology to assess and map the potential development of forest ecosystems exposed to climate change and atmospheric nitrogen deposition: a pilot study in Germany. *Sci Total Environ* 521–522:108–122
  45. Schröder W, Nickel S, Völksen B, Dreyer A, Wosniok W (2019) Nutzung von Bioindikationsmethoden zur Bestimmung und Regionalisierung von Schadstoffeinträgen für eine Abschätzung des atmosphärischen Beitrags zu aktuellen Belastungen von Ökosystemen. B1:1–189, Bd. 2:1–296. UBA-Texte 91/2019
  46. Schröder W, Pesch R (2007) Synthesizing bioaccumulation data from the German metals in mosses surveys and relating them to ecoregions. *Sci Tot Environ* 374:311–327
  47. Schröder W, Pesch R, Schmidt G (2007) Statistical classification of terrestrial and marine ecosystems for environmental planning. *Landscape Online* 2:1–22. <https://www.landscapeonline.de/103097lo200702>. Accessed 24 June 2019
  48. Schröder W, Schmidt G, Hornsman I (2006) Landschaftsökologische Raumgliederung Deutschlands. In: Fränze O; Müller, F; Schröder W (Hrsg) *Handbuch der Umweltwissenschaften. Fundamentals and applications of ecosystem research*. Landsberg am Lech, Munich, Zurich, Chapter V-1.9, 17 Erg.Lfg., pp. 1–100
  49. Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples). *Biometrika* 52(3–4):591–611
  50. Siewers U, Herpin U, Strassburger S (2000) Heavy metal entries in Germany. Moos-monitoring 1995 part 2: geological yearbook, special issues, issue SD 3, Stuttgart: Bornträger
  51. Simpson D, Benedictow A, Berge H, Bergström R, Emberson LD, Fagerli H, Flechard CR, Hayman GD, Gauss M, Jonson JE, Jenkin ME, Nyiri A, Richter C, Semeena VS, Tsyro S, Tuovinen J-P, Valdebenito A, Wind P (2012) The EMEP MSC-W chemical transport model; technical description. *Atmos Chem Phys* 12(16):7825–7865
  52. Steinnes E (1995) A critical evaluation of the use of naturally growing moss to monitor the deposition of atmospheric metals. *Sci Total Environ* 160/161:243–249
  53. Steinnes E, Rühling A, Lippo H, Mäkinen A (1997) A Reference materials for large-scale metal deposition surveys. *Accredit Qual Assur* 2:243–249
  54. Thompson B (2006) Foundations of behavioral statistics: an insight-based approach. Guilford Press, New York
  55. Travnikov O, Ilyin I (2005) Regional model MSCE-HM of heavy metal trans-boundary air pollution in Europe. EMEP/MSCE Technical Report 6/2005, p 59. [http://en.msceast.org/reports/6\\_2005.pdf](http://en.msceast.org/reports/6_2005.pdf). Accessed 24 June 2019
  56. Williams G (2011) Data mining with rattle and R. The art of excavating data for knowledge discovery. Springer, New York, p 374
  57. Zechmeister HG, Grodzinska K, Szarek-Lukaszewska GH (2003) Bryophytes. In: Markert B et al (eds) *Bioindicators and Biomonitors*. Elsevier, Amsterdam, pp 329–375

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)