Nickel, Stefan; Schröder, Winfried; Wosniok, Werner; Harmens, Harry; Frontasyeva, Marina V.; Alber, Renate; Aleksiayenak, Julia; Barandovski, Lambe; Blum, Oleg; Danielsson, Helena; de Temmerman, Ludwig; Dunaev, Anatoly M.; Fagerli, Hilde; Godzik, Barbara; Ilyin, Ilia; Jonkers, Sander; Jeran, Zvonka; Karlsson, Gunilla Pihl; Lazo, Pranvera; Leblond, Sebastien; Liiv, Siiri; Magnússon, Sigurður H.; Mankovska, Blanka; Martínez-Abaigar, Javier; Piispanen, Juha; Poikolainen, Jarmo; Popescu, Ion V.; Qarri, Flora; Radnovic, Dragan; Santamaria, Jesus Miguel; Schaap, Martijn; Skudnik, Mitja; Špirić, Zdravko; Stafilov, Trajce; Steinnes, Eiliv; Stihi, Claudia; Suchara, Ivan; Thöni, Lotti; Uggerud, Hilde Thelle; Zechmeister, Harald G.. 2017. **Modelling and mapping heavy metal and nitrogen concentrations in moss in 2010 throughout Europe by applying Random Forests models**.

Research paper

Modelling and mapping heavy metal and nitrogen concentrations in moss in 2010 throughout Europe by applying Random Forests models

Stefan Nickel*, Winfried Schröder*, Werner Wosniok**, Harry Harmens***, Marina V. Frontasyeva****, Renate Alber, Julia Aleksiayenak, Lambe Barandovski, Oleg Blum, Helena Danielsson, Ludwig de Temmermann, Anatoly M. Dunaev, Hilde Fagerli, Barbara Godzik, Ilia Ilyin, Sander Jonkers, Zvonka Jeran, Gunilla Pihl Karlsson, Pranvera Lazo, Sebastien Leblond, Siiri Liiv, Blanka Mankovska, Javier Martínez-Abaigar, Juha Piispanen, Jarmo Poikolainen, Ion V. Popescu, Flora Qarri, Dragan Radnovic, Jesus Miguel Santamaria, Martijn Schaap, Mitja Skudnik, Zdravko Špirić, Trajce Stafilov, Eiliv Steinnes, Claudia Stihi, Ivan Suchara, Lotti Thöni, Hilde Thelle Uggerud, Harald G. Zechmeister

*Chair of Landscape Ecology, University of Vechta, POB 15 53, 49377 Vechta, Germany, Tel.: +49(0) 4441 15 420 (Secretary), Fax: +49(0) 4441 15 583, E-mail: stefan.nickel@uni-vechta.de, winfried.schroeder@uni-vechta.de

** University of Bremen, POB 330 440, 28334 Bremen, Tel.: +49(0) 421 218 63780 (Secretary), Fax: +49(0) 421 218 63799, E-mail: wwosniok@math.uni-bremen.de

*** ICP Vegetation Programme Coordination Centre, Centre for Ecology and Hydrology, Bangor, Gwynedd LL57 2UW, United Kingdom, Tel.: +44(0)1248 374512/374500, Fax: +44(0)1248 362133, E-mail: hh@ceh.ac.uk,

**** Moss Survey Coordination Centre, Frank Laboratory of Neutron Physics, Moscow Region, Russian Federation, Tel.: +7(49621)65609, Fax: +7(49621)65085, E-mail: mfrontasyeva@jinr.ru

Abstract

Objective. This study explores the statistical relations between the concentration of nine heavy metals (HM) (arsenic (As), cadmium (Cd), chromium (Cr), copper (Cu), mercury (Hg), nickel (Ni), lead (Pb), vanadium (V), zinc (Zn)), and nitrogen (N) in moss and potential explanatory variables (predictors) which were then used for mapping spatial patterns across Europe. Based on moss specimens collected in 2010 throughout Europe, the statistical relation between a set of potential predictors (such as the atmospheric deposition calculated by use of two chemical transport models (CTM), distance from emission sources, density of different land uses, population density, elevation, precipitation, clay content of soils) and concentrations of HMs and nitrogen (N) in moss (response variables) were evaluated by the use of Random Forests (RF) and Classification and Regression Trees (CART). Four spatial scales were regarded: Europe as a whole, ecological land classes covering Europe, single countries participating in the European Moss Survey (EMS), and moss species at sampling sites. Spatial patterns were estimated by applying a series of RF models on data on potential predictors covering Europe. Statistical values and resulting maps were used to investigate to what extent the models are specific for countries, units of the Ecological Land Classification of Europe (ELCE), and moss species.

Results. Land use, atmospheric deposition and distance to technical emission sources mainly influence the element concentration in moss. The explanatory power of calculated RF models varies according to elements measured in moss specimens, country, ecological land class, and moss species. Measured and predicted medians of element concentrations agree fairly well while minima and maxima show considerable differences. The European maps derived from the RF models provide smoothed surfaces of element concentrations (As, Cd, Cr, Cu, N, Ni, Pb, Hg, V, Zn), each explained by a multivariate RF model and verified by CART, and thereby more information than the dot maps depicting the spatial patterns of measured values.

Conclusions. RF is an eligible method identifying and ranking boundary conditions of element concentrations in moss and related mapping including the influence of the environmental factors.

Keywords. Atmospheric deposition, biomonitoring, Ecological Land Classification Europe, spatial reference systems

1 Introduction

Enhanced atmospheric deposition and correlated concentrations of HM and N may cause serious problems for human health and ecosystem integrity (Bobbink et al. 2010). The degree of pollution may be explored by determining element concentrations in the air, water, soil, or sediments. Alternatively, or complementarily, monitoring organisms (bioindicators, biomonitors) are used for monitoring and mapping spatial patterns of element concentrations (Markert at al. 2003) or further analysis, e.g. by use of multivariate analysis (Factor analysis (FA) and / or Principal component analysis (PCA)) (Špirić et al 2013). Such organisms might accumulate many elements to measurable concentrations indicating an average degree of pollution over time. The concentration in the monitoring organisms reflects the element fraction available for uptake by organisms (Bjerregaard 2015). Mosses used for this study are ectohydric and absorb water over the plant surface. Mosses receive and accumulate elements directly from the atmosphere via wet, occult and dry atmospheric deposition (Glime 2006). Therefore, chemical analyses of moss specimens provide a surrogate, time-integrated measure of the spatial patterns of element deposition. Biomonitoring using mosses is easier and cheaper than deposition sampling with technical devices so that a much higher spatial sampling density can be achieved. Especially for HM, experimental data for occult and dry deposition are hardly available, and also data for dry deposition of N are very limited. Concentrations of various key metals in moss have been successfully calibrated versus atmospheric deposition levels of the same metals (Berg and Steinnes 1997). Although the moss concentration data provide no direct quantitative measurement of total deposition, this information can be derived by statistical approaches relating element concentrations in mosses to measured element concentrations in atmospheric deposition (Harmens et al. 2010, 2011, 2015). This way, the spatial resolution of atmospheric HM and N deposition maps can be enhanced (Schröder et al. 2011 a, 2011 b, 2012, 2014). Since 1990, every five years the European Moss Surveys (EMS) have been providing data on concentrations of HM, and since 2005 concentrations of N in moss (Harmens et al. 2015). Sampling, chemical analyses and quality control of data were performed according to a standardized protocol (Moss Manual ICP Vegetation for EMS 2010).

The EMS 2010 provided data on concentrations of aluminium (Al), arsenic (As), cadmium (Cd), chromium (Cr), copper (Cu), iron (Fe), lead (Pb), mercury (Hg), nickel (Ni), antimony (Sb), vanadium (V), zinc (Zn) and nitrogen (N) in moss, collected at up to 4499 sample sites across 24 European countries. In the present study data, on As, Cd, Cr, Cu, Hg, Ni, Pb, V, Zn, and N concentrations were used (Harmens et al. 2015) because modelled atmospheric deposition data were available for these compounds. As holds true for the EMS, the Co-operative Programme for Monitoring and Evaluation of Long-range Transmission of Air Pollutants in Europe (EMEP) is a part of the United Nations Economic Commission for Europe (UNECE) in the framework of the Convention on Long-range Transboundary Air Pollution. EMEP uses emission data from the European countries to model atmospheric transport and deposition of Cd, Hg, N, and Pb with a grid size of 50 km by 50 km. The modelling results are validated against measurements of Cd, Hg, Pb, and N concentrations in atmospheric particulate matter and wet deposition collected with technical devices at up to 70 sites across Europe (Tørseth et al. 2012).

The aim of this study was to analyze the multivariate statistical relations between concentrations of As, Cd, Cr, Cu, Hg, Ni, Pb, Zn and N in moss and potential explanatory variables. To reach this aim, following objectives were investigated by:

- identifying and ranking the selected explanatory variables which are most important in explaining the spatial variation of element concentration in mosses using RF;
- comparing the results of between RF and CART and, by this, derive conclusions on how the method may influence the results (this is a matter of quality control);
- preparing and evaluating the maps using regression prediction based on RF results;
- exploring by use of RF models and the resulting maps, to what extent the statistical relations between element concentration in moss and selected explanatory variables are specific for elements, moss species, countries and ecological land classes.

Using the above mentioned data from Europe, the statistical relations of potential predictors such as modelled atmospheric deposition, distance from respective emission sources, elevation, density of various land uses, population density, precipitation, and clay content of soils with response variables (HM and N concentration in moss, respectively) were evaluated using Random Forests (Breiman 2001; Liaw and Wiener 2002). Areas of RF application are, amongst others, astronomy, autopsy, transport planning, medicine, and environmental sciences (Fawagreh et al. 2014). Examples for the latter category were given by Cianci et al. (2015), Evans and Cushman (2009), Howard et al. (2014) and Magness et al. (2008) predicting species, Deloncle et al. (2007) predicting weather regimes, Pal (2005) classifying forests and Thums et al. (2008) marine species, Rothwell et al. (2008) evaluating the key environmental drivers controlling N leaching from European forests, Spekkers et al. (2015) predicting flood damage, and Mascaro et al. (2014) mapping forest carbon. Based on the data obtained across Norway, Meyer et al. (2015 a) have shown the predictive relevance for a similar set of regional factors for Cd, Hg, and Pb concentrations in moss by using CART (Breiman et al. 1984). Contrary to CART, RF in conjunction with the Geographic Information System (GIS) used here were additionally applied for regression mapping, i.e. transforming spatial information of the independent variables to continuous surfaces and respective maps of HM and N concentration in mosses. RF models explain the spatial patterns of HM concentrations in moss and, together with the maps, were used to investigate to what extent the statistical relations between element concentration in moss and selected explanatory variables are specific for elements, moss species, countries and ecological land classes of Europe (ELCE, **Figure S1** and **Table S1** in the supplement). The latter describes the spatial pattern of 40 land classes, defined by characteristic values of 48 ecological attributes (Hornsmann et al. 2008; Schröder and Pesch 2007).

2 Material and Methods

2.1 Data

Statistical relations between atmospheric deposition of HM and N derived from the numeric chemical transport models (CTMs) LOTOS-EUROS (LE; Builtjes et al. 2016; Schaap et al. 2008) and EMEP (Simpson et al. 2012, Travnikov and Ilyin 2005) and respective data on element concentrations in moss were examined by use of RF with data compiled in **Table 1**. The EMEP CTM provides Europe-wide atmospheric deposition data of Pb, Cd, Hg and N calculated on a grid of 50 km by 50 km. Following Harmens et al. (2012) we used the three year sum of HM deposition modelled by EMEP as a corresponding parameter to the HM concentration in the sampled 3-year old shoots of the mosses (here the period of 2008-10 represents the base year of 2010). Furthermore, the three year sum of modelled deposition rates for $NH_4^+$and NO_3^- were used as total atmospheric deposition. Also 3-year sums of deposition from the CTM LOTOS-EUROS were used, only available for the time period 2009-11. LE provides deposition rates for As, Cd, Cr, Cu, Ni, Pb, V, and Zn on a 25 km by 25 km grid covering Europe. Additional information about the CTM is given in the supplement (**Table S2**). For examining further influences of spatial relations between emission sources and EMS sites, distances were calculated by means of Geographic Information System (GIS) based on element-specific data from the European Pollutant Release and Transfer Register (E-PRTR; EEA 2016 a). With regard to influences of different land uses pattern around the moss sampling sites, percentages of agricultural, forestry and urban land uses within a radius of 1, 5, 10, 25, 50, 75, and 100 km, derived from CORINE Land Cover 2006 (EEA 2016 b) and Global Land Cover (EEA 2016 c), were calculated. Population density was integrated using grid data at a resolution of 100 km by 100 m (SEDAC 2016). Elevation was included from the Digital Elevation Model (DEM, 90 m by 90 m) of the Shuttle Radar Topography Mission (SRTM 2016) and, respectively, precipitation from New et al. (2002) with a grid size of 20 km by 20 km. Besides that clay content (FAO 2009) was added due to its significance for HM binding capacity of soil.

Table 1. Potential predictors for HM and N concentration in moss in 2010

2.2 Modelling and mapping

The statistical relations between element concentrations in moss and potential explanatory variables were modelled by the use of RF (Breiman 2001) from which, then, surface maps of element concentrations were derived. These maps were compared to the site-specific measurements of element concentrations in moss. The RF regarded four spatial scales: Europe as a whole, ecological land classes covering Europe, single countries participating in the EMS, and moss species at sampling sites. RF are used to construct a prediction rule and to assess and rank variables with respect to their ability to predict the response variable. If the RF minimizes a squared error, normal distribution is not an essential requirement. But extremely asymmetric error distributions reduce the quality of predictions and make e.g. the difference between mean and median prediction important.The ranking is done by considering variable importance measures computed for each predictor. These relative measures as pure numbers without unit identify and rank predictors. After validation, the resulting prediction rule can then be applied, e.g. for mapping element concentrations in environmental compartments such as soil or moss. RF can cope with high dimensional data and can even be applied to highly correlated predictors, is not based on a particular stochastic model and can also capture nonlinear association patterns between predictors and the response. RF is a classification and regression technique aggregating a large number of decision trees. Several trees constructed from a training data set yield a prediction of the response. Variants of RF are characterized by the procedure used to generate the modified data sets on which each individual tree is constructed, and the way the predictions of each individual tree are aggregated to produce a unique consensus prediction. In the original RF method (Breiman 2001), each tree is a standard classification or regression tree (CART) (Breiman et al. 1984) using the decrease of Gini impurity, i.e. the degree of heterogeneity of a variable measured by the Gini index as a splitting criterion and selecting the splitting predictor from a randomly selected subset of predictors. Each tree is constructed using a bootstrap sample from the original data set, and the predictions of all trees are finally aggregated. This version of RF is implemented in most of the available software. Boulesteix et al. (2012) and Fawagreh et al. (2014) compiled and reviewed RF implementations and their features. Internal validation is calculated in terms of the out-of-bag (OOB) error: Each observation is an OOB observation for some of the trees, i.e., it was not used to construct them. The OOB error is the average error frequency obtained when the observations from the data set are predicted using the trees for which they are OOB. Thus, Random Forests are ensembles of multiple decision trees combined into a single model. Compared with single decision trees, like CART, RF tends to be more robust to outliers and overfitting (Williams 2011; Ziegler and König 2014). Verikas et al. (2011) surveyed respective literature and presented comparatively several tests. CART models are prone to overfitting data, which can lead to predictive errors. RF models reduce the over-fitting problem. Instead of building a single predictive tree model from all available data, RF builds typically 500 to 2000 trees (Prasad et al. 2006), using randomized subsets of data and explanatory variables to build each tree. The number of predictors used to find the best split at each node is a randomly chosen subset of the total number of predictors. The RF trees are grown to maximum size without pruning, and aggregation is performed by averaging the trees. Out-of-bag samples can be used to calculate an unbiased error rate and variable importance. Because a large number of trees are grown, there is limited generalization error (i.e., the true error of the population opposed to the training error only). The impossibility of overfitting is a very useful feature for prediction. By growing each tree to maximum size without pruning and selecting only the best split among a random subset at each node, RF tries to maintain some prediction strength while inducing diversity among trees (Breiman 2001). Random predictor selection diminishes correlation among unpruned trees and keeps the bias low. By taking an ensemble of unpruned trees, variance is also reduced. Another advantage of RF is that the predicted output depends only on one user-selected parameter, with the number of predictors to be chosen randomly at each node. This process of internal cross-validation prevents from over-fitting inherent to a single CART model (Breiman 2001).In this investigation RF for the first time was used to explain and map the geographical distribution of atmospheric deposition accumulated in moss throughout Europe.

2.3 Workflow

All statistical analysis and modelling were implemented in R (R Core Team 2013, Williams 2011). Analyses with HM and N concentrations in moss as response variable were based on a reasonably large sample size of at least 2154 and at maximum 3664 out of 4499 sampling points. Observations were partitioned into training datasets (90%), which were used to build the RF models, and independent test datasets (10 %) for measuring the quality of the RF models. Some observations from the training datasets had to be excluded from the analyses due to missing information on predictor variables. Based on the null hypothesis principle, the Shapiro-Wilk-test (Shapiro and Wilk 1965) was used to assess whether concentrations of HM and N in moss as response variables were normally or lognormally distributed. In all cases, target variables were log-transformed due to a non-normal distribution. For deciding the number of trees to build, plots of error rates progressively calculated against the number of trees were used. Observations with missing values were removed from the dataset. The number of variables to consider at each split was defined as one-third of the number of predictors (Williams 2011, p. 263). Models were then optimized using measures for relative variable importance as Increased Node Purity and model accuracy as Pseudo R Squared. Increased Node Purity represents the total increase of decision treenode's purity when splitting the dataset. It is measured for a specific variable as the mean increase of the Gini index over all trees according to Equation 1 (Louppe et al. 2013).

$$\text{Impurity}(X_m) = \frac{1}{N_T} \sum_{T} \sum_{t \in T: v(s_t) = X_m} p(t) \Delta i(s_t, t)$$

Equation 1

with: Xm = predictor variable; where X_m is used; i(t) = impurity measure (here: Gini index);
t = nodes, N_T = trees in the forest; $v(s_t)$ = variable used in the split; s_t = split;
T = Tree; p(t) = proportion No. of N_T / No. of samples reaching t

Pseudo R^2 were calculated as the square of the correlation between the predicted and observed values (**Equation 2**), thus, measuring the quality of the model (Liaw and Wiener 2002).

$$\text{Pseudo } R^2 = \left(\frac{S_{xy}}{\sqrt{S_x * S_y}}\right)^2 \qquad \textbf{Equation 2}$$

with: x = observed values; y = predicted values; S_{xy} = Covariance of x and y;
S_x = Standard deviation of x; S_y = Standard deviation of y

Non-significant predictors were stepwise eliminated from the models using a top-down approach. The statistical measures were used for comparison of the full and reduced models to find the optimum model including only those independent variables which explained a higher proportion of the variance in the data. Predicted Versus Observed plots including Pseudo R^2 (R Core Team 2013) were inspected for deciding which model yields the best fit to the data. To minimize limitations of the use of Pseudo R^2 (**Equation 2**) for comparing different models measure was preferably compared for the same outcome variables (specifically for elements) and number of observations in the test datasets were set to a minimum of N = 30. Following Liu et al. (2014), the goodness of fit of the predictions modelled to the test data were evaluated by use of mean and standard deviation of Pseudo R^2 based on multiple (in this investigation 10) runs. Since a Pseudo R^2 does not rely on linear relationships between predictors and the response, it could not be tested for significance similar to R^2, for which the p-value is usually calculated by means of F-statistics testing whether the null hypothesis R^2=0 (Wood 1990).

Finally, optimized RF models were applied on available spatial information yielding regression maps as results. All geographic information on the predictor variables such as HM and N deposition, emission, climate, altitude, population and land use features available with blanket coverage of participating countries or regions, respectively, were combined by means of classical GIS functions (overlay, spatial join) as implemented in ESRI's ArcGIS 10.2. Based on this, RF predictive models were applied to calculate a corresponding number of predictive maps, which result from reasonable combinations of HM and N concentration in moss as dependent variable with relevant predictor variables (**Table 2**) covering Europe at a spatial resolution of 10 km by 10 km for the year 2010. For some regions, element concentrations in moss could not be calculated due to missing data on predictor variables (e.g. Russia, Iceland, and Belarus).

3 Results and discussion

Since for the atmospheric deposition different data sources were used, i.e. EMEP and LOTOS-EUROS modelling results, the R^2 values of the respectively different RF models were compared for Pb and Cd for which both deposition models produce data. The comparison yielded higher R^2 values for RF models based on EMEP deposition values than those RF models relying on LE results (Pb_{EMEP}: $R^2 = 0.68$; Pb_{LE}: $R^2 = 0.63$; Cd_{EMEP}: $R^2 = 0.61$; Cd_{LE}: $R^2 = 0.58$). This corresponds to higher correlation coefficients (Spearman) between HM deposition and respective HM concentration in moss (Pb_{EMEP}: $r = 0.70$; Pb_{LE}: $r = 0.64$; Cd_{EMEP}: $r = 0.66$; Cd_{LE}: $r = 0.65$, $p < 0.01$). Thus, **Table 2 and 3** provide only information on modelled deposition as predictor in RF models with the highest R^2 for Cd and Pb. The lower rank of LE fields could be explained partly the lower quality of the emission data for this group of elements. Since both models have used different emission data (Builtjes et al. 2016), the comparison does not provide an indication for model quality.

Table 2. Relative importance of predictors for measured element concentrations in moss sampled across Europe as quantified by Increased Node Purity

The highest importance for concentrations of HM and N at the European level, measured as Increased Node Purity, could be found for land use within a 100 km radius around the sample sites, atmospheric deposition, distance from emission sources and precipitation (**Table 2, Figure 1**). Land use influences the regional emission and the site-specific atmospheric deposition which, according to the respective vegetation cover, may further be influenced by canopy drip. This regional spatial trend is in agreement with the auto-correlation range detected by use of variogram analysis and suggests that the large-scale variation (100 km) exceeds the small-scale variation (1 km), regardless of site-specific considerations (Meyer et al. 2015 b). For this reason, predictors with lower radii have been excluded from further modelling specified for different spatial levels (**Tables 4-6**). As shown in the maps for the whole Europe, variation of observed HM and N concentrations in moss are generally wider than the ranges of the predicted values (**Figures 2-5, Figures S2-S7)**, i.e., the RF models cause a smoothing effect on the respective response variable. The models for the European level (**Table 3**) correspond to and explain the maps for Cd, Hg, Pb and N depicted in **Figures 2-5** and for As, Cr, Cu, Ni, V, and Zn in **Figures S2-S7** (supplementary materials). Reasonable numbers of trees were between 200 and 300. Further increases had no significant influence on the results. From **Tables 2-3** it is obvious that land use and atmospheric deposition are the most meaningful predictors for the element concentrations measured in moss. Adding Increased Node Purity for all elements, highest values can be found for atmospheric deposition derived from EMEP for Cd, Pb and N, density of agricultural land use within a 75 (Cr) and 100 km (all elements except Cr) radius, density of forestry land use within a 100 km radius (all elements) and distance between sampling site and the nearest HM or N emission source followed by atmospheric deposition that is derived from LOTOS-EUROS for other metals, density of agricultural land use within a 50 km radius, density of forestry land use within a 75 km radius, density of urban land use within a 100 km radius, elevation and precipitation. Accordingly, their relevance for mapping element concentrations for unsampled locations (**Figures 2-5** and **S2-S7**) is particularly high. Contrary to surface maps derived by interpolation techniques such as Kriging, **Figures 2-5** and **S2-S7** do not rely on statistical modelling of the autocorrelation of measured values disregarding the relations between measured element concentrations in moss and predictors. The RF models calculated for Europe as a whole explain up to 68 % (for Pb) of the variance of the element concentration in moss (**Table 3**). The lowest variance was explained for Hg and Zn, which both showed a rather homogenous spatial distribution of concentrations in moss across Europe and thus were less explained by the spatial distributed predictors. The respective RF models for single countries participating in the EMS reach 78 % for Ni in Iceland (**Table 4**), for ecological land classes 83 % for Cr (class D_17) (**Table 5**) and for moss species at sampling sites 73 % for N (*Hylocomium splendens*) (**Table 6**). Thus, the ecological landscape classification seems to integrate characteristics which are meaningful for the complex deposition / bio-accumulation phenomenon monitored. **Tables 3-6** show that the R^2, measuring the quality of the respective RF model, differ element- and scale-specifically. Element-specific mean values for country-specific RF models with R^2 higher than 0.25 were calculated for Pb (0.34) followed by Cu (0.30) and Cd (0.28) (**Table 4**). The highest country-specific accuracies of RF models were found for Iceland, Sweden, Norway and Finland with R squareds between 0.40 and 0.49 averaged over all elements. RF models with the highest R^2 could be built for Iceland (Ni: R^2 = 0,78, Pb: R^2 = 0,70) and Norway (Pb: R^2 = 0,74). However, the standard deviation of R^2 in Iceland appears to be relative high. **Table 5** shows that R^2 of RF models are significantly landscape-specific. Again RF models reveal highest R^2 for Pb (0.40) averaged over all ELCE classes. R^2 came out to be the highest for ELCE classes C_0 (The Alps, Iceland, western and northern Scandinavia, Kola Peninsula, northwest Russia, Caucasus), D_17 (Scandinavia, western Russia) and F4_2 (Western/central and southern Europe, including southern Great Britain, eastern France, southern Belgium, Luxembourg, the Alps, Italy, eastern and southeast Europe, including the Carpathian Mountains, and the Balkans. This is approximately in line with the findings for the country-specific RF models, whereas quality are predominantly higher compared to the country-specific models. **Table 6** suggests that element concentrations in *Pleurozium schreberi*, *Hylocomium splendens* and *Hypnum cupressiforme* could be best explained by calculated RF models, with R^2 between 0.40 and 0.56 averaged over all elements. Hence, these moss species indicate best the environmental conditions modelled by use of RF. For comparison of R^2 values it should be noted that the training datasets used for series of predictions contained different sample sizes (Bergtold et al. 2011, UCLA 2011). R^2 are related to the sample size with correlation coefficients (Pearson) of r = 0.46 (country-specific models), r = 0.17 (ELCE-specific models) and r = 0.75 (moss-specific models). To enhance the explicative power of such models, there is a need to include more information on potential predictors at the regional scale derived from maps and data bases and, as integral part of the EMS, site-specific information. Such a design was realized in

Slovenia (Skudnik et al. 2015) and in the German contribution to the EMS 2000 and 2005 and will be part of the German moss survey to be conducted in 2016.

Table 3. Characteristics of optimized RF models calculated for Europe

Figure 1. Predictive importance of land use within a 1, 5, 10, 25, 50 75 and 100 km radius around the sites where moss was sampled in 2010, calculated by RF for Europe as a whole [Increased Node Purity] (**Table 2**)

Table 4. Pseudo R^2 of country-specific RF models

Table 5. Pseudo R^2 of landscape-specific RF models

Table 6. Pseudo R^2 of moss-specific RF models

The maps (**Figures 2-5, S2-S7**) show lowest element values in Fennoscandia. Thereby, the concentrations of Cu, Hg, and Zn in moss are spatially rather homogeneous, while other HM such as Cd, Pb and V vary across space. Cd and Pb concentrations in Eastern and Southeastern Europe and V in Southeastern Europe are elevated compared to Western and Northern Europe. Astonishingly, in the North of Fennoscandia a Ni hot spot was detected, differing noticeably from the respective LOTOS-EUROS model calculation. At the Kola Peninsula, at a close distance from the Norwegian and the Finnish border, one of the largest metallurgic smelters of the world is located near the town Nikel. The Nikel smelters were constructed for the processing of locally mined Nickel ores and have been in operation since 1932. Since 1971, the smelters have also processed copper and nickel ores from Norilsk, Central Siberia (Dauvalter 1994; Kashulin et al. 2001). Out-of-date equipment and technology for metal smelting make this enterprise a serious pollution source in the region (Lukin et al. 2003). Unfortunately, the location and size of the emission of these smelters have not been incorporated correctly in several emission inventories (Prank et al. 2011). Other studies in biotic and a-biotic environments confirm this hotspot in northern Fennoscandia (Amundsen et al. 2011; Kashulin et al. 2001). Regarding the spatial pattern of N it should be noted that neither Denmark, Germany, Great Britain and the Netherlands participated in the EMS 2010, because high values can be expected in (part of) these countries.

Figure 2. Maps of Cd concentration in moss 2010 (left = observed, right = predicted by RF)

Figure 3. Maps of Pb concentration in moss 2010 (left = observed, right = predicted by RF)

Figure 4. Maps of Hg concentration in moss 2010 (left = observed, right = predicted by RF)

Figure 5. Maps of N concentration in moss 2010 (left = observed, right = predicted by RF)

The results yielded by this study were based on a combination of geostatistics and multivariate tree-based models applied to areas of different spatial extent. The CART and RF models help explaining spatial patterns of HM and N concentrations in moss by identifying and ranking (inter)correlated boundary conditions such as land use and atmospheric deposition. Furthermore, the CART and RF models verify the outcomes of the geostatistical analyses in terms of spatial autocorrelation. Using both CART and RF models provide cross-validated insights into the complex interrelations between atmospheric deposition of HM and N and related accumulation in moss on different spatial scales. Unlike classical regression techniques for which the relationship between the response and predictors is pre-specified, e.g. linear, quadratic, CART does not assume such a relationship. It constructs decision rules on the predictor variables by partitioning the data into successively smaller groups with binary splits based on a single predictor. Splits for all of the predictors are examined and the best split is chosen. For regression trees, the selected split is the one that maximizes the homogeneity in each of the two resulting groups with respect to the response variable. The output is a tree diagram with the branches determined by the splitting rules and a series of terminal nodes containing the response (Breiman et al. 1984, Nisbet et al 2009).

One of the strengths of a single CART is that it is simple to interpret: The relevant predictors are included in the tree and the earlier a variable appears in a tree, the more important it is (Loh 2011). With RF, this simplicity is lost because many trees (here: 200 - 300 trees) have to be considered simultaneously (Ziegler and König 2014). Even if Random Forests are not so easy to understand compared to CART because individual trees cannot be examined separately, it provides several metrics supporting the interpretation of results (Williams 2011). Variable importance is evaluated based on how much worse the prediction would be if the data for that predictor were permuted randomly. The resulting tables can be used to compare relative importance among predictors. Ferree and Anderson (2013) and Grossman et al. (2010) applied based models for mapping ecoregions as done by Hornsmann et al. (2008) mapping Ecological Land Classes of Europe (Hornsmann et al. 2008) which were used in this investigation for spatially stratifying the RF models and, based on this, mapping and explaining spatial patterns of atmospheric deposition accumulated in moss specimens sampled across Europe.

Some of the explanatory variables have been examined earlier for 3 metals (Cd, Hg, Pb) based on data from Norway (Meyer et al. 2015 a, Nickel et al. 2015). The set of potential predictors has been enlarged (LOTOS-EUROS modelling besides Cd and Pb also for As, Cr, Cu, Ni, V and Zn, wider ranges for land use, distance to emission sources) and statistical relations were examined based on European data. Contrary to site related maps (**Figures 2-5**) and CART (Meyer et al. 2015 a), RF in conjunction with the Geographic Information System (GIS) allows transforming spatial information of the independent variables with blanket coverage to continuous surfaces and respective maps of HM and N concentration in moss explained by the RF models. Europe-wide predictions by use of RF is new and can be compared with predictions by use of kriging (Johnston et al. 2003, Schröder et al. 2012). Contrary to linear regression modelling used in previous studies (Nickel et al. 2015), the residuals of RF models presented did not show any spatial autocorrelation, i.e. must be characterized as spatially discontinuous. Thus, mapping by use of Regression Kriging with interpolations of continuous surfaces that represents the residuals was not recommended. This is in line with similar results for residuals in linear models for N and δ15N concentrations in moss in Slovenia (Skudnik et al. 2015). Model accuracy could be improved through an inclusion of categorical variables (e.g. country, ELCE, moss species, and analysis method), but lead to maps with less continuous surfaces (e.g. at borders of countries). Therefore, in this investigation, predictors were limited to those with continuous data. However, several models including categorical variables were tested with different combinations of predictors included. They were tested based on independent samples, compared regarding the Pseudo R^2, applied to surface covering maps of predictors and compared to the spatial patterns of measured element values. The explanatory power of these models including categorical predictors such as country, analytical technique, moss species or ecological land class slightly enhanced the explanatory power of the models compared to those given in **Table 3**. When interpreting the results, inaccuracies of the predictor data have to be generally taken into account. Due to missing quality assessment, we have considered values of influencing factors as correct without relevant error. EMEP and LE deposition models were both used to investigate whether both show high or low correlation and whether the strength of correlation varies across the concentration of elements. Strictly, the model quality cannot be measured by this design since we do not know whether the models used the same emission data. Further analyses should aim at clarifying why some RF models are significant for some countries but not for others. The same should be done with regard to the results of the moss-specific RF modelling.

In variogram analysis, the major range describes the distance in between point measurements showing high spatial auto-correlation. It could be assumed, that land use patterns observed within these ranges should be more relevant compared to other ranges, when taking the land use as a split criterion for generating homogenous subsets in a RF model. In the current study the density of different land use were examined in a range of between 1 and 100 km. The high importance of the 100 km radius (**Figure 1**, **Table 2**) may indicate that the spatial trends on such distances obscure those on smaller ones. Moss data from the EMS 2005 revealed ranges of 59.3 km for Cd, 255.0 km for Pb, and 209.0 km for N (Schröder et al. 2012). In **Table 2** the slightly raised value of 90.66 for the relative importance for Cd in a

50 km radius and agricultural land use could be interpreted as an obvious relation between the radius and the geostatistical range of 59.3 km.

Since RF was used for the first time for mapping geographical patterns of pollutants in terrestrial ecosystems, the results of the regionalizing at hand cannot be compared with results of other RF studies Meyer et al. (2015 a) found, based on data obtained across Norway, comparable predictive relevance for a similar set of regional factors for Cd, Hg, and Pb concentrations in moss by using CART. In addition, Meyer (2015 b) investigated by application of RF the relevance of site and regional factors for HM and N concentrations in moss sampled across Germany in 2005, 2012, and 2013. The results support the findings of the current study. Highest R^2 values of RF models were observed for central and northern European countries. It should further be examined whether this phenomenon is due to low element concentrations, small variability of explanatory variables or due to sampling density (Harmens et al. 2010, Ilyin et al. 2011). Country-specific correlations between modelled EMEP deposition and Cd, Hg and Pb concentrations in moss for previous EMS were reported by Harmens et al. (2012). In this context it should be further investigated why the explanatory power of RF models is generally higher for ELCE classes than for countries. Is this due to a coincidence that RF models can be best explained for moss species that are sampled the most? Further, it should be investigated in detailed studies, why the variance explained is low for elements such as Hg and Zn. Hg is a global pollution with low spatial variability (Schröder et al. 2013), Zn an essential nutrient for moss, so is metabolised (Harmens 2009). This might both contribute to a more homogenous spatial distribution of their concentration in moss (Harmens et al. 2015). However, one should not forget uncertainty in deposition modelling, contributing to variation. Highest variance explained was found for Pb, so moss seems to be very suitable as monitors of Pb deposition (Aboal et al. 2010). The variance explained for N is higher than expected, considering the N is a macronutrient and being metabolised in moss tissue (Harmens et al. 2011).

Regarding the predictor identification and ranking, this study indicates that the radius for examining the the influence of different spatial land use density around the sampling sites could be even more enlarged to find possible maxima (e.g. 150, 200 or 250 km). Precipitation at a higher spatial resolution and a time period corresponding to the 3-year sum of deposition (here: 2008–2010) should be included. Additional, population density could be examined in extended buffers around the sampling sites. The temporal heterogeneity of the deposition data (here: time lag of one year) could affect the significances of the importance metrics for the atmospheric deposition and, respectively, the model accuracy, measured by the Pseudo R^2. The extent, to which this helds true, should also be investigated in a further study.

4 Conclusions

This investigation yielded for four different spatial scales the identification and ranking of explanatory variables which are most important in explaining the spatial variation of element concentration in mosses. Thereby, the application of multivariate correlation modelling by use of RF and CART allowed deriving conclusions on how the methods might influence the results and subsequent mapping derived by regression prediction based on RF results. The multivariate models and the resulting maps allow defining the statistical relations between element concentration in moss and selected explanatory variables are specific for elements, moss species, countries and ecological land classes. To enhance the explicative power of such models we suggest to include more information on potential predictors at the regional scale derived from maps and data bases. This kind of site-specific and region-specific metadata should be collected along with the EMS and should be, together with the measurements of HM and N concentrations, analysed integratively by means of multivariate spatial statistics. Such a design was realized in the German part of the EMS 2000 and 2005 (Pesch and Schröder 2006; Schröder and Pesch 2005) and will be a part of the German moss survey to be conducted in 2016.

Acknowledgements

This research paper was only possible through the help and support of the Federal Environmental Agency, Dessau, Germany, the Meteorological Synthesizing Centre - East (MSC-E), Moscow, Russia and West (MSC-W), Oslo, Norway, and the ICP Vegetation Programme Coordination Centre, Centre for Ecology & Hydrology, Bangor, UK. We thank the United Kingdom Department for Environment, Food and Rural Affairs (Defra; contract AQ0810 and AQ0833), the UNECE (Trust Fund) and the Natural Environment Research Council (NERC) for funding the ICP Vegetation Programme Coordination Centre.

Authors‘ contributions

The participants of the European Moss Survey supplied the data. WS headed the computations executed by Stefan Nickel. Winfried Schröder and Stefan Nickel participated in writing the article. All Authors have read and commented on the draft manuscript.

References

Aboal J.R., Fernandez J.A., Boquete T., Carballeira A. 2010. Is it possible to estimate atmospheric deposition of heavy metals by analysis of terrestrial mosses? Science of the Total Environment 408, 6291-6297.

Amundsen P. A., Kashulin N. A., Terentjev P., Gjelland K. Ø., Koroleva I. M., Dauvalter V. A., Sandimirov S., Kashulin A., Knudsen R. 2011. Heavy metal contents in whitefish (*Coregonus lavaretus*) along a pollution gradient in a subarctic watercourse. Environmental Monitoring and Assessment, 182, 1-4.

Berg T., Steinnes E. 1997. Use of Mosses (*Hylocomium splendens* and *Pleurozium schreberi*) as Biomonitors of Heavy Metal Deposition: From Relative to Absolute Deposition Values. Environmental Pollution, 98, 1, 61-71.

Bergtold J. S., Yeager E. A., Featherstone A. 2011. Sample Size and Robustness of Inferences from Logistic Regression in the Presence of Nonlinearity and Multicollinearity. Published online: ageconsearch.umn.edu/bitstream/103771/2/Bergtold%20et%20al.%20Logit%20Bias%20Paper.pdf

Bjerregaard, P., Andersen, C. B. I., Andersen, O. 2015. Ecotoxicology of metals - sources, transport, and effects on the ecosystem. In: Nordberg G.F., Fowler B.A., Nordberg M. (Eds.) Handbook on the toxicology of metals. Academic Press, 425-459.

Bobbink R., Hicks K., Galloway J., Spranger T., Alkemade R., Ashmore M., Bustamante M., Cinderby S., Davidson E., Dentener F. 2010. Global assessment of nitrogen deposition effects on terrestrial plant diversity: a synthesis. Ecol App, 20, 30-59.

Boulesteix A. L., Janitza S., Kruppa J., König I. R. 2012. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2, 493-507.

Breiman L., Friedman J., Olshen R., Stone C. 1984. Classification and regression trees. Wadsworth, Belmont, CA.

Breiman L. 2001. Random Forests. Machine Learning, 45, 1, 5-32.

Builtjes P., Schaap M., Jonkers S., Nagel H. D., Schlutow A., Nickel S., Schröder W. 2016. Impacts of heavy metal emissions on air quality and ecosystems in Germany. 5th Progress Report on behalf of the German Federal Environmental Agency, Dessau. Unpublished work, November 2016.

Cianci D., Hartemink N., Ibáñez-Justicia A. 2015. Modelling the potential spatial distribution of mosquito species using three different techniques. International Journal of Health Geographics, 14, 10, 1-10 [DOI 10.1186/s12942-015-0001-0].

Dauvalter V. A. 1994. Heavy metals in lake sediments of the Kola Peninsula, Russia. Science of the Total Environment, 158, 51–61.

Deloncle A., Berk R., D’Andrea F., Ghil M. 2007. Weather regime prediction using statistical learning. Journal of Atmospheric Science, 64, 1619–1635.

EEA 2016 a. European Pollutant Release and Transfer Register (E-PRTR), Member States reporting under Article 7 of Regulation (EC) No 166/2006. http://www.eea.europa.eu/data-and-maps/data/member-states-reporting-art-7-under-the-european-pollutant-release-and-transfer-register-e-prtr-regulation-9 (09.02.2016)

EEA 2016 b. Corine Land Cover 2006. http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-raster-2 (09.02.2016).

EEA 2016 c. Global Land Cover 2000 – Europe (GLC2000). http://www.eea.europa.eu/data-and-maps/data/global-land-cover-2000-europe (09.02.2016).

Evans JH. S., Cushman S. A. 2009. Gradient modeling of conifer species using random forests. Landscape Ecology, 24, 5, 673-683.

FAO (Food and Agriculture Organization of the United Nations) / IIASA (International Institute of Applied Systems Analysis) / ISRIC-World Soil Information / ISS-CAS (Institute of Soil Science – Chinese Academy od Science) / JRC (Joint Research Centre of the European Commission) (2009): Harmonized World Soil Database (version 1.1). FAO, Rome, Italy and IIASA, Laxenburg, Austria.

Fawagreh K., Gaber M. M., Elyan E. 2014. Random forests: from early developments to recent advancements, Systems Science & Control Engineering, 2, 1, 602-609.

Ferree, C., Anderson M. G. 2013. A map of terrestrial habitats of the Northeastern United States: Methods and approach. The Nature Conservancy, Eastern Conservation Science, Eastern Regional Office. Boston, MA, pp. 1-79.

Glime, J. M. 2006. Bryophyte Ecology. Michigan Technological University and the International Association of Bryologists, p. 631.

Grossmann E., Ohmann J., Kagnan J., May H., Gregory M. 2010. Mapping ecological systems with a Random Forest Model: Tradeoffs between errors and bias. Gap Analysis Bulletin, 17, 16-22.

Harmens H. 2009. Review of the European moss survey (1990 - 2005): deliverables and use of data. NERC/Centre for Ecology and Hydrology, 18 p.

Harmens H., Norris D. A., Cooper D. M., Mills G., Steinnes E., Kubin E., Thöni L., Aboal J. R., Alber R., Carballeira A., Coșkun M., De Temmerman L., Frolova M., Frontasyeva M., Gonzáles-Miqueo L., Jeran Z., Leblond S., Liiv S., Maňkovská B., Pesch R., Poikolainen J., Rühling Å., Santamaria J. M., Simonèiè P., Schröder W., Suchara I., Yurukova L., Zechmeister H. G. 2011. Nitrogen concentrations in mosses indicate the spatial distribution of atmospheric nitrogen deposition in Europe. Environmental Pollution, 159, 2852-2860.

Harmens H., Norris D. A., Koerber G. R., Buse A., Steinnes E., Pisspanen J., Alber L., Aleksiayenak Y., Blum O., Coşkun M., Dam M., de Temmerman L., Fernández J.A., Frolova M., Frontasyeva M., González-Miqueo L., Grodzińska K., Jeran Z., Korzekwa S., Krma M., Kvietkus K., Leblond S., Liivt S., Magnússon S. H., Maňkovská B., Pesch R., Rühling Å., Santamaria J. A., Schröder W., Spiric Z., Suchara I., Thönia L., Urumov V., Yurukova L., Zechmeister H. G. 2010. Mosses as biomonitors of atmospheric heavy metal deposition: Spatial patterns and temporal trends in Europe. Environmental Pollution, 158, 3144-3156.

Harmens H., Ilyin I., Mills G., Aboal J.,R., Alber R., Blum O., Coskun M., De Temmerman L., Fernandez J. A., Figuera R., Frontasyeva M., Godzik B., Goltsova N., Jeran Z., Korzekwa S., Kubin E., Kvietkus K., Leblond S., Liiv S., Magnusson S. H., Mankovska B., Nikodemus O., Pesch R., Poikolainen J., Radnovic D., Rühling A., Santamaria J. M., Schröder W., Spiric Z., Stafilov T., Steinnes E., Suchara I., Tabors G., Thöni L., Turcsanyi G., Yurukova L., Zechmeister H. G. 2012. Country-specific correlations across Europe betwen modelled atmospheric cadmium and lead deposition and concentration in mosses. Environmental Pollution, 166, 1–9.

Harmens H., Norris D. A., Sharps K., Mills G., Alber R., Aleksiayenak Y., Blum O., Cucu-Man S.-M., Dam M., De Temmerman L., Ene A., Fernández J.A., Martinez-Abaigar J., Frontasyeva M., Godzik B., Jeran Z., Lazo P., Leblond S., Liiv S., Magnússon S. H., Maňkovská B., Pihl Karlsson G., Piispanen J., Poikolainen J., Santamaria J. M., Skudnik M., Spiric Z., Stafilov T., Steinnes E., Stihi C., Suchara I., Thöni L., Todoran R., Yurukova L., Zechmeister H. G. 2015. Heavy metal and nitrogen concentrations in mosses are declining across Europe whilst some “hotspots” remain in 2010. Environmental Pollution 200: 93-104.

Hernandez P. A., Franke I., Herzog S. K., Pacheco V., Paniagua L., Quintana H. L., Soto A., Swenson J. J., Tovar C., Valqui T. H., Vargas J., Young B. E. 2008. Predicting species distributions in poorly-studied landscapes. Biodiversity and Conservation, 17, 1353-1366.

Hornsmann I., Pesch R., Schmidt G., Schröder W. 2008. Calculation of an Ecological Land Classification of Europe (ELCE) and its application for optimising environmental monitoring networks. In: Car A., Griesebner G., Strobl J. (Eds.), Geospatial Crossroads @ GI_Forum '08: Proceedings of the Geoinformatics Forum Salzburg. Wichmann, Heidelberg, pp. 140-151.

Howard C., Stephens P. A., Pearce-Higgins J. W., Gregory R. D., Willis S. G. 2014. Improving species distribution models: the value of data on abundance. Methods in Ecology and Evolution, 5, 506–513.

ICP Vegetation 2010. Heavy Metals in European Mosses: 2010 Survey. Monitoring Manual, International Cooperative Programme on Effects of Air Pollution on Natural Vegetation and Crops, ICP Coordination Centre, CEH Bangor, pp. 1-16.

Ilyin, I., Rozovskaya, O., Travnikov, O., Varygina, M., Aas, W., Uggerud, H.T. 2011. Heavy Metals: Transboundary Pollution of the Environment. EMEP Status Report 2/2011. Meteorological Synthesizing Centre - East, Moscow, Russian Federation, http://www.msceast.org.

Johnston K., Ver Hoef J. M., Krivoruchko K., Lucas N., 2003. Using ArcGIS geostatistical analyst. ESRI, – Redlands, Calif. 300 p.

Kashulin, N. A., Ratkin, N. E., Dauvalter, V. A., Lukin, A. A. 2001. Impact of airborne pollution on the drainage area of subarctic lakes and fish. Chemosphere, 42, 51–59.

Liaw A., Wiener M. 2002. Classification and Regression by randomForest. R News, 2, 3, 18-22.

Liu S., Dissanayake S., Patel S., Dang X., Mlsna T, Chen Y, Wilkins D. 2014. Learning accurate and interpretable models based on regularized random forests regression. BMC Systems Biology, 8, Suppl 3, S5, 1-9 [doi:10.1186/1752-0509-8-S3-S5].

Loh W.-Y. 2011. Classification and regression trees. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1, 14-23.

Louppe 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 (30.11.2015).

Lukin A. A., Dauvalter V. A., Kashulin N. A., Yakovlev V., Sharov A., & Vandysh O. 2003. Assessment of copper-nickel industry impact on a subarctic lake ecosystem. Science of the Total Environment, 306, 73–83.

Magness D. A., Huettmann F., Morton J. M. 2008. Using Random Forests to provide predicted species distribution maps as a metric for ecological inventory & monitoring programs. In: Applications of Computational Intelligence in Biology. Studies in Computational Intelligence, 122, 209-229.

Markert B. A., Breure A. M., Zechmeister H. G. 2003. Bioindicators & Biomonitors. Elsevier, Amsterdam.

Mascaro J., Asner G.P., Knapp D. E., Kennedy-Bowdoin T., Martin R. E., Anderson C., Higgins M., Chadwick K. D. 2014. A tale of two "forests": Random forest machine learning aids tropical forest carbon mapping. PLoS ONE, 9, 1, 1-9: e85993. doi:10.1371/journal.pone.0085993.

Meyer M., Schröder W., Pesch R., Steinnes E., Uggerud H.T. 2015 a. Multivariate association of regional factors with heavy metal concentrations in moss and natural surface soil sampled across Norway between 1990 and 2010. In: Journal of Soils and Sediments, 15, 410-422.

Meyer M. 2015 b. Standortspezifisch differenzierte Erfassung atmosphärischer Stickstoff- und Schwermetalleinträge mittels Moosen unter Berücksichtigung des Traufeffektes und ergänzende Untersuchungen zur Beziehung von Stickstoffeinträgen und Begleitvegetation. Dissertation Universität Vechta, 218 pp. New M., Lister D., Hulme M., Makin I. 2002. A high-resolution data set of surface climate over global land areas. Climate Research 21, 1–25.

Nickel S., Hertel A., Pesch R., Schröder W., Steinnes E., Uggerud H. T. 2014. Modelling and mapping spatio-temporal trends of heavy metal accumulation in moss and natural surface soil monitored

1990-2010 throughout Norway by multivariate generalized linear models and geostatistics. Atmospheric Environment, 99, 85-93.

Nickel S, Schröder W 2017. Umstrukturierung des deutschen Moos-Monitoring-Messnetzes für eine regionalisierte Abschätzung atmosphärischer Deposition in terrestrische Ökosysteme. In: Schröder, W.; Fränzle, O.; Müller, F. (Hg.): Handbuch der Umweltwissenschaften. Grundlagen und Anwendungen der Ökosystemforschung. Kap. V-1.8. 24. Erg.Lfg., Wiley-VCH, Weinheim:1-46.

Nisbet R., Elder J., Miner G. 2009. Handbook of Statistical Analysis and Data Mining Applications. Elsevier, 864 p.

Pal M. 2005. Random forest classifier for remote sensing classification. International Journal of Remote Sensing, 26, 217-222.

Pesch R., Schröder W. 2006. Integrative exposure assessment through classification and regression trees on bioaccumulation of metals, related sampling site characteristics and ecoregions. Ecological Informatics 1, 1, 55-65.

Prank M., Sofiev M., Denier Van Der Gon H. A. C., Kaasik M., Ruuskanen T. M., Kukkonen J. 2010. A refinement of the emission data for Kola Peninsula based on inverse dispersion modelling, Atmospheric Chemistry and Physics, 10, 22, pp. 10849-10865.

Prasad A. M., Iverson L. R., Liaw A. 2006. Newer classification and regression tree techniques: Bagging and Random Forests for ecological prediction Ecosystems, 9, 181-199.

R Core Team 2013: R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna.

Rothwell J. J., Futter M. N., Dise N. B. 2008. A classification and regression tree model of controls on dissolved inorganic nitrogen leaching from European forests. Environmental Pollution 156, 544-552.

Schaap M., Sauter F., Timmermans R. M. A., Roemer M., Velders G., Beck J., Builjes P. J. H. 2008. The LOTOS–EUROS model: description, validation and latest developments. International Journal of Environmental Pollution, 32, 2, 270-290.

Schröder W., Holy M., Pesch R., Zechmeister H., Harmens H., Ilyin I. 2011 a. Mapping atmospheric depositions of cadmium and lead in Germany based on EMEP deposition data and the European Moss Survey 2005. Environmental Sciences Europe, 23, 19, 1-14.

Schröder W., Holy M, Pesch R., Harmens H., Fagerli H. 2011 b. Mapping background values of atmospheric nitrogen total depositions in Germany based on EMEP deposition modelling and the European Moss Survey 2005. Environmental Sciences Europe, 23, 18, 1-9.

Schröder W, Pesch R 2005. Time series of metals in mosses and their correlation with selected sampling site-specific and ecoregional characteristics in Germany. Environmental Science and Pollution Research 12, 3, 159-167.

Schröder W., Pesch R. 2007. Synthesizing bioaccumulation data from the German Metals in Mosses Surveys and relating them to ecoregions. Science of the Total Environment, 374, 311-327.

Schröder W., Pesch R., Harmens H., Fagerli H., Ilyin I. 2012. Does spatial auto-correlation call for a revision of latest heavy metal and nitrogen deposition maps? Environmental Sciences Europe, 24, 20, 1-15.

Schröder W., Pesch R., Hertel A., Schönrock S., Harmens H., Mills G., Ilyin I. (2013): Correlation between atmospheric deposition of Cd, Hg and Pb and their concentrations in mosses specified for ecological land classes covering Europe. Atmospheric Pollution Research 4, 267-274.

Schröder W., Pesch R., Schönrock S., Harmens H., Mills G., Fagerli H. 2014. Mapping correlations between nitrogen concentrations in atmospheric deposition and mosses for natural landscapes in Europe. Ecological Indicators, 36, 563-571.

SEDAC 2016. Socioeconomic Data and Application Center. Gridded Population of the world. http://sedac.ciesin.columbia.edu/data/collection/gpw-v3 (09.02.2016)

Shapiro S. S., Wilk M. B. 1965. An analysis of variance test for normality (complete samples). Biometrika, 52, 3-4, 591-611.

Simpson D., Benedictow A., Berge H., Bergstrøm R., Emberson L. D., Fagerli H., Flechard C. R., Hayman G. D., Gauss M., Jonson J. E., Jenkin M. E., Nyiri A., Richter C., Semeena V. S., Tsyro S., Tuovinen J.-P., Valdebenito A., Wind P. 2012. The EMEP MSC-W chemical transport model; technical description. Atm. Chem. Phys. 12 (16), 7825-7865.

Skudnik M., Jeran Z., Batič F., Simončič P., Kastelec D., 2015. Potential environmental factors that influence the nitrogen concentration and δ15N values in the moss Hypnum cupressiforme collected inside and outside canopy drip lines. Environmental Pollution, 198, 78-85.

Spekkers M. H., Clemens F. H. L. R., ten Veldhuis J. A. . 2015. On the occurrence of rainstorm damage based on home insurance and weather data. Natural Hazards and Earth System Science, 15, 261-272.

Špirić Z., Vučković I., Stafilov T., Kušan V., Frontasyeva M. 2013. Air pollution study in Croatia using moss biomonitoring and ICP-AES and AAS analytical techniques. Archives of Environmental Contamination and Toxicology, 65, 1, 33-46.

SRTM 2016. Shuttle Radar Topography Mission (SRTM). Digital Elevation Model (DEM). http://www2.jpl.nasa.gov/srtm/cbanddataproducts.html (09.02.2016).

Thums M., Bradshaw C. J. A., Hindell M.A. 2008. A validated approach for supervised dive classification in diving vertebrates. Journal of Experimental Marine Biology and Ecology, 363, 75-83.

Tørseth K., Aas W., Breivik K., Fjæraa A. M., Fiebig M., Hjellbrekke A. G., Lund Myhre C., Solberg S., Yttri K. E. 2012. Introduction to the European Monitoring and Evaluation Programme (EMEP) and observed atmospheric composition change during 1972 and 2009. Atmospheric Chemistry and Physics, 12, 5447-5481.

Travnikov O., Ilyin I. 2005. Regional Model MSCE-HM of Heavy Metal Transboundary Air Pollution in Europe. EMEP/MSC-E Technical Report 6/2005, 59 p.

Verikas A., Gelzins A., Bacauskiene M. 2011. Mining data with random forests: A survey and results of new tests. Pattern Recognition, 44, 330-349.

Williams G. 2011. Data mining with rattle and R. The art of excavating data for knowledge discovery. Springer, 374 p.

Wood P. 1990. Applications of Scaling to Development Research. Von Eye A. (eds.) 1990. Statistical Methods in Longitudinal Research: Principles and Structuring Change. Academic Press, Boston, 256 p.

UCLA - Statistical Consulting Group, 2011. FAQ: What are pseudo R-squareds?

Ziegler A., König I. R. 2014. Mining data with random forests: current options for real-world applications. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 2014, 4, 55-63.

Table 1. Potential predictors for HM and N concentration in moss in 2010

Variables	Comment & Source	Unit
Total atmospheric deposition (HM, N)	Modelled atmospheric deposition of As, Cd, Cr, Cu, Ni, Pb, V, Zn over three years (LOTOS-EUROS 2009-2011)	µg / m²
	Modelled atmospheric deposition of Cd, Hg, Pb, N over three years (EMEP 2008-2010) [1]	µg / m²
Distance from emission sources (HM, N)	Derived from European Pollutant Release and Transfer Register (E-PRTR 2008-2010)	Km
Density of agricultural land use within a 1, 5, 10, 25,50, 75 and 100 km radius around the moss sampling sites	Derived from CORINE Land Cover (CLC 2006) and Global Land Cover 2000 (GLC 2000) for Russia, Ukraine and Belarus	%
Density of forestry land use within a 1, 5, 10, 25,50, 75 and 100 km radius around the moss sampling sites	Derived from CLC 2006 and GLC 2000	%
Density of urban land use within a 1, 5, 10, 25,50, 75 and 100 km radius around the moss sampling sites	Derived from CLC 2006 and GLC 2000	%
Population density	Gridded Population of the World (GPW 2010)[2]	Inhabitants / km²
Elevation	World digital elevation model (ETOPO5)	m. a. s. l
Precipitation	1991-2002 (New et al. 2002)	mm / month
Clay content	Proportion of grain size (FAO 2009)	%

[1] HM data provided by MSC-East (November 2013); N data downloaded from http://emep.int/mscw/index_mscw.html (ca July 2014)

[2] SEDAC 2016. Socioeconomic Data and Application Center. Gridded Population of the world. http://sedac.ciesin.columbia.edu/data/collection/gpw-v3 (09.02.2016)

Table 2. Relative importance of predictors for measured element concentrations in moss sampled across Europe as quantified by Increased Node Purity

Predictor	As_	Cd	Cr	Cu	Hg	Ni	Pb	V	Zn	N	Rank
LE_dep	**55.02**	---	**126.88**	**44.84**	**26.81**	**71.04**	---	**58.85**	**30.71**	---	**3**
EMEP_dep	---	**258.69**	---	---	---	---	**238.35**	---	---	**15.38**	**1**
den_agr_01	25.15	18.43	25.25	6.72	10.25	23.65	14.62	19.00	14.50	2.45	26
den_agr_05	28.52	25.23	44.70	8.31	13.03	31.76	19.70	26.17	14.29	3.23	24
den_agr_10	33.09	26.39	55.46	10.21	13.72	36.17	22.72	31.35	14.89	4.35	21
den_agr_25	55.15	46.96	78.19	13.23	16.01	46.26	24.76	60.64	16.43	5.85	13
den_agr_50	59.37	90.66	91.35	16.83	19.56	58.94	54.93	59.13	19.40	9.03	10
den_agr_75	77.02	87.08	**134.15**	16.60	23.07	59.49	81.74	60.88	20.30	9.92	**6**
den_agr_100	**126.04**	**100.76**	119.94	**23.98**	**26.5**	**78.13**	**90.94**	**72.73**	**25.4**	**14.48**	**2**
den_for_01	46.67	16.7	63.26	6.24	9.48	56.95	13.07	51.96	13.35	3.04	22
den_for_05	47.08	25.11	65.41	9.33	13.03	67.65	16.77	45.36	15.79	3.53	18
den_for_10	46.66	32.01	54.71	9.94	15.92	65.82	20.2	37.35	17.18	4.33	19
den_for_25	46.83	33.27	58.1	10.92	17.01	51.17	25.44	34.23	21.57	4.69	17
den_for_50	42.47	39.36	62.69	11.90	22.03	84.6	28.91	40.06	21.19	5.07	13
den_for_75	49.10	48.23	87.12	14.58	23.20	99.45	32.84	48.26	22.17	5.00	11
den_for_100	**70.72**	**50.07**	**107.91**	**17.12**	**26.19**	**120.63**	**36.75**	**71.16**	**26.74**	**6.00**	**5**
den_urb_01	21.45	7.68	24.94	4.78	4.88	15.56	6.94	16.14	8.82	1.04	27
den_urb_05	30.28	16.36	39.86	8.23	9.39	25.62	15.73	24.19	13.02	2.36	25
den_urb_10	28.45	22.97	44.76	11.86	12.59	32.86	25.11	26.28	15.92	3.10	23
den_urb_25	39.69	35.10	49.68	16.48	14.21	38.47	30.49	32.82	18.48	4.13	20
den_urb_50	51.49	45.83	56.64	24.16	14.90	58.52	47.83	35.65	20.32	4.33	15
den_urb_75	54.41	51.77	64.08	**26.66**	17.83	55.56	56.66	40.9	24.71	**6.43**	12
den_urb_100	**65.64**	**69.77**	**81.54**	23.99	**21.90**	**67.10**	**66.86**	**43.20**	**24.85**	5.81	**8**
Distance	**72.34**	**52.92**	**141.77**	**51.44**	**19.57**	**160.36**	**51.02**	---	**34.34**	5.1	**4**
Clay content	24.46	6.57	**34.22**	2.57	5.73	13.34	5.13	14.31	5.95	0.87	28
Elevation	**83.32**	**37.33**	**77.84**	**15.6**	**22.82**	**67.2**	28.37	**69.73**	**25.31**	**11.53**	**9**
Population dens.	39.95	**45.09**	**60.58**	**17.66**	14.97	**43.52**	**56.57**	42.99	19.25	8.86	15
Precipitation	**65.11**	**47.08**	**65.81**	**18.24**	**31.02**	**68.78**	**28.93**	**61.38**	**24.73**	**15.49**	**7**

Explanation: LE_dep = Total atmospheric deposition (LOTOS-EUROS); EMEP_dep = Total atmospheric deposition (EMEP); den_agr = Density of agricultural land use within a 1, 5, 10, 25,50, 75 and 100 km radius; den_for = Density of forestry land use within a 1, 5, 10, 25,50, 75 and 100 km radius; den_urb = Density of urban land use within a 1, 5, 10, 25,50, 75 and 100 km radius; distance = Distance from HM and N emission sources; Rank = Rank of Mean of Ranks of relative importance measure; Predictors with relative high importance for building the RF models are in bold print

Table 3. Characteristics of optimized RF models calculated for Europe

Variable	As	Cd	Cr	Cu	Hg	Ni	Pb	V	Zn	N
Predictor										
LE_dep	149.97	---	276.54	**75.11**	---	166.44	---	142.41	77.33	---
EMEP_dep	---	**373.67**	---	---	72.10	---	**363.32**	---	---	**53.30**
den_agr_100	**285.11**	223.34	**348.70**	58.74	73.64	208.29	207.90	242.37	72.23	51.25
den_for_100	254.19	140.22	339.59	39.30	**89.18**	**345.69**	105.88	**298.07**	77.87	22.10
den_urb_100	191.45	194.20	249.04	75.39	66.08	---	162.57	176.81	**78.47**	---
Distance	157.80	141.57	156.51	68.47	57.69	282.57	137.61	---	75.52	---
Elevation	151.11	85.56	162.24	31.91	60.26	151.62	---	162.97	65.20	---
Population dens.	---	134.06	179.64	43.07	---	135.52	---	127.49	---	27.01
Precipitation	160.89	103.50	133.40	35.30	68.39	166.34	74.70	160.21	65.29	15.13
Parameter										
No. of observations	3010	3499	3526	3192	3057	3524	3397	3538	3664	2154
No. of variables	2	2	2	2	2	2	2	2	2	2
No. of trees	200	200	200	200	300	200	200	200	200	200
Var. explained [%]	53.54	60.38	63.22	55.35	39.36	57.98	67.14	59.04	27.56	60.11
MSE	0.3685	0.2644	0.3324	0.1057	0.1766	0.3099	0.1976	0.2656	0.1908	0.0538
Pseudo R^2	0.55 (0.05)	0.61 (0.04)	0.64 (0.04)	0.54 (0.03)	0.39 (0.03)	0.61 (0.03)	0.68 (0.03)	0.61 (0.04)	0.32 (0.04)	0.55 (0.05)

Explanation: Predictor: e.g. 149.97 = Relative predictor importance (dimensionless number), the highest relative predictor importance, measured as Increased Node Purity, is depicted in bold; MSE = Mean of squared residuals; Var. explained [%]: Percent variance explained based on training dataset; Pseudo R^2 based on test dataset; For example: 0.55 = Arithmetic mean of Pseudo R^2, (0.0x) = Standard deviation of Pseudo R^2; the workflow for model optimization is described in section 2.3

Table 4. Pseudo R^2 of country-specific RF models

Country	n	As	Cd	Cr	Cu	Hg	Ni	Pb	V	Zn	N	Mean
Albania	31	0.19 (0.15)	0.06 (0.06)	0.23 (0.12)	0.22 (0.07)	0.27 (0.10)	0.12 (0.09)	0.05 (0.06)	0.03 (0.04)	0.32 (0.06)	---	0.17 [0.19]
Austria	191	0.16 (0.05)	0.28 (0.17)	0.08 (0.06)	0.11 (0.15)	0.11 (0.11)	0.10 (0.06)	0.29 (0.11)	0.09 (0.12)	0.19 (0.12)	0.13 (0.08)	0.16 [0.12]
Belarus	46	0.05 (0.07)	---	0.20 (0.15)	---	---	0.24 (0.17)	---	0.13 (0.08)	0.60 (0.14)	---	0.24 [0.20]
Bulgaria	99	0.11 (0.10)	0.08 (0.07)	0.04 (0.04)	0.25 (0.12)	---	0.03 (0.05)	0.22 (0.10)	0.09 (0.07)	0.11 (0.10)	0.07 (0.05)	0.11 [0.09]
Croatia	91	0.07 (0.07)	0.09 (0.06)	0.16 (0.11)	0.17 (0.11)	0.03 (0.04)	0.21 (0.09)	0.19 (0.10)	0.34 (0.08)	0.10 (0.09)	**0.47** (0.11)	0.18 [0.17]
Czech Republic	242	**0.46** (0.10)	**0.51** (0.10)	0.38 (0.21)	**0.45** (0.11)	0.18 (0.13)	0.23 (0.12)	**0.53** (0.12)	**0.45** (0.14)	0.35 (0.16)	0.21 (0.14)	0.37 [0.42]
Estonia	69	0.00 (0.00)	0.02 (0.03)	0.03 (0.02)	0.14 (0.07)	0.03 (0.03)	0.04 (0.05)	0.03 (0.03)	0.02 (0.02)	0.04 (0.04)	0.06 (0.05)	0.05 [0.03]
Finland	396	0.21 (0.22)	**0.57** (0.06)	0.32 (0.15)	**0.67** (0.10)	0.08 (0.08)	**0.64** (0.20)	**0.60** (0.10)	0.27 (0.12)	0.10 (0.11)	**0.51** (0.11)	**0.40 [0.42]**
France	412	0.08 (0.06)	0.31 (0.12)	0.13 (0.13)	0.25 (0.13)	0.15 (0.08)	0.23 (0.10)	0.35 (0.19)	0.38 (0.13)	0.22 (0.18)	0.29 (0.15)	0.24 [0.24]
Iceland	114	**0.66** (0.44)	0.26 (0.10)	**0.48** (0.40)	0.35 (0.33)	0.10 (0.08)	**0.78** (0.30)	**0.70** (0.12)	---	**0.61** (0.37)	---	**0.49 [0.55]**
Macedonia	42	0.04 (0.05)	0.09 (0.06)	0.30 (0.07)	0.05 (0.05)	0.07 (0.07)	0.35 (0.16)	0.28 (0.10)	0.10 (0.05)	0.18 (0.13)	0.04 (0.05)	0.15 [0.10]
Norway	433	**0.45** (0.17)	**0.64** (0.15)	0.30 (0.17)	**0.66** (0.21)	0.10 (0.09)	**0.43** (0.25)	**0.74** (0.08)	**0.50** (0.14)	0.29 (0.10)	---	**0.45 [0.45]**
Poland	290	---	0.55 (0.19)	0.05 (0.04)	0.27 (0.21)	0.27 (0.14)	0.13 (0.13)	**0.43** (0.24)	0.17 (0.14)	0.18 (0.11)	0.11 (0.09)	0.24 [0.18]
Romania	295	0.21 (0.12)	---	0.13 (0.12)	---	---	0.14 (0.13)	---	0.18 (0.09)	0.12 (0.13)	---	0.16 [0.14]
Russia (Ivanovo, Kostromskaya, Tikhvin-Leningradskaya)	60	0.06 (0.06)	0.26 (0.18)	**0.44** (0.18)	---	---	0.07 (0.08)	---	0.14 (0.18)	0.03 (0.03)	---	0.17 [0.11]
Slovakia	37	---	0.08 (0.05)	---	0.04 (0.04)	---	---	0.10 (0.09)	0.13 (0.06)	---	0.05 (0.04)	0.08 [0.08]
Slovenia	72	0.06 (0.05)	0.14 (0.09)	0.25 (0.07)	0.36 (0.12)	0.19 (0.13)	0.12 (0.12)	0.23 (0.09)	0.19 (0.14)	0.21 (0.1)	0.30 (0.09)	0.20 [0.20]
Spain (Galicia, Navarra, Rioja)	181	0.04 (0.03)	0.17 (0.09)	0.26 (0.31)	0.36 (0.36)	0.16 (0.13)	0.10 (0.08)	0.25 (0.21)	0.20 (0.32)	0.37 (0.10)	0.30 (0.15)	0.22 [0.23]
Sweden	572	0.20 (0.20)	**0.57** (0.11)	**0.51** (0.11)	**0.58** (0.14)	0.28 (0.15)	0.31 (0.14)	**0.65** (0.11)	**0.66** (0.12)	0.28 (0.16)	---	**0.45 [0.51]**
Switzerland	126	0.10 (0.07)	0.35 (0.19)	0.14 (0.09)	0.24 (0.19)	0.13 (0.11)	0.18 (0.20)	0.18 (0.15)	0.24 (0.09)	0.23 (0.09)	0.07 (0.07)	0.19 [0.18]
Mean		0.19 [0.11]	0.28 [0.26]	0.23 [0.23]	0.30 [0.25]	0.14 [0.13]	0.23 [0.18]	0.34 [0.28]	0.23 [0.18]	0.24 [0.21]	0.20 [0.13]	

Explanation: n = number of observations in training dataset; countries or regions with less than 30 observations in training or test dataset (Kosovo, Denmark(Faroe Islands), Belgium, Ukraine (Donetsk) and Italy (Bolzano) were excluded; Pseudo R^2 >= 0.4 are depicted in bold; For example: 0.19 = Arithmetic mean of Pseudo R-Squareds; (0.15) = Standard deviation of Pseudo R^2; Mean = Pseudo R^2 averaged over all rows / columns (median is displayed in box brackets); --- = No Data.

Table 5. Pseudo R^2 of landscape-specific RF models

ELCE	n	As	Cd	Cr	Cu	Hg	Ni	Pb	V	Zn	N	Mean
B_1	43	0.18 (0.09)	0.34 (0.15)	0.23 (0.08)	0.32 (0.11)	0.17 (0.12)	0.4 (0.11)	0.47 (0.17)	0.33 (0.12)	0.23 (0.18)	**0.59** (0.35)	0.33 [0.33]
B_2	83	0.10 (0.10)	0.22 (0.09)	0.32 (0.20)	0.22 (0.26)	0.11 (0.11)	**0.60** (0.14)	**0.67** (0.08)	0.25 (0.19)	0.04 (0.09)	0.35 (0.36)	0.29 [0.24]
C_0	230	**0.49** (0.18)	**0.52** (0.11)	0.35 (0.18)	**0.41** (0.15)	0.16 (0.10)	**0.44** (0.24)	**0.58** (0.09)	**0.40** (0.14)	0.15 (0.24)	**0.68** (0.37)	**0.42 [0.42]**
D_7	156	0.31 (0.27)	0.21 (0.10)	0.43 (0.15)	**0.60** (0.21)	0.25 (0.16)	**0.72** (0.11)	0.30 (0.21)	**0.69** (0.10)	0.13 (0.13)	0.09 (0.08)	0.37 [0.31]
D_13	127	**0.44** (0.17)	**0.55** (0.12)	0.37 (0.11)	0.35 (0.11)	**0.40** (0.20)	**0.46** (0.09)	**0.44** (0.12)	0.29 (0.16)	0.32 (0.2)	0.31 (0.18)	0.39 [0.39]
D_14	107	0.17 (0.17)	0.28 (0.22)	**0.51** (0.17)	0.06 (0.05)	0.27 (0.17)	**0.44** (0.16)	0.35 (0.21)	**0.49** (0.09)	0.06 (0.05)	**0.52** (0.17)	0.32 [0.32]
D_17	124	0.39 (0.19)	0.35 (0.22)	**0.83** (0.04)	**0.63** (0.11)	0.22 (0.17)	**0.70** (0.09)	0.37 (0.17)	0.34 (0.13)	0.08 (0.09)	0.22 (0.18)	**0.41 [0.36]**
D_18	225	**0.69** (0.16)	0.28 (0.17)	0.31 (0.11)	0.35 (0.09)	0.31 (0.17)	0.30 (0.16)	0.34 (0.08)	**0.49** (0.17)	0.22 (0.08)	0.38 (0.23)	0.37 [0.33]
D_19	228	0.06 (0.07)	0.37 (0.17)	0.39 (0.18)	0.35 (0.15)	0.39 (0.16)	0.28 (0.24)	0.35 (0.13)	0.27 (0.11)	0.07 (0.08)	0.25 (0.15)	0.28 [0.32]
D_22	141	0.30 (0.24)	0.33 (0.15)	**0.52** (0.18)	**0.52** (0.1)	0.07 (0.06)	0.38 (0.17)	**0.50** (0.13)	**0.45** (0.11)	0.13 (0.08)	---	0.37 [0.31]
F1_1	65	**0.48** (0.25)	**0.67** (0.15)	0.14 (0.11)	0.29 (0.13)	0.33 (0.11)	0.31 (0.10)	**0.55** (0.10)	0.25 (0.11)	0.24 (0.10)	0.10 (0.10)	0.34 [0.30]
F1_2	278	0.27 (0.18)	0.19 (0.14)	0.21 (0.16)	0.34 (0.19)	0.34 (0.13)	0.22 (0.13)	0.32 (0.16)	0.23 (0.23)	0.18 (0.15)	0.22 (0.20)	0.25 [0.23]
F2_5	48	**0.52** (0.32)	**0.64** (0.12)	0.09 (0.08)	0.03 (0.04)	0.06 (0.09)	0.08 (0.07)	0.33 (0.21)	0.22 (0.08)	0.05 (0.05)	0.04 (0.04)	0.21 [0.09]
F2_6	283	**0.48** (0.17)	0.28 (0.12)	**0.51** (0.11)	0.31 (0.15)	0.33 (0.13)	**0.45** (0.16)	**0.49** (0.14)	**0.50** (0.17)	0.18 (0.15)	0.21 (0.16)	0.38 [0.39]
F3_1	174	0.27 (0.14)	0.17 (0.16)	0.19 (0.14)	0.06 (0.06)	0.36 (0.1)	0.24 (0.12)	0.23 (0.15)	**0.47** (0.18)	0.10 (0.11)	0.05 (0.07)	0.21 [0.21]
F3_2	85	0.13 (0.09)	**0.40** (0.11)	0.08 (0.09)	0.22 (0.09)	0.15 (0.08)	0.03 (0.02)	0.24 (0.12)	0.09 (0.07)	0.21 (0.14)	0.22 (0.13)	0.18 [0.18]
F4_2	551	**0.45** (0.11)	**0.44** (0.16)	**0.59** (0.10)	0.37 (0.21)	**0.47** (0.17)	0.34 (0.13)	**0.49** (0.13)	**0.48** (0.18)	0.21 (0.16)	0.22 (0.11)	**0.41 [0.45]**
G1_0	164	0.18 (0.13)	0.07 (0.05)	0.31 (0.13)	0.31 (0.16)	0.13 (0.13)	0.23 (0.13)	0.32 (0.18)	0.27 (0.16)	0.26 (0.19)	0.08 (0.06)	0.22 [0.25]
G2_0	159	0.20 (0.12)	0.29 (0.08)	**0.51** (0.10)	0.25 (0.14)	**0.50** (0.20)	**0.61** (0.12)	0.39 (0.15)	**0.47** (0.14)	**0.49** (0.12)	0.24 (0.20)	0.39 [0.43]
J_2	30	0.05 (0.07)	0.20 (0.10)	0.29 (0.10)	0.25 (0.10)	**0.62** (0.07)	**0.44** (0.09)	**0.43** (0.09)	0.31 (0.12)	0.15 (0.08)	0.05 (0.08)	0.28 [0.27]
S_0	32	0.13 (0.09)	**0.53** (0.08)	0.21 (0.16)	0.38 (0.17)	0.02 (0.02)	0.29 (0.19)	**0.42** (0.09)	0.15 (0.16)	0.11 (0.07)	**0.57** (0.19)	0.28 [0.25]
U_2	75	0.13 (0.11)	0.20 (0.14)	0.26 (0.13)	0.19 (0.13)	0.18 (0.07)	0.29 (0.12)	0.11 (0.10)	0.30 (0.10)	0.06 (0.09)	0.22 (0.09)	0.19 [0.20]
Mean		0.29 [0.27]	0.34 [0.31]	0.35 [0.32]	0.31 [0.32]	0.27 [0.26]	0.38 [0.36]	**0.40 [0.38]**	0.35 [0.32]	0.17 [0.15]	0.27 [0.22]	

Explanation: ELCE = Ecological Landscape Classes of Europe (Hornsmann et al. 2008); n = number of observations in training dataset; ELCE with less than 30 observations in training or test dataset (L2, M5, D_10, U_1, D_8) were excluded; Pseudo R^2 > 0.4 are depicted in bold; For example: 0.18 = Arithmetic mean of Pseudo R^2 (0.09) = Standard deviation of Pseudo R^2; Mean = Pseudo R^2 averaged over all rows / columns (median is displayed in box brackets); --- = No Data.

Table 6. Pseudo R^2 of moss-specific RF models

Moss species	n	As	Cd	Cr	Cu	Hg	Ni	Pb	V	Zn	N	Mean
Homalothecium lutescens	34	0.07 (0.06)	0.10 (0.08)	0.25 (0.11)	0.16 (0.11)	0.22 (0.07)	0.24 (0.14)	0.16 (0.08)	0.05 (0.07)	0.05 (0.03)	0.04 (0.06)	0.13 [0.13]
Hylocomium splendens	1027	**0.55** (0.21)	**0.65** (0.13)	**0.62** (0.16)	**0.50** (0.22)	0.34 (0.14)	**0.65** (0.17)	**0.72** (0.08)	**0.41** (0.21)	0.27 (0.19)	**0.73** (0.16)	**0.54 [0.59]**
Hypnum cupressiforme	850	0.34 (0.16)	0.34 (0.19)	**0.42** (0.16)	**0.42** (0.21)	**0.55** (0.17)	0.27 (0.16)	**0.46** (0.20)	**0.53** (0.15)	**0.43** (0.17)	0.28 (0.12)	**0.40 [0.42]**
Pleurozium schreberi	1861	**0.63** (0.19)	**0.69** (0.11)	**0.57** (0.17)	**0.65** (0.08)	**0.47** (0.15)	**0.47** (0.19)	**0.68** (0.18)	**0.58** (0.14)	0.25 (0.16)	**0.63** (0.12)	**0.56 [0.61]**
Pseudosclero-podium purum	318	0.32 (0.12)	0.16 (0.10)	0.17 (0.09)	**0.46** (0.16)	0.27 (0.15)	0.19 (0.13)	**0.44** (0.13)	0.36 (0.17)	0.28 (0.16)	**0.45** (0.19)	0.31 [0.30]
Thuidium tamariscinum	32	0.14 (0.10)	0.35 (0.08)	0.21 (0.12)	0.07 (0.06)	0.18 (0.10)	0.08 (0.09)	**0.47** (0.07)	0.15 (0.08)	0.22 (0.12)	0.09 (0.09)	0.20 [0.17]
Mean		0.34 [0.33]	0.38 [0.35]	0.37 [0.34]	0.38 [0.44]	0.34 [0.31]	0.32 [0.26]	**0.49 [0.47]**	0.35 [0.39]	0.25 [0.26]	0.37 [0.37]	

Explanation: n = number of observations in training dataset; moss species with less than 30 observations in training or test dataset were excluded; Pseudo R^2 >= 0.4 are depicted in bold; For example: 0.07 = Arithmetic mean of Pseudo R^2; (0.06) = Standard deviation of Pseudo R^2; Mean = Pseudo R^2 averaged over all rows / columns (median is displayed in box brackets); --- = No Data

Figure 1. Predictive importance of land use within a 1, 5, 10, 25, 50 75 and 100 km radius around the sites where moss was sampled in 2010, calculated by RF for Europe as a whole [Increased Node Purity] (**Table 2**)

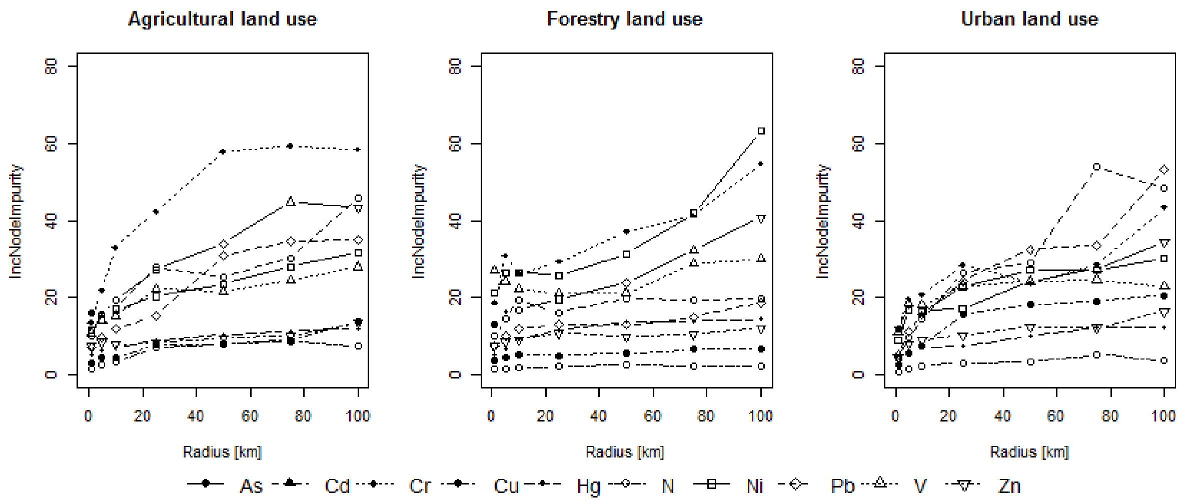


Figure 2. Maps of Cd concentration in moss 2010 (left = observed, right = predicted by RF)

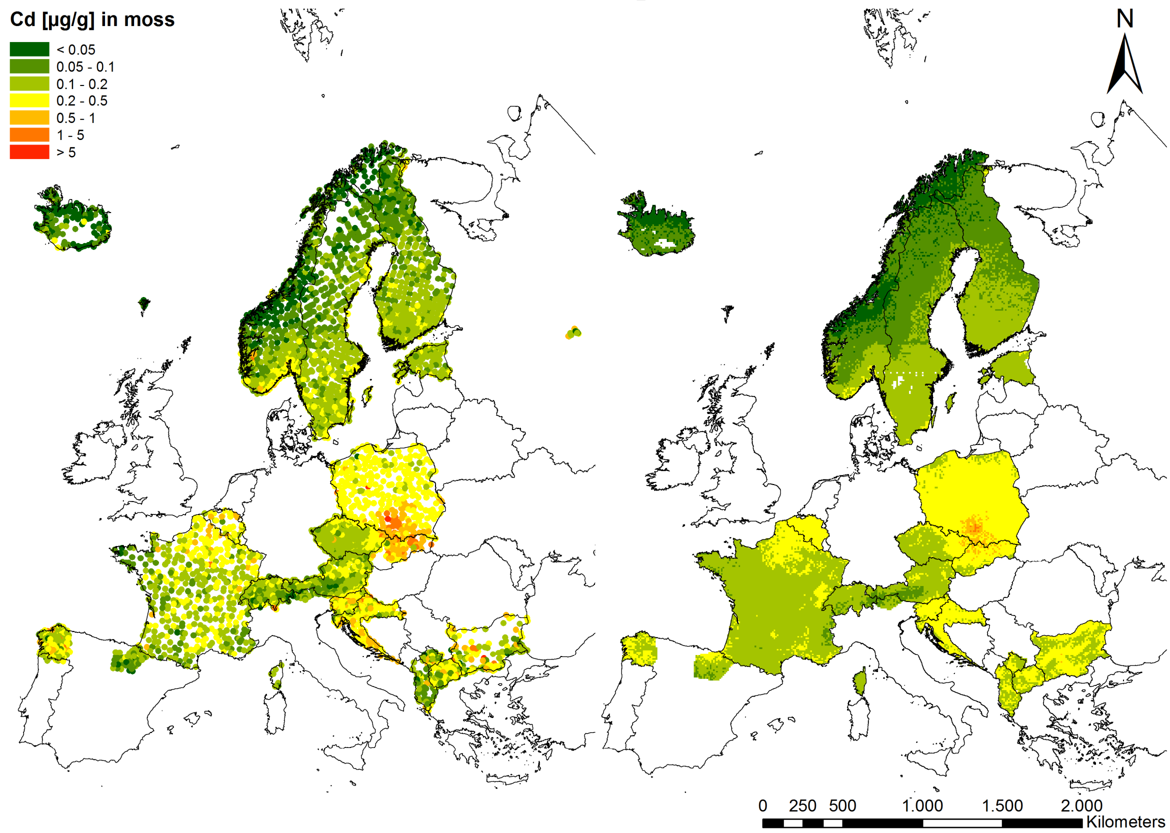


Figure 3. Maps of Pb concentration in moss 2010 (left = observed, right = predicted by RF)

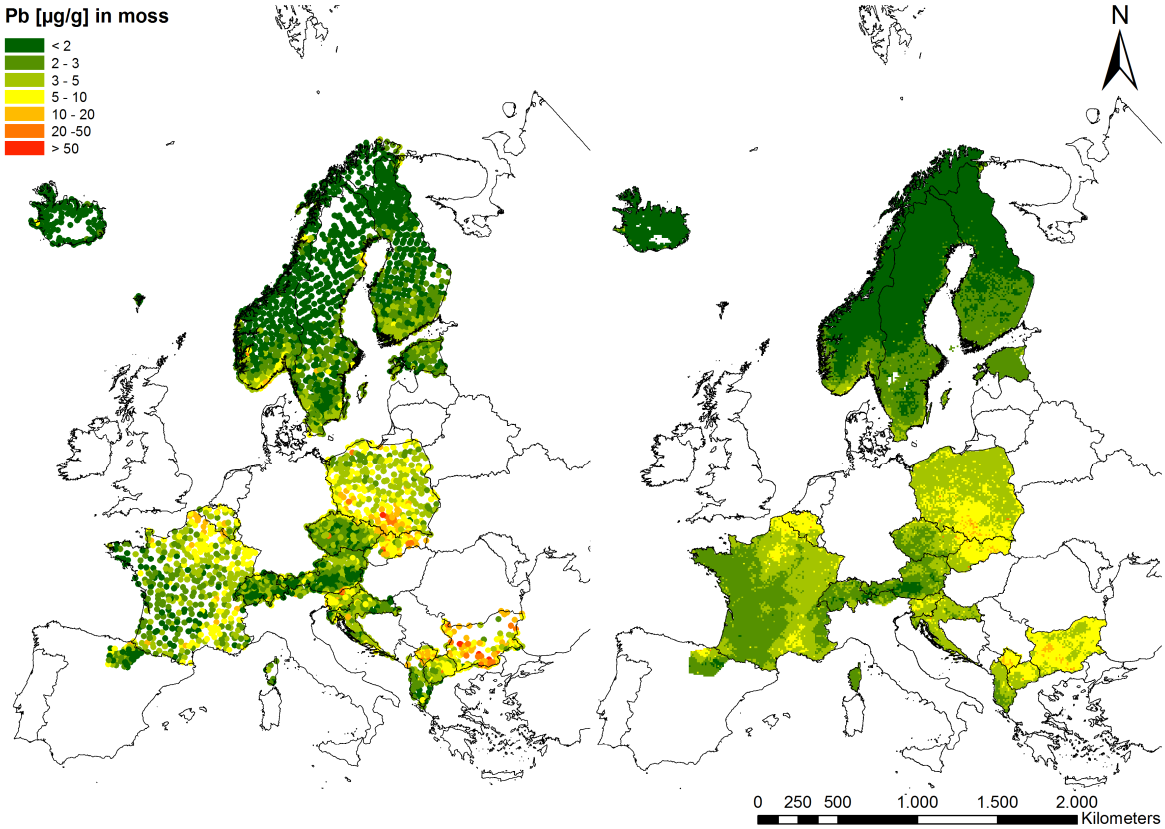


Figure 4. Maps of Hg concentration in moss 2010 (left = observed, right = predicted by RF)

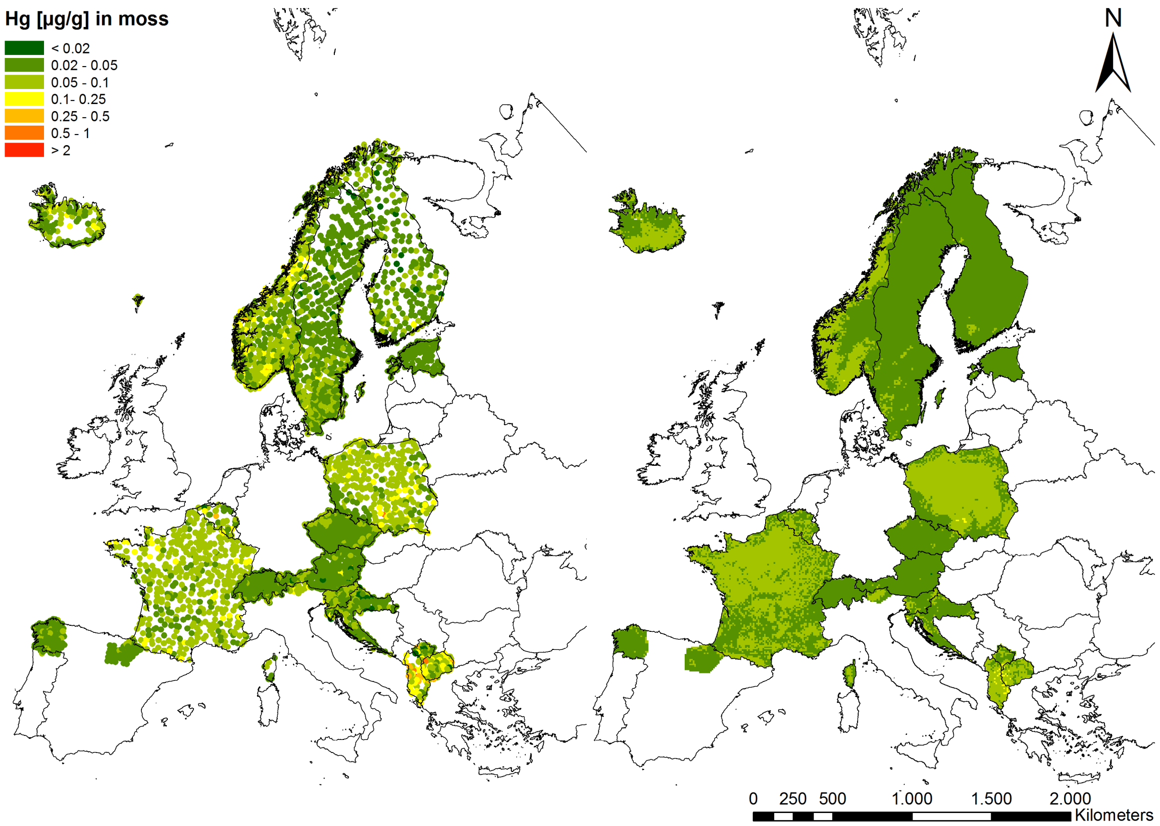
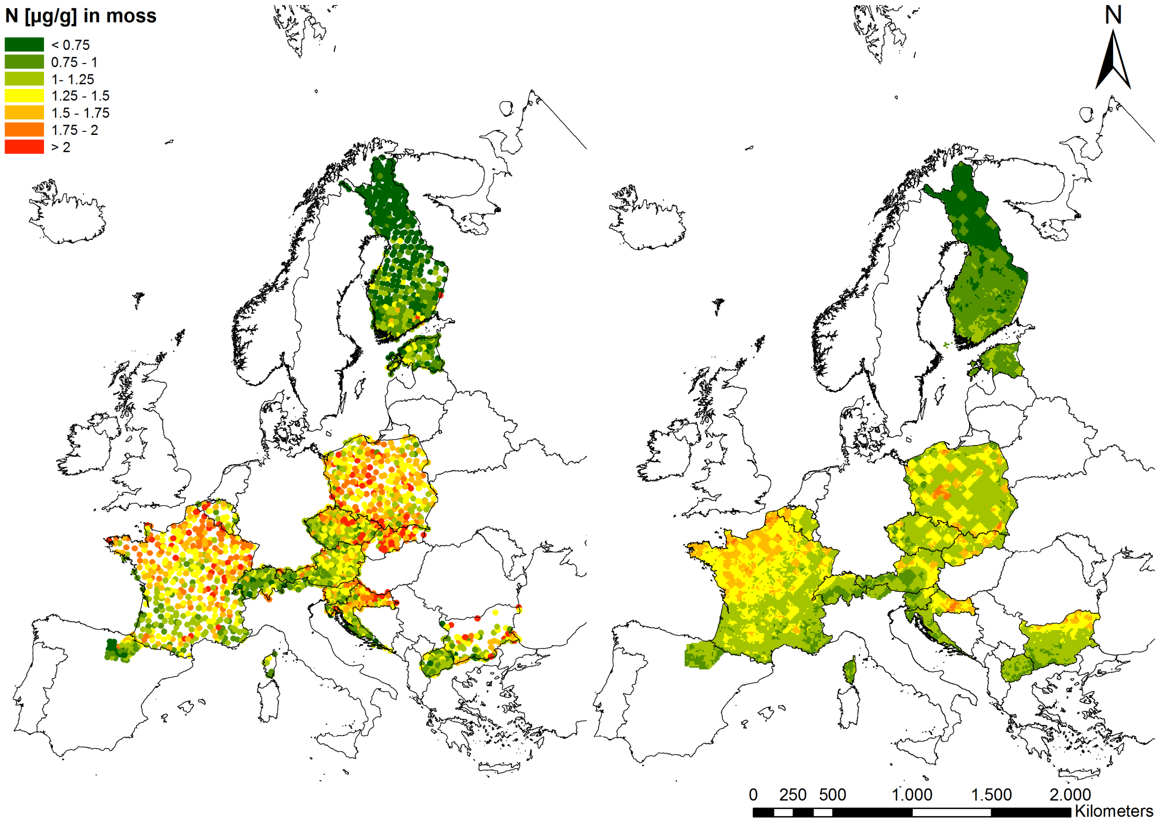


Figure 5. Maps of N concentration in moss 2010 (left = observed, right = predicted by RF)



Renate Alber, Environmental Agency of Bolzano, Laives, Italy
Renate.Alber@provinz.bz.it

Julia Aleksiayenak, International Sakharov Environmental University, Minsk, Belarus
beataa@gmail.com

Lambe Barandovski, Ss. Cyril and Methodius University, Skopje, Macedonia
lambe@pmf.ukim.mk

Oleg Blum, National Botanical Garden, Academy of Science of Ukraine, Kiev, Ukraine
blum@nbg.kiev.ua

Helena Danielsson, IVL Swedish Environmental Research Institute, Göteborg, Sweden
helena.danielsson@ivl.se

Ludwig de Temmermann, Veterinary and Agrochemical Research Centre CODA-CERVA, Tervuren, Belgium
ludwig.detemmerman@coda-cerva.be

Anatoly M. Dunaev, Ivanovo State University of Chemistry and Technology, Ivanovo, Russia
amdunaev@ro.ru

Hilde Fagerli, Meteorological Synthesizing Centre-West, Oslo, Norway
hilde.fagerli@met.no

Barbara Godzik, W. Szafer Institute of Botany, Polish Academy of Sciences, Kraków, Poland
b.godzik@botany.pl

Ilia Ilyin, Meteorological Synthesizing Centre East, Moscow, Russia
Ilia.ilyin@msceast.org

Sander Jonkers, TNO, Utrecht, Netherlands
sander.jonkers@tno.nl

Zvonka Jeran, Jožef Stefan Institute, Ljubljana, Slovenia
zvonka.jeran@ijs.si

Gunilla Pihl Karlsson, IVL Swedish Environmental Research Institute, Göteborg, Sweden
gunilla.pihl.karlsson@ivl.se

Pranvera Lazo, University of Tirana, Tirana, Albania
pranveralazo@gmail.com

Sebastien Leblond, National Museum of Natural History, Paris, France
sleblond@mnhn.fr

Siiri Liiv, Tallinn Botanic Garden, Tallinn, Estonia
siiri.liiv@botaanikaaed.ee

Sigurður H. Magnússon, Icelandic Institute of Natural History, Garðabær, Iceland
sigurdur@ni.is

Blanka Mankovska, Institute of Landscape Ecology, Slovak Academy of Sciences, Bratislava, Slovak Republic
bmankov@stonline.sk

Javier Martinez-Abaigar, University of La Rioja, Logroño, Spain
javier.martinez@unirioja.es

Juha Piispanen, Natural Resources Institute Finland (Luke), Oulu, Finland
juha.piispanen@luke.fi

Jarmo Poikolainen, Natural Resources Institute Finland (Luke), Oulu, Finland
jarmo.poikolainen@gmail.com

Ion V. Popescu, Valahia University of Targoviste, Targoviste, Romania
ivpopes@yahoo.com

Flora Qarri, University of Vlora, Vlorë, Albania
flora.qarri@gmail.com

Dragan Radnovic, University of Novi Sad, Faculty of Sciences, Novi Sad, Serbia
dragan.radnovic@dbe.uns.ac.rs

Jesus Miguel Santamaria, University of Navarra, Navarra, Spain
chusmi@unav.es

Martijn Schaap, TNO, Utrecht, Netherlands
martijn.schaap@tno.nl

Mitja Skudnik, Slovenian Forestry Institute, Ljubljana, Slovenia
mitja.skudnik@gozdis.si

Zdravko Spiric, Green Infrastructure Ltd, Zagreb, Croatia
zdravko_spiric@hotmail.com

Trajce Stafilov, Ss. Cyril and Methodius University, Skopje, Macedonia
trajcest@pmf.ukim.mk

Eiliv Steinnes, Norwegian University of Science and Technology, Trondheim, Norway
eiliv.steinnes@ntnu.no

Claudia Stihi, Valahia University of Targoviste, Targoviste, Romania
stihi@valahia.ro

Ivan Suchara, Silva Tarouca Research Institute for Landscape and Ornamental Gardening
suchara@vukoz.cz

Lotti Thoni, FUB-Research Group for Environmental Monitoring, Rapperswil, Switzerland
lotti.thoeni@fub-ag.ch

Hilde Thelle Uggerud, Norwegian Institute for Air Research, Kjeller, Norway
hilde.thelle.uggerud@nilu.no

Harald G. Zechmeister, University of Vienna, Vienna, Austria
harald.zechmeister@univie.ac.at

Highlights

- Comprehensive analysis of relations between atmospheric deposition and accumulation

- Random Forests (RF) allows for multiple regression analysis

- Atmospheric deposition, land use and distance to emission sources are relevant factors

- Measured elements, countries and ecological land classes determine the models accuracy

- RF enables predictive mapping of element concentrations in moss