Reviewer 1

General Comments: Ion exchange resin precipitation collectors (IERs) have been used to quantify element/ion deposition in remote locations, but there remain questions about the reliability of IERs. This study examined different laboratory approaches (wet vs dry resins; drip vs drip-shake extraction; molarity of extraction) to processing IERs and also compared IERs to more traditional wet precipitation collectors in a forest with four different canopy covers. The study addresses important questions. However, the experimental design is a major weakness of the work. In both the laboratory and the field part of the work, complete statistical analyses are not shown. Given the unbalanced design and low replication (n=1 for some treatment combinations in the lab study), I read, but did not comment on the Discussion. I was not convinced that the laboratory study especially provided convincing new insights into questions about IER processing.

Below we answer two main aspects, i.e., low replication and experimental design versus statistical analysis.

Low replication

- The reviewer is correct; our sample size is indeed small. For the extraction tests conducted in the laboratory, we were limited to $N = 2{-}3$ due to constraints on time and budget. We acknowledge this limitation and have removed any tests conducted with only one sample from the manuscript.
- Duplo samples are often utilized in the early stages of research where the primary goal is to assess the feasibility of a method. This was the objective of our study. To provide insight into the variation between these samples, we have included the standard error in Table 5 (Table 3 in the original manuscript), where we report the recovery efficiency following extractions. While larger sample sizes would certainly reduce standard errors further, we noted that the standard errors tend to be lower when the recovery approaches 100%, indicating minimal differences between duplicate and triplicate samples in cases of high recovery efficiency.
- We have added six lines to our discussion (section 4.2) to address our low sample size and recommend larger sample sizes for future testing of the resin's recovery efficiency. The added lines are as follows: "For this test, we used generally duplicate or triplicate samples, which can be considered a low sample size. However, because these tests were performed under controlled laboratory conditions, a small sample set can be justified. When using the ion exchange resin method for field studies, we recommend testing the extraction fluid with a larger number of samples to reduce the standard error (Table 5). Nonetheless, we are confident that our conclusions are justified, given the controlled circumstances of the laboratory tests and the relatively low standard errors."
- In summary, while our sample sizes were small, the controlled laboratory conditions and low standard errors support the reliability of our conclusions.

Experimental design and statistical analysis

- Our objective was not to test differences between treatments, but rather to identify treatments that consistently yield good and reliable results for the IER method. Furthermore, we want to emphasize that (1) the adsorption capacity, stability and recovery efficiency of the resin must be tested when using the ion exchange resin method and (2) to provide a way how to test this adsorption capacity, the stability and the recovery efficiency of the resin. In our effort to show the different ways how to extract the resin we ended up with a design with different sample sizes for the different groups. To address this, we excluded groups with only one sample and focused on treatments with two or more samples to ensure the robustness of our method's results. Furthermore, as explained in more detail by the specific comments regarding the extraction efficiency, the unbalanced design does not necessarily pose problems when using Anova as long as the type of the Sum of Squares is reported and as long as the interaction effects are significant: see Smith and Cribbie (2014) for a comprehensive comparison of unbalanced Anova using type I, II and type III sum of squares. Furthermore, with the unbalanced design of this study, we demonstrate that the IER method can be effectively employed with various extractants, resin types, and extraction equipment. We show that reliable results can be obtained using self-made equipment which makes the method well applicable all around the world.

Specific comments

Lines 107ff – IER construction is not the same across all published IER studies. Please specify the diameter and length of the resin tubes themselves.

- ➢ The diameter and length of the resin tubes are indeed important to include in the manuscript. In the methods section of the original manuscript, we first described the preparation of the resin columns before we described the deposition samplers in more detail, thus including the requested information in line 198. In the revised manuscript, we added this information right after line 107. The revised text is as follows: "We prepared 45 resin columns for the laboratory tests of elemental adsorption and recovery (including the blanks), followed by the preparation of 30 column's for the field test of the IER-method. The resin columns had a volume of 15.7ml and an inner diameter and length of respectively 12.4 and 130mm. First, the resin column's were cleaned using 0.2M HCl and demineralized water to remove weakly attached chemicals from the column walls."

Line 115. Change weighted to weighed.

- ➢ We changed this word, comment was followed.

Line 127 – Existing deposition data for where?

- ➢ The deposition data was taken from nearby measurement stations. The locations of these measurement stations are now included in line 127-128. We used the data of 2015 from all these measurement stations in which the wet-only deposition is sampled weekly or biweekly. We were not able to use more recent data as these were not yet available.

Lines 128-135. This is a long sentence – and one that I do not fully understand. "…representative stations" – representative of what? What does "for the funnel surface" mean?

- The phrase ‘representative stations’ is changed to ‘the nearby weather stations’, which are now also explicitly mentioned. Line 128-129 is now changed into: “First, based on existing wet-deposition data from nearby measurement stations located in Biest Houtakker, Speuld, De Zilk and Vredepeel (NL) (RIVM, 2015), we estimated the bulk deposition amounts (kg ha^{-1}) for different elements, and then used those to determine the needed molarity of the solution that was used to test the adsorption capacity of the resin”.
- With the funnels surface we meant that the seasonal concentrations in the wet-deposition, which were expressed in mg L^{-1}, were multiplied by the precipitation that would be captured by the funnel (in L) by multiplying the recorded precipitation (in mm or L m^{-2}) with the horizontal surface of the funnel (in m^2).We now removed the phrase “funnels surface” and split the sentence in three different sentences for better understandability: “To estimate the maximum bulk deposition values, the monthly measurements of existing bulk deposition data of the nearby weather stations (umol l^{-1}) (RIVM, 2015), were summed to seasonal concentrations, expressed in mg L^{-1}. Then, the stations were selected with the highest seasonal deposition, occurring during summer, for both macro- and micronutrients, based on the total molarity of the rainwater. These seasonal concentrations were then multiplied by the precipitation (in L) that would be captured by the funnel, by multiplying the recorded precipitation (in mm or L m^{-2}) with the horizontal surface of the funnel (in m^2) to estimate the total deposition captured by a funnel.”.

Lines 135-136 – Table S1 is referenced in text, but when I read Table S1 I had a hard time understanding what the Table was showing or how it related to the sentence in the manuscript.

- We agree that the heading of table 1 was confusing as we suggested that the reported values were used to calculate the total deposition instead of the throughfall. In Table S1 we presented “Ratios of throughfall (mostly also including stemflow, SF) to bulk deposition reported in literature” as now clearly mentioned in the table heading. We used this information to derive an average multiplication factor of 2 to convert bulk deposition to throughfall, based on the reported values for the ratio throughfall/bulk deposition of the tracer Na. This has now been rephrased in this way in the main text (Table S1).

Lines 138-142.This information could be put in to a small table. It would be useful to give the recipe for these solutions (how many grams of each salt?).

- We placed the information of line 138 to 142 in a new table (shown below). Therefore, we changed the text in line 138 to: “The total elemental content of this throughfall flux, multiplied by 4 (assumed that the summer values are representative of the entire year, which is a precautionary approach), was dissolved in a 1 L solution separately for macro and micronutrients using stock solutions resulting in an extraction solution containing values reflecting the maximum annual total deposition in the Netherlands (Table 1)” .

Table 1: The throughfall flux used to test the adsorption capacity and recovery efficiency of the ion exchange resin. We used stock solutions with known molarity to make the macro and the micro solution used to drip through the resin. The total volume of the used stock solution (in ml L^{-1}) and the concentration in umol per element are given.

Stock solution		Type	Total	Ca	Cu	Cl	Fe	K	Mg	Mn	Na	PO_4	SO_4	Zn	NH_4	NO_3
Code	Mol		*mL L^{-1}*	*umol*												
Na_2SO_4	0.5	Macro	0.90								450		450			
NaCl	1	Macro	1.40			1400					1400					
KNO_3	1	Macro	0.18					180								180
KH_2PO_4	1	Macro	0.02					20				20				
NH_4NO_3	1	Macro	1.82												1820	1820
NH_4Cl	1	Macro	2.18			2180									2180	
$MgSO_4$	1	Macro	0.3						300				300			
$CaCl_2$	0.5	Macro	0.8	400		400										
$FeCl_2$	0.1	Micro	6.0			600	600									
$Cu(NO_3)2.3H_2O$	0.275	Micro	0.036		9.9											9.9
$Zn(NO_3)2.6H_2O$	0.267	Micro	0.075											20		20
$Mn.SO_4.H_2O$	0.01	Micro	15.0							150			150			
			total	400	9.9	4580	600	200	300	150	1850	20	900	20	4000	2030

Line 145 – Should be column or column’s, not columns

- Changed.

Line 147 – Should be resin or resin’s, not resins

- Changed.

Lines 150-155 – Were the leachate and demineralized water samples filtered prior to analysis? Hopefully, yes.

- Filtering water samples before chemical analysis is typically done for four main reasons: removing particles, preventing contamination, ensuring accuracy and precision, and protecting analytical instruments. However, in this specific laboratory test, filtering was not necessary for several reasons. First, the resin was thoroughly flushed (500 grams of resin flushed with 8 liters of demineralized water), ensuring no small particles remained. Additionally, all materials that came into contact with the nutrient-containing fluids, which simulate total deposition, were cleaned according to standard laboratory protocols, minimizing the risk of contamination from large particles. The nutrient solution, representing the total annual deposition, was well mixed, with no undissolved particles visible. Therefore, filtering was not needed to prevent contamination or protect analytical instruments. Importantly, unnecessary filtering can introduce contaminants, potentially reducing the accuracy and precision of the analysis. To maintain sample integrity, filtering was avoided unless absolutely necessary which was not the case for this study. Finally, the accuracy and precision of the analysis were validated by including standard samples with known concentrations, a routine procedure in the lab. This ensured the reliability and accuracy of the results without the need for additional filtering.

Line 163 – "…previous studies…" but only one reference. I do not see in Fenn et al. (2018) a discussion of the molarity of the extraction solution. That paper does have a section of using solutions other than KCl or KI is a researcher wants to quantify K+ deposition.

- This source did indeed not justify our claim. We changed the text of the manuscript to "since a higher recovery of the base cations was found with a 1M HCl extraction compared to a 0.5M HCl extraction". This can be found on page 53 of Fenn et al., 2018.

Table 1 – These are very low sample sizes.

- This point has been answered in our reply to the main comment on low replication.

Line 185 - Table 2 – I think I can see why there is not an equal number of samplers in bulk deposition and throughfall in each treatment, but statistically, I do not see how the design presented in Table 2 would work.

- In response to the reviewers comments, we changed the statistics, now using another aspect of the design. Initially we compared the IER-method and the original water method for each harvest intensity treatment thereby ignoring the funnels placement. The harvest intensity treatment is shown in table 2. This test was set up to result in 4 groups containing 7 paired samplers. However, upon reflection, we realized our main focus was comparing the water-based deposition method with the ion exchange resin method in forest gaps (bulk deposition) and beneath the canopy (throughfall). To clarify, we updated the Table 2 header to specify that we only tested both methods in throughfall and bulk deposition. The original design included 13 paired throughfall samplers and 15 paired bulk deposition samplers, but due to bird feces contamination, we ended up with 9 pairs of throughfall samples and 9 pairs of bulk deposition samples. Initially, we reported different treatments expecting higher deposition in small forest gaps, influenced by surrounding trees. However, this expectation is only relevant if the IER method performs differently than the water method for throughfall and bulk deposition which was not the case (see revised version of figure 4). If no such difference exists, there's no need to control for gap size in comparing the IER method to the water method. Our results indicate that IER and water method covary consistently independent of forest gap versus crown cover (bulk vs. throughfall).
- In line with this, we revised the field study statistics to exclude the treatments (control, high-thinning, shelterwood, and clearcut), focusing instead on the funnel's position (throughfall or bulk deposition). We ran linear models for all elements with the funnel position as a random structure. These models were constructed as $yij=\beta 0+\beta 1xij+uj+\epsilon ij$, where y is the result using the IER method, x is the result using the original method, u represents the random structure, and ϵ the residuals. The random structure was included only if it improved model performance by $\Delta 2$, following Zuur et al. (2009). None of the models showed improved performance

with the funnel position as a random variable. The results are visualized in the revised Figure 4. The statistics are added to section 2.4 of the revised manuscript.

Line 192 – What is a "common" deposition collector?

- This are the bulk and throughfall deposition samplers that are commonly used in other studies as well (Bleeker et al., 2003) and are designed to capture and store the precipitation. This precipitation is then sampled for volume and nutrient concentration to calculate the bulk or throughfall deposition. To clarify this, we added the phrase 'collecting the precipitation next to and below the forest' between brackets right after mentioning this common deposition collector.

Line 193 – Maybe add "collectively" after "The 7 collectors per plot"

- Added

Line 222 – Change send to sent

- Changed

Lines 222ff – Were the samples filtered?

- Yes, the samples of the original water method were classified as surface waters and are therefore filtered following the standard procedure of the CBLB laboratory. We now added this.
- The samples of the IER-method were not filtered as explained in the reply to the comments on line 150-155. We added this information.

Line 223 – Change contents to concentrations

- Changed.

Lines 246-254 – Calculating adsorption capacity and recovery efficiency based on analyte concentrations assumes that the volume of the added solution is the same as the volume of the extract solution. Was this always the case?

- The volume of the added solution matched the extract solution's volume because we let the fluid fully drain from the resin. Thus, the subsample taken from this extractant is assumed to represent the entire extract solution. This assumption is valid as the resin was wet and gravity-drained before adding the nutrient solution, representing the total annual deposition, making the change in wetness negligible.

Lines – 263ff – The results section suffers from not considering results in light of the ANOVA and subsequent Tukey's tests. As an example, statistical results are not considered at all regarding the results shown in Figure 3. And how can it be that overloading of the cation and/or anion exchange capacity results in 100% adsorption capacity?

The reviewer addresses two points in this remark, namely the statistical tests for the adsorption capacity and the 100% adsorption for some elements when the resin is loaded beyond its capacity.

- We indeed did no statistical testing for differences in adsorption capacity since the small sample sizes strongly limit the statistical power of such tests.
- In our revised manuscript, we added a new Table 4 to replace Figure 3. In this table, we report the mean and the standard error of the mean. These standard errors are generally quite small, indicating a high precision of our measurements. Note that unlike the adsorption tests; we used ANOVA for the recovery efficiency and field tests (lines 255-259) since these data showed more variability (larger standard errors) and was less intuitive to understand. In response to your feedback, we conducted a generalized test per treatment group for the adsorption data, as individual element-specific tests were impractical due to low sample sizes and generally low standard errors, resulting in essentially constant data that lacks statistical evaluability. For elements where mean adsorption was less than 100%, we used the Wilcoxon signed rank test to assess the hypothesis of adsorption equality to 100%. To address ties (values equal to 100%), we implemented appropriate corrections within the Wilcoxon test methodology. However, we believe that the differences between the tests are readily apparent when comparing the means and standard errors of the data. Given the very small errors, we consider the observed differences as real, and not as a result of random variation.
- Overloading the cation and/or anion exchange capacity did not result in 100% adsorption capacity for all elements. We observed that only certain elements, which strongly bind to the resin, achieved (nearly) 100% adsorption capacity when the resin was overloaded beyond its exchange capacity. This suggests that the resin has a high affinity for these specific elements. In contrast, other elements, which have lower affinity for the resin, showed lower adsorption capacity under similar conditions. We discuss this phenomenon in detail in the manuscript, specifically in lines 349–362. For clarity we add these lines here: "To further test the affinity of the resin for the studied elements, the resin was loaded to approximately 160% and 240% of its capacity. Based on the adsorption capacity beyond the resins capacity, we found that the cation bed has an affinity of Ca = Fe > Cu = Mn = Zn > Mg > K > NH_4 > Na which is in line with the previous reported resin affinity (Skogley and Dobermann, 1996). The anion bed has an affinity of S > NO_3 > P which agrees with earlier studies (Skogley and Dobermann, 1996; Park et al., 2014). The resins affinity and the adsorption capacity for different levels of loading beyond the resins capacity is of importance for resin columns under suspicion of overloading. We did not find lower adsorption of Ca and Fe and only slightly lower adsorption of Cu, Mg, Mn and Zn, indicating that, when columns are slightly overloaded, these estimates are still reliable. When columns are loaded > 100% of the capacity, the estimates for K, Na, P, S, NH_4 and NO_3 are not reliable. Therefore, in case of suspicion of ion exchange overload, tests are recommended to check if stoichiometry between any element of Ca, Cu, Mg, Mn and Zn with K, Na, P, S, NH_4 and NO_3 falls within the stochiometric range of natural deposition estimates. We strongly recommend collecting the resin columns prior to resin saturation as adsorption of Na and P can further decrease when saturating the resin up to 90 or 100%. The time period that the resin can stay in the field depends on the total atmospheric deposition and the volume of resin used. For remote areas with low deposition levels and low risk of sample contamination (e.g. by bird feces) the resin can stay for multiple months up to a year in the field as long as adequate resin volumes are used".

Lines 290ff – Table 3 does not show statistical interactions. The bottom row shows mean recovery efficiencies for each element, but there were element interactions with pretreatment, molarity, and extraction type, so averaging element means across all treatments ignores the ANOVA interactions. Table 3 clearly illustrated the unbalanced ANOVA design and exceptionally minimal replication. Table S3 shows part of an ANOVA, but it only shows 2-way interactions (and not all of the 2-way interactions). A complete ANOVA would have 3-way interactions and one 4-way interactions. It is hard to tell what complete ANOVA would look like (it might not even run because of the unbalanced design with low replication).

- The reviewer addresses here multiple concerns. We have considered all these concerns point by point:
    - Table 3 (table 5 in the revised version) indeed does not show statistical interactions. We provided the arithmetic means and the arithmetic average of these means. To clarify this, we changed the average recovery to average arithmetic recovery to clearly indicate that we chose to give the arithmetic means here. The reviewer is right that it is necessary to give an overview of the interaction effects. The effects of the ANOVA interactions, following a Tukey's post-hoc test, are now added to the supplements in table S6-S9.The ANOVA test results are still in table S3.
    - Table 3 (table 5 in the revised version) indeed show the unbalanced design consisting of duplo measurements and one group with 6 samples. We now removed the groups who had only one sample as these groups did not add to the research output of this paper and caused confusion, especially regarding the statistics. The choice for duplo measurements is already argued in the reply to the previous comment regarding table 1. The unbalanced design does not necessarily pose problems when using Anova as long as the type of the Sum of Squares is reported and as long as the interaction effects are significant: see Smith and Cribbie (2014) for a comprehensive comparison of unbalanced Anova using type I, II and type III sum of squares.
    - In Table 3 (table 5 in the revised version), we presented the complete ANOVA results, which include the elements, pre-treatment, molarity, and extraction type as explanatory variables, along with all two-way interactions involving the elements. However, we cannot test for interactions between pre-treatment and molarity or between pre-treatment and extraction type because we did not test the two extraction types for each molarity, nor did we test the pre-treatment for each molarity. We did not pursue a full factorial design involving molarity, pre-treatment, and extraction type because we believe the effect of pre-treatment (drying the resin prior to extraction) does not vary with different molarities of the extractant. Additionally, while the extraction method (drip versus shake-drip) could affect the contact time of the fluid with the resin and potentially interact with different molarities, our goal was not to determine the specific molarity at which the shake-drip method becomes superior to the drip method. Because we had the data, we expanded the Anova and now added the interaction between the pre-treatment and the extraction type in the revised table S3.
    - Finally, we intentionally did not include three- and four-way interactions in our analysis. These higher-order interactions significantly increase the complexity of the analysis and can be difficult to interpret. In this manuscript, our goal is to identify an appropriate extraction method for the ion exchange resin method using tested adsorption and recovery percentages. We are not focused on

specific combinations of molarity, pre-treatment, and extraction type that may lead to higher recovery percentages. Our aim is to determine general patterns, such as whether increasing molarities is beneficial or whether a shake-drip method is preferable to a drip method. While a four-way interaction might reveal that a specific combination works best for a particular element, this does not necessarily improve the overall estimates of deposition using the IER method, as long as the adsorption and recovery percentages are reliable and have a small standard error. Reliable estimates for adsorption and recovery are far more critical than finding the perfect extraction method. Including all possible interaction terms can lead to overfitting, capturing noise rather than true underlying relationships. It also increases the risk of multicollinearity and, in cases of sparse data, can result in unreliable and unstable estimates of interaction effects. We believe in maintaining simplicity in statistical models while adequately explaining the data. Including unnecessary higher-order interactions violates the principle of parsimony and can obscure the main effects and lower-order interactions that are more important and interpretable.

Lines 310 ff – In Figure 4 and Table 5, if the goal is to compare corrected, blank corrected and recovery corrected regressions, it would be appropriate to use ANCOVA for homogeneity of slopes. This would be a better approach than simply showing the highest R2 value in bold. One cannot compare log-transformed regression coefficients to non-transformed regression coefficients (for S).

- In figure 4, the goal is to compare the data from the IER collectors and the water collectors using the adsorption capacity and the recovery efficiency for the different elements as determined with the laboratory test. Therefore, there is no need to change the statistics here as we do not want to compare corrected, blank corrected and recovery corrected regressions in this figure.
- We indeed aim to compare the corrected, blank corrected, and recovery corrected regressions in Table 5. However, we believe that using ANCOVA for homogeneity of slopes is not appropriate for this situation because we do not have separate blanks for each column nor separate recovery efficiencies for each column. This forces us to use the average contamination in blanks and the average recovery efficiency as a covariate variable in comparison to the uncorrected data of the IER collectors. ANCOVA requires the covariate to vary among the individuals in the dataset. A single average value for the covariate is not suitable and will not provide meaningful adjustments in the analysis. Without variability, a single value cannot explain any variance in the dependent variable. Including a covariate that is constant across all observations is redundant and does not contribute to the model, as it fails to adjust the dependent variable based on individual differences. Furthermore, it is not possible to use the uncorrected IER data as a covariate for the corrected IER data as an ANCOVA as there should be perfect correlation between the independent variables. As the IER data can only be corrected by averages, the covariate and the independent variable will be perfectly correlated violating the ANCOVA assumptions making this test not suitable for this data.
- It is indeed true that log-transformed regressions cannot be compared to non-transformed regression coefficients. We have revised the complete statistics in response to your comment on Table 2. This revision involved adding the funnel position (indicating bulk deposition or throughfall deposition) as a random factor,

but only if it would improve the regression model AIC by Δ2. Similar to the models shown in Figure 4, this was not the case for the IER data, whether corrected for blanks or for recovery efficiency. In the revised statistics, we have now moved away from using data transformation, resulting in comparable models.

- Finally, in order to increase the comparability between the models, we added the mean absolute error of each model to table 5.

Line 326 – I do not see that Fig. S1 shows canopy openness results.

- The canopy openness treatment related to the forest harvest intensity treatments as shown in table 2. In response to the comments on table 2 we changed the entire statistics and moved away from these canopy openness treatments (consisting of the control forest, thinned forest, shelterwood cut and clearcut) and focused now only on the throughfall and bulk deposition types. We therefore changed line 326 and removed the statement regarding canopy openness results.

Literature

Bleeker, A., Draaijers, G., van der Veen, D., Erisman, J.W., Mols, H., Fonteijn, P., Geusebroek, M., 2003. Field intercomparison of throughfall measurements performed within the framework of the Pan European intensive monitoring program of EU/ICP Forest. Environ Pollut 125, 123-138.

Park, S.-C., Cho, H.-R., Lee, J.-H., Yang, H.-Y., YANG, O.-B., 2014. A study on adsorption and desorption behaviors of 14C from a mixed bed resin. Nuclear Engineering and Technology 46, 847-856.

Skogley, E.O., Dobermann, A., 1996. Synthetic ion-exchange resins: Soil and environmental studies. J Environ Qual 25, 13-24.

Smith, C.E., Cribbie, R., 2014. Factorial ANOVA with unbalanced data: a fresh look at the types of sums of squares. Journal of Data Science 12, 385-403.

Zuur, A., Ieno, E.N., Walker, N., Saveliev, A.A., Smith, G.M., 2009. Mixed effects models and extensions in ecology with R. Springer Science & Business Media.