



# VIDEREUDVIKLING OG KLARGØRING AF METALSTAT FOR LANDSDÆKKENDE BEREGNINGER

Videnskabelig rapport fra DCE – Nationalt Center for Miljø og Energi

nr. 606

2024



AARHUS  
UNIVERSITET

DCE – NATIONALT CENTER FOR MILJØ OG ENERGI



# VIDEREUDVIKLING OG KLARGØRING AF METALSTAT FOR LANDSDÆKKENDE BEREGNINGER

---

Videnskabelig rapport fra DCE – Nationalt Center for Miljø og Energi

nr. 606

2024

Peter Borgen Sørensen<sup>1</sup>  
Christian Frølund Damgaard<sup>1</sup>  
Jesper Leth Bak<sup>1</sup>  
Hans Estrup Andersen<sup>1</sup>  
Denitza Voutchkova<sup>2</sup>  
Lærke Thorling<sup>2</sup>  
Poul Løgstrup Bjerg<sup>3</sup>  
Peter Engelund Holm<sup>4</sup>

<sup>1</sup>Aarhus Universitet, Institut for Ecoscience

<sup>2</sup>De Nationale Geologiske Undersøgelser for Danmark og Grønland

<sup>3</sup>Danmarks Tekniske Universitet, Department of Environmental and Resource Engineering

<sup>4</sup>Københavns Universitet, Institut for Plante og Miljøvidenskab



AARHUS  
UNIVERSITET

DCE – NATIONALT CENTER FOR MILJØ OG ENERGI

# Datablad

Serietitel og nummer:	Videnskabelig rapport fra DCE - Nationalt Center for Miljø og Energi nr. 606
Kategori:	Rådgivningsrapporter
Titel:	Videreudvikling og klargøring af MetalStat for landsdækkende beregninger
Forfatter(e):	Peter Borgen Sørensen <sup>1</sup> , Christian Frølund Damgaard <sup>1</sup> , Jesper Leth Bak <sup>1</sup> , Hans Estrup Andersen <sup>1</sup> Denitza Voutchkova <sup>2</sup> , Lærke Thorling <sup>2</sup> Poul Løgstrup Bjerg <sup>3</sup> , Peter Engelund Holm <sup>4</sup>
Institution(er):	<sup>1</sup> Aarhus Universitet, Institut for Ecoscience <sup>2</sup> De Nationale Geologiske Undersøgelser for Danmark og Grønland <sup>3</sup> Danmarks Tekniske Universitet, Department of Environmental and Resource Engineering <sup>4</sup> Københavns Universitet, Institut for Plante og Miljøvidenskab
Udgiver:	Aarhus Universitet, DCE – Nationalt Center for Miljø og Energi ©
URL:	<a href="https://dce.au.dk">https://dce.au.dk</a>
Udgivelsesår:	2024
Redaktion afsluttet:	Juni 2024
Faglig kommentering:	Martin Mørk Larsen
Kvalitetssikring, DCE:	Iben Boutrup Kongsfelt
Sproglig kvalitetssikring:	Charlotte Kler
Ekstern kommentering:	Kommentarerne findes her: <a href="https://dce.au.dk/fileadmin/dce.au.dk/Udgivelser/Videnskabelige_rapporter_600-699/KommentarerSR/SR6xx_komm.pdf">https://dce.au.dk/fileadmin/dce.au.dk/Udgivelser/Videnskabelige_rapporter_600-699/KommentarerSR/SR6xx_komm.pdf</a>
Finansiel støtte:	Miljøstyrelsen
Bedes citeret:	Sørensen, P.B., Damgaard, C.F., Bak, J.L., Andersen, H.E., Voutchkova, D., Thorling, L., Bjerg, P.L., Holm, P.E. 2024. Videreudvikling og klargøring af MetalStat for landsdækkende beregninger. Aarhus Universitet, DCE – Nationalt Center for Miljø og Energi, 68 s. - Videnskabelig rapport nr. 606 Gengivelse tilladt med tydelig kildeangivelse
Sammenfatning:	Denne rapport indgår i et større arbejde med udvikling af statistiske modeller til estimering af koncentrationer af miljøfarlige stoffer i overfladevand for Miljøstyrelsen ved at videreudvikle og anvendeliggøre en pilotudgave af en statistisk model (MetalStat) for forekomsten af 5 metaller som opløst koncentration i vandløbsvand. Rapporten tager således udgangspunkt i afrapportering af MetalStats principper givet i Sørensen et al., 2022 og 2023. Modellen er sat op, så metalkoncentrationen i vandløbet regnes for at være bestemt af det ID15 opland, som målestationen er placeret i, hvor et ID15 opland har en varierende udstrækning på typisk 10-15 km <sup>2</sup> . Den primære opgave for 2023 er at få pilotudgaven af MetalStat gjort brugbar med levering af resultater, der kan indgå i Miljøstyrelsens forvaltning.
Emneord:	metaller, vandløb, koncentrationsniveau, model, MetalStat
Foto forside:	Lystrup Å, Peter Borgen Sørensen
ISBN:	978-87-7156-873-8
ISSN (elektronisk):	2244-9981
Sideantal:	68
Supplerende oplysninger:	Rapportens bilag er skrevet af projektdeltagerne fra GEUS, Denitza Voutchkova, Lærke Thorling, med en lettere redigering af Peter Borgen Sørensen

# Indhold

<b>Forord</b>	<b>5</b>
<b>Sammenfatning</b>	<b>6</b>
<b>Summary</b>	<b>9</b>
<b>1 Indledning</b>	<b>12</b>
<b>2 Opdatering med måledata</b>	<b>14</b>
<b>3 Forbedret beskrivelse af grundvandsbidrag</b>	<b>16</b>
3.1 Korrelationsanalyse mellem grundvandsmodel og simuleringer i MetalStat	16
3.2 Screening for sammenhæng mellem grundvandskoncentration og vandløbskoncentration	18
3.3 Opdateret X matrice i MetalStat	27
<b>4 Resultater fra MetalStat</b>	<b>31</b>
4.1 Generelt model output	31
4.2 Landsdækkende koncentrationsestimater	35
<b>5 Udarbejdelse af beslutningsgrundlag til Miljøstyrelsen</b>	<b>44</b>
<b>6 Udvikling af fremtidig MetalStat driftsmodel, samt fremadrettet arbejde med modellering</b>	<b>47</b>
6.1 Udvikling af fremtidig MetalStat driftsmodel	47
6.2 Fremadrettet arbejde med modellering for andre miljøfarlige stoffer	48
<b>7 Konklusion</b>	<b>49</b>
<b>8 Referencer</b>	<b>52</b>
<b>Bilag, Beregning af grundvandskoncentration (Engelsk)</b>	<b>53</b>
Introduction	53
Input data and methods	54
Defining aquifer/groundwater type	55
GVF classification	56
Results	59
Concentrations of Cd, Cu, Ni, Pb and Zn at the groundwater body level	61
Groundwater contribution of Cd, Cu, Ni, Pb and Zn to the ID15 catchments	62
References	67



## Forord

Denne rapport beskriver resultatet af et projekt, hvis formål var at videreudvikle modellen "MetalStat", så modellen kan levere beslutningsgrundlag til en vurdering af overholdelse af miljøkvalitetskrav for metallerne bly, cadmium, kobber, nikkel og zink i vandløb. Projektet er udført for Miljøstyrelsen med henblik på at kunne understøtte vandplanlægningen med modellering af koncentrationen af miljøfarlige stoffer i overfladevand bl.a. i forhold til, om der er risiko for forekomst af koncentrationer over miljøkvalitetskravene.

Projektgruppen består af: Peter Borgen Sørensen og Christian Frølund Damgaard, der primært har haft ansvar for den matematiske konstruktion af modellen samt levering af valideringsdata til Miljøstyrelsen. Lærke Thorling og Denitza Voutchkova har primært udviklet et nyt grundvandsmodul, der er blevet diskuteret i en faglig arbejdsgruppe bestående af Hans Estrup Andersen, Poul L. Bjerg, Peter Engelund Holm og Jesper Leth Bak. Projektet har været fulgt af en gruppe medrepræsentanter fra Miljøstyrelsen, primært bestående af Rasmus Dyrmosø Nørregaard, Maj-Britt Bjergager samt Kasper Lauge Thomassen. Martin Mørk Larsen har været fagfællebedømmer for Institut for Ecoscience. For DCE er rapporten kvalitetssikret af Iben Boutrup Kongsfelt. Projektet har været drøftet i denne gruppe. Miljøstyrelsen har haft et udkast af rapporten til kommentering.

Projektet er udført som et udviklingsprojekt, og den primære projektperiode var fra juni 2023 til december 2023 med afrapportering i 2024. Aarhus Universitet har bistået med indkøb af regneserver.

## Sammenfatning

Formålet med MetalStat er at udnytte den viden om vandkoncentrationer, der er indsamlet gennem målinger af et udsnit af danske vandløb, til bedst muligt at forudsige tilstanden i alle danske vandløb. For at sikre mest mulig viden om metalkoncentrationerne i de forskellige typer vandområder er det således vigtigt, at betydningen af alle hovedtyper af emissionskilder er dækket ind med de udvalgte målestationer. Disse vilkår dækker over forskelle i kilder til metallerne fra menneskelige aktiviteter og geogene kilder. MetalStat skal ses som en generalisering og kausal forståelse af de indsamlede monitoringsdata og således som en optimeret brug af de ressourcer, der er brugt i monitoringsprogrammet.

MetalStat er en statistisk model, hvilket betyder, at modellen beregner sandsynligheden for at måle forskellige koncentrationer på et givent sted. Nogle koncentrationsniveauer vil derved blive tillagt en større sandsynlighed end andre, og dette giver et interval af realistiske koncentrationsniveauer. Beregningen af dette interval afhænger af placering i landet, sæson og år. MetalStat beskriver forekomsten af de fem metaller: bly, cadmium, kobber, nikkel og zink i vandløbsvand (Sørensen m.fl., 2023). De beregnede intervaller af realistiske koncentrationsniveauer blev valideret mod måledata, der ikke var blevet brugt til kalibrering af MetalStat i Sørensen m.fl., 2023. På baggrund af disse resultater blev det i 2023 besluttet at fortsætte med udvikling af MetalStat hen mod at kunne understøtte vurderinger af, om miljøkvalitetskrav er overholdt i danske vandområder for de fem metaller.

I 2023 blev MetalStat opdateret med målinger for 2020 og 2021 med de prøve-tagninger, hvor mindst tre metaller er målt i samme prøve. Disse prøver fordeles sig ret heterogent mellem årene, hvor der er en tendens til, at de seneste år er bedst repræsenteret med prøver. Dette kan øge usikkerheden ved estimering af den generelle betydning af år på tværs af alle ID15-oplande, da forskelle mellem enkelte år kan være en konsekvens af forskelle i valg af opland, der indsamles prøver fra. Dette forhold gælder dog primært ved sammenligning mellem enkelte år, mens en generel tidstrend mellem flere år ikke er nær så følsom for en heterogen fordeling af prøveindsamlingen.

Da grundvandsbidraget kan være betydeligt for visse metaller, er MetalStat i denne rapport udvidet med et potentielt grundvandsbidrag af hvert metal blandt de beskrivende variable for vandløbskoncentrationen ( $X$ -matricen). MetalStat er således blevet udvidet med en grundvandsbeskrivelse for hvert ID15-opland, der kombinerer grundvandsdata med en empirisk model for at få en landsdækkende beskrivelse. Grundvandsforekomsternes koncentrationniveauer anvendes til at beregne grundvandets middelkoncentration for hvert ID15-opland for de underliggende grundvandsforekomster. MetalStat har i forvejen en beregnet tilførsel af grundvand til hvert ID15-opland fra DK-modellen, så denne vandmængde bliver multipliceret med grundvandskoncentrationen for at få et mål for potentielt massebidrag til hvert ID15-opland. Det potentielle massebidrag indgår i  $X$ -matricen på lige fod med de øvrige variable, der søger at forklare forekomsten af metallerne.

Sammenlignes de forudsagte grundvandskoncentrationer i grundvandsforekomsterne med landsdækkende og målte vandløbskoncentrationers median værdier for ID15-oplande, ses det, at grundvandskoncentrationerne ofte er høje sammenlignet med vandløbskoncentrationerne, især for cadmium og



zink og for grundvandsmagasiner med iltede forhold og lav pH. Derfor overrasker det, at en simpel korrelation mellem målt vandløbskoncentration og tilhørende beregnet/målt grundvandskoncentration placeret under det målte ID15-opland kun udviser en svag, men dog positiv, korrelation for metallerne bly, cadmium, nikkel og zink, mens kobber ikke viser nogen korrelation. MetalStat tager dog langt mere detaljerede og geografisk fordelte hensyn til forholdene i oplandene sammenlignet med denne simple analyse. Der kan således godt være enkelte områder af Danmark, hvor et specifikt grundvandsbidrag har stor betydning, mens denne effekt drukner i en generel simpel data-analyse for hele Danmark.

Pilotversionen af MetalStat, svarende til den testede version i Sørensen m.fl. (2022), er udviklet til en forløbelig driftsversion i forhold til beregningstid og opdateret med nye målinger og målestationer frem til og med 2021 og er opdateret med den forbedrede grundvandsbeskrivelse. Resultaterne i denne rapport svarer til udtræk fra MetalStat den 14/12-2023.

Med forbehold for at korrelationskoefficienten  $R^2$  er et groft og forsimplet mål for  $X$ -matricens evne til at beskrive forskellen i forekomsten af metallerne, så opnås følgende  $R^2$  med stigende rækkefølge bly (0,13) -> nikkel (0,23) -> cadmium (0,31) -> zink (0,35) -> kobber (0,37). Denne liste viser, at bly har en noget lavere  $R^2$  værdi sammenlignet med de øvrige metaller, hvilket enten kan skyldes, at der mangler en beskrivende variabel i  $X$ -matricen for bly, eller at bly simpelthen er mere tilfældigt fordelt over landet pga. en historisk gammel og generel kontaminering uden større og regelbaserede geografiske forskelle. Specifikt er zink blevet reguleret i landbruget i de seneste år, så her kunne det f.eks. være relevant med en opdateret  $X$  matrice, så kalibreringen af MetalStat kan adskille betydningen af landbrugsdrift hhv. før 2020 og fra 2020 og fremefter. MetalStat kan med fordel blive underlagt en grundig revidering af  $X$ -matricen for at optimere den og inddrage tidseffekter samt tolke de enkelte variable i  $X$ -matricen og derigennem identificere nøglekilder til metaller i vandløbene.

Der er et rumligt sammenfald mellem vandløbskoncentrationerne med de beregnede grundvandskoncentrationer fra grundvandsmodellen hos især cadmium, nikkel og zink, mens bly og kobber ikke udviser så tydelig tendens, hvilket også passer med, at bly og kobber netop var de to metaller, der viste sig mindst relateret til grundvandskoncentrationer. Det er især de vestjyske vandløbsoplande omkring Skjern Å og Varde Å for cadmium, nikkel og zink, som har et sammenfald mellem modellerede vandløbskoncentrationer fra MetalStat og beregnede grundvandskoncentrationer.

Denne rapport leverer beslutningsgrundlaget til Miljøstyrelsen i forhold til, om resultater fra MetalStat skal anvendes i deres sagsbehandling. Et udtræk fra MetalStat den 14/12-2023 for 84 ID15-oplande er således leveret til Miljøstyrelsen. For disse 84 oplande er der målte værdier, som ikke er taget med i kalibrering af modellen, som Miljøstyrelsen verificerer udtrækket op imod. Disse 84 oplande er udvalgt til at udgøre en repræsentativ stikprøve både i forhold til geografisk placering og i forhold til type af opland, som den er beskrevet gennem  $X$ -matricen.

Der er et arbejde i gang med at mindske beregningstiden for MetalStat med brug af Laplace approksimationer, og første indledende test-beregninger er meget lovende, men det er dog for tidligt at komme med helt endelige konklusioner omkring den reelle formindskelse i beregningstid, da der forestår

en mere præcis vurdering af, hvor god approksimationen er. Dette arbejde regnes for afsluttet i 2024, og hvis det lykkes at reducere beregningstiden, som det tyder på, bør MetalStat inddrage massebalancen mellem tilstødende oplande direkte i løsning af de styrende ligninger, så den nuværende efterjustering med massebalancekorrektion helt kan undgås. Dette vil udvide modellen med flere latente variable og har derfor indtil nu været anset som urealistisk af regnetekniske grunde.

Arbejdet med MetalStat har gjort det tydeligt, at modelleringsaktiviteten skal betragtes opdelt mellem hhv. en statistisk model og en forklarende model ( $X$  matrice), hvor den statistisk model med fordel kan anvendes mere generelt på måledata for stort set alle miljøfarlige stoffer, mens den forklarende model anvendes over for de stofgrupper, hvor der hersker tilstrækkeligt med viden om årsagen til stoffernes forekomst.

## Summary

The purpose of MetalStat is to utilize the knowledge of water concentrations gathered through measurements collected from a fraction of Danish running waters to predict the condition of all Danish running waters. To ensure the best possible knowledge about metal concentrations in the different types of water courses, it is, thus, important that all main categories of emission sources are covered with the selected measured stretches of running waters. These categories cover differences in sources of the metals from human activities and geogenic sources. MetalStat must be seen as a generalization and causal understanding of the collected monitoring data and, thus, as an optimized use of the resources applied in the monitoring program.

MetalStat is a statistical model that calculates the probability of measuring different concentrations at a given location. Some concentration levels will thereby be assigned a greater probability than others, and this gives a range of realistic concentration levels. The calculation of this interval depends on the location in the country, seasonality and year. MetalStat describes the occurrence of the five metals: lead, cadmium, copper, nickel and zinc (Sørensen et al., 2023). The calculated intervals of realistic concentration levels were validated against measurement data that had not been used for the calibration of MetalStat in Sørensen et al., 2023. Based on these results, it was decided in 2023 to continue with the development of MetalStat towards being able support assessments of whether environmental quality criteria have been met in Danish running water areas for the five metals.

In 2023, MetalStat was updated with measurements for 2020 and 2021 with samplings, where at least three metals were measured in the same sample. These samples are distributed quite heterogeneously between the years, where there is a tendency for the most recent years to be best represented with samples. This may increase the uncertainty when estimating a general significance of changes between years across all ID15 catchments, as differences between individual years may be a consequence of differences in the choice of catchment from which samples are collected. However, this relationship primarily applies to comparisons between individual years, while a general time trend between several years is not nearly as sensitive to a heterogeneous distribution of the sample collection.

As the groundwater contribution may be important for certain metals, MetalStat is extended to include a possible groundwater contribution of each metal among the descriptive variables for the running water concentration. MetalStat has, thus, been expanded with a groundwater description for each ID15 catchment, which combines groundwater data with an empirical model to obtain a country-wide description. The groundwater body concentration levels are used to calculate the average groundwater concentration for each ID15 catchment between the underlying groundwater bodies. MetalStat has a calculated supply of groundwater to each ID15 catchment from the DK model in advance, so this water volume is multiplied by the groundwater concentration to determine the potential mass contribution to each ID15 catchment. The potential mass contribution is included in the  $X$  matrix together with the other variables that seek to explain the occurrence of the metals.

A simple comparison of the predicted groundwater concentrations in the groundwater bodies with the median values of national and measured stream concentrations for ID15 catchments, discloses a groundwater concentration level comparable to the level of running water, especially for cadmium and zinc and for groundwater reservoirs with oxic conditions and low pH. Therefore, it is surprising that a simple correlation between measured stream concentration and corresponding calculated/measured groundwater concentration located below the measured ID15 catchment only shows a weak, but positive, correlation for the metals lead, cadmium, nickel, and zinc, while copper does not show any correlation. Nevertheless, MetalStat includes more detailed and geographically distributed considerations compared to this simple analysis. Individual areas of Denmark can be identified, where a specific groundwater contribution is of great importance, while this effect is masked in a general simple data analysis for the whole of Denmark.

Similarly to the version tested in Sørensen et al. (2022), the pilot version of MetalStat has been developed into a temporal operating version in relation to calculation time and has been updated with new measurements and measuring stations, including year 2021, as well as updated with the improved groundwater description. The results in this report correspond to results from MetalStat delivered 14/12-2023.

Bearing in mind that the correlation coefficient  $R^2$  is a rough and simplistic measure of the ability of the  $X$  matrix to describe the difference in the occurrence of the metals, the following  $R^2$  is obtained with increasing order: lead (0.13) -> nickel (0.23) -> cadmium (0.31) -> zinc (0.35) -> copper (0.37). This list shows that lead has a somewhat lower  $R^2$  value compared to the other metals, which can either be due to a missing descriptive variable in the  $X$  matrix for lead, or to the fact that lead is simply more randomly distributed across the country due to a historically old and general contamination without any major or rule-based geographical differences. Specifically, zinc has been regulated in agriculture in recent years, so here it could e.g. be relevant with an updated  $X$  matrix, so that the calibration of MetalStat can separate the contribution from agricultural operations taking place before 2020 from activities at year 2020 or later. MetalStat can benefit from a reassessment of the  $X$  matrix to optimize the collection of variables and to include time effects as well as interpret the individual variables in the  $X$  matrix and thereby identify key sources of metals in the watercourses.

There is a spatial coincidence between the running water concentrations and the calculated groundwater concentrations from the groundwater model, especially for cadmium, nickel and zinc, while lead and copper do not show such a clear trend, which also fits with the fact that precisely lead and copper were two metals that appeared least related to groundwater concentrations. This especially applies to the West Jutland catchments around Skjern Å and Varde Å for cadmium, nickel, and zinc, which have a coincidence between modeled running water concentrations from MetalStat and calculated groundwater concentrations for these three metals.

This report provides the decision-making basis for the Danish Environmental Protection Agency in relation to whether results from MetalStat should be used in their case processing. An extract from MetalStat on December 14, 2023, for 84 ID15 catchments has, thus, been delivered to the Environmental Agency. For those 84 catchments, concentration values were measured that were not included in the calibration of the model, and these measurements

were thus used by the Environmental Protection Agency in the verification. These 84 catchments have been selected to form a representative sample both in relation to geographical location and in relation to the type of catchment, as described through the  $X$  matrix.

Work is in progress to improve the calculation time for MetalStat using Laplace approximations, and first initial test calculations are very promising, but it is still too early to make definitive conclusions about the real improvement in calculation time. This work is expected to be completed in 2024, and if, as expected, it succeeds in reducing the calculation time, MetalStat should include the mass balance between adjacent catchments directly in solving the governing equations, so that the current readjustment with mass balance correction can be completely avoided. This will expand the model with several latent variables and has therefore previously been considered unrealistic for computational reasons.

The work with MetalStat has made it clear that the modeling activity must be considered to be divided between a statistical model and an explanatory model ( $X$  matrix), where the statistical model can advantageously be applied more generally to measurement data for virtually all environmentally hazardous substances, while the explanatory model is used for those substance groups where there is sufficient knowledge of the cause of the substance's occurrence.

# 1 Indledning

I 2020 tog Miljøstyrelsen initiativ til at få lavet en udredning, der skulle afklare potentialet for at modellere koncentrationsniveauer for miljøfarlige stoffer (MFS) i vandløb (Sørensen m.fl., 2021). Denne udredning opstillede en række mulige modelleringsaktiviteter og anviste forskellige modelleringsprincipper, som var tilpasset situationer med mere eller mindre manglende viden og dermed forskellige grader af usikkerhed. På baggrund af denne udredning besluttede Miljøstyrelsen at igangsætte arbejdet med at udvikle en statistisk model med fokus på fem metaller: bly, cadmium, kobber, nikkel og zink. Disse metaller blev vurderet til at være dem med de mest komplette landsdækkende data og med velbeskrevet geokemi som led i en strategi med at opstille statistiske modeller for MFS i vandløb. I 2021 blev der udviklet en statistisk model (MetalStat), der kunne anvende monitoringsdata til bestemmelse af opløste metalkoncentrationer i danske vandløb (Sørensen m.fl., 2022). Modellen udnytter et forudgående arbejde med modeller for næringsstoffer i danske vandløb, der inddeler Danmark i såkaldte ID15-oplande med varierende udstrækning på typisk 10-15 km<sup>2</sup>.

MetalStat bygger på massebalancer, der kobler emission, geokemi samt transport til og i vandløbet. Udfordringen for en landsdækkende model er at opnå en systematisk og tilstrækkelig kvantificering af usikkerheden. For at kunne levere en tilstrækkelig troværdig usikkerhedsanalyse er MetalStat derfor bygget op som en statistisk model, der kombineres med en massebalance for metallerne, både i forhold til tilført mængde ( $X$  matricen) og i forhold til massetransporten mellem oplandene. Dette er en ny tilgang i forhold til den traditionelle modellering af metaller, men inden for folkesundhed, hvor der udvikles modeller for udbredelse af sygdomme, er den statistiske tilgang velkendt (Lawson, 2018). MetalStat kan derfor ses som en metodemæssig synergi mellem konventionel modellering af koncentrationsniveauer af metaller og de statistisk hierarkiske modeller, som bliver anvendt til f.eks. at modellere forekomst af sygdomme.

Formålet med MetalStat er at udnytte den viden om vandkoncentrationer, der er indsamlet gennem målinger af et udsnit af danske vandløb, til bedst muligt at forudsige tilstanden i alle danske vandløb. Det betyder, at nye målinger fra f.eks. NOVANA også i fremtiden vil kunne forbedre MetalStats evne til at estimere mere præcise forudsigelser, og MetalStat bør derfor løbende opdateres med de nyeste målinger i det omfang, sådanne bliver tilgængelige. Hvis der er nogle vilkår, der ikke er dækket ind med målinger i de monitoringsprogrammer, der danner input til MetalStat, vil disse vilkår heller ikke være afspejlet i modellens forudsigelser. Det er således vigtigt, at betydningen af alle hovedtyper af emissionskilder er dækket ind med de udvalgte målestationer. Disse vilkår dækker over forskelle i kilder til metallerne fra menneskelige aktiviteter og geogene kilder. Det er derfor vigtigt, at måleprogrammer løbende evalueres for at sikre, at alle vilkår er dækket ind. MetalStat skal ses som en generalisering og kausal forståelse af de indsamlede monitoringsdata og således som en optimeret brug af de ressourcer, der er brugt i monitoringsprogrammet.

MetalStat er en statistisk model, hvilket betyder, at modellen beregner sandsynligheden for at måle forskellige koncentrationer på et givent sted. Nogle koncentrationsniveauer vil derved blive tillagt en større sandsynlighed end

andre, og dette giver et interval af realistiske koncentrationsniveauer. Beregningen af dette interval afhænger af placering i landet, sæson og år. Der er mange detaljer bag beregningerne, der gør, at regnearbejdet er omfattende, så de første beregninger i 2021 blev foretaget for de tre metaller bly, cadmium og nikkel.

I 2022 blev MetalStat udvidet til i alt fem metaller: bly, cadmium, kobber, nikkel og zink (Sørensen m.fl., 2023). Desuden blev de beregnede intervaller af realistiske koncentrationsniveauer valideret mod målinger, der ikke var blevet brugt til kalibrering af MetalStat (Sørensen m.fl., 2023). Resultatet af dette arbejde var så lovende, at det i 2023 blev besluttet at fortsætte med udvikling af MetalStat hen mod at kunne understøtte vurderinger af, om miljøkvalitetskrav er overholdte i danske vandområder for de fem metaller. Den endelige beslutning om at anvende resultater fra MetalStat ligger uden for denne rapports kompetence, men rapporten vil levere beslutningsgrundlaget for en sådan beslutning. MetalStat omfatter et stort beregningsarbejde med mange variable, så for at gøre modellen mere brugbar bør der udføres yderligere tiltag for at mindske behovet for cpu timer til beregningerne. Desuden er beregningsprincipperne bag MetalStat lovende i forhold til at beregne koncentrationsniveauer for andre stoffer end de fem metaller, hvorfor en generel optimering af modellen til at kunne omfatte andre stoffer også vurderes ønskelige. En omkodning af MetalStat er dog ikke blevet prioriteret som en primær aktivitet i 2023, da Sørensen m.fl., 2023, også konkluderede, at MetalStat burde udvides fagligt til eksplicit at inddrage et muligt grundvandsbidrag, da det geogene bidrag af metaller for områder i Danmark potentielt kan have betydning for vandløbskoncentrationen. På den baggrund blev det besluttet, at arbejdet i 2023 skulle prioritere grundvandsbidraget over en egentlig omkodning af MetalStat, hvorfor den primære modifikation af MetalStat i 2023 bestod af en optimering af beregningstiden af den eksisterende kode, en opdatering med flere måledata samt opstart af en beregningsmæssig optimeret omkodning.

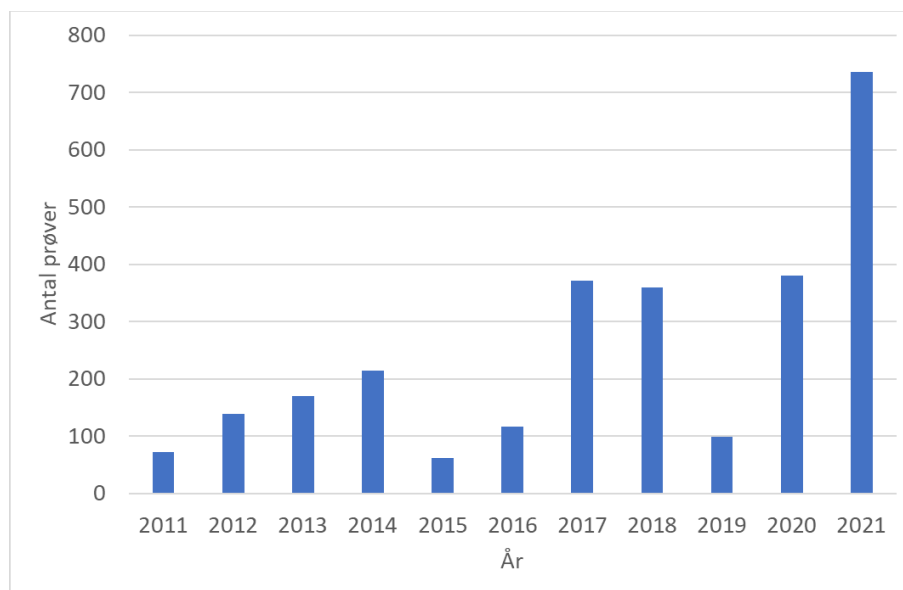
Følgende aktiviteter er beskrevet i denne rapport:

- Opdatering af MetalStat med nye måledata
- Opdatering af MetalStat med en bedre beskrivelse af grundvandsbidraget
- Opstart af arbejdet med at omkode MetalStat til en beregningsoptimeret version
- Levering af beregning som beslutningsgrundlag til, om MetalStat skal understøtte en vurdering omkring miljøkvalitetskravs overholdelse
- anbefalinger til fremtidigt arbejde med modellering af MFS, inklusiv metaller i vandløb.

## 2 Opdatering med måledata

I 2023 blev MetalStat opdateret med indsamlede vandprøver for 2020 og 2021, og det samlede antal prøver for hvert år er vist i Figur 2.1. "Fordeling af antal prøver for metallerne mellem de enkelte år". En prøve er defineret som en prøvetagning, hvor der mindst er målt for tre ud af de fem metaller, der er i modellen. I denne sammenhæng tæller en måling under detektionsgrænsen med som en målt værdi. Det ses på Figur 2.1. "Fordeling af antal prøver for metallerne mellem de enkelte år", at antallet af prøver varierer en del mellem årene, og der er en tendens til, at de seneste år er bedst repræsenteret.

**Figur 2.1.** Fordeling af antal prøver for metallerne mellem de enkelte år



Prøverne fordeler sig mellem de enkelte metaller som vist i Tabel 2.1. "Fordeling af analyser for de fem metaller", og det ses, at der for bly er en del målinger under detektionsgrænsen, mens der for kobber og zink er en del manglende værdier, hvor der ikke er blevet analyseret for metallet.

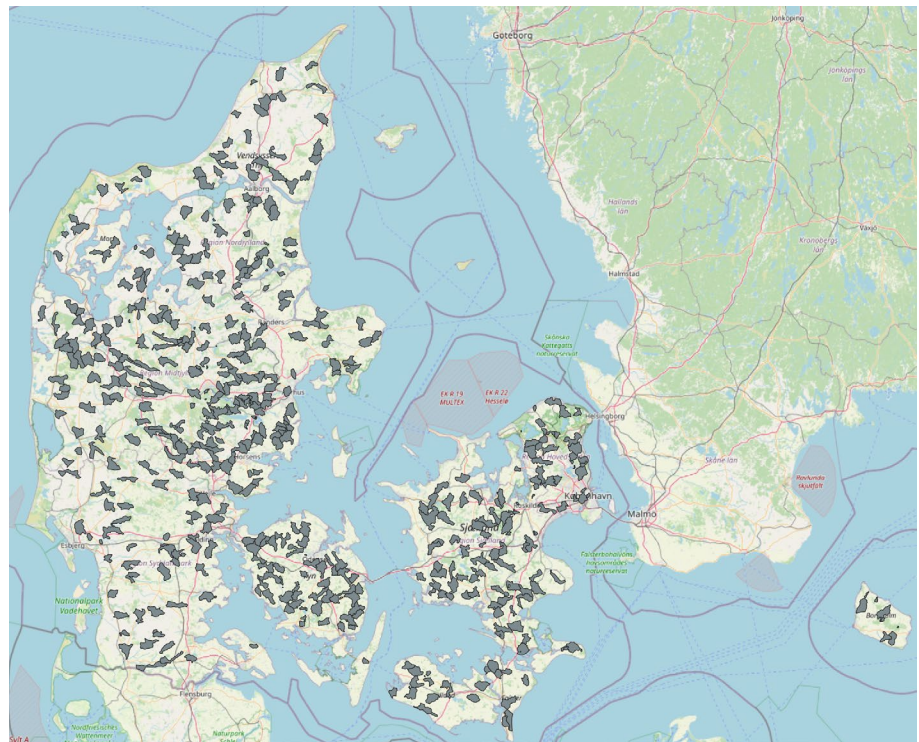
**Tabel 2.1.** Fordeling af analyser for de fem metaller

Metal	Fundet	Under detektionsgrænsen	Ikke analyseret
Bly	1917	803	0
Cadmium	2398	321	1
Kobber	1673	14	1033
Nikkel	2696	24	0
Zink	1437	126	1157

Alle prøver, der er anvendt, er tilegnet et ID15opland svarende til det opland, der afvander direkte til det sted, hvor prøven er indsamlet. På Figur 2.2 ses en oversigt over de oplande, der er tilegnet målinger i større eller mindre omfang, og viser således de vandløbsoplande, som MetalStat anvender til at forudsige koncentrationsniveauer i hele landet.



**Figur 2.3.** Angivelse af de ID15-oplande, der har tildelt prøver med målte koncentrationsniveauer for mindst tre metaller.



### 3 Forbedret beskrivelse af grundvandsbidrag

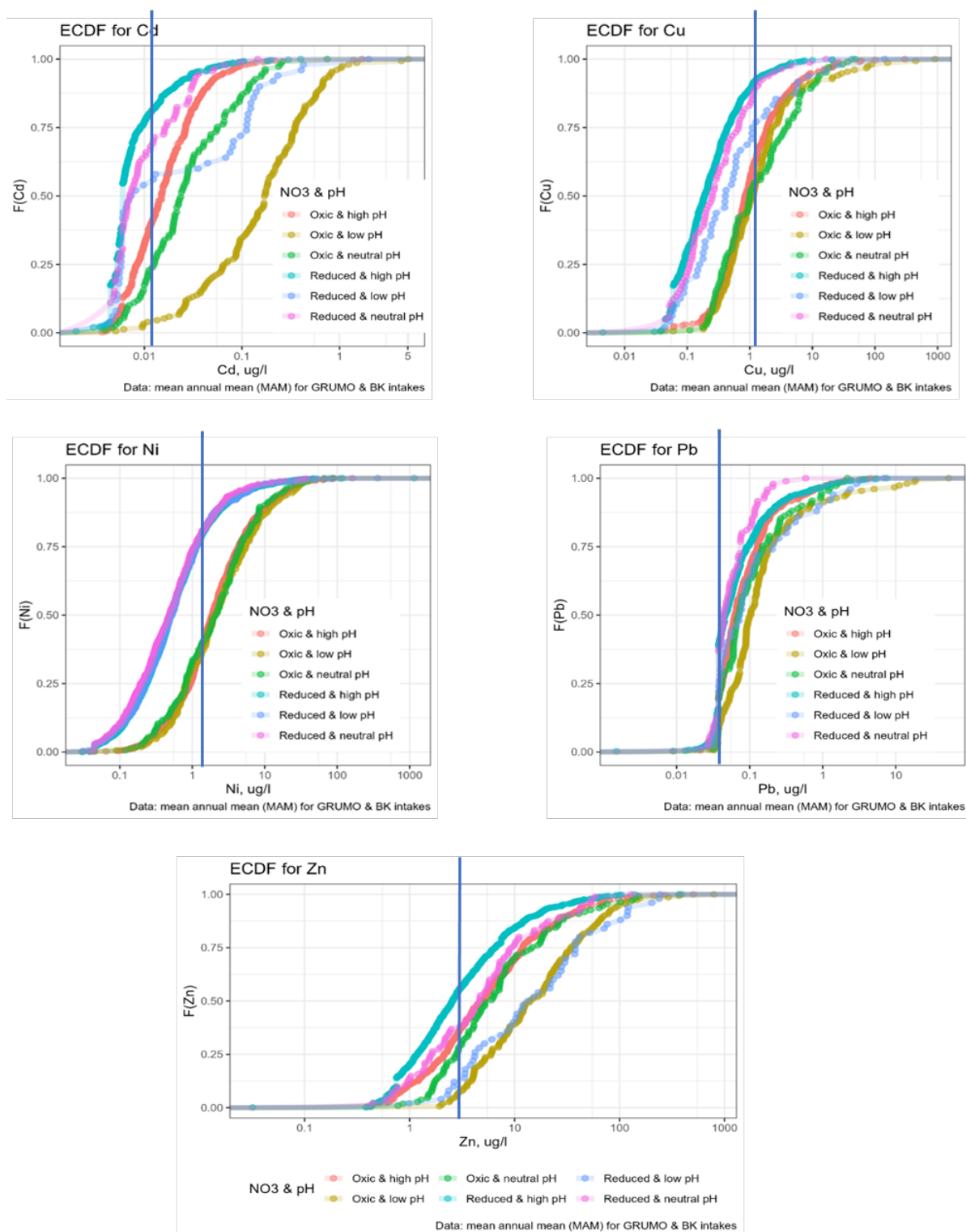
I pilotversionen af MetalStat er en mulig geogen kilde til metal fra grundvand alene inddraget indirekte, f.eks. gennem mængden af tilstrømmet grundvand, i form af en middel årsvolumen af tilført grundvand til vandløbet i hvert ID15-opland. Så i det omfang der er en landsdækkende konstant baggrundskoncentration af metal i grundvand, vil tilførslen til vandløb være proportional med grundvandstilførslen og derfor allerede være beskrevet i pilotversionen af MetalStat. Hvis der derimod er geografiske forskelle på grundvandskoncentrationen for de fem metaller, vil pilotversionen ikke kunne beskrive disse forskelle som en forskel i grundvandstilførsel. For at sikre størst mulig inddragelse af geogene kilder til metallerne er MetalStat derfor udvidet med en simpel grundvandsmodel, der er udviklet med brug af grundvandsmålinger.

Sammenhængen mellem geokemiske parametre, geokemisk oprindelse og forekomsten af metaller i grundvandet er undersøgt af Mortensen m.fl., 2021. En hovedkonklusion fra denne undersøgelse er, at geogen forekomsten af de fem metaller primært afhænger af pH og redox-forhold, og at store geografiske forskelle for nogle af metallerne er knyttet til forskellige reservoirbjergarter. Med det udgangspunkt anbefalede Sørensen m.fl., 2022, at udbygge X matricen til bedre at kunne beskrive grundvandsbidraget. Sørensen m.fl., 2022, anbefalede, at der blev taget udgangspunkt i eksisterende undersøgelser, der relaterer metaller i grundvand til geologi, pH og redoxforhold, hvor der foretages en indledende analyse af sammenhænge mellem metalkoncentration i grundvand, pH, redox og geologi, som derefter skal danne input til beskrivelsen med X matricen i MetalStat. Med det som udgangspunkt blev GEUS i 2023 inddraget i arbejdet med den opgave at udvide X matricens variable med brug af en grundvandsmodel. Redox-forhold og pH bliver kombineret med geologien i det grundvandsmagasin, der forudsættes at levere vand til vandløbet i det pågældende ID15-opland. En sådan udvidelse af MetalStat er indeholdt i nærværende rapportering, hvor der her i Kapitel 3 samt bilag udvikles en model for grundvandsbidrag, der kobles til MetalStat ved at indføre beregninger fra grundvandsmodellen som en del af X matricen. Princippet er at multiplicere en grundvandskoncentration for hvert ID15-opland med den beregnede mængde grundvand, der tilstrømmer oplandet, for derved at få en potentiel massetilførsel af metal fra grundvandet. Denne tilførsel regnes for potentiel, og MetalStat bestemmer ved kalibrering den nationale vægtningsfaktor (konstant), som skal tillægges den potentielle tilførsel for bedst muligt at beskrive variationer i vandløbskoncentrationerne. I Kapitel 4 sammenholdes visuelt landsdækkende beregninger med MetalStat efter kalibrering til grundvandskoncentrationer med landsdækkende beregninger fra grundvandsmodellen for at undersøge, om der er noget umiddelbart aftryk af grundvandskoncentrationer i vandløbsmålingerne.

#### 3.1 Korrelationsanalyse mellem grundvandsmodel og simuleringer i MetalStat

Der er udviklet en model for grundvandskoncentrationer, som udvider X-matricen i MetalStat med en grundvandskomponent for hvert metal. Denne model med tilhørende dataanalyse er detaljeret beskrevet i bilag til denne rapport. På Figur 3.1 ses en opsummering af de målte grundvandskoncentrationer, der indgår i modellen for grundvandsbidrag, fordelt på styrende redox og pH, som beskrevet i bilaget. I Figur 3.2 angiver den lodrette streg for hvert

metal den beregnede globale median (hele Danmark, alle år og sæsoner) af opløst vandløbskoncentration jf. MetalStat. Sammenlignes grundvandskoncentrationerne med median for vandløbskoncentrationerne, ses det, at grundvandskoncentrationerne ofte er høje sammenlignet med vandløbskoncentrationerne, især for cadmium og zink under iltede forhold med lav pH. Derfor synes det rimeligt at se grundvandet som værende en potentiel kilde til metallerne, og at grundvand ikke blot er en fortyndende faktor, der reducerer koncentrationsniveauet i vandløbet.



**Figur 3.3.** Opsummering af data grundlag for grundvandsmodellen (Bilag, Figur B7) med angivelse af global median værdi (lodret streg) for vandløbskoncentration fra MetalStat, hvor metallerne oppe fra og ned fra højre er cadmium, kobber, nikkel, bly, og zink

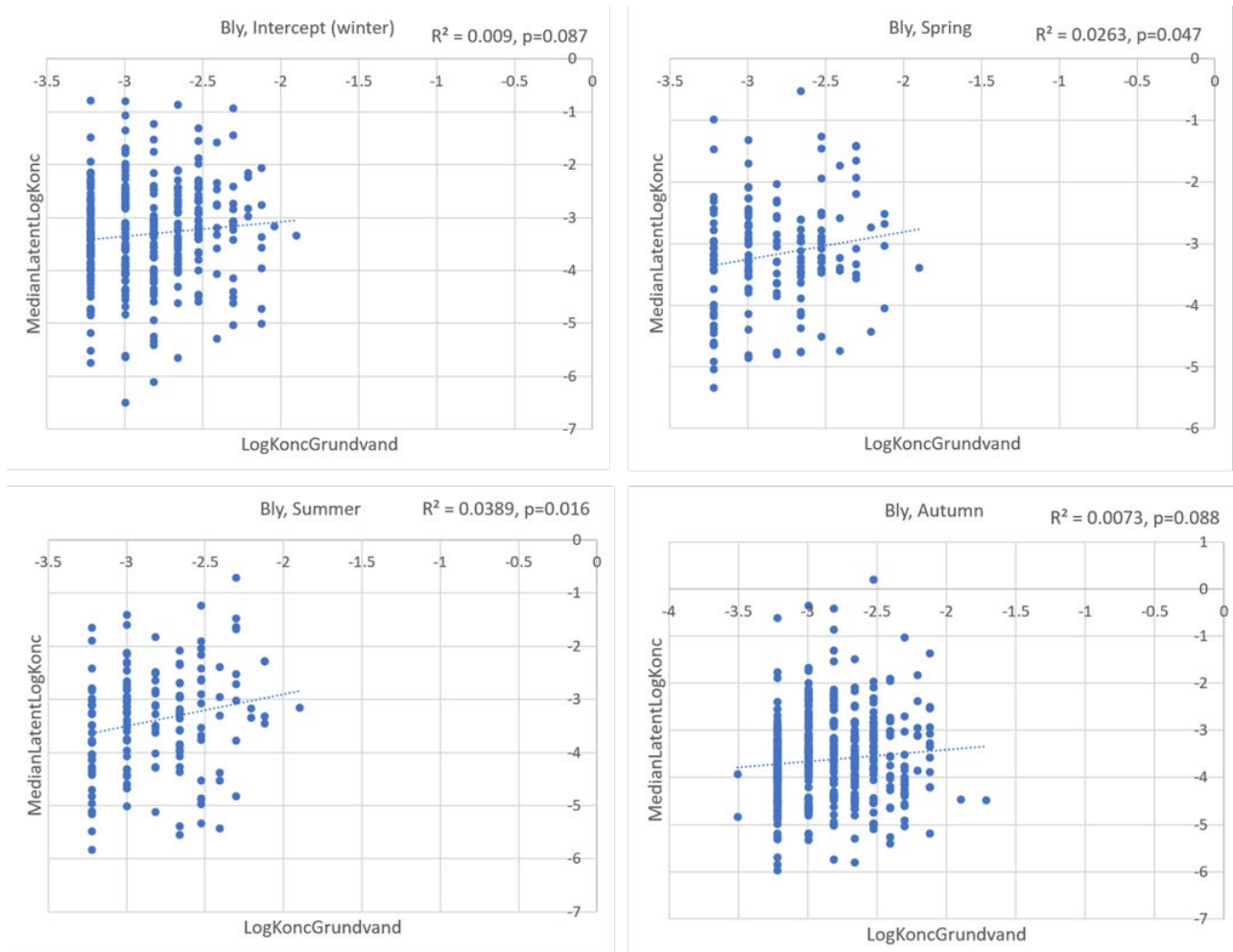
### 3.2 Screening for sammenhæng mellem grundvandskoncentration og vandløbskoncentration

Da over halvdelen af danske vandløbs vandføring typisk kommer fra grundvand, og da koncentrationsniveauet for især bly og cadmium typisk er størst i grundvandet sammenlignet med vandløb (se Figur 3.4), virker det oplagt, at grundvandets større eller mindre indhold af metaller bør kunne have en afsmittende effekt på metallernes vandløbskoncentration. For senere at kunne evaluere de prognosticerede koncentrationsniveauer fra MetalStat, udføres derfor i dette afsnit en sammenligning mellem målte vandløbskoncentrationer på den ene side, og de grundvandskoncentrationer, grundvandsmodellen forudsiger for de målte oplande på den anden.

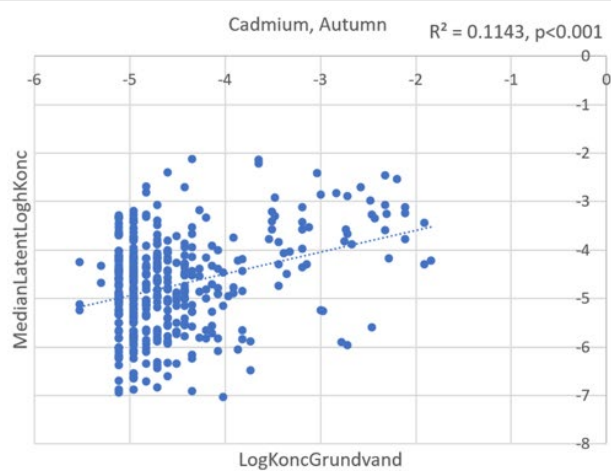
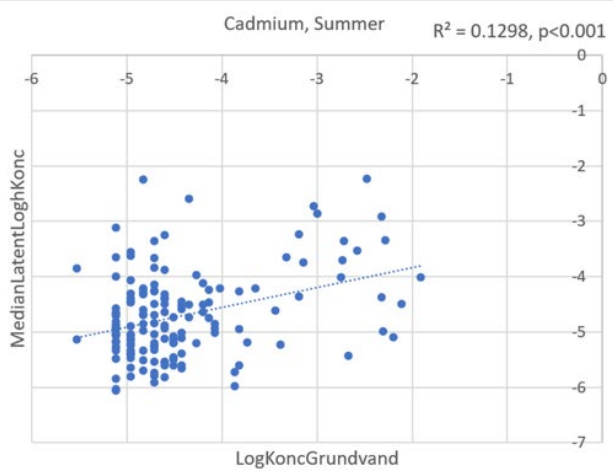
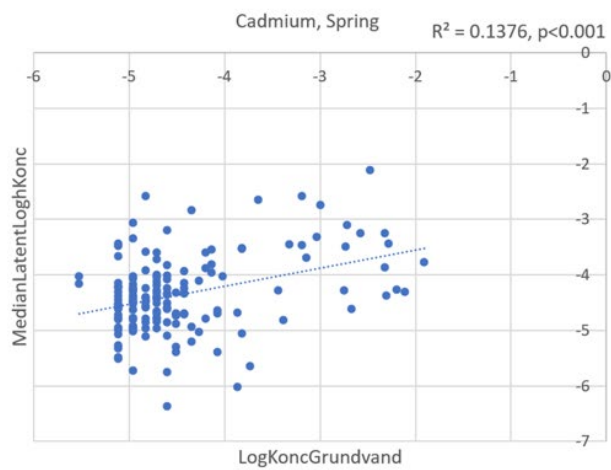
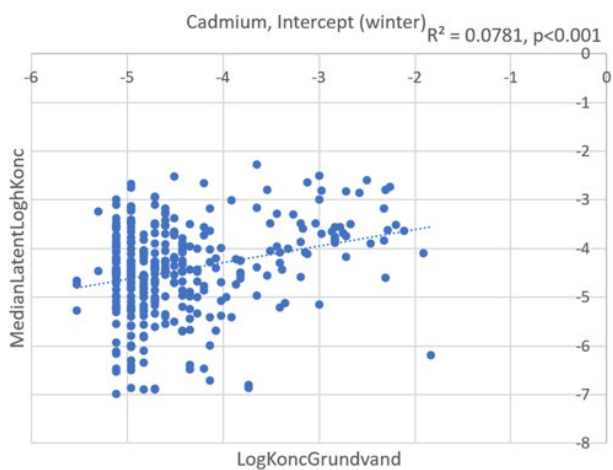
Et ID15-opland får typisk vand opstrøms fra andre ID15-oplande og er derfor ikke alene påvirket af grundvandsmagasinet under oplandet. Udstrækningen af de styrende pH og redox forhold er dog så stor (Bilag, Figurer B4-B6), at det ofte dækker et område med en udstrækning langt ud over de enkelte ID15-oplande. Derfor vil det grundvand, som løber til et ID15-opland fra de andre opstrøms ID15-oplande, også typisk komme fra de samme pH og redox forhold, som befinder sig lige under oplandet. En simpel analyse vil således være at sammenholde de målte ID15-oplande med de tilsvarende beregnede grundvandskoncentrationer for det grundvand, som befinder sig under det pågældende opland.

Andelen af grundvand i vandløbene er typisk ikke konstant over året. Om foråret og sommeren vil der ofte løbe en større andel af grundvand sammenlignet med efterår og vinter. Det skyldes dels, at den overfladenære afstrømning til vandløbene mindskes pga. den øgede fordampning fra rodzonen i den varmere del af året, og dels at grundvandsbidraget primært opbygges gennem efterår, vinter og tidligt forår og således toppe om foråret og først på sommeren. Det betyder, at den bedste korrelation mellem vandløbskoncentration og grundvandskoncentration forventes at ses om sommeren. Den latente koncentration i MetalStat er estimeret med brug af vandløbsmålinger og er derfor et godt koncentrationsestimat ud fra de målte koncentrationer, og da MetalStat estimerer en latent koncentration (logaritmisk værdi) for hver sæson, bliver hver sæson derfor analyseret separat, som om der er bedst korrelation mellem grundvandskoncentration og vandløbskoncentration i forår- og sommersæsonen.

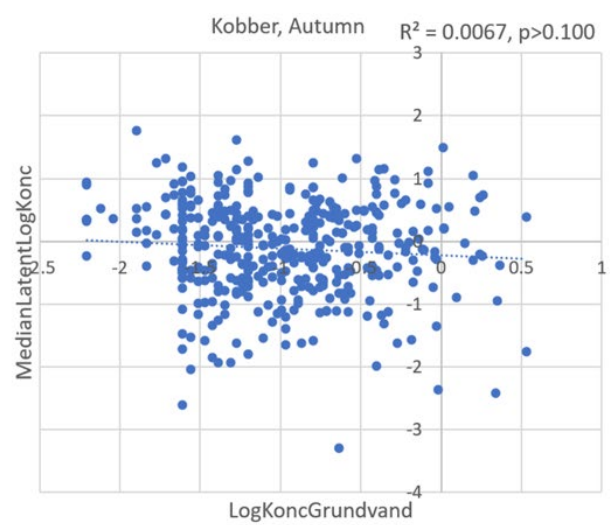
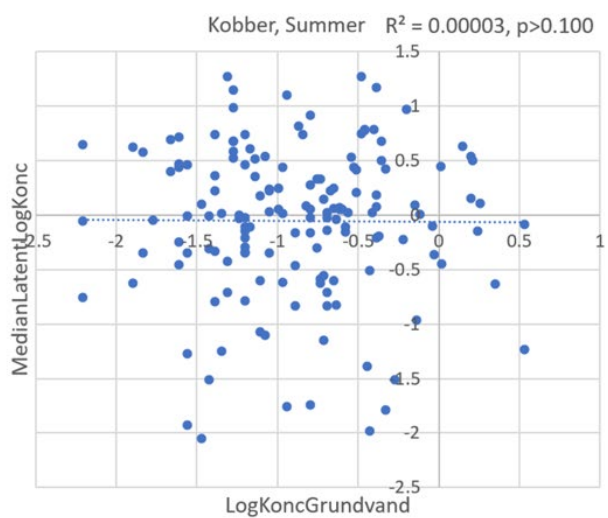
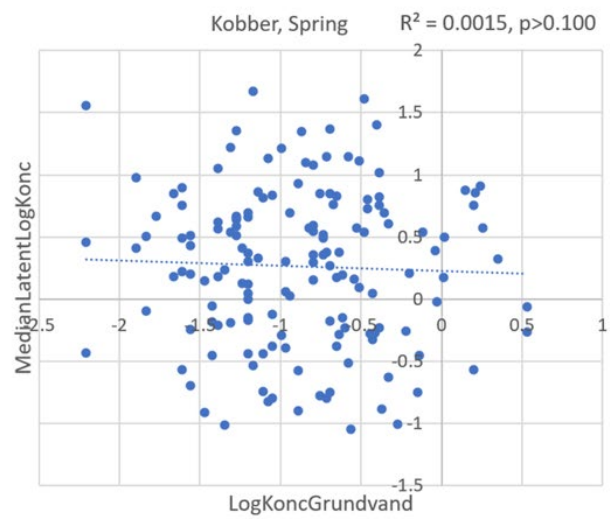
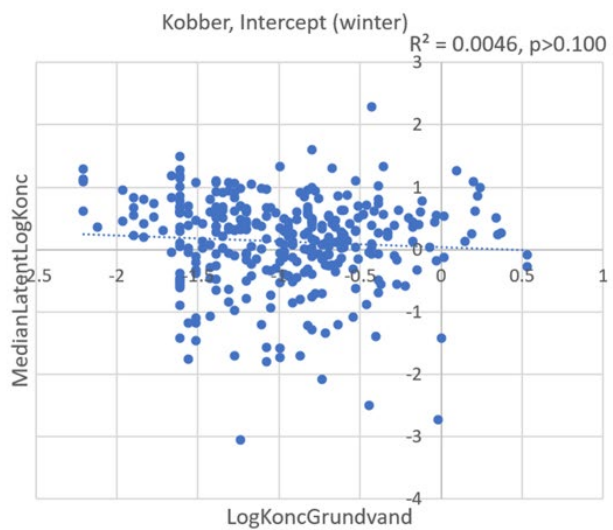
Fra Figur 3.5 til Figur 3.6 ses punktdiagrammer, hvor logaritmen til den målte grundvandskoncentration vises på x-aksen, og den latente koncentration (log værdi) fra MetalStat vises på y-aksen. Grundvandskoncentrationen svarer til median af målte koncentrationer for de grundvandsforekomster, der ligger under det pågældende ID15-opland, for hvilket den latente vandløbskoncentration er beregnet med MetalStat. De viste punktdiagrammer kan altså tolkes som grundvandsforekomsters koncentrationer på x-aksen mod vandløbskoncentrationer på y-aksen for de vandløb, der på overfladen løber hen over de pågældende grundvandsforekomster. På hvert punktdiagram er angivet den log lineære  $R^2$  værdi samt p-værdien for dennes signifikans ( $H_0: R^2=0$ ).



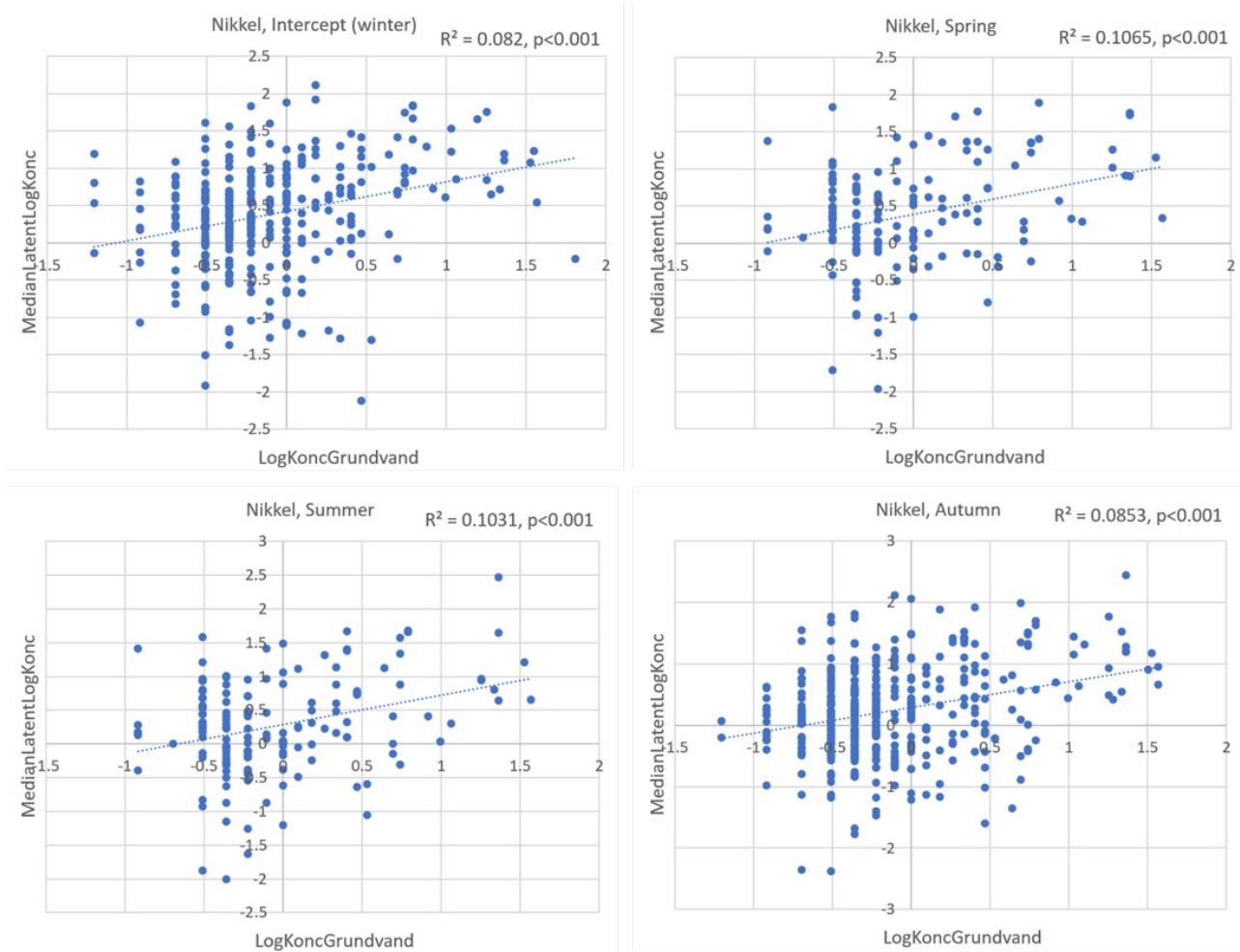
**Figur 3.7.** Punktdiagram af relationen mellem logaritmen til koncentrationsniveauerne af bly i hhv. grundvand (x akse) og latent variabel for vandløbsvand (logaritmisk værdi) i ID15-området (y-akse) over grundvandsmagasinet.



**Figur 3.8.** Punktdiagram af relationen mellem logaritmen til koncentrationsniveauerne af cadmium i hhv. grundvand og latent variabel for vandløbsvand (logaritmisk værdi) i ID15-området over grundvandsmagasinet.

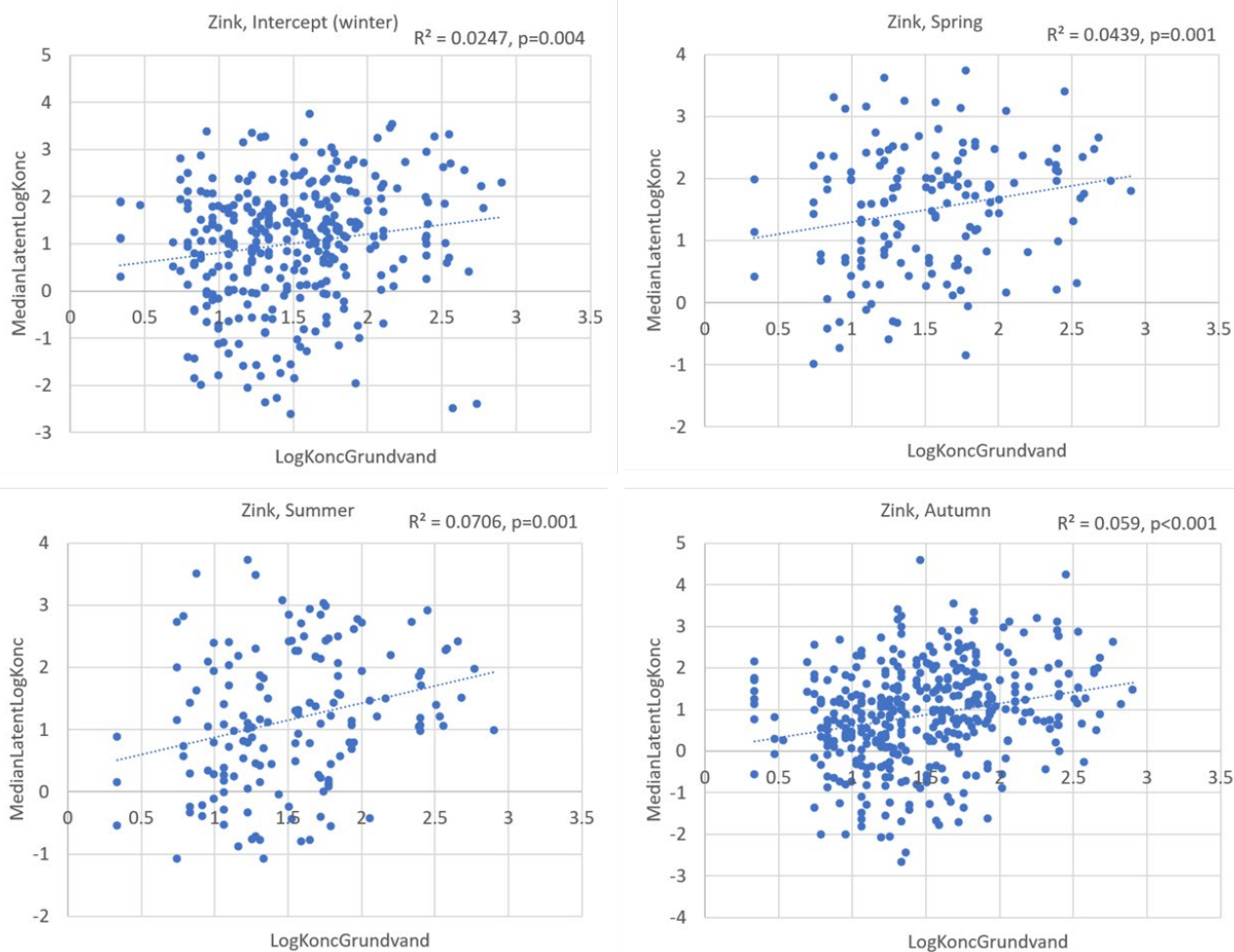


**Figur 3.9.** Punktdiagram af relationen mellem logaritmen til koncentrationsniveauerne af kobber i hhv. grundvand og latent variabel for vandløbsvand (logaritmisk værdi) i ID15-området over grundvandsmagasinet.



**Figur 3.10.** Punktdiagram af relationen mellem logaritmen til koncentrationsniveauerne af nikkel i hhv. grundvand og latent variabel for vandløbsvand (logaritmisk værdi) i ID15-området over grundvandsmagasinet.





**Figur 3.11.** Punktdiagram af relationen mellem logaritmen til koncentrationsniveauerne af zink i hhv. grundvand og latent variabel for vandløbsvand i ID15-oplandet over grundvandsmagasinet.

Der ses generelt en svag, men dog positiv, korrelation for metallerne bly, cadmium, nikkel og zink, mens kobber ikke viser nogen korrelation. Formodningen om, at korrelationen er størst om sommeren, synes at holde stik for de metaller, der udviser mest tydelig korrelation, hvilket antyder, at der faktisk er et aftryk af grundvandsbidrag i vandløbskoncentrationerne. Den svage korrelation overrasker set i lyset af det relativt høje koncentrationsniveau i grundvandet og kunne tyde på, at der sker vigtige udfældningsmekanismer i overgangen mellem grundvand og vandløb og indikerer, at der ikke vil være stor gevinst ved at indføre grundvandkoncentrationer i MetalStat. Dette er dog kun en indikation, da MetalStat tager langt mere detaljerede hensyn til forholdene i oplandene sammenlignet med denne simple analyse. Analysen peger på, at cadmium og nikkel vil være de metaller, hvor modellen styrkes mest ved indførslen af et grundvandsbidrag.

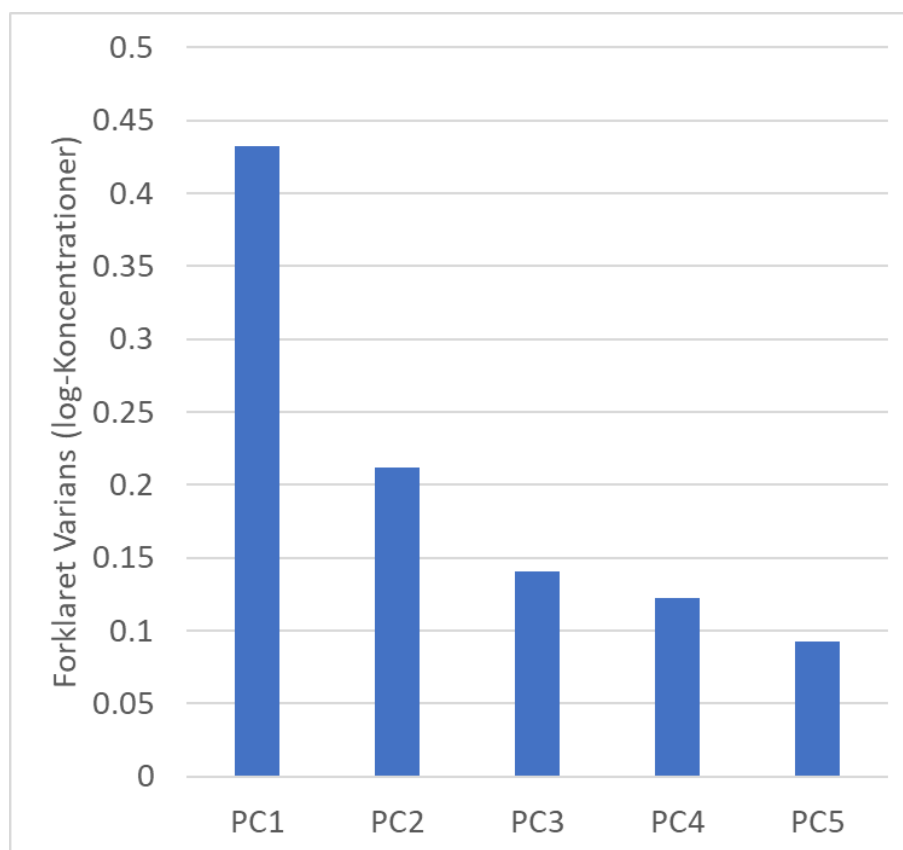
### 3.2.1 Sammenlignende Principal Komponent Analyse for hhv. vandløb og grundvand

Som en indledende analyse af de mulige sammenhænge mellem koncentrationsniveauer i hhv. vandløb og grundvand udføres neden for en sammenlignende Principal Komponent Analyse (PCA) for hhv. vandløbskoncentrationer og grundvandskoncentrationer, der er baseret på korrelationsmatricen. Målinger forudsættes at være log-normal fordelte i analysen.

### PCA af vandløbskoncentrationer

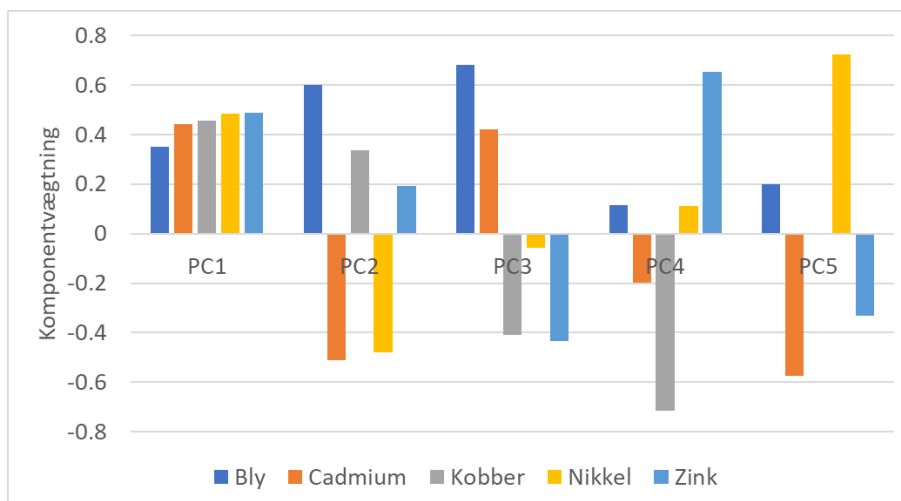
Der udføres en PCA for de vandløbsprøver, der opfylder, at alle 5 metaller er målt til at være over detektionsgrænsen. Det giver et reduceret data set på i alt 942 prøver. Variansen beskrevet ved hver principalkomponent (PC1-PC5) er vist i Figur 3.12.

**Figur 3.13.** Forklaret varians (log koncentrationer) for hver af principalkomponenterne, (PC1-PC5), hvor der kun indgår prøver, hvori alle metaller er målt, og disse er målt til at være over detektionsgrænsen (svarende til 942 prøver).



I Figur 3.14 ses alle fem principalkomponenter fra PCA i rækkefølge mod højre efter faldende forklaret varians. Der vil i alt være fem principalkomponenter, svarende til at der er analyseret fem metaller, hvor andelen af forklaret varians afgør, hvor vigtig den pågældende principalkomponent er i den efterfølgende analyse af komponentvægtning. Det ses, at første principalkomponent (PC1) kan beskrive knap halvdelen af den samlede varians for de fem metaller, hvilket indikerer, at der er en betydelig generel korrelation mellem metallerne. I Figur 3.15 ses et betydeligt fald i forklaret varians fra PC1 til PC2, hvorefter den forklarede varians falder mere jævnt fra PC2 til PC5, dog med et betydeligt fald fra PC2 til PC5. Variationen i værdien for en principalkomponent er mere eller mindre korreleret til en variation i log-koncentration for et metal. Dette beskrives som en komponentvægtning mellem metal og principalkomponent, hvor vægtning=-1 afslører en 100 % omvendt korrelation, vægtning=0 afslører en fuldkommen mangel på korrelation, og en vægtning på +1 afslører en 100 % positiv korrelation. De enkelte metalleres vægtning til principalkomponenterne ses i Figur 3.16.

**Figur 3.17.** Metallernes komponentvægtning (loading) for hver af principalkomponenterne, svarende til den PCA, der er vist på Figur 3.7.

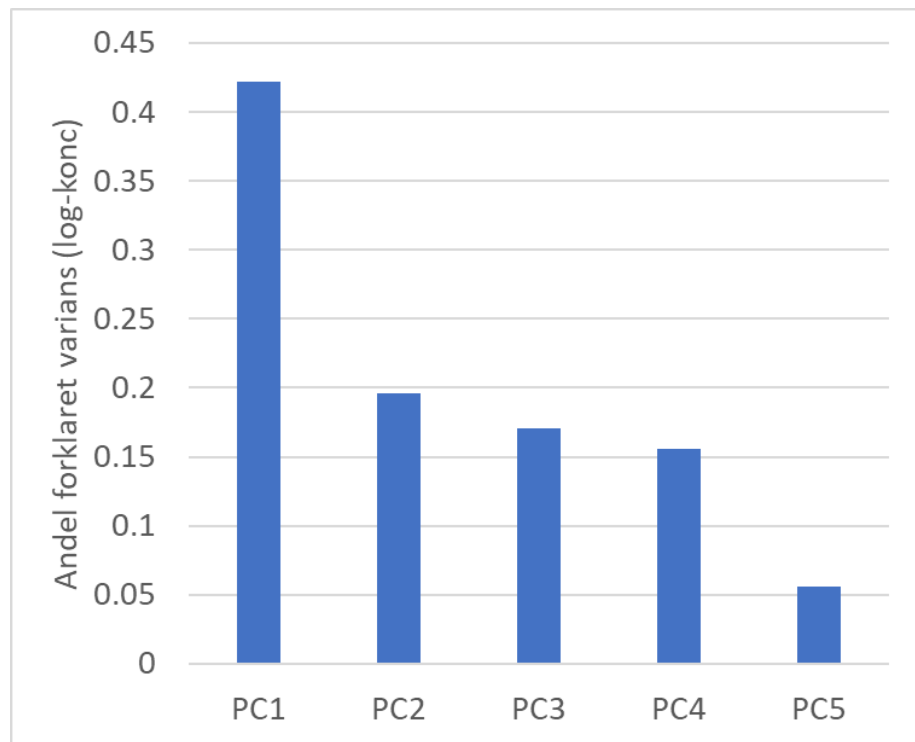


I Figur 3.8 ses, at komponentvægtningen for alle fem metaller er positiv for første komponent (PC1), og derfor er der en generel positiv korrelation mellem de fem metalleres koncentrationsniveauer. Dette sammenfald mellem metallernes koncentrationsniveau udnyttes af den multivariate beskrivelse i MetalStat til at mindske usikkerheden i modellens forudsigelser. Her er det dog vigtigt at påpege, at de prøver, der indgår i denne analyse, har det tilfældes, at alle metaller er målt over detektionsgrænsen, så de er med andre ord plukket ud af den samlede datamængde på en måde, så der er tendens til at skabe en positiv korrelation mellem alle metaller. Den reelle betydning af denne bias er svær at bestemme, og denne PCA skal derfor primært ses som en indledende analyse. Det ses i Figur 3.8, at bly har en lidt mindre vægtning til PC1 sammenlignet med de andre metaller, og det antyder, at bly-kontaminering af vandløbsvandet har en lidt anden generel årsag end de andre fem metaller samlet set. Det passer også med, at bly kommer ud med stor vægtning på både PC2 og PC3, hvor PC2 viser, at der er prøver, hvor især bly, men også delvist kobber, er høje, mens cadmium og nikkel er lave. PC3 viser, at bly for nogle prøver varierer sammen med cadmium, modsat kobber og zink. PC4 er primært en adskillelse af zink fra kobber, mens PC5 adskiller nikkel, og i mindre omfang bly fra lave cadmium og i mindre omfang zink. Denne indledende PCA kunne sammenholdes med resultater fra MetalStat i en større undersøgelse af mulige årsager til metallernes forekomst i vandløb, men det vurderes at ligge ud over denne rapporters rammer at lave en sådan afdækning af årsager.

#### PCA af grundvandskoncentrationer

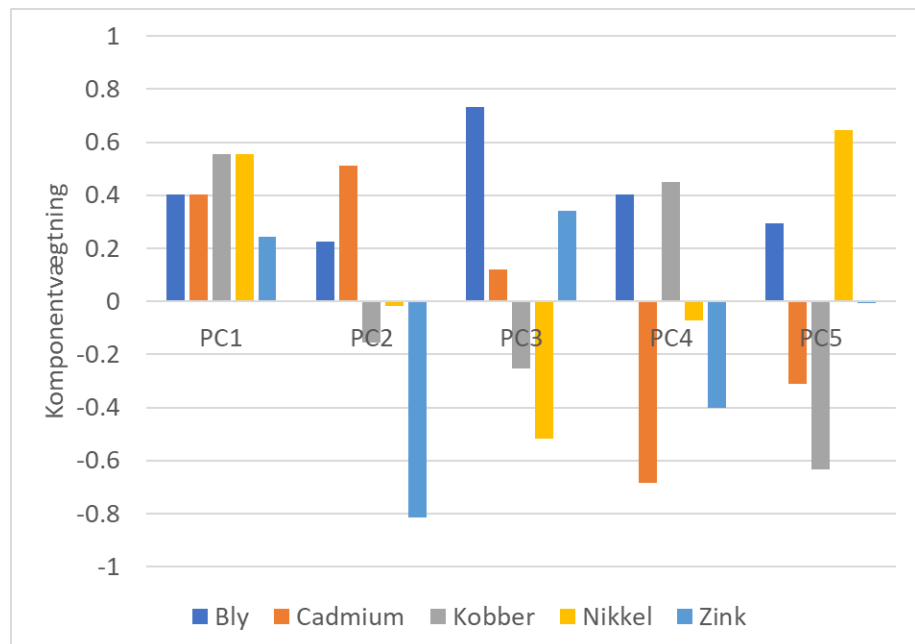
Der er udført en PCA på grundvandsmodellens forudsagte grundvandskoncentrationer (Log værdier), hvor Figur 3.9 viser andel varians forklaret ved hver principalkomponent. Det ses på figuren, at PC1 kan forklare ca. 40 % af den samlede varians med et større fald i forklaret varians fra PC1 til PC2. Komponenterne PC2 til PC4 har næsten samme andel forklaret varians, mens der sker et betydeligt fald til PC5.

**Figur 3.18.** Andel varians beskrevet for hver principalkomponent af de modellerede grundvandskoncentrationer.



Metallernes komponentvægtning af grundvandskoncentrationerne er vist i Figur 3.10. Her ses en positiv komponentvægtning for alle metaller til PCA1, hvor zink dog viser en noget mindre vægtning end de øvrige metaller. PCA2 domineres af en meget kraftig negativ komponentvægtning på zink, der især adskiller zink fra cadmium. PCA3 adskiller især bly og zink fra kobber og nikkel. PCA4 adskiller bly og kobber fra cadmium og zink. PCA5 ses i Figur 3.9 ikke at have stor betydning, med en ringe andel af forklaret varians, hvorfor denne komponents vægtning til metallerne ikke kan tillægges samme betydning som for de øvrige komponenter.

**Figur 3.19.** Komponentvægtning (Loading) for hvert metal af hver principalkomponent af de modellerede grundvandskoncentrationer.



#### Sammenligning af PCA-resultater for hhv. vandløb og grundvand

I dette afsnit sammenlignes Figur 3.7 og Figur 3.8 for vandløbsvand med Figur 3.9 og Figur 3.10 for grundvand. Sammenlignes andelen af forklaret varians, ses den samme tendens for PC1, hvor den første principalkomponent kan

forklare knap halvdelen af den samlede varians (Figur 3.7 og Figur 3.9). Derimod ses en forskellig relation mellem vandløb og grundvand for de øvrige principalkomponenter, for mens vandløb viser en jævnt faldende andel varians fra PC2 til PC5, så viser PC2, 3 og 4 næsten samme andel for grundvand, mens PC5 falder drastisk (Figur 3.9). Denne forskel mellem Figur 3.7 og Figur 3.9 kan skyldes, at mens alle vandløbsmålingerne er fastlagt på reelt målte værdier, så er grundvandskoncentrationerne fastlagt som en kombination af både målte og beregnede værdier ud fra pH og redox forhold. Det, at nogle værdier er beregnede, skaber som en matematisk konsekvens en korrelation mellem metallerne, som kommer ud i principalkomponenterne, før den sidste komponent, som typisk består af den variation, som ikke har nogen korrelation mellem metallerne. Derfor bliver bidraget til den sidste komponent (PC5) for grundvand begrænset af, at det kun er den målte andel af oplandene, som bidrager til PC5.

For både vandløb og grundvand ses en positiv komponentvægtning for alle metaller til den første principalkomponent. Der er dog nogle klare afvigelser mellem hhv. vandløb og grundvand for især bly og zink, hvor bly havde mindre vægtning til PC1 i vandløb end de andre metaller, mens dette var tilfældet for zink i grundvand. Især zink synes at have et andet mønster mellem komponentvægtningerne, når vandløb og grundvand sammenlignes, som konsekvens af, at zink har flyttet noget variation fra PC1 ud på de andre principalkomponenter. De andre metaller kommer ud med komponentvægtninger, der er mere sammenlignelige mellem vandløb og grundvand. Det ses ved, at de andre metaller kommer ud med en komponentvægtning af cirka samme værdi, der dog flytter plads mellem principalkomponenterne, hvor f.eks., den negative komponentvægtning for cadmium flytter plads fra PCA2 for vandløb til PC4 for grundvand. For de fire metaller bly, cadmium, kobber og nikkel er der et vist sammenfald mellem PC2 for vandløb og PC4 for grundvand.

### 3.3 Opdateret X matrice i MetalStat

Den beregnede grundvandskoncentration for hvert metal anvendes i X matricen, så der indføres fem nye variable. Hver af disse koncentrationsniveauer multipliceres med grundvandstilstrømningen for det pågældende opland for at beregne potentiel grundvandstilførsel til oplandet efter samme princip som de øvrige variable i X matricen.

MetalStat er multi-variabel ved at modellere alle metaller med indbyrdes korrelation som en vektor. Derfor vil en analyse af korrelationen mellem de enkelte metaller grundvandskoncentration give et præj om, hvordan MetalStat vil være i stand til at udnytte de ekstra variable i X matricen.

Grundvandsmodellens forudsagte grundvandskoncentrationer er for hvert opland multipliceret med den mængde grundvand, som tilføres vandløbet i oplandet, for at få en variabel i X-matricen, der angiver den potentielle tilførsel af metal fra grundvandet jvf. grundvandsmodellen. Den potentielle grundvandstilførsel vil således i MetalStat blive kalibreret separat for hvert metal mod målte vandløbskoncentrationer for at estimere den reelle tilførsel af metal. Den samlede X matrice er opsummeret i Tabel 3.1.

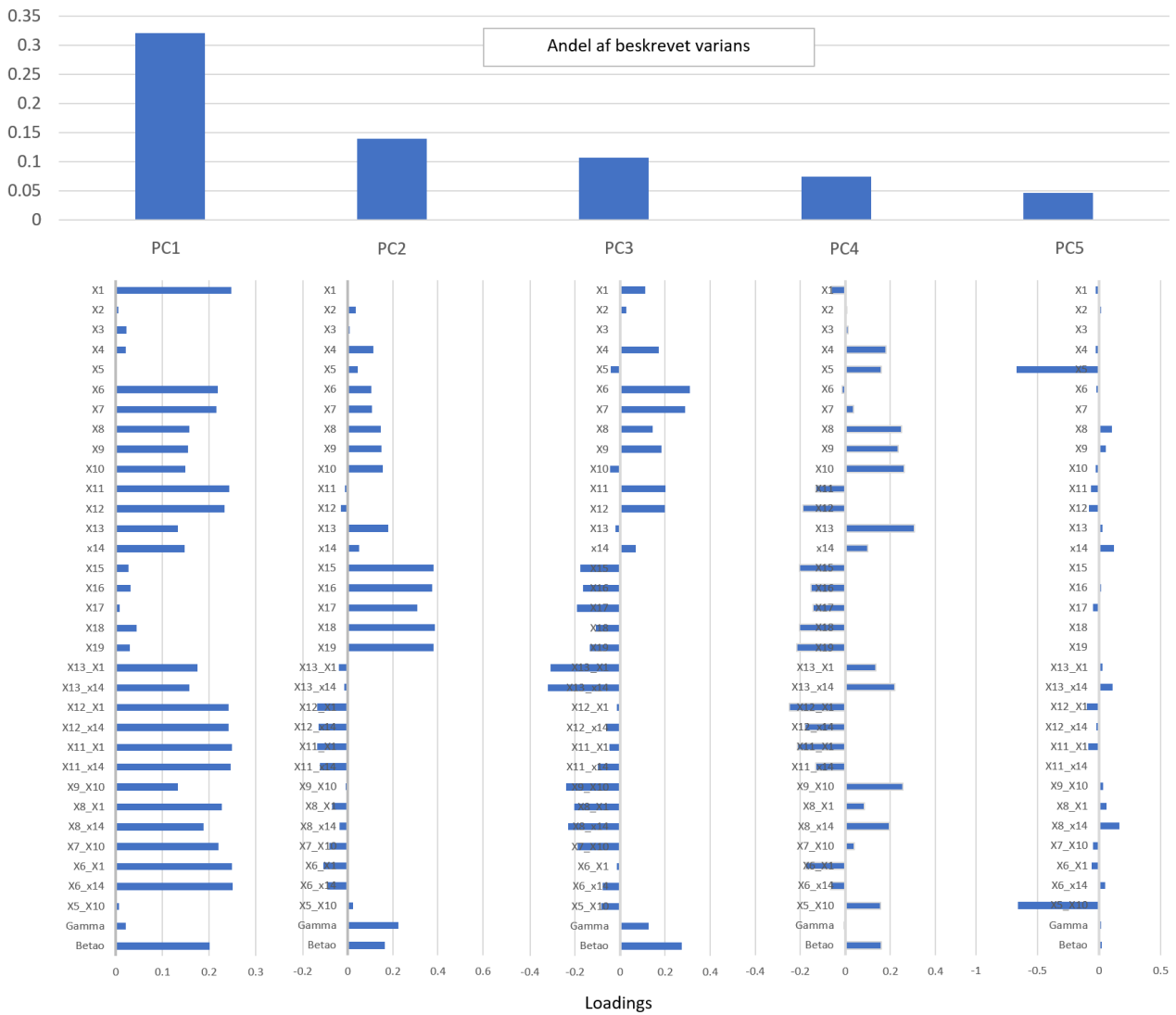
**Tabel 3.1.** Opsummering af X matricen, hvor de nye variable X15-19 er beskrevet i denne rapport, mens de øvrige er beskrevet i Sørensen et al, 2021.

Variabel	Forklaring
Gamma	Retention af metal i vandsystemet
Beta0	Generelt arealbidrag
X1	Areal Landbrug
X2	Regnvandsbetinget udløb
X3	Spildevand fra offentlige rensningsanlæg
X4	H <sup>+</sup> (pH) rodzone
X5	H <sup>+</sup> (pH) 1-2 m dybde
X6	Lerindhold a horisont
X7	Lerindhold b horisont
X8	Kulstofindhold a horisont
X9	Kulstofindhold b horisont
X10	Grundvandsbidrag
X11	Drænet areal
X12	Makropore indeks
X13	Nettonedbør
X14	Mængde svinegylle
X15	Potentiel grundvandstilførsel af nikkel
X16	Potentiel grundvandstilførsel af kobber
X17	Potentiel grundvandstilførsel af cadmium
X18	Potentiel grundvandstilførsel af bly
X19	Potentiel grundvandstilførsel for zink

X matricen er sammensat af mange datakilder, der kan korrelere mere eller mindre. Der vil f.eks. typisk være en positiv korrelation mellem areal med landbrug og areal med højt makropore indeks, mens regnvandsbetingede udløb kunne tænkes ofte at være placeret i samme oplande som dem, der også modtager spildevand. Det er også muligt, at grundvandskoncentrationer kan korrelere med andre variable i X matricen, fordi grundvandet rent faktisk er påvirket af overfladenære kilder til metaller og således ikke alene kan tolkes som geogene "naturlige" kilder til metallerne i vores vandløb. Den slags korrelation kan bevirke, at MetalStat let kommer til at tilegne en variabel i X matricen en betydning, som helt eller delvist er et artefakt, da den reelle betydning skyldes en anden korrelerende variabel. Det er derfor nyttigt at kende den generelle korrelations-struktur i X matricen for at vurdere, om en effekt af f.eks. grundvandsbidrag kan beskrives med andre variable end dem, der specifikt er beregnet med grundvandsmodellen. En sådan korrelations-struktur er vist for de første fem principalkomponenter på Figur 3.11.

Figur 3.11 viser resultatet for de fem mest betydende principalkomponenter fra en PCA, der har analyseret værdisætningen af attributterne vist i Tabel 3.1, der er brugt i MetalStat til kalibrering mod målte koncentrationer. Denne PCA er baseret på korrelationsmatricen. Øverst og vandret er angivet den andel af den samlede variation, der er beskrevet ved de fem første principalkomponenter. Under hver af disse søjler er et lodret søjlediagram, der viser komponentvægtning til den pågældende principalkomponent for hver variabel i X matricen, inklusivt de relevante interaktioner. Samlet set beskriver de første fem principalkomponenter 69 % af den samlede varians.

Generelt ses det, at mange variable har en betydelig komponentvægtning til de første fire principalkomponenter. Da alle  $X$  matricer variable, på nær gamma, er skaleret med vandvolumenet i hver ID15-opland, som beskrevet i Sørensen m.fl., (2021), vil der være en stor fælles positiv korrelation, der kommer ud som første komponent. Den femte principalkomponent er derimod styret af komponentvægtning for en enkelt variabel ( $X_5$ ) og dennes interaktion med  $X_{10}$ . Det ses således, at de fire mest betydende principalkomponenter har vægtning til mange variable i  $X$  matricen, mens principalkomponenter med lavere betydning (fra PCA5) typisk vil vægtes fra et mindre antal variable fra  $X$  matricen. Det ses, at den potentielle grundvandstilførsel for alle fem metaller kommer ud samlet med komponentvægtning til anden principalkomponent. Desuden kommer grundvandstilstrømningen ( $X_{10}$ ) ud med positiv komponentvægtning sammen med det potentielle grundvandsbidrag, hvilket ikke overrasker, da grundvandsbidraget er bestemt som produktet mellem potentiel grundvandstilførsel og de modellerede grundvandskoncentrationer. Desuden understøtter den fælles positive komponentvægtning til første principalkomponent for grundvandskoncentrationer (Figur 3.10), at der også vil være et tilsvarende sammenfald i komponentvægtning for metallerne for deres variable i  $X$  matricen. Da en række variable i  $X$  matricen også har en betydelig komponentvægt på PC2, indikerer det, at MetalStat i pilotversionen delvist kan tage højde for grundvandstilførslen, selvom dette ikke specifikt er indført som variabel for hvert metal som et bidrag fra grundvandstilførslen generelt med ens betydning over hele landet.



**Figur 3.20.** Kortlægning af datastruktur af X matricen med brug af PCA. Øverst og vandret ses de første 5 principalkomponenters andel af beskrevet varians, hvor 69 % af den samlede variation er beskrevet. Lodret er de enkelte af X matricens elementers komponentvægtning til hver principalkomponent vist som søjlediagram. Interaktion mellem attributter vises som f.eks. "X13\_X1", der svarer til interaktionen mellem X1 og X13.



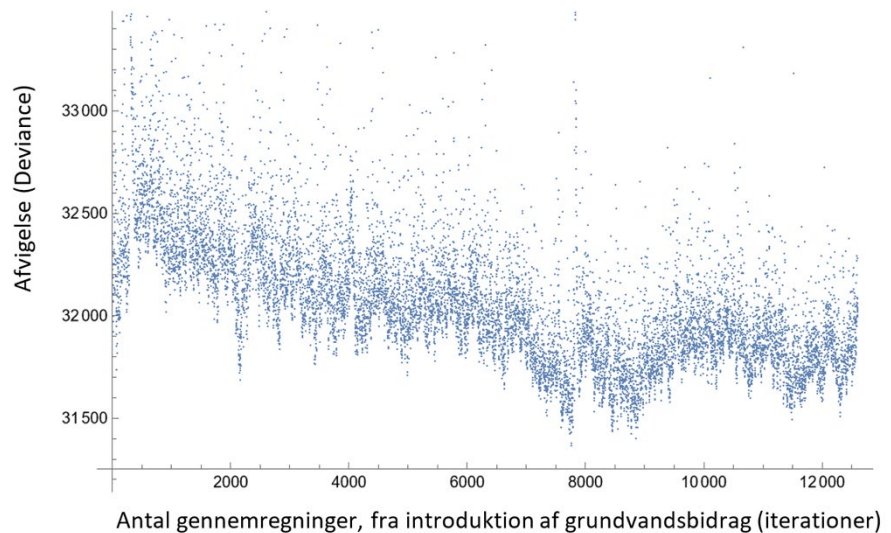
## 4 Resultater fra MetalStat

I løbet af 2023 blev det klart, at en egentlig programmering af driftsversionen af MetalStat ikke kunne færdiggøres i 2023. Derfor er udviklingsversionen udviklet til en forløbellig driftsversion med mindsket beregningstid og opdateret med nye målinger og målestationer til og med 2021 samt opdateret med resultater fra en forbedret grundvandsbeskrivelse. Resultaterne i denne rapport beskriver udtræk fra MetalStat den 14/12-2023.

### 4.1 Generelt model output

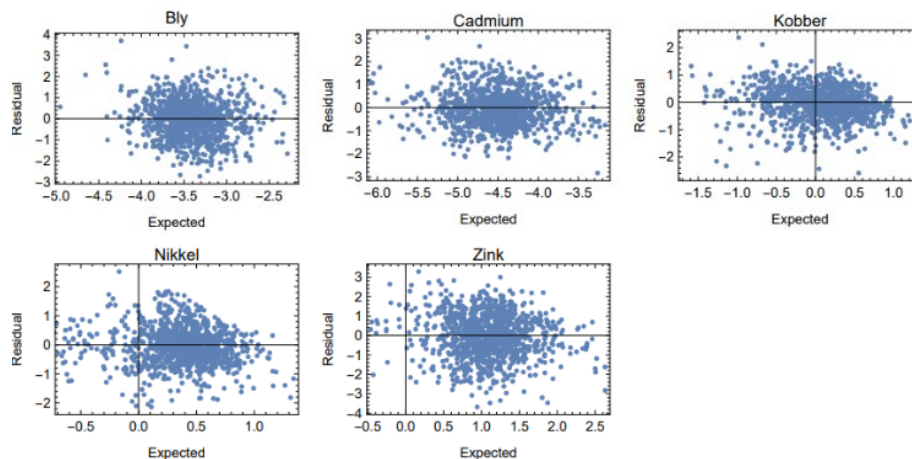
I den foreløbige driftsversion af MetalStat er modelparametre estimeret ved en iterativ Metropolis-Hasting algoritme. En nærmere beskrivelse af metoden kan ses i Sørensen m.fl. (2022) samt Sørensen m.fl. (2023). Ved hver iteration beregnes en afvigelse til målingerne (Deviance), og på Figur 2.1 ses afvigelsen for hver iteration med start på det tidspunkt, hvor den udvidede X matrice med grundvandsbidrag blev introduceret. Der er foretaget ca. 13.000 iterationer siden introduktion af grundvandsbidrag, og det ses af Figur 4.1, at opstarten med den nye X matrice øgede afvigelsen lidt, men i løbet af de følgende ca. 6000 iterationer rettede modellen sig ind, og der ses et svagt faldt mod en mindre afvigelse. De sidste 2000 iterationer er brugt til afrapportering af resultater i denne rapport (udtræk fra MetalStat den 14/12-2023). Indførslen af grundvandskoncentrationer har ikke givet markante forbedringer af modellens afvigelse, hvilket enten kan skyldes, at der ikke er nogen større korrelation til grundvandskoncentrationen, eller at MetalStat i den forrige version har kalibreret grundvandsbidraget til en af de andre attributter i X matricen, og at der således er en redundant beskrivelse af grundvandstilførslen. En mulig attribut med relevans for grundvandsbidraget i den forrige version af MetalStat kan f.eks. være grundvandsindstrømning (X10 i Tabel 3.1), som har en betydelig komponentvægtning positivt til PCA2, hvilket er sammenfaldende med grundvandsbidragets fem nye attributter jf. Figur 3.9, og her kunne indløb til oplandet af grundvand (X10) være en oplagt mulighed, som nævnt i forrige afsnit. En redundant beskrivelse af grundvandstilførslen er ikke umiddelbart noget problem i bestræbelse på at forudsige koncentrationsniveauer, og det er først nu efter grundvandsmodellen er inddraget i MetalStat, at det er muligt at se, om der er redundante variable i forhold til, at modellen rent faktisk kan beskrive grundvandsbidraget. I det løbende udviklingsarbejde af MetalStat bør X matricen undersøges og justeres for, om antallet af variable kan reduceres. En sådan justering er mulig med udgangspunkt i, at MetalStat kan levere en vurdering af relevansen for værd variable i X matricen som deskriptor for den opløste vandløbskoncentration. Formålet med denne rapport er at levere forudsigelser af koncentrationsniveauer for ikke målte vandområder, hvilket er muligt med den anvendte X matrice, mens en optimering og tolkning af de enkelte variable i X matricen regnes for at være en kommende aktivitet.

**Figur 4.1.** Udviklingen i afvigelsen for MetalStat gennem rækken af iterationer fra det tidspunkt, hvor grundvandsbeskrivelsen blev indført i X matricen.



For de ID15-oplande, hvor der er indsamlet prøver, kan de latente koncentrationer (log værdi) estimeres ud fra målte koncentrationer, mens X matricen kan bruges til at forudsige koncentrationerne i hvert ID15-opland (log værdi). Forskellen mellem den estimerede latente koncentration (log værdi) og modellens forudsagte koncentration (log værdi) kan derfor beregnes for de ID15-oplande, hvorfra der er indsamlet prøver, og denne forskel definerer et residual. Dette residual er et mål for usikkerheden ved at forudsige koncentrationerne med brug af X matricen og vises som residualplot i Figur 4.2, hvor residualerne er afbilledet som funktion af forudsagt værdi. Et sådan residualplot er nyttigt til at undersøge, om der hersker en systematisk variation i modellens usikkerheder. Det, at residualerne er logaritmiske værdier, betyder, at de skal tolkes som relative forskelle mellem latente koncentrationer og forudsagte koncentrationer, så hvis f.eks. et residual har værdien "1", betyder det, at den latente koncentration er  $\text{Exp}(1)=2.7$  gange højere end den forudsagte koncentration. De anførte residualer er altså udtryk for en relativ usikkerhed. Generelt ses en tilfældig fordeling af residualerne i Figur 4.2, men dog antydes en svag systematik for nikkel, hvor residualerne for de høje forudsagte værdier (i højre side af plottet) synes mindre end for de lavere forudsagte værdier, hvilket indikerer, at X matricen har større relativ usikkerhed ved estimat af lavere nikkel koncentrationer sammenlignet med de højere niveauer. En sådan asymmetri er ikke særlig kritisk set i forhold til, at det er de høje koncentrationer, som er vigtigst at forudsige med modellen, og at disse jf. Figur 4.2 har mindre usikkerhed sammenlignet med de lavere koncentrationer.

**Figur 4.2.** Residualplot for MetalStat.



Den samlede andel af forklaret varians for alle oplande ( $R^2$ ) indikerer, hvor specifik X matricen er til at adskille oplande med forskellige koncentrationsniveauer og er angivet for hvert metal i Tabel 4.1. De latente koncentrationer (log værdi) fra målemodellen anvendes til at beregne  $R^2$  ved at sammenligne de latente koncentrationer med de forudsagte koncentrationer ved brug af X matricen, og  $R^2$  afspejler derfor, hvor godt X matricen kan forudsige de latente koncentrationer for de oplande, der er målte. En  $R^2$  værdi på én svarer til, at X matricen leverer et matematisk identisk estimat af de latente koncentrationer, hvilket reelt er umuligt, men dog ideelt set ønskeligt. Jo tættere  $R^2$  kommer på værdien én, jo bedre kan X matricen forklare de forskellige koncentrationsniveauer for forskellige ID15-oplande. Da den latente koncentration er udtrykt som logaritmen til den virkelige koncentration, så beskriver  $R^2$  en relativ forskel mellem målt koncentration og beregnet koncentration. Det betyder, at mindre forskelle mellem lave koncentrationsniveauer bidrager til  $R^2$  med samme vægt som større forskelle mellem højere koncentrationsniveauer, hvis den relative (procentvise) forskel er den samme for de to niveauer. Derfor kan mange små og ubetydelige afvigelser mellem lave koncentrationsniveauer godt skjule en models evne til at finde de høje og betydelige koncentrationsniveauer i en relativt beskeden  $R^2$  værdi. Hvis et metal f.eks. findes tilfældigt fordelt ved lave koncentrationer i de fleste målte oplande, men hvor enkelte oplande har høje værdier, så kan en model med ret lav  $R^2$  alligevel godt være i stand til at identificere de få oplande med forhøjet koncentration i det omfang, X matricen kan beskrive den pågældende og sjældne variation. Med disse forbehold er  $R^2$  især et indbyrdes mål for, hvor godt metallerne beskrives i forhold til hinanden. En indbyrdes sammenligning af  $R^2$  mellem metallerne i Tabel 4.1 viser, at bly har en noget lavere  $R^2$  værdi sammenlignet med de øvrige metaller. Det kan enten skyldes, at der mangler en beskrivende variabel i X matricen, eller at bly simpelthen er relativt tilfældigt fordelt over landet uden større og regelbaserede forskelle. En forklaring kunne være, at bly primært hidrører fra en historisk kontaminering i f.eks. blyholdig benzin, der har andre årsager end beskrevet med X matricen, som f.eks. befolkningstæthed, hvilket ikke er med i den nuværende version af X matricen.

**Tabel 4.1.** Andel af varians forklaret for de latente log-koncentrationer

Stof	Bly	Cadmium	Kobber	Nikkel	Zink
$R^2$ (IC)	0.13	0.31	0.37	0.23	0.35

Korrelationen mellem de latente koncentrationer (log-værdi) er vist i en korrelationsmatrice i Tabel 4.2. Der ses en generel positiv korrelation mellem alle metaller, hvilket stemmer overens med, at alle vandløbsmålinger har positive komponentvægtninger til den først principalkomponent (Figur 3.8).

**Tabel 4.2.** korrelationsmatrice mellem de latente log-koncentrationer

	Bly	Cadmium	Kobber	Nikkel	Zink
Bly	1	0.30	0.44	0.15	0.30
Cadmium	0.30	1	0.31	0.24	0.30
Kobber	0.44	0.31	1	0.20	0.40
Nikkel	0.15	0.24	0.20	1	0.140
Zink	0.30	0.30	0.40	0.140	1

Korrelationsmatricen mellem de prædikterede log-koncentrationer ved brug af X matricen er vist i Tabel 4.3, og her ses en forstærket positiv korrelation mellem metallerne sammenlignet med Tabel 4.2.

**Tabel 4.3.** korrelationsmatrice mellem de forudsagte log-koncentrationer

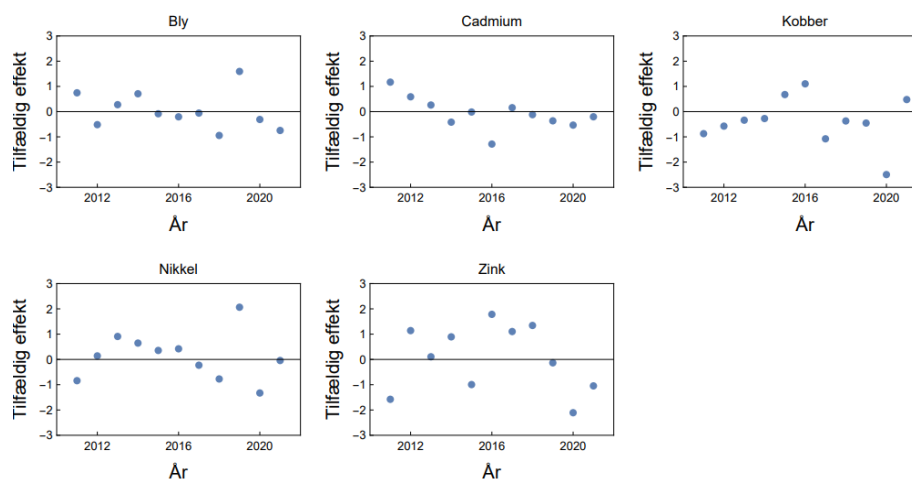
	Bly	Cadmium	Kobber	Nikkel	Zink
Bly	1	0.52	0.54	0.50	0.46
Cadmium	0.52	1	0.66	0.78	0.59
Kobber	0.54	0.66	1	0.70	0.62
Nikkel	0.50	0.78	0.70	1	0.62
Zink	0.46	0.59	0.62	0.62	1

År er med som tilfældig effekt i målemodellen i MetalStat, hvor der estimeres en effekt for hvert år, der er målt. Det betyder, at MetalStat giver et landsdækkende tidsforløb for de målinger, som er brugt i modellen, og Figur 4.3 viser disse estimater for hvert metal. Forskellige miljøfaktorer, f.eks. nedbør, kombineret med ændret emissionsmønster kan have betydning for effekten af år, men forskelle kan også afspejle en ændret strategi for dataindsamlingen. Forskelle i indsamlingsstrategi for prøveindsamling mellem år kan desuden medføre en forskel mellem år, der hverken skyldes ændringer i miljøfaktorer eller emission. Det varierende antal prøver fra år til år kan også påvirke udsvingene mellem år, da færre prøver øger risikoen for, at større tilfældige udsving kan påvirke det gennemsnitlige niveau det pågældende år. Dette kan f.eks. være årsagen til, at bly i Figur 4.3 udviser det største udsving i år 2019, som samtidig er et af de år med færrest indsamlede prøver (Figur 2.1).

MetalStat er robust over for ubeskrevne forskelle fra år til år, da sådanne forskelle er modelleret ved tilfældige effekter, der kan tolkes som "Ukendte forskelle mellem år". Hvis der er større systematisk variation (faldende eller stigende) i tilfældig effekt af år, bør modellen dog granskes for at vurdere, om den skal justeres pga. ændrede forhold. Hvis X matricens sammensætning af variable justeres for bedre at kunne tage højde for tidsmæssige ændringer, vil det reducere de tilfældige effekter af år, disse forskelle vil nu i stedet være forklaret med X matricen, hvilket i øvrigt vil gøre modellen stærkere i dens forudsigelser af forskelle mellem ID15-oplande. Effekten mellem år for zink og kobber ses at være stærk negativ i 2020, og mens kobber retter sig tilbage, beholder zink en negativ effekt i 2021, og en opdatering med nye måledata for 2022 og 2023 vil kunne afgøre, om MetalStat skal opdateres ved f.eks. at op-

datere med en  $X$  matrice, der afhænger af tid. Zink er blevet reguleret i landbruget i de seneste år, så her synes det relevant med en opdateret  $X$  matrice, der kan beskrive skiftet i brugen af zink i landbruget.

**Figur 4.3.** Tilfældige effekter på år, målt relativt mellem år.



## 4.2 Landsdækkende koncentrationsestimater

Koncentrationsniveauet for hvert metal er forudsagt for hvert ID15-opland for et ukendt år og som en sandsynlighedsfordeling. Variansen af denne sandsynlighedsfordeling kan betragtes som usikkerheden på koncentrationsniveauet i det specifikke opland. Alternativt kunne forudsigelserne dække et bestemt historisk år, f.eks. år 2016, og i det tilfælde vil året være kendt, så her er effekter af året estimeret specifikt for år 2016, og variansen af forudsigelsen vil være mindre.

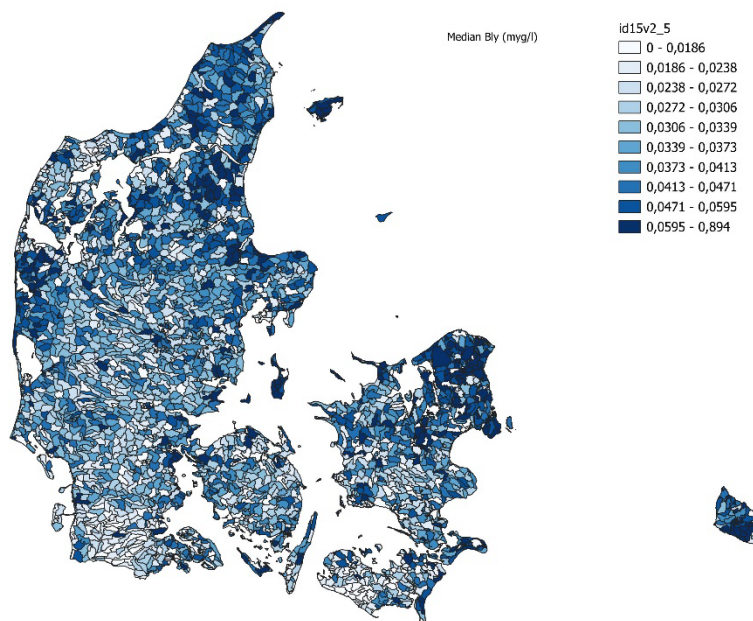
De målte koncentrationsniveauer er anvendt til at estimere  $X$  matrixens parametre samt de øvrige effekter (sæson, opland, år), men MetalStat forudsiger koncentrationsniveauer efterfølgende for alle oplande uden at tage hensyn til, om der er målt i nogle af oplandene. De oplande, der indeholder målestationer, har altså dels en estimeret latent koncentration fra MetalStat baseret på målinger, og dels en forudsagt koncentration baseret på MetalStat's landsdækkende modellering ved brug af  $X$  matrixen. Selvom den landsdækkende modellering er kalibreret med brug af de latente koncentrationer, så vil den estimerede værdi for den latente koncentration i et opland ikke være identisk med den forudsagte værdi ved brug af den landsdækkende model. Der er mindst usikkerhed på estimatet af den latente koncentration sammenlignet med det forudsagte koncentrationsniveau fra den landsdækkende model.

Hvis det forudsættes, at en latent koncentration fra måledata er repræsentativ for det ID15-opland, hvori målingerne er fortaget, så kan de estimerede latente koncentrationer med fordel erstatte de forudsagte koncentrationsniveauer med brug af  $X$  matrixen i de oplande, for hvilke der er målinger. Desuden kan det udnyttes, at der er vandudveksling mellem oplande, hvorved et koncentrationsniveau i et opland, hvorfra der strømmer vand ned til et nedstrøms opland, vil påvirke oplandskoncentrationen nedstrøms afhængig af metallets retention samt af, hvor stor en andel af vandet i nedstrøms opland, der hidrører fra oplandene opstrøms. I de massebalancekorrigerede resultater fra MetalStat er koncentrationsniveauerne justeret, så de passer med vandudveksling mellem oplande samt estimeret retention, og den latente koncentration for de oplande, der har målinger, erstatter den forudsagte koncentration

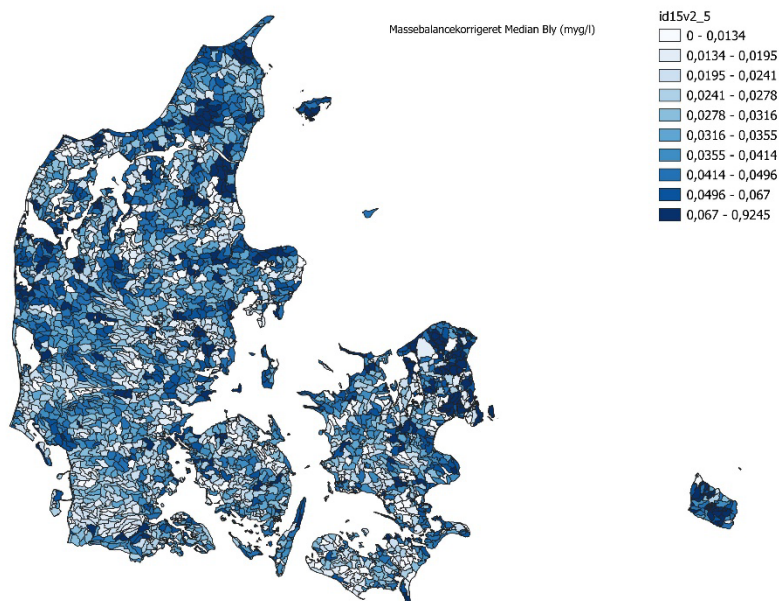
med brug af X matricen. I de massebalancekorrigerede koncentrationer er effekter af år taget væk, så disse værdier svarer til de forventede middel-års-værdier. Dette er gjort, fordi det ikke giver mening at sammenligne en koncentration i opstrøms oplande for ét år med koncentrationerne nedstrøms for et andet år. Princippet bag massebalancekorrigering er beskrevet, diskuteret og evalueret med uafhængige målinger i Sørensen m.fl., (2023). Det er vigtigt at påpege, at massebalancekorrektionen ikke ændrer på værdien for de latente koncentrationer og dermed ikke ændrer på de ID15-oplande, der har målte værdier.

I Figur 4.4 ses forudsagt koncentration alene med brug af X matricen (median) for bly svarende til den forventede koncentration i det pågældende opland. Intervalinddelingen er valgt til at have samme antal oplande i hvert interval fordelt på 10 intervaller. Det betyder, at intervalgrænserne svarer til fraktilværdier af beregnet median mellem oplande med 10 % interval. Så f.eks. for bly er grænsen her mellem nederste og næstnederste niveau 0.0186 µg/l, hvilket så svarer til 10 % fraktilen af medianer mellem ID15-oplande for bly, mens 20 % fraktilværdien for bly er 0.0238 µg/l. De tilsvarende massebalancekorrigerede koncentrationer for bly ses i Figur 4.5. Ved at sammenligne Figur 4.4 og Figur 4.5 ses, at det massebalancekorrigerede resultat synes mindre spraglet mellem oplande med mindre spring i koncentrationstværdier mellem nabo-oplande. Denne forskel er en konsekvens af, at koncentrationstværdierne for den del oplande er justeret til at matche massebalancen med nabo-oplande som beskrevet i Sørensen m.fl., 2023. I Figur 4.6 ses beregnet grundvandskoncentration fordelt på ID15-oplande svarende til figur B13 for bly (se Bilag). En visuel sammenligning mellem Figur 4.5 og Figur 4.6 afslører et vist sammenfald mellem høje koncentrationer i Vestjylland og på Nordsjælland, mens de høje grundvandskoncentrationer især på Nord- og Sydfyn samt sydvest Sjælland ikke synes afspejlet på Figur 4.5. Generelt set er der ikke nogen stærk sammenhæng mellem Figur 4.5 og Figur 4.6, hvilket understøttes af den svage korrelation, der ses i Figur 3.2.

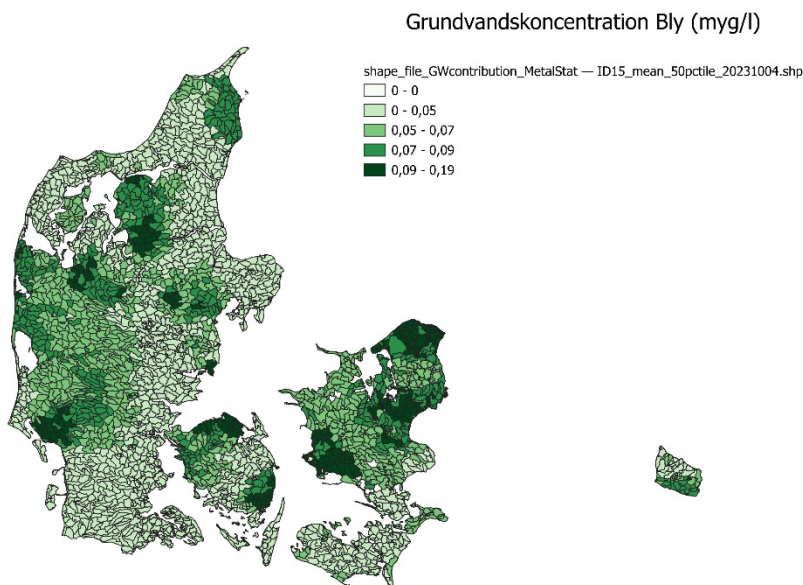
**Figur 4.4.** Forudsagte koncentrationstværdier som median for bly.



**Figur 4.5.** Massebalancekorrektion af det forudsagte værdier samt substitution med latente koncentrationer i de målte oplande for bly.

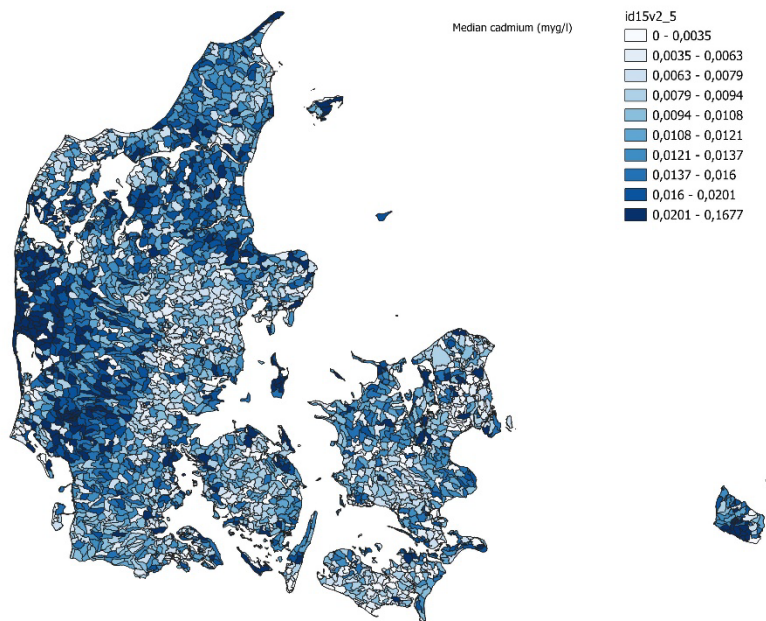


**Figur 4.6.** Grundvandskoncentrationer beregnet jf. bilag for hver ID15-opland for bly. Denne figur svarer til figur B13 i bilag for bly.

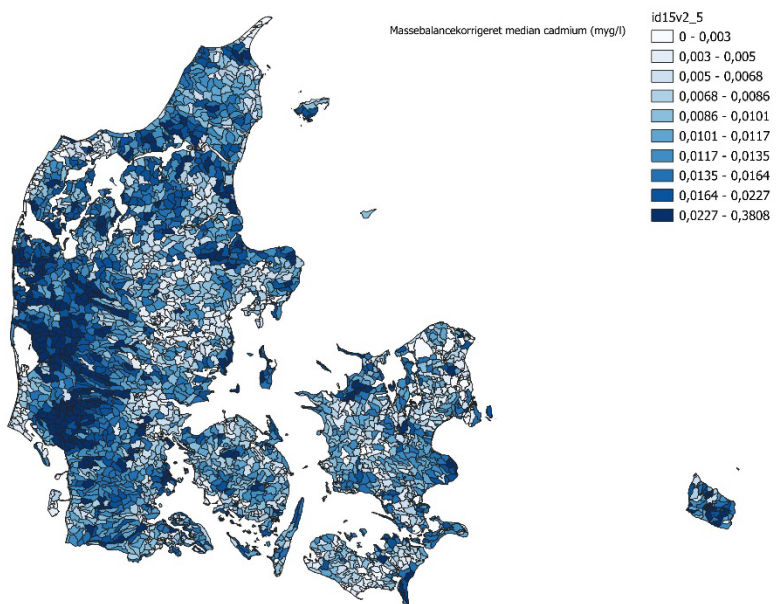


I Figur 4.7 ses forudsagt koncentration alene med brug af X matricen (median) for cadmium med en afbildning ligesom bly oven over. De massebalancekorrigerede koncentrationer for cadmium ses i Figur 4.8, og ved sammenligning med Figur 4.7 ses, at den massebalancekorrigerede version synes noget mindre spraglet mellem oplande med mindre spring i koncentrationstværdier mellem nabo-oplande. Denne forskel er især udtalt i Vestjylland. I Figur 4.9 ses beregnet grundvandskoncentration fordelt på ID15-oplande, svarende til figur B13 for cadmium (se Bilag). En visuel sammenligning mellem Figur 4.8 og Figur 4.9 afslører et tydeligt sammenfald mellem høje vandløbs- og grundvandskoncentration i Vestjylland. Dette tydelige sammenfald bekræfter den relativt tydelige korrelation, der ses på Figur 3.3.

**Figur 4.7.** Forudsagte koncentrationsværdier som median for cadmium.

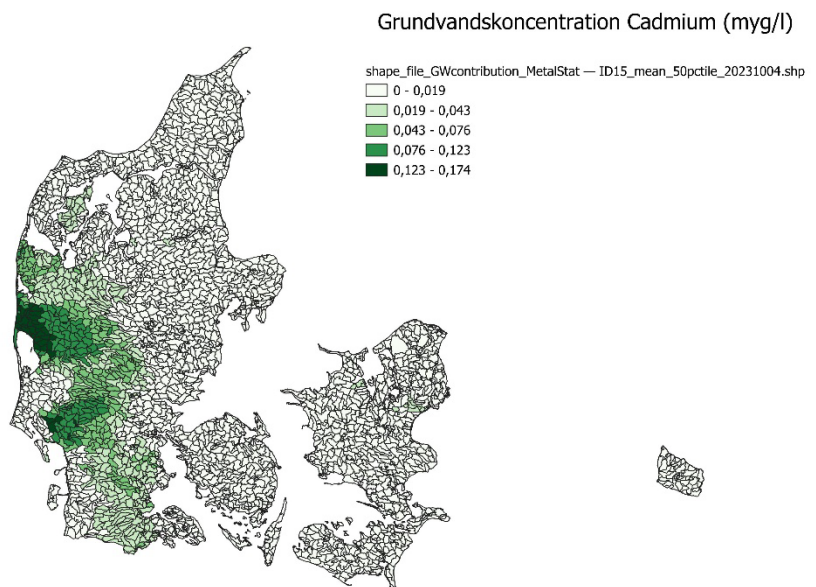


**Figur 4.8.** Massebalancekorrektion af det forudsagte værdier samt substitution med latente koncentrationer i de målte oplande for cadmium.



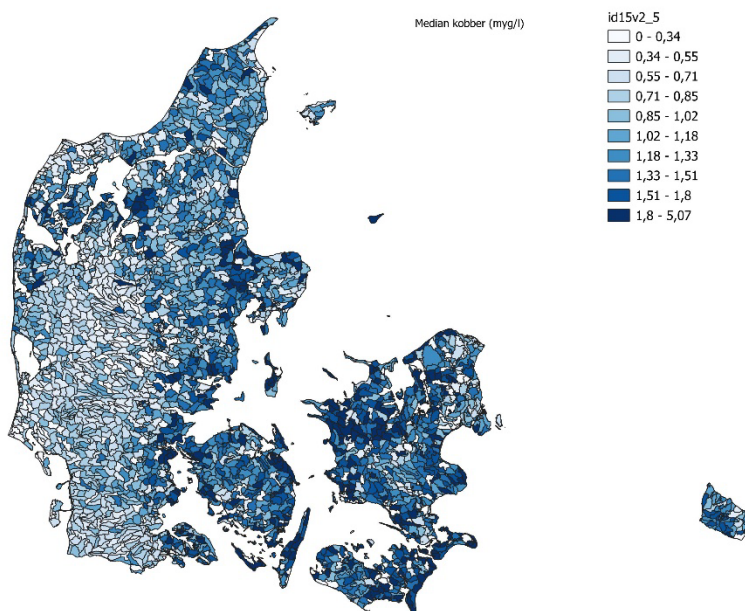


**Figur 4.9.** Grundvandskoncentrationer beregnet jf. bilag for hver ID15-opland for cadmium. Denne figur svarer til figur B13 i bilag for cadmium.

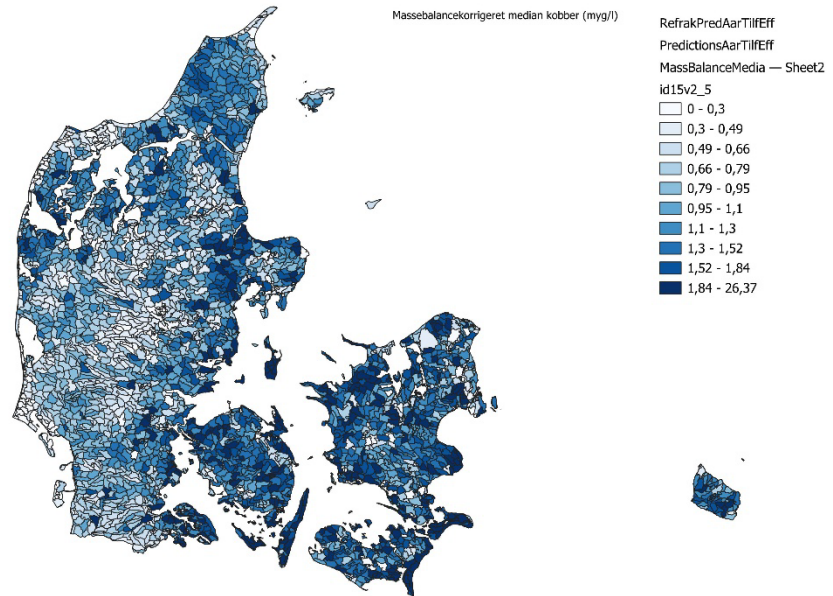


I Figur 4.10 ses forudsagt koncentration alene med brug af X matricen (median) for kobber med en afbildning ligesom bly oven over. De tilsvarende massebalancekorrigerede koncentrationer for kobber ses i Figur 4.11. Ved at sammenligne Figur 4.10 og Figur 4.11 ses, at det massebalancerede korrigerede resultat synes noget mindre spraglet mellem oplande, især for Nord- og Østjylland. I Figur 4.12 ses beregnet grundvandskoncentration fordelt på ID15-oplande, svarende til figur B13 for kobber (se Bilag). En visuel sammenligning mellem Figur 4.11 og Figur 4.12 afslører intet umiddelbart sammenfald mellem høje vandløbs- og grundvandskoncentration. Dette manglende sammenfald passer sammen med den manglende korrelation, der ses i Figur 3.4.

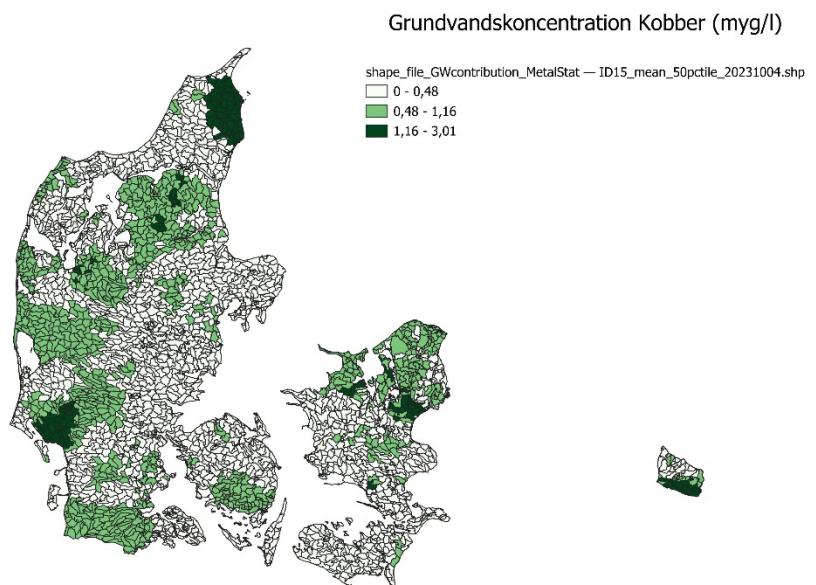
**Figur 4.10.** Forudsagte koncentrationer som median for kobber.



**Figur 4.11.** Massebalancekorrektion af det forudsagte værdier samt substitution med latente koncentrationer i de målte oplande for kobber.

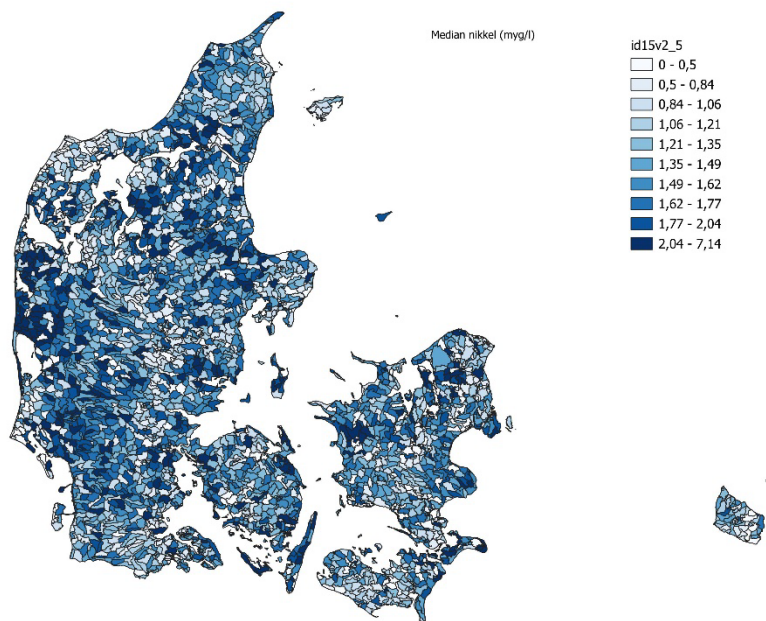


**Figur 4.12.** Grundvandskoncentrationer beregnet jf. bilag for hver ID15-opland for kobber. Denne figur svarer til figur B13 i bilag for kobber.

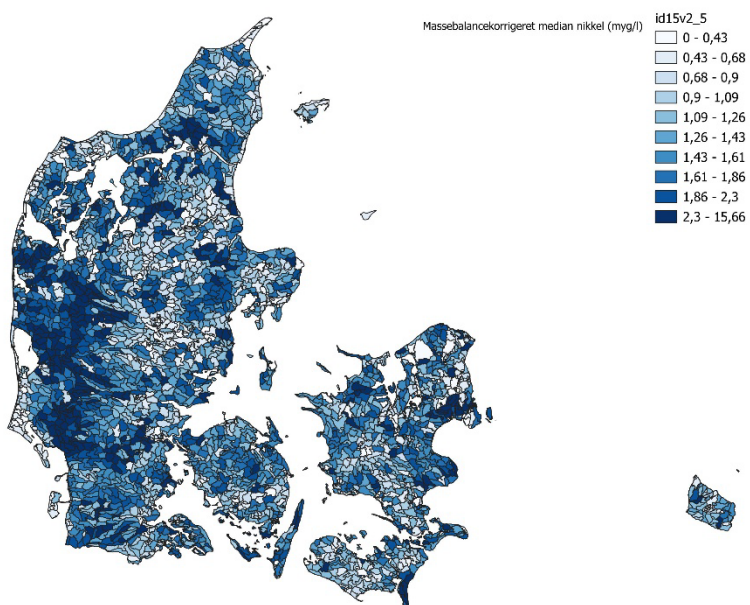


I Figur 4.13 ses forudsagt koncentration alene med brug af X matricen (median) for nikkel med en afbildning ligesom bly oven over. De tilsvarende massebalancekorrigerede koncentrationer for nikkel ses i Figur 4.14. Ved at sammenligne Figur 4.13 og Figur 4.14 ses, at det massebalancerede korrigerede resultat også for nikkel synes noget mindre spraglet mellem oplande, især for Vestjylland. I Figur 4.15 ses beregnet grundvandskoncentration fordelt på ID15-oplande, svarende til figur B13 for nikkel (se Bilag). En visuel sammenligning mellem Figur 4.14 og Figur 4.15 afslører et tydeligt sammenfald mellem høje vandløbs- og grundvandskoncentration, især i Vestjylland. Dette sammenfald passer sammen med den tydelige korrelation, der ses i Figur 3.5.

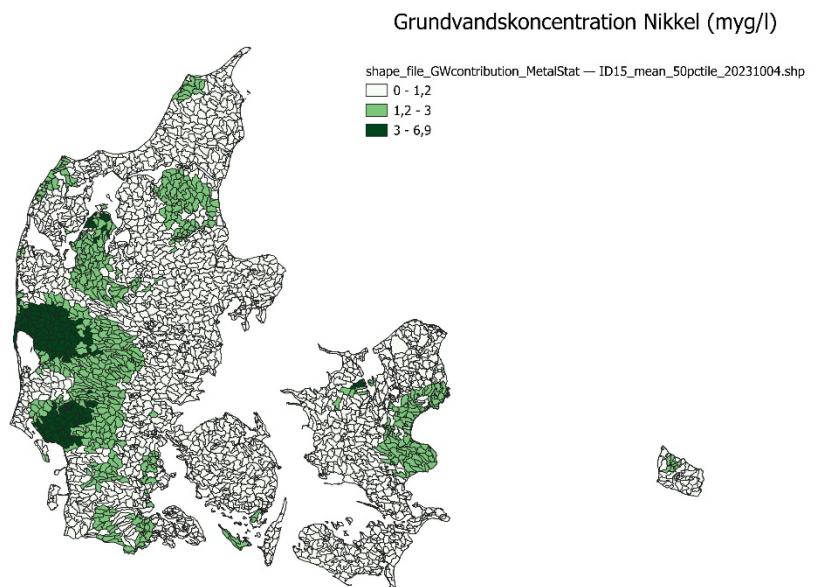
**Figur 4.13.** Forudsagte koncentrationsværdier som median for nikkel.



**Figur 4.14.** Massebalancekorrektion af det forudsagte værdier samt substitution med latente koncentrationer i de målte oplande for nikkel.

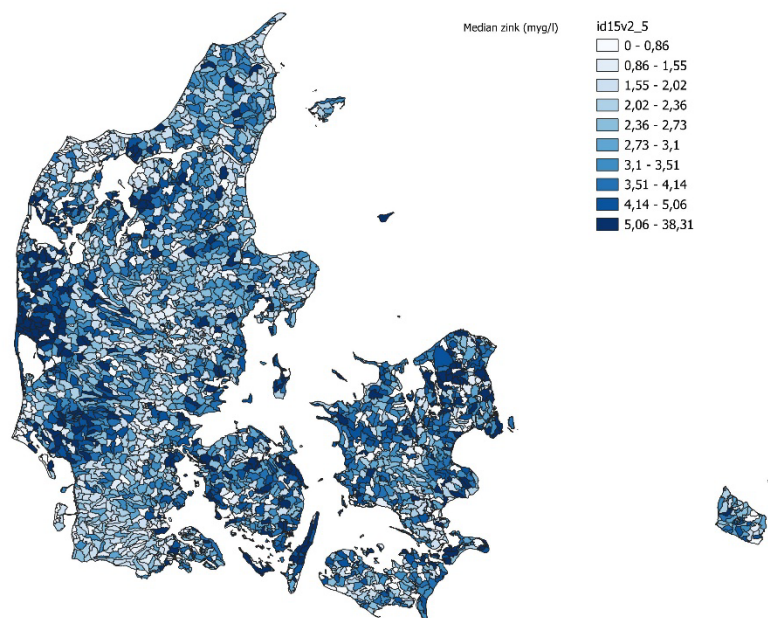


**Figur 4.15.** Grundvandskoncentrationer beregnet jf. bilag for hver ID15-opland for bly. Denne figur svarer til figur B13 i bilag for nikkel.

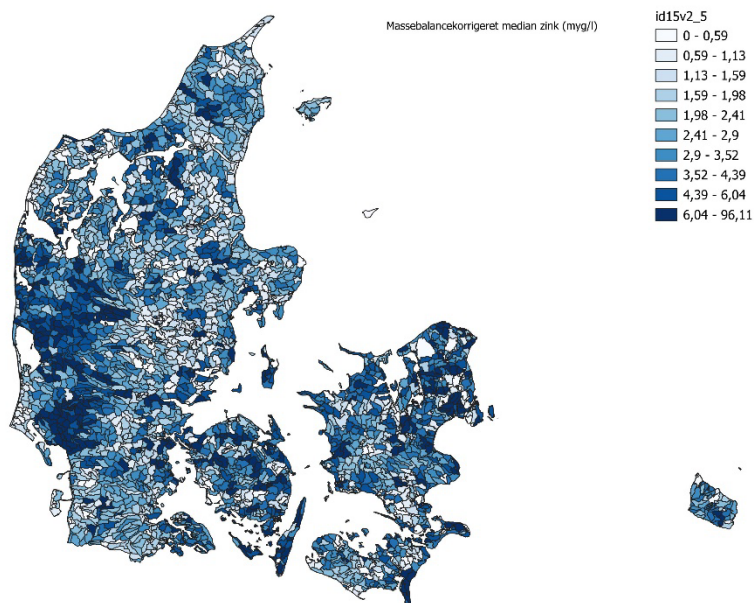


I Figur 4.16 ses forudsagt koncentration alene med brug af X matricen (median) for zink med en afbildning ligesom bly oven over. De tilsvarende massebalancekorrigerede koncentrationer for nikkel ses i Figur 4.17. Ved at sammenligne Figur 4.16 og Figur 4.17 ses, at det massebalancerede korrigerede resultat også for nikkel synes noget mindre spraglet mellem oplande, især for Vestjylland. I Figur 4.18 ses beregnet grundvandskoncentration fordelt på ID15-oplande, svarende til figur B13 for zink (se Bilag). En visuel sammenligning mellem Figur 4.17 og Figur 4.18 afslører et tydeligt sammenfald mellem høje vandløbs- og grundvandskoncentration, især i Vestjylland. Dette sammenfald passer sammen med den tydelige korrelation, der ses i Figur 3.6.

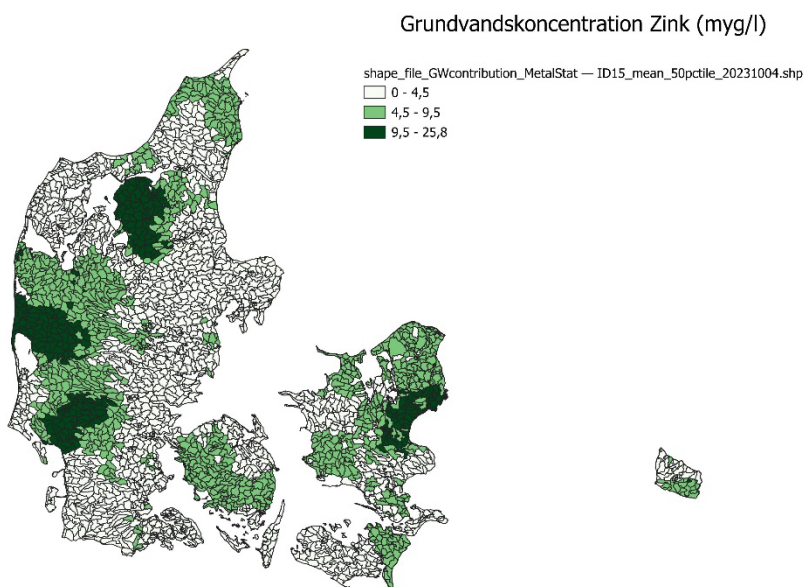
**Figur 4.16.** Forudsagte koncentrationsværdier som median for zink.



**Figur 4.17.** Massebalancekorrektion af det forudsagte værdier samt substitution med latente koncentrationer i de målte oplande for zink.



**Figur 4.18.** Grundvandskoncentrationer beregnet jf. bilag for hver ID15-opland for bly. Denne figur svarer til figur B13 i bilag for zink.



## 5 Udarbejdelse af beslutningsgrundlag til Miljøstyrelsen

MetalStat har leveret et beslutningsgrundlag til Miljøstyrelsen, så styrelsen kan beslutte, om resultater fra MetalStat med fordel kan indgå i vurdering af, om miljøkvalitetskravene er overholdt for ikke målte vandområder. Princippet i valideringen er, at Miljøstyrelsen udfører vurderinger efter gældende retningslinjer, både med brug af målinger fra valideringsstationerne samt med brug af forudsagte koncentrationer fra MetalStat. Da resultater fra MetalStat skal anvendes til at udtale sig om ikke målte vandområder, vil de målinger, der indgår i valideringsstationerne, ikke blive anvendte i kalibrering af MetalStat. Ved parvis sammenligning af konklusionerne fra målte værdier og forudsigelserne fra MetalStat, vil Miljøstyrelsen beslutte, om resultater fra MetalStat skal indgå i Miljøstyrelsens sagsbehandling. MetalStat leverer forudsigelser og kan kvantificere usikkerhed og naturlig variation på forskellige måde alt efter det specifikke spørgsmål, der søges besvaret. Foreløbigt vælges det at levere to typer af forudsigelser til beslutningsgrundlag:

- Forudsigelse af latent koncentration alene med brug af X matricen uden massebalancekorrektion, hvor betydningen af variation mellem år er kvantificeret ved tillægge et bidrag fra et tilfældigt år.
- Massebalance korrigerede prognoser jf. metoden i Sørensen et al, 2023. Årsmiddel af latente koncentrationer korrigeres, så massebalancen passer mellem oplande. Der anvendes årsmiddel, da det kun giver mening at justere en massebalance mellem to oplande, der ikke afhænger af tidspunktet. Desuden forudsætter massebalancekorrektionen, at den latente koncentration i et opland er repræsentativ for hele oplandet. I det omfang der er estimeret en latent koncentration for et ID15-opland, så substituerer dette estimerede forudsigelsen med brug af X matricen.

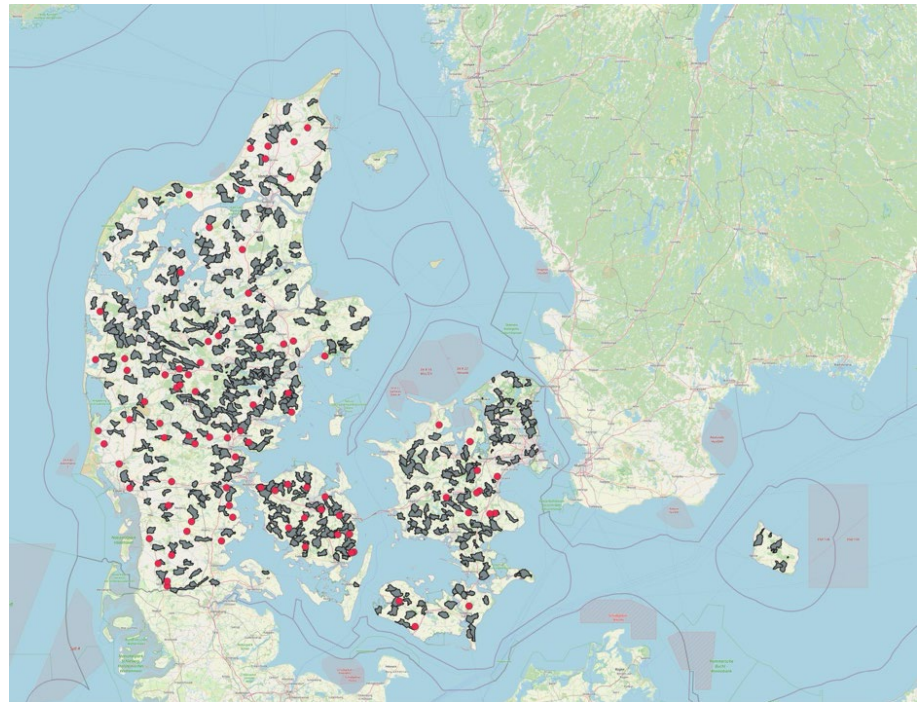
Brug af forudsigelser alene med X matricen tillagt et bidrag fra et tilfældigt år har den fordel, at usikkerheden ikke underestimeres, men ulempen er, at forskellen mellem oplande udviskes, og modellen derved får sværere ved at adskille oplande og inddrage lokale målinger.

Brug af massebalancerede korrigerede forudsigelser har den fordel, at der tages hensyn til de målinger, der forligger i nogle ID15-oplande, samt deres betydning for ID15-oplande i samme hovedopland. Det øger modellens evne til at tage lokale hensyn og adskille et opland fra et andet. Ulempen er, at i det omfang at målingerne i samme vandssystem ikke er så repræsentative som forudsat, vil der være en ukendt usikkerhed i prognoserne, som ikke er modelleret.

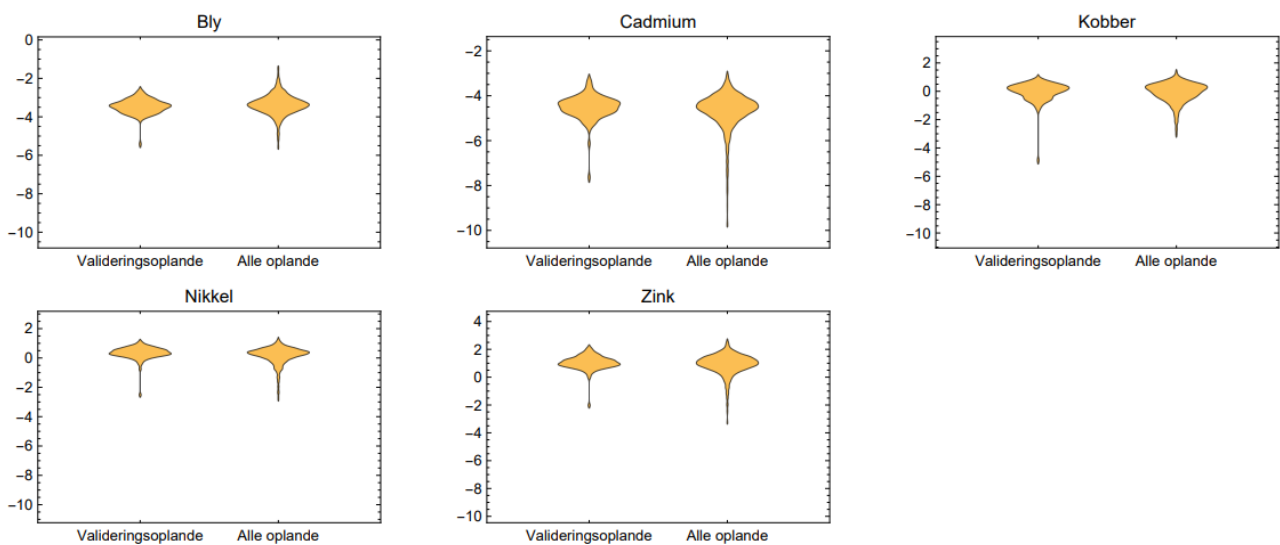
Der er udvalgt 84 ID15-oplande, for hvilke målinger er holdt ude ved kalibreringen af MetalStat, og som derfor bliver forudsagt af modellen uden kendskab til de målte værdier. Der udføres vurdering af om miljøkvalitetskravene er overskredet eller ej for disse oplande både med brug af målte koncentrationer og med brug af forudsigelser fra MetalStat for de 84 oplande.

Valideringsstationer er udvalgt på en måde, så de ligger relativt geografisk spredt, se Figur 5.1.

**Figur 5.1.** Placering af valideringsstationer (røde punkter), hvor demarkerede felter svarer til de ID15-oplande, der har andre målestationer med i MetalStat.



I en sammenligning af medianværdierne af prognoserne mellem de 84 ID15-oplande fra beslutningsgrundlaget og alle ID15-oplande var der ikke nogen tydelig systematisk afvigelse jf. Figur 5.2, og det blev derfor konkluderet, at de 84 ID15-oplande kan antages at udgøre en statistisk repræsentativ stikprøve af alle oplande.

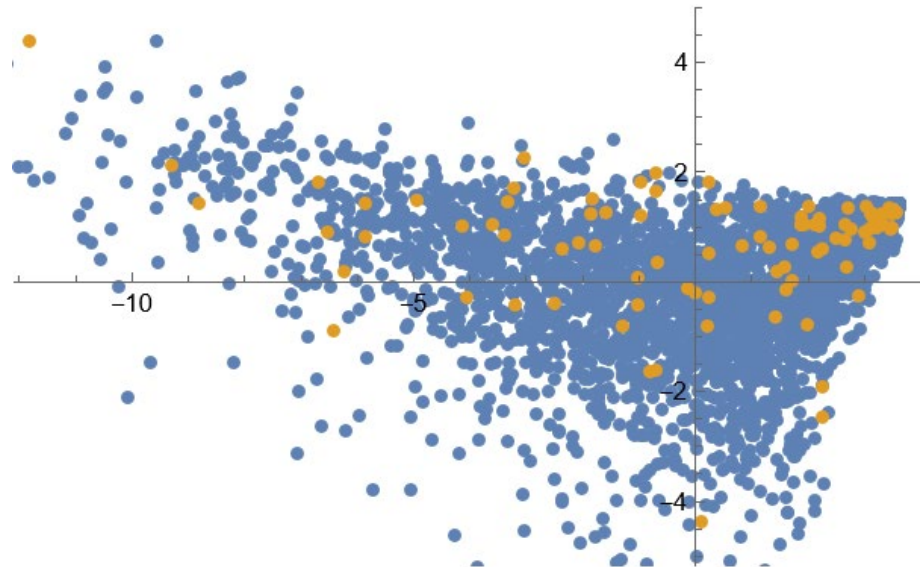


**Figur 5.2.** Sammenligning mellem målinger, der er brugt til validerings målinger (beslutningsgrundlag), samt alle målinger brugt som input til MetalStat (Alle oplande). På y akse vises logaritmen til koncentrationen.

På Figur 5.2 blev det vist, at de målte koncentrationer i beslutningsgrundlaget ikke afveg fra de målte koncentrationer, som indgår i MetalStats kalibrering. På tilsvarende vis kan det undersøges, om X matricen for de ID15-oplande, der er udvalgt til beslutningsgrundlag, er repræsentativ for X matricen generelt i Danmark. Som input er udført en generel dimensionsanalyse af X matricen for at kontrollere, at de 84 valgte oplande spreder sig tilfældigt i det multidimensionale X rum. I en ordination af de forklarende variable (X matricen), hvor man reducerer dimensionen af datasættet til de to mest betydende dimensioner, svarende til PC1 og PC2 i Figur 3.11, fandt vi, at de 84 ID15-op-

lande fra valideringsdatasættet (gule punkter i Figur 5.3) ikke systematisk afveg fra alle ID15-oplande (blå og gule punkter i Figur 5.3). Det blev derfor konkluderet, at de 84 ID15-oplande kan antages at udgøre en repræsentativ stikprøve af alle oplande.

**Figur 5.3.** Fordelingen af de 84 valgte ID15-oplande (gule punkter) sammenlignet med alle øvrige ID15-oplande.





## 6 Udvikling af fremtidig MetalStat driftsmodel, samt fremadrettet arbejde med modellering

### 6.1 Udvikling af fremtidig MetalStat driftmodel

Den nuværende version af MetalStat bygger på en MCMC (Marco Chain Monte Carlo) beregning, hvor mange gentagende beregninger afsøger variable for mulige værdier. Det er en meget fleksibel, og derfor attraktiv, metode til kortlægning af usikkerheder for komplekse og sammensatte modeller, såsom MetalStat, hvilket er baggrunden for, at udviklingsversionen af MetalStat er bygget op omkring MCMC metoden. Ulempen ved MCMC-metoden er, at den er beregningskrævende, når der er mange variable i modellen, og der er et behov for at reducere beregningstiden i udviklingsversionen af MetalStat.

Når beregningstiden for en model skal reduceres, er der i princippet tre muligheder:

1. Hardware optimering. Dette indebærer brug af flere CPU'er eller GPU'er, så det er muligt at udføre så mange beregning per tid som muligt.
2. Software optimering. Dette indebærer brug af koder, der ligger så tæt på maskinen som muligt, og som regner variable ud på den mest optimale måde i forhold til maskinens arkitektur.
3. Matematisk optimering. Dette indebærer omformulering af ligningerne i modellen, så kun de mest nødvendige beregninger foretages i forhold til at opnå et tilstrækkeligt nøjagtigt resultat. Det kan også omfatte mere matematiske "smarte" ligninger, der udtrykker modellen på en måde, der ikke kræver det samme beregningsarbejde.

I 2023 blev der på ECOS indkøbt en 24 kerner regneserver for at sikre en hardware optimering, så MetalStat kunne levere de lovede resultater til Miljøstyrelsen. Ligeledes i 2023 blev kodningen i MetalStat paralleliseret som led i en software optimering for at kunne udnytte alle kernerne i regneserveren. Denne aktivitet har mindsket regntiden til 1/10 i forhold til den version af MetalStat, der blev brugt til afrapportering af Sørensen m.fl., 2023. Som opstart til at gøre koden mere maskinnær blev der udført en indledende test med brug af Python, som dog ikke indtil videre har givet resultater, der kan konkluderes på i forhold til en kommende drift version af MetalStat.

I løbet af 2023, og i tæt diskussion med DTU Aqua ved Professor Anders Nielsen, viste det sig lovende at forsøge med en optimering af estimeringsproceduren i MetalStat. Denne diskussion er ikke en del af kravspecifikationerne fra Miljøstyrelsen, da den først opstod i sidste del af 2023 og ind i 2024. Princippet er, at MetalStat beholder hele sin oprindelige ligningsstruktur uændret, men at ligningerne løses ved integration som approksimation til den fælles posteriori sandsynlighedsfordeling af parametrene. Approksimationens princip er først at anvende en numerisk differentiering for at finde toppunkter for de statistiske fordelingsfunktioner og derefter anvende Laplace approksimation til at integrere disse fordelinger omkring deres toppunkter. Metoden er

operationaliseret i en software pakke i R, som udvikles under DTU Aqua (Kristensen m.fl., 2016).

Ved hjælp af disse approksimationer bliver det muligt at integrere sandsynlighedsfunktionerne i modellen som alternativ til at afsøge parametrene fordeling med brug af MCMC. De første indledende test-beregninger tyder på, at regeringstiden kan reduceres til under 1/1000, hvilket tyder på, at denne approksimation kan eliminere problemet med lang beregningstid. Det er dog for tidligt at komme med endelige konklusioner, da der forestår en vurdering af, hvor god approksimationen er.

## **6.2 Fremadrettet arbejde med modellering for andre miljøfarlige stoffer**

En gennemgang af potentiale og udfordringer for modellering af miljøfremmede stoffer i vandløb kan ses i Sørensen et al, (2021). Siden udgivelsen af denne rapport har udviklingen af MetalStat dog løftet de metodemæssige muligheder for modellering, der kan danne skabelon for modeller for andre miljøfarlige stoffer. Det er især tilfældet set i lyset af de muligheder, der er åbnet op for reduktion af beregningstiden.

MetalStat er sammensat af to undermodeller: (1) En statistisk model, der tolker målinger; (2) En forklarende model (oplandsmodel), der søger at forklare resultaterne fra den statistiske model gennem variable ( $X$ -matricen). Det er meget vigtigt at tage udgangspunkt i denne opdeling, når potential for modellering af miljøfarlige stoffer skal vurderes. Basalt set er en statistisk model i en eller anden form uundgåelig, lige så snart data skal tolkes, og det er alene et spørgsmål om at være bevidst om dette og bruge den bedst mulige statistiske model til formålet med de data, der er til rådighed. Den forklarende model er derimod kun mulig med et vist vidensniveau og med et vist volumen af data, både i form af målte koncentrationer og i form af forklarende variable. Det betyder, at der bør udarbejdes en statistisk model for hver gruppe af miljøfarlige stoffer, der tolker de målte data og angiver retningslinjer for den bedste strategi, hvis der ønskes udført yderligere målinger. Den statistiske model kan så relativt let udvides med en forklarende model, hvis der er viden nok til dette. På denne måde kan modellering blive en integreret del af databehandlingen generelt for miljøfarlige stoffer.



Pilotversionen af MetalStat, svarende til den testede version i Sørensen m.fl. (2022), er udviklet til en forløbelig driftsversion i forhold til beregningstid og opdateret med nye målinger og målestationer frem til og med 2021 og er opdateret med resultater fra en forbedret grundvandsbeskrivelse. Resultaterne i denne rapport svarer til udtræk fra MetalStat den 14/12-2023.

X-matricens evne til at forklare forekomsten af metal i vandløbene, udtrykt som log-koncentrationer, kan kvantificeres gennem korrelationskoefficienten  $R^2$ . Ideelt set ønskes en  $R^2$  værdi tæt på en, men en  $R^2$  værdi, der bygger på log-koncentrationer, påvirkes af de relative forskelle mellem model og måling, så små afvigelser mellem lave koncentrationer kan bidrage til  $R^2$  værdien med samme vægt som større forskelle mellem høje koncentrationer. Så en relativ lille  $R^2$  betyder ikke nødvendigvis, at modellen ikke kan identificere de høje koncentrationsniveauer med god sikkerhed. Med dette forbehold opnås følgende  $R^2$  med stigende rækkefølge bly (0,13) -> nikkel (0,23) -> cadmium (0,31) -> zink (0,35) -> kobber (0,37). Denne liste viser, at bly har en noget lavere  $R^2$  værdi sammenlignet med de øvrige metaller, hvilket enten kan skyldes, at der mangler en beskrivende variabel i X-matricen for bly, eller at bly simpelthen er mere tilfældigt fordelt over landet pga. en historisk gammel og generel kontaminering uden nogle større og regelbaserede geografiske forskelle.

En tilfældig effekt fra år viser, at zink og kobber har en stærk negativ tilfældig effekt i 2020, hvor zink gentager den negative effekt i 2021, og en opdatering med nye måledata for 2022 og 2023 vil kunne afgøre, om X-matricen i MetalStat skal justeres i forhold til at beskrive tidseffekter. Zink er blevet reguleret i landbruget i de seneste år, så her kunne det f.eks. være relevant med en opdateret X matrice, så kalibreringen af MetalStat kan adskille betydningen af landbrugsdrift hhv. før 2020 og fra 2020 og fremefter.

Massebalancekorrigerede resultater kan ses som en huludfyldning af årsmiddelværdier for umålte oplande med brug af målte oplande og understøttet af X-matricen. Det er vigtigt at påpege, at massebalancekorrektionen ikke ændrer på værdien for de latente koncentrationer og dermed ikke ændrer på de ID15-oplande, der har målte værdier. Massebalancekorrektionen giver rummeligt set mere homogene resultater uden så store ændringer i koncentrationens værdien mellem nabolande sammenlignet med ukorrigerede resultater. Denne tendens er ikke helt så udtalt for bly som for de andre metaller, hvilket kan skyldes, at de målte koncentrationer varierer mere tilfældigt mellem ID15-oplande, og at de latente koncentrationer derfor varierer uafhængigt af lokaliteten, hvilket vil betyde, at massebalancekorrektionen har svært ved at finde et fælles koncentrationsniveau for umålte oplande mellem de målte.

Der er et rumligt sammenfald mellem vandløbskoncentrationerne med de beregnede grundvandskoncentrationer fra grundvandsmodellen hos især cadmium, nikkel og zink, mens bly og kobber ikke udviser så tydelig tendens, hvilket også passer med, at bly og kobber netop var de to metaller, der viste sig mindst relateret til grundvandskoncentrationer. Det er især de Vestjyske vandløbsoplande omkring Skjern Å og Varde Å for cadmium, nikkel og zink, som har et sammenfald mellem modellerede vandløbskoncentrationer fra MetalStat og beregnede grundvandskoncentrationer for disse tre metaller. Der er også en tendens til, at det er de massebalance-korrigerede resultater fra MetalStat, som udviser mest tydelig sammenhæng mellem vandløbskoncentrationer og grundvandskoncentrationer, hvilket kan forklares ved, at grund-

vandspåvirkningen kan give et øget koncentrationsniveau for målte ID15-oplande, som massebalancekorrektionen justerer over på tilstødende ikke målte ID15-oplande.

Et udtræk fra MetalStat den 14/12-2023 for 84 ID15-oplande er leveret til Miljøstyrelsen, for hvilke der er målte værdier, som ikke er taget med i kalibrering af modellen. Disse 84 oplande er udvalgt til at udgøre en repræsentativ stikprøve både i forhold til geografisk placering og i forhold til type af opland, som den er beskrevet gennem  $X$ -matricen. Dette udtræk kan bruges af Miljøstyrelsen som beslutningsgrundlag til at beslutte, om resultater fra MetalStat skal anvendes i deres sagsbehandling. Ved parvis sammenligning af konklusionerne fra målte værdier og forudsigelser fra MetalStat vil Miljøstyrelsen beslutte, om resultater fra MetalStat skal indgå i Miljøstyrelsens arbejde med vandplanerne.

Ved hjælp af Laplace approksimationer bliver det muligt at integrere sandsynlighedsfunktionerne i modellen som alternativ til at afsøge parametrenes fordeling med brug af MCMC. De første indledende test-beregninger er meget lovende, men det er dog for tidligt at komme med helt endelige konklusioner omkring den reelle forbedring i beregningstid, da der forestår en vurdering af, hvor god approksimationen er. Hvis det lykkes at reducere beregningstiden, bør MetalStat inddrage massebalancen mellem tilstødende oplande direkte i løsning af de styrende ligninger, så den nuværende efterjustering med massebalancekorrektion helt kan undgås. Dette vil udvide modellen med flere latente variable, og det har derfor indtil nu været anset som urealistisk af regnetekniske grunde. Desuden vil en regnetidsreduktion gøre det muligt at lave en mere grundig revurdering af  $X$ -matricen for at optimere den og inddrage tidseffekter. I denne revurdering indgår også en mulighed for at tolke de enkelte variable i  $X$ -matricen og derigennem identificere nøglekilder til metaller i vandløbene.

Arbejdet med MetalStat har gjort det tydeligt, at modelleringsaktiviteten skal betragtes op delt mellem hhv. en statistisk model og en forklarende model ( $X$  matrice), hvor den statistisk model med fordel kan anvendes mere generelt på måledata, mens den forklarende model anvendes over for de stofgrupper, hvor der hersker tilstrækkeligt med viden om årsagen til stoffernes forekomst.

## 8 Referencer

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., & Bell, B. M. (2016). TMB: Automatic Differentiation and Laplace Approximation. *Journal of Statistical Software*, 70(5), 1-21. <https://doi.org/10.18637/jss.v070.i05>

Lawson, A.B., 2018, *Bayesian Disease Mapping, Hierarchical modelling in spartial epidemiology*. Thord Edition, Taylor & Francis Group, LLC, ISBN-13:978-1-138-57542-4 (hbk).

Sørensen, P.B., Andersen, H.E., Fauser, P., Bjerg, P.L., Bro, R., Holm, P.E. & Abrahamsen 2021 Muligheder for modellering af miljøfarlige forurenende stoffer i overfladevand. Aarhus Universitet, DCE – Nationalt Center for Miljø og Energi, 60 s. - Videnskabelig rapport nr. 414  
<https://dce2.au.dk/pub/SR414.pdf>

Sørensen P.B., m.fl., 2022: Peter Borgen Sørensen, Christian Frølund Damgaard, Poul Løgstrup Bjerg, Hans Estrup Andersen, Peter Engelund Holm, Henrik Tornberg, Jesper Leth Bak, Goswin Johann Heckrath, Ane Kjeldgaard, Patrik Fauser. 2022. Udvikling af model til understøttelse af vandplanlægningen i fastlæggelsen af koncentrationsniveauer af metaller i danske vandløb. Aarhus Universitet, DCE – Nationalt Center for Miljø og Energi, 70 s. - Videnskabelig rapport nr. 484. <https://dce2.au.dk/pub/SR484.pdf>.

Sørensen, P.B., Damgaard, C.F, Bjerg, P.L., Andersen, H.E., Holm, P.E., Bak, J.L., Rasmussen, D., Kjeldgaard, A., 2023. Videreudvikling af model for metaller i vandløbsvand, MetalStat. Aarhus Universitet, DCE – Nationalt Center for Miljø og Energi, 107 s. - Videnskabelig rapport nr. 522.  
<https://dce2.au.dk/pub/SR522.pdf>

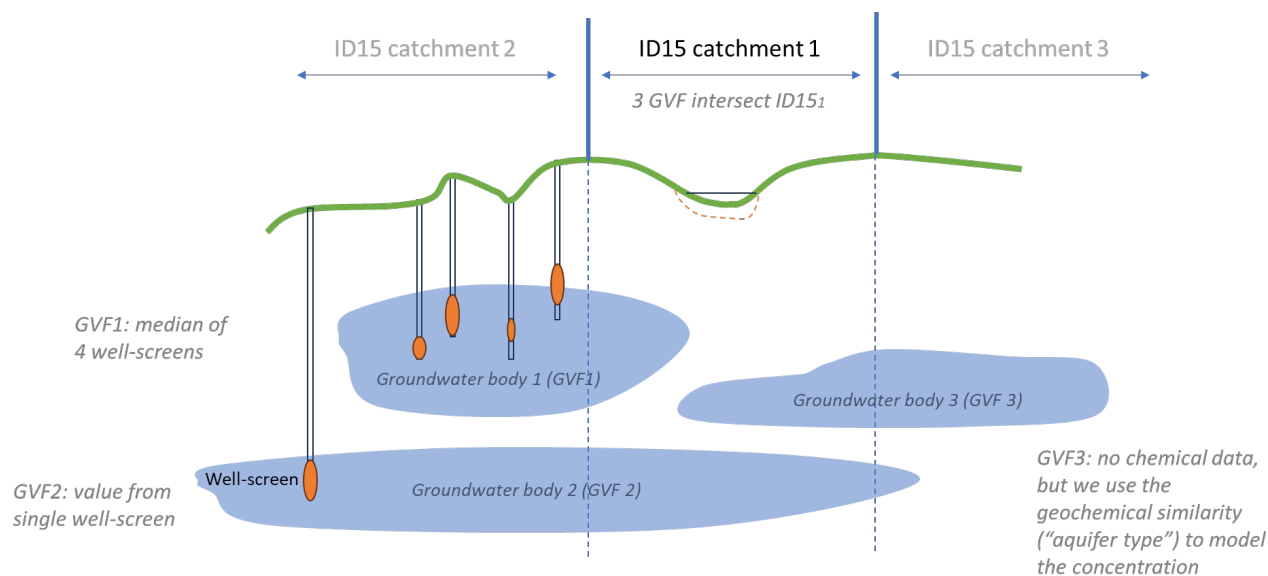
## Bilag, Beregning af grundvandskoncentration (Engelsk)

Dette bilag er udarbejdet af Denitza Voutchkova og Lærke Thorling.

### Introduction

The purpose of this work is to provide nationwide estimates on the potential groundwater input of Cd, Cu, Ni, Pb and Zn to the surface water bodies. The desired output is at ID15 catchment level (shapefile, polygon) with typical concentrations for the five metals for each ID15 catchment. The actual groundwater input of trace elements to the surface waters is unknown. Currently, it is challenging to estimate at the national scale, as the groundwater-surface water interactions have not been resolved at this scale yet.

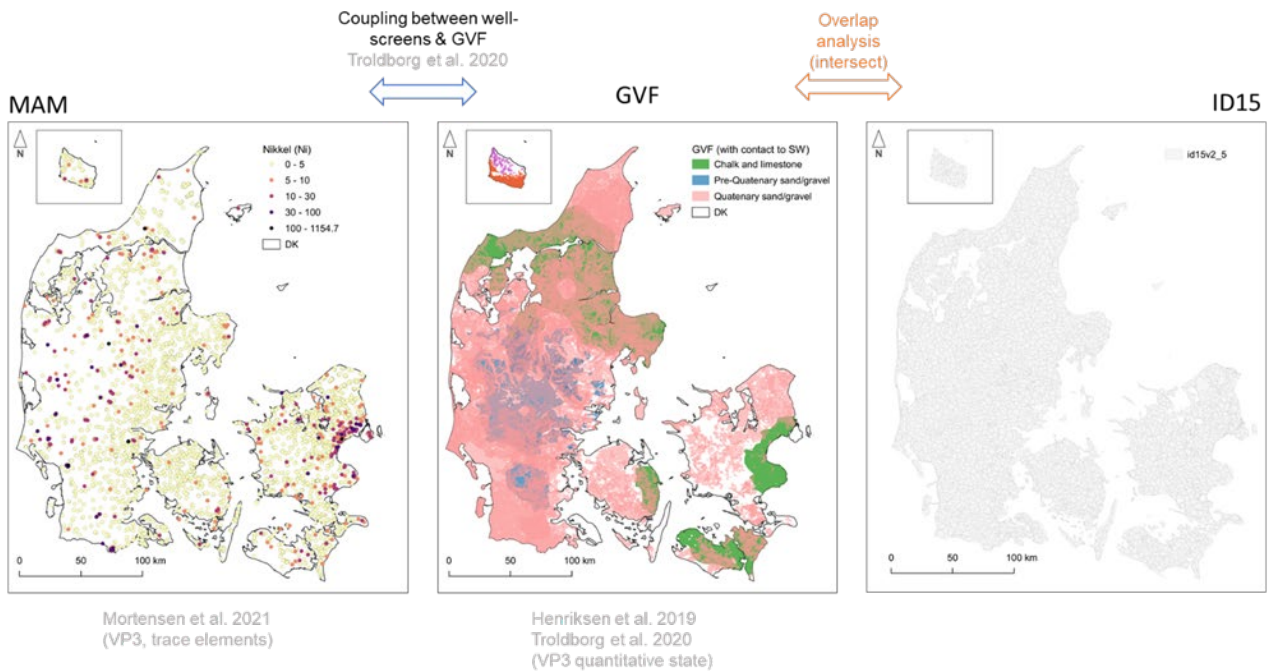
Figure B1 shows a conceptual model of our system, where the ID15 catchments represent the water flow at the surface. Underneath the ID15 catchments, there could be multiple groundwater bodies (GVF) that are potentially in contact with the surface water. In this hypothetical example, there are three GVFs with varying levels of information about the metal concentrations in the groundwater. For example, GVF1 has four well-screens with data, GVF2 has only one well-screen with data, and GVF3 has no well-screens (unknown groundwater concentrations).



**Figure B1.** Conceptual model of our system showing the ID15 catchments at the surface and hypothetical groundwater bodies intersecting spatially the ID15 catchments

The methodology has three major steps (Figure B2):

1. Estimating typical concentration **at the well-screen level**,
2. Estimating/modelling the typical groundwater concentrations of **at the GVF/aquifer type level**,
3. Estimating the potential groundwater input to the surface waters **at the ID15 catchment level**.

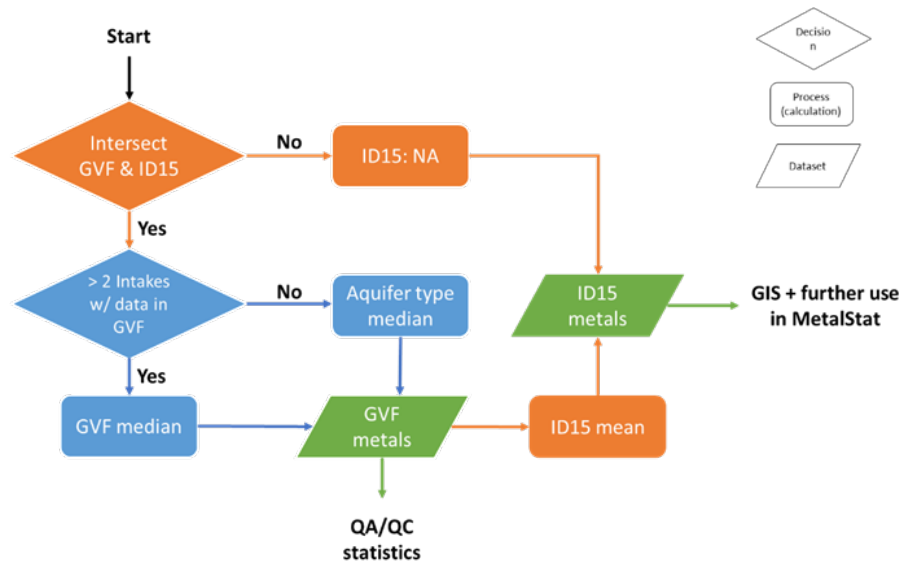


**Figure B2.** Visualisation of the three methodological steps: well-screen level (left), groundwater body (mid), ID15 catchment (right); abbreviations: MAM – mean annual mean, GVF – groundwater body, ID15 – surface water catchments

### Input data and methods

The described three-stepped methodology is formalised with the flow-chart (a decision tree) shown on Figure B3.

**Figure B3.** Flow chart for deriving the metal influence from groundwater to the ID15 catchments





**Table B1** Data source

Data	Reference
Mean annual mean concentration (MAM) of Cd, Cu, Ni, Pb, Zn in well-screens from the national monitoring network (GRUMO) and waterworks wells (BK) for the period 2000-2018 (incl.). Well-screen meta-data also available.	Mortensen m.fl. (2021)
Groundwater bodies (GVF, Grundvandsforekomster) <i>in contact with surface waters</i> from the DK-model 2019	Henriksen m.fl. (2019, 2021) Troldborg m.fl. (2020)
Coupling between well-screens and GVF	Troldborg m.fl. (2020)
ID15 catchments, v.2	provided by Peter B. Sørensen

The following calculations were done:

- **At ID15 catchment level:** mean of the typical metal concentrations at the GVF intersecting a specific ID15. It was considered whether a weighted mean should be used instead, where the weight was the GVF volume. It was decided not to do this because there is not enough information to conclude that larger GVF contribute to the surface waters more than smaller GVF.
- **At GVF level:** calculating typical metal concentrations based on median of the typical concentrations at the well-screens within the GVF/aquifer type. The median was used because it is more robust against outliers and the non-normality of the data. Two different calculations were done at this level, based on how many well-screens with data were there in the GVF:
  - **If there were > 2 well-screens in the GVF** – a median of the MAMs at these well-screens was calculated. These are referred to as “estimated” concentrations.
  - **If there were ≤ 2 well-screens in the GVF** – the typical concentration for the specific aquifer type was used instead. The typical concentration was modelled as the median of all well-screens that are linked to the specific aquifer type (see below for more details). These are referred to as “modelled” concentrations because information from other locations is transferred to the GVF, where there is no or limited information.
- **At well-screen level:** mean annual mean (MAM) for the period 2000-2018 (incl.). This part of the work was done in Mortensen et al. (2021). MAM was used because the time-series were highly irregular (different sampling frequencies in different years) and this way each year had equal weight.
- **In GIS:** an intersect between ID15 and GVF polygons to determine which GVF were overlapping spatially with which ID15 catchments.

### Defining aquifer/groundwater type

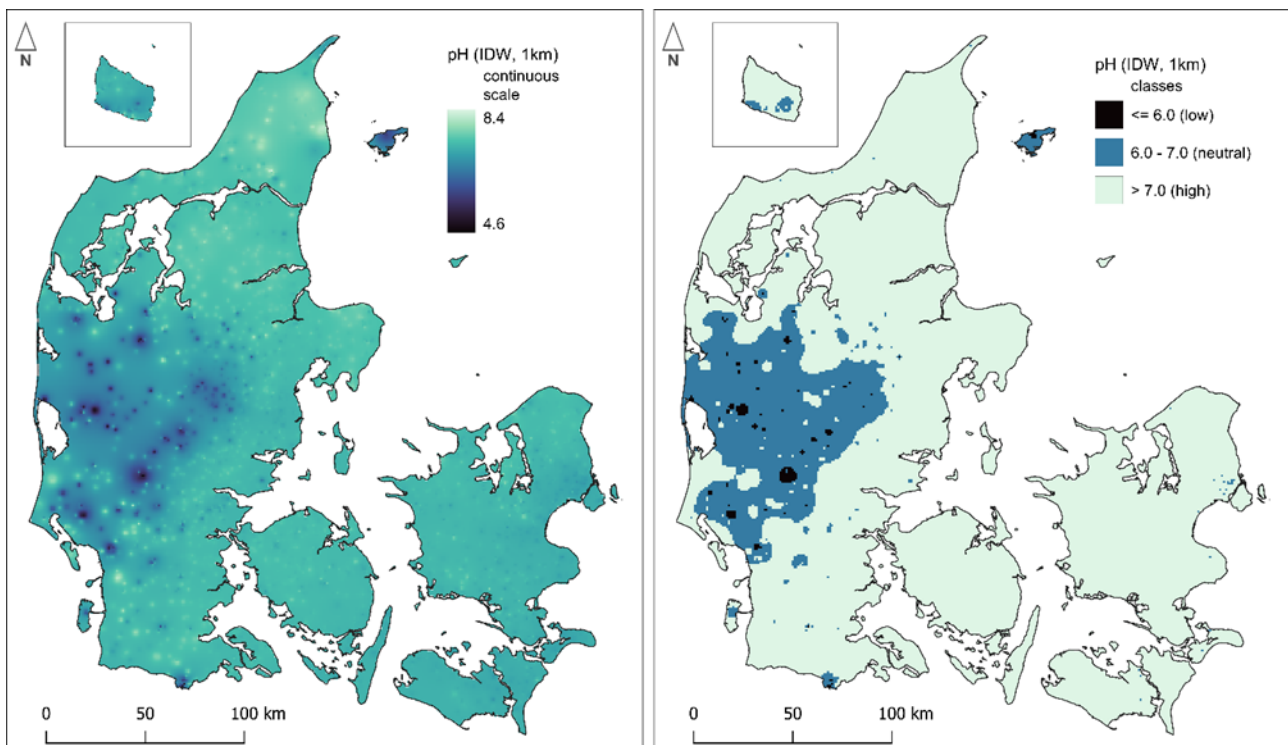
The definition of aquifer types follows the methodology developed for the natural background levels for Vandplan III (Mortensen et al., 2021). The definition combines classification based on geology, geography, pH and redox. Both redox and pH have either direct or indirect importance for the mobility/sorption of the five metals. The same pH and redox classes were used for all five metals. Both the well-screens and the GVF were classified, so that they could be matched together. The classification is as follows:

At a well-screen level:

- Reduced/Oxic based on nitrate concentration ( $\text{NO}_3 \leq 2$  and  $> 2$  mg/L).
- Low, neutral, high pH ( $\leq 6$ , 6-7,  $> 7$ , respectively). This differs from Mortensen m.fl. (2021), as one additional class was added here, " $> 7$ ", while there were only two pH classes. The classes were redefined based on Figure 4 (right).
- nbl\_unit, a combination of geography ("dkms" - Sjælland, "dkmf" - Fyn, "dkmj" - Jylland, "dkmb" - Bornholm) and geology ("ks" quaternary sand, "ps" pre-Quaternary sand, "kalk" - carbonate aquifer, "uu" - various units on Bornholm), e.g. dkmj\_ks - Quaternary sand on Jutland)

At a GVF level:

- Oxidic if the volumetric portion of the GVF ("OxProcent") was  $> 50\%$ , otherwise - reduced.
- Same classes as at the well-screen level. The GVFs were not classified based on pH previously, so here we calculated a median pH at the GVF (QGIS zonal statistics) based on an IDW interpolated map (Figure A4)
- nbl\_unit (same as at well-screen level).
- The typical metal concentration at each type of aquifer was calculated based on a median of the specific well-screen types (from the corresponding classification).



**Figure B4.** Interpolated groundwater pH map (inverse distance weighted, IDW, 1km resolution) based on 10y-mean for 5,673 well-screens

### GVF classification

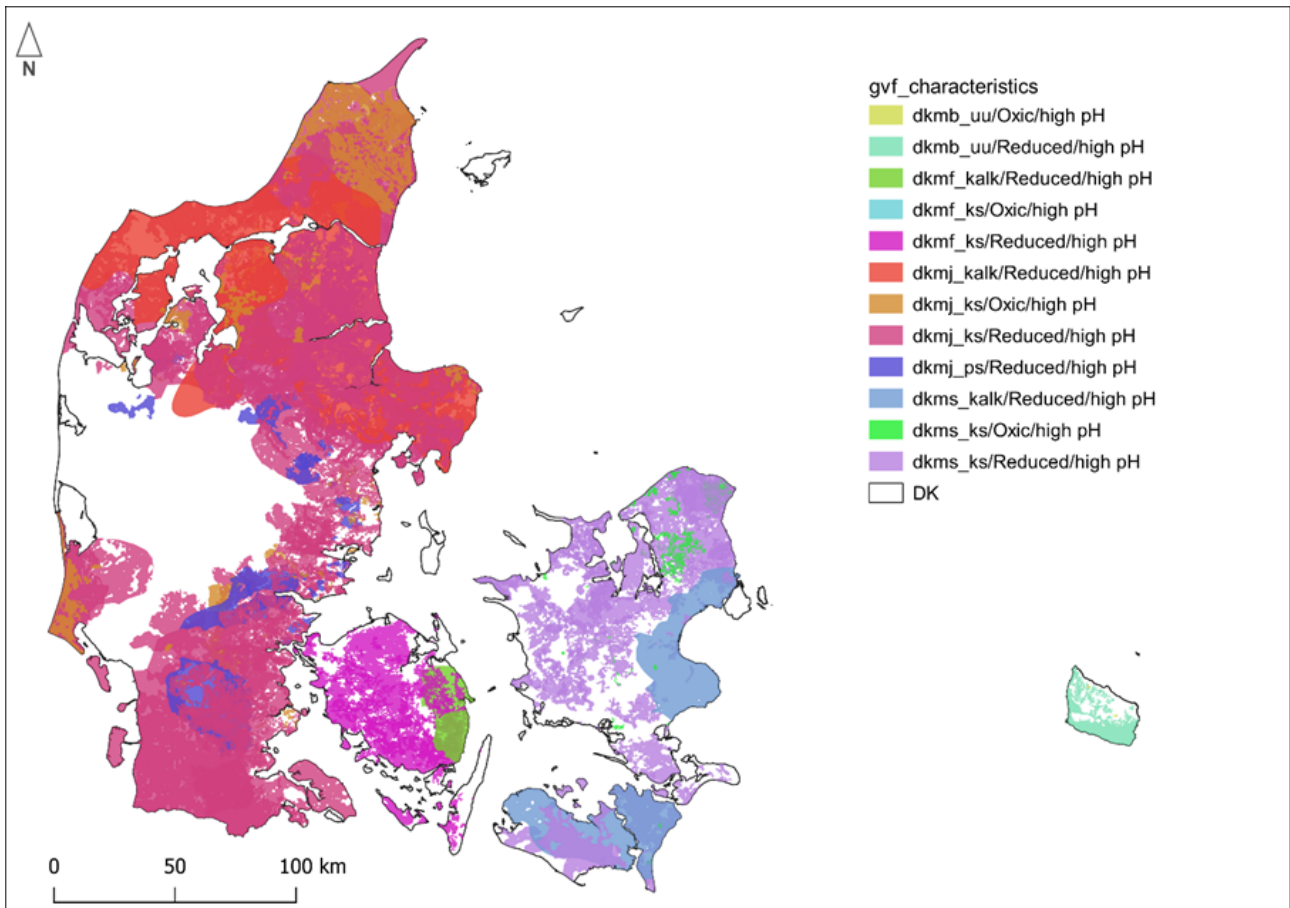
There were 740 GVF with a link to surface water. Based on the pH and redox conditions, these GVF were classified into 17 different aquifer types (Table B2, Figure B5 and Figure B6).

Most of the GVF (85.7%) were shallow (“Terrænnær”) and the rest were regional (“Regional”). Based on the mean depth of the GVF, the median depth for the 740 GVF included here was 4.6 m and the interquartile range (Q25-Q75) was 0.67-13.9m.

Nearly 10% of the GVF were layered. A 70% were in the first three Quaternary sand layers of the DK-model (ks1, ks2, ks3), while 15 were in the carbonate aquifer layer.

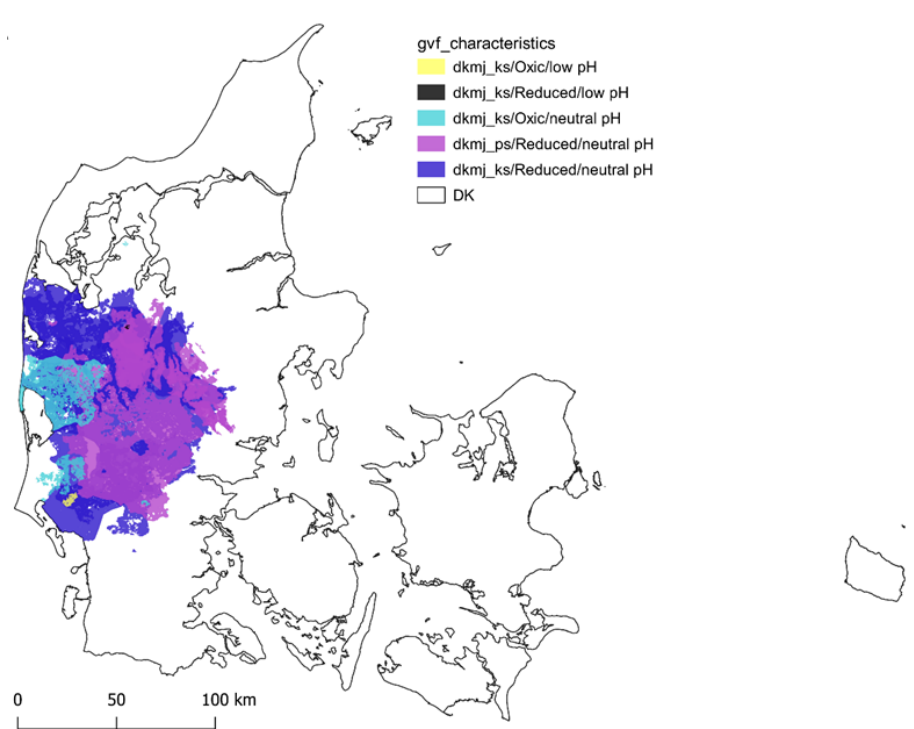
**Table B2** Aquifer type classification of each groundwater body (GVF)

Aquifer type	Geo-graphy	Geology	pH	NO3 (mg/L)	Number of GVF
dkmj_ks/Oxic/low pH	Jylland	Quaternary sand	≤ 6	> 2	1
dkmj_ks/Reduced/low pH	Jylland	Quaternary sand	≤ 6	≤ 2	1
dkmb_uu/Oxic/high pH	Bornholm	Various	>7	> 2	2
dkmf_kalk/Reduced/high pH	Fyn	Carbonates	>7	≤ 2	2
dkmf_ks/Oxic/high pH	Fyn	Quaternary sand	>7	> 2	3
dkms_kalk/Reduced/high pH	Sjælland	Carbonates	>7	≤ 2	4
dkmj_kalk/Reduced/high pH	Jylland	Carbonates	>7	≤ 2	9
dkmj_ks/Oxic/neutral pH	Jylland	Quaternary sand	6-7	> 2	11
dkmb_uu/Reduced/high pH	Bornholm	Various	>7	≤ 2	17
dkmj_ks/Reduced/neutral pH	Jylland	Quaternary sand	6-7	≤ 2	25
dkmj_ps/Reduced/neutral pH	Jylland	Pre-Quaternary sand	6-7	≤ 2	25
dkmj_ps/Reduced/high pH	Jylland	Pre-Quaternary sand	>7	≤ 2	32
dkms_ks/Oxic/high pH	Sjælland	Quaternary sand	>7	> 2	47
dkmf_ks/Reduced/high pH	Fyn	Quaternary sand	>7	≤ 2	75
dkmj_ks/Oxic/high pH	Jylland	Quaternary sand	>7	> 2	97
dkms_ks/Reduced/high pH	Sjælland	Quaternary sand	>7	≤ 2	173
dkmj_ks/Reduced/high pH	Jylland	Quaternary sand	>7	≤ 2	216



**Figure B5.** Classification of groundwater bodies (GVF) based on pH and redox conditions. Only the GVF with high pH (> 7) are shown here; transparency is used so the spatial overlapping of the GVF can be visualized.

**Figure B6.** Classification of groundwater bodies (GVF) based on pH and redox conditions. Only the GVF with low or neutral pH ( $\leq 6$  and 6-7, respectively) are shown here; transparency is used so the spatial overlapping of the GVF can be visualized.



## Results

### Concentrations of Cd, Cu, Ni, Pb, and Zn at the well-screen level

The data at the well-screen level is from Mortensen et al. (2021). Only groundwater wells from the national monitoring program (GRUMO) and the waterworks wells for drinking water purposes (BK) were included. In total, there were 5673 well-screens (GRUMO & BK), but the data-availability varies (Table B3). The nationwide concentration distributions stratified by pH and redox classes are shown on Figure B7.

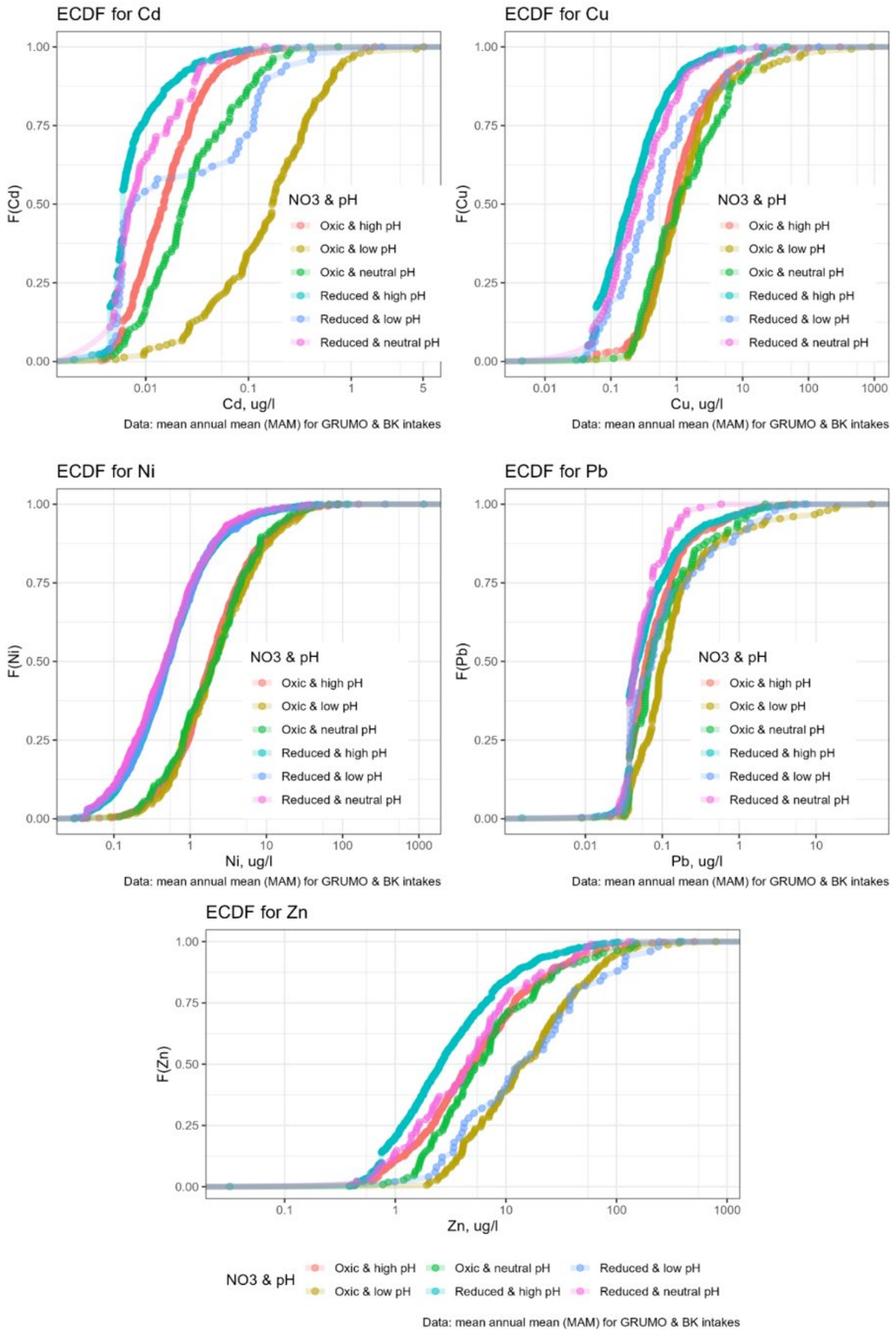
**Table B3.** Trace element data availability at well-screen and groundwater body level

Trace element	Well screens (n)	GVF (n)	GVF in contact with SW <sup>[1]</sup> (n)	GVF with estimated concentration <sup>[2]</sup> (n)	GVF with modelled concentration <sup>[3]</sup> (n)
Cd	1666	317	196	112	628
Cu	1424	285	186	106	634
Ni	5672	484	251	179	561
Pb	1667	320	197	115	625
Zn	1689	320	198	114	626

[1] number of GVF in contact with surface waters (SW) that have well-screens with data.

[2] number of GVF where the concentration was estimated as the median from the well screens in the GVF (with more than two well-screens).

[3] number of GVF where the concentration was modelled based on the median for the aquifer type (this was done because there were no well-screens within the GVF or there were 1 or 2 well screens).



**Figure B7.** Empirical cumulative distribution function (ECDF) for Cd, Cu, Ni, Pb and Zn in Denmark; each point represents the MAM concentration at a well screen

## Concentrations of Cd, Cu, Ni, Pb and Zn at the groundwater body level

A “typical” concentration was assigned to each GVF in a different way, depending on the data availability. In cases, where there were more than two well-screens with data, the typical concentrations were estimated as the median of the well-screens’ MAMs. In cases, where there were less or no well-screens with data, the typical concentration was modelled by using the aquifer-type median.

Table B4 provides a summary statistic about the depth of the two types of GVF. In general, the GVF with modelled concentrations were much shallower, with a median range 3.0-3.6 m, while for the GVF with data, the median range was 14.9-18.4 m (Table B2). This means that there were no direct observations for the GVF near the terrain, which are most probably affected to a larger degree by anthropogenic activities. Using modelled concentrations based on the aquifer types would then most probably result in underestimation of the typical concentration in those GVF. The lack of direct observations reflects the pre-selection of well-screens for the purposes of setting natural background levels (Mortensen et al., 2021), where all shallow wells monitoring point-source pollution were excluded from the dataset.

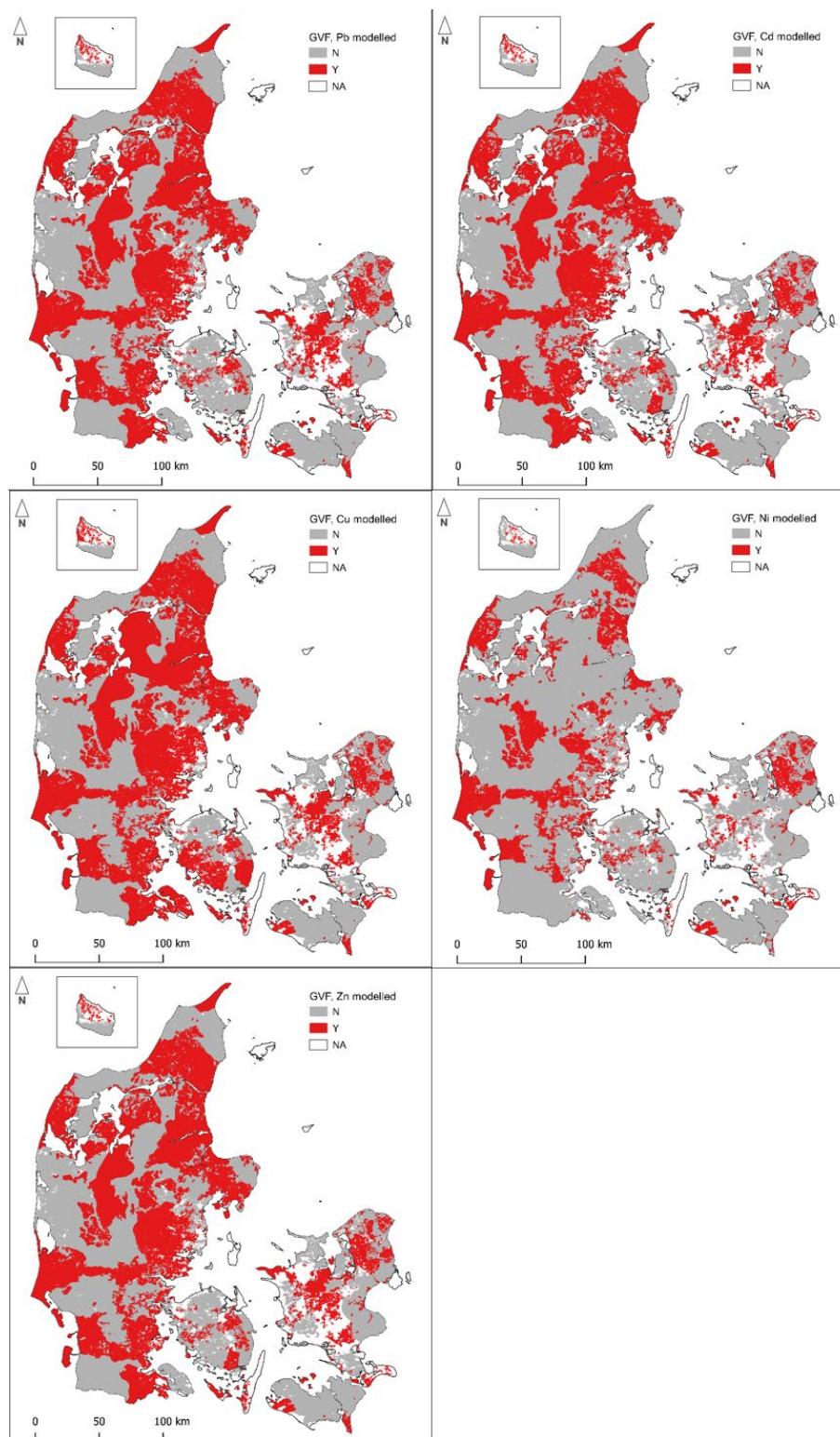
**Table B4.** Depth statistics for the groundwater bodies that were with estimated or modelled typical concentrations. IQR – interquartile range (25th percentile to 75th percentile, Q25 – Q75)

Trase element	Depth of GVF with <i>estimated</i> concentrations (m)		Depth of GVF with <i>modelled</i> concentrations (m)	
	median	IQR	median	IQR
<b>Cd</b>	15.8	5.8–35.1	3.6	0.4–10.2
<b>Cu</b>	14.9	5.5–35.0	3.6	0.4–10.5
<b>Ni</b>	18.4	8.1–34.8	3.0	0.2–7.6
<b>Pb</b>	15.9	6.0–35.2	3.5	0.4–10.0
<b>Zn</b>	15.8	5.9–35.2	3.6	0.4–10.2

Figure B8 shows the typical concentrations at GVF as a function of the number of well-screens with data. It was decided that GVF with two or less well-screens would be treated as GVF without data (so the typical concentrations were modelled) in order to avoid the influence of outliers (extremely high values at single location). This decision was based on assessing the typical concentrations based on similar plots as on Figure B8, with all GVF with at least one well-screen with data were also estimated and not modelled. The range of modelled and estimated values after adjusting the methodology were more comparable, as it can be seen in Figure B8.

Figure B6 shows the location of the GVF with modelled typical concentrations for the five metals (with red color). It should be noted that there are multiple layers of GVF, so here the GVF with modelled concentrations are shown on top, but there are other GVF with estimated typical values deeper. Ni is the metal with best data-availability out of these five, so this is also visible on the maps on Figure B9.

**Figure B9.** Groundwater bodies (GVF), where the typical concentrations for Cd, Cu, Ni, Pb were modelled (Y – yes, N – no) based on the typical values for the aquifer type (as opposed to estimation based on direct observations from well-screens within the GVF)



### Groundwater contribution of Cd, Cu, Ni, Pb and Zn to the ID15 catchments

The groundwater contribution to the ID15 catchment was estimated as outlined in the Methods section (see Figure B3 for flowchart). There were three scenarios possible when estimating the groundwater contribution at ID15 catchment (Figure B10):

- a) mix of modelled and estimated values at the GVF level



b) only modelled values at the GVF level

c) only estimated values at the GVF level.

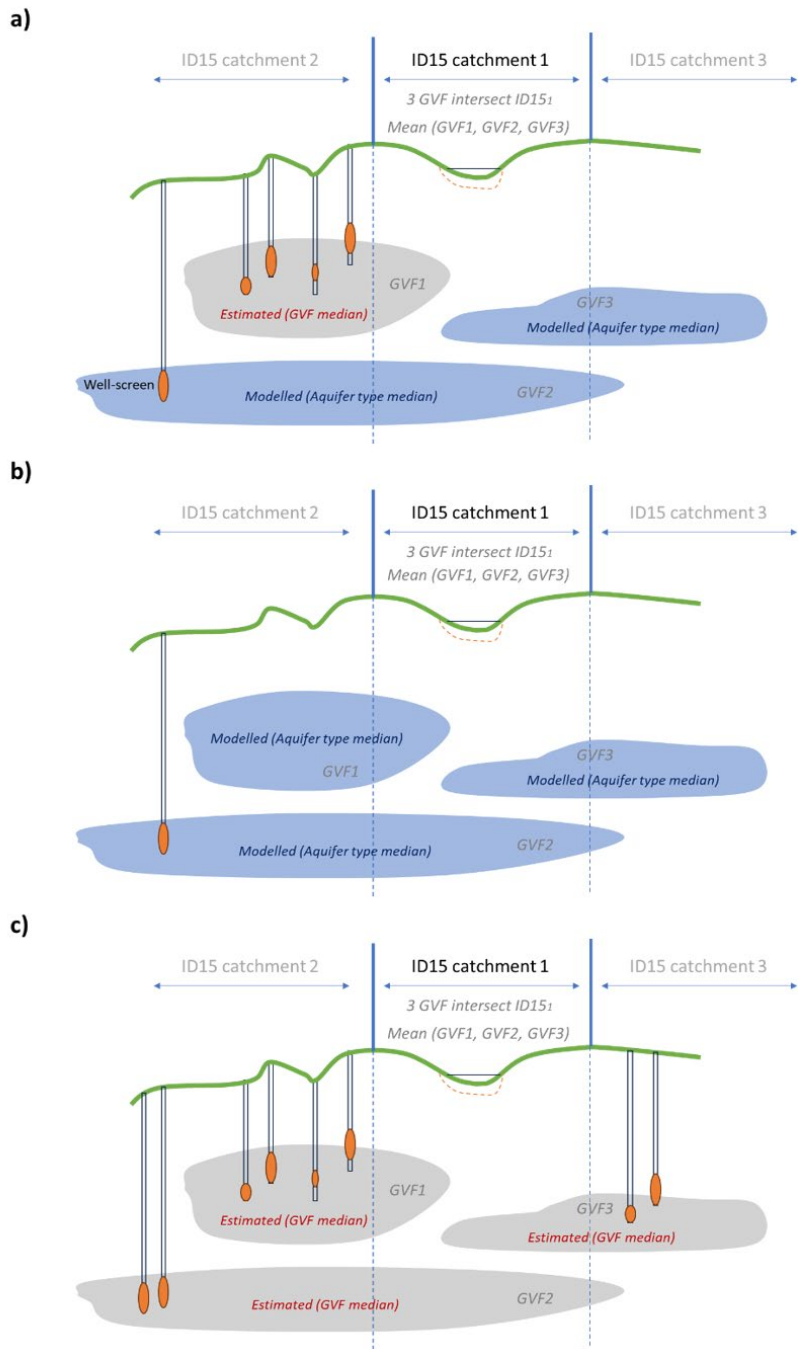
Figure B11 shows the number of GVF intersecting the ID15 catchments (from 0 to 13). A histogram with distribution of ID15 catchments based on the number of intersecting GVF can be seen on Figure B11.

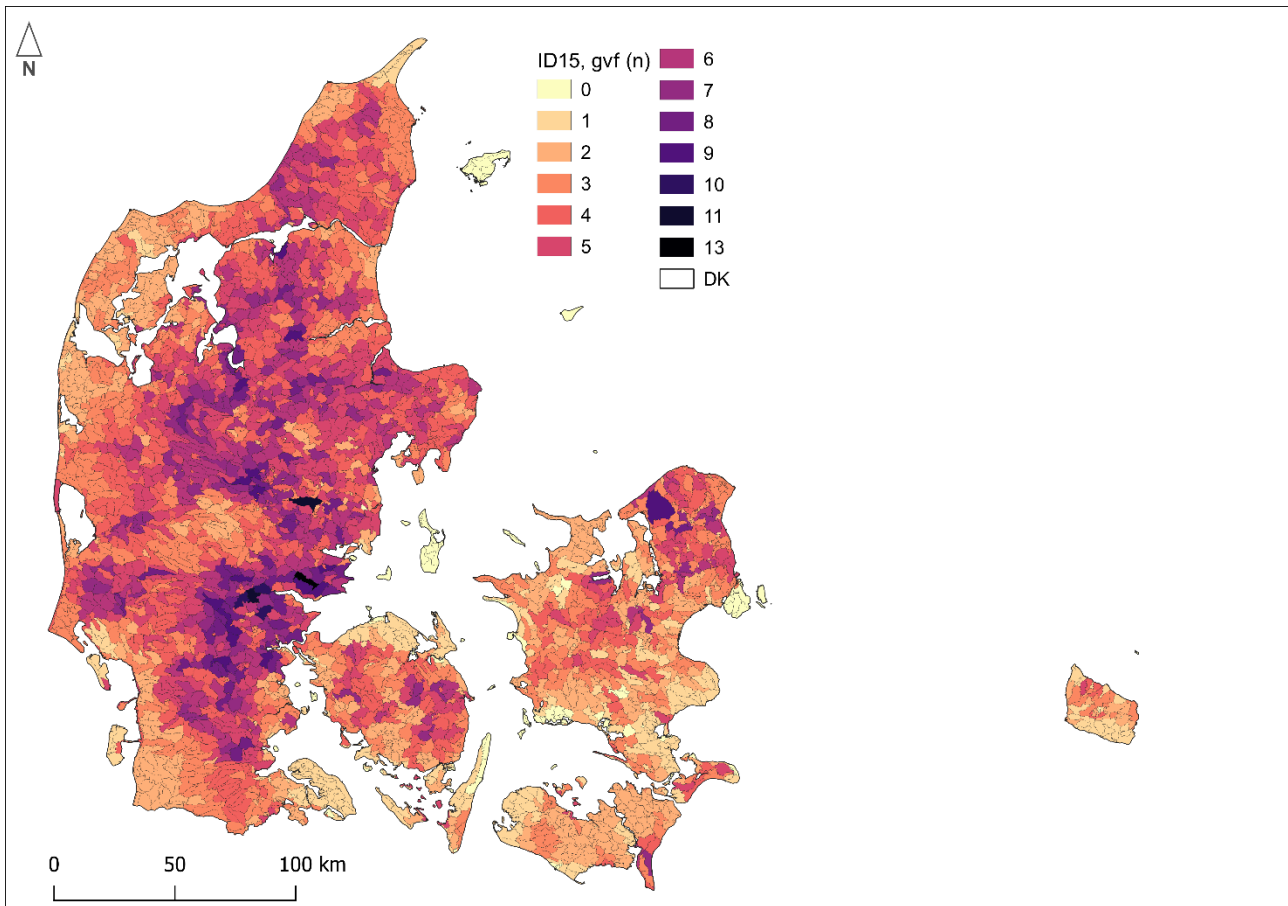
In 117 ID15 catchments, it was not possible to model (or estimate) typical concentrations because there were no intersecting GVF or because these were outside of the modelling domain of the DK-model (e.g. some small islands), see Figure B13 and Table B5.

Histograms with the groundwater contribution for each of the metals to the ID15 catchments can be seen on Figure B12, while Table B5 provides an overview of the number and percentage of ID15 catchments with mix, modelled or estimated values (see a, b, c, scenarios above and in Figure A11).

Figure B13 shows maps with the groundwater contribution to the ID15 catchments for each of the metals.

**Figure B10.** The three scenarios for estimating the groundwater contribution to the ID15 catchments: a) mix of modelled and estimated values at the GVF level; b) only modelled values at the GVF level; only estimated values at the GVF level.



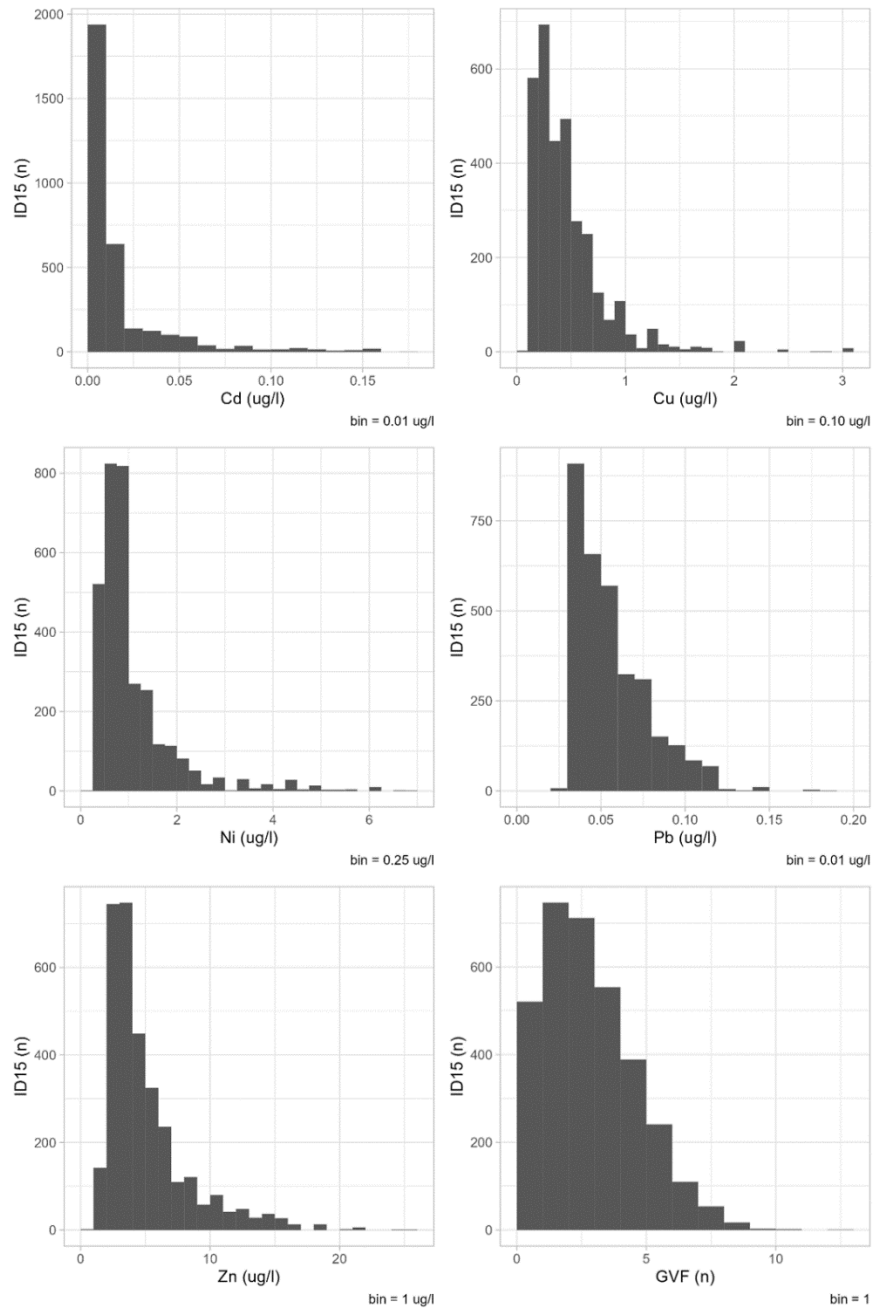


**Figure B11.** Number of groundwater bodies that intersect the ID15 catchments; the analysis was done in QGIS (“intersect”)

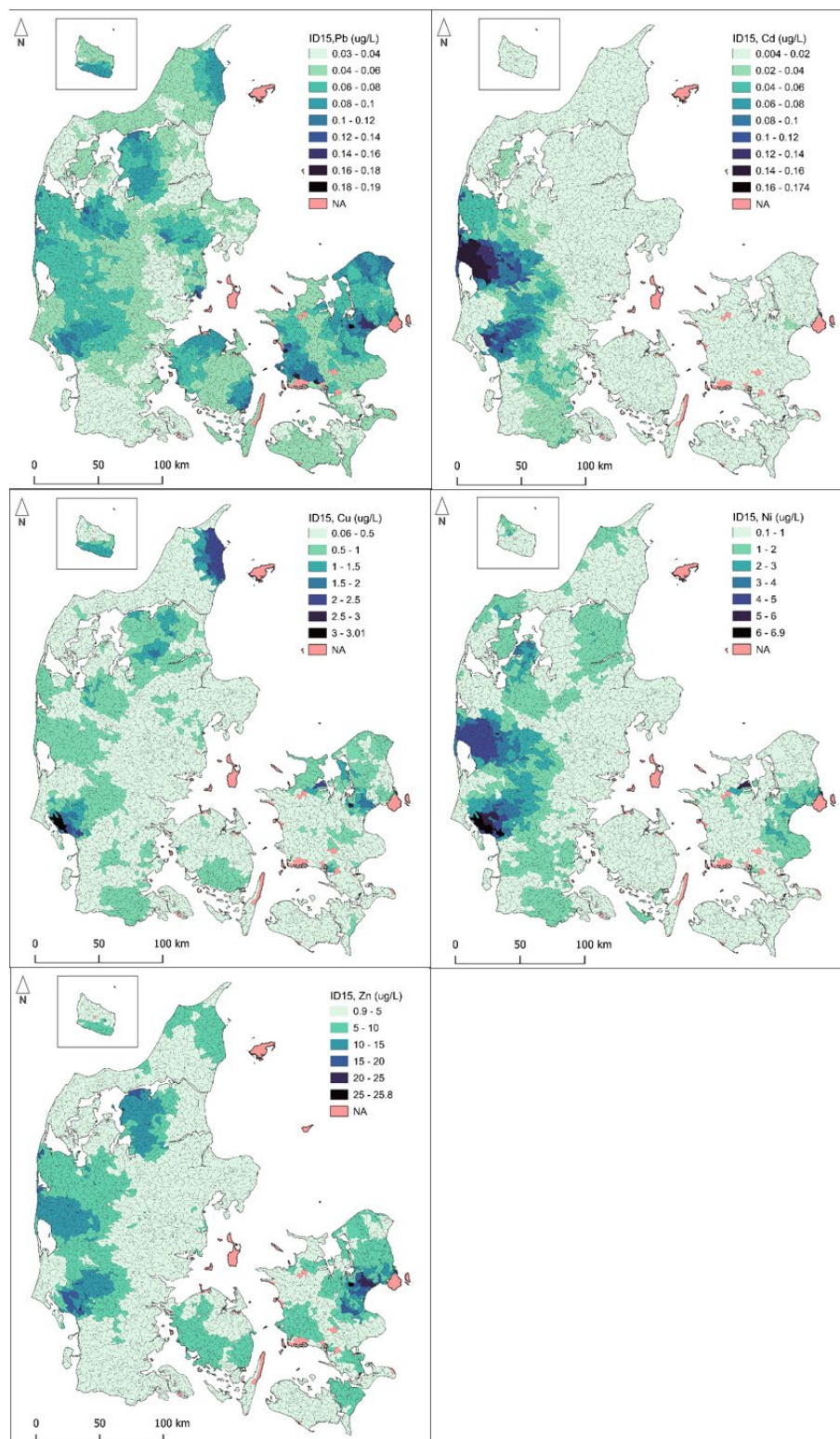
**Table B5.** Number of ID15 catchments by type for each trace element; in total there were 3351 ID15 catchments

Parameter	Estimated (n,%)	Modelled (n)	Mix (n)	NA (n)
<b>Cd</b>	854 (25.5%)	507 (15.1%)	1873 (55.9%)	117 (3.5%)
<b>Cu</b>	768 (22.9%)	592 (17.7%)	1874 (55.9%)	117 (3.5%)
<b>Ni</b>	1419 (42.3%)	141 (4.2%)	1674 (50.0%)	117 (3.5%)
<b>Zn</b>	869 (27.4%)	489 (14.6%)	1876 (56.0%)	117 (3.5%)
<b>Pb</b>	873 (26.1%)	467 (13.9%)	1894 (56.5%)	117 (3.5%)

**Figure B12.** Histograms of the metal concentrations at the ID15 level and the number of GVF at ID15 catchment



**Figure B13.** Groundwater contribution of metals (Cd, Cu, Ni, Pb) at the ID15 catchment



## References

Henriksen, H. J., Voutchkova, D., Troldborg, L., Ondracek, M., Schullehner, J., & Hansen, B. (2019). National Vandressource Model. Beregning af udnyttelsesgrader, afsækning og vandløbspåvirkning med DK model 2019. (Danmarks og Grønlands Geologiske Undersøgelse Rapport; Vol. 2019, No. 32). GEUS. <https://doi.org/10.22008/gpub/32632> and Data: <https://doi.org/10.22008/FK2/P88XCU>

Henriksen, H. J., Voutchkova, D., Ondracek, M. C., Troldborg, L., & Thorling, L. (2021). Konsolidering af kvantitativ tilstandsvurdering for danske grundvandsforekomster i potentiel ringe tilstand på basis af ekspertvurdering. Supplerende vurderinger af kvantitativ tilstand for 90 grundvandsforekomster med modelberegnet udnyttelsesgrad større end 30%. (Danmarks og Grønlands Geologiske Undersøgelse Rapport; Vol. 2021, No. 2). GEUS. <https://doi.org/10.22008/gpub/34562>

Mortensen, M. H., Ernstsén, V., Voutchkova, D., & Thorling, L. (2021). Dokumentationsrapport. Udvikling af metode til vurdering af grundvandsforekomsters kemiske tilstand for udvalgte uorganiske sporstoffer og salte. Udarbejdet af GEUS for Miljøstyrelsen. (Danmarks og Grønlands Geologiske Undersøgelse Rapport; Vol. 2021, No. 19). GEUS. <https://doi.org/10.22008/gpub/34579>

Troldborg, L. (2020). Afgrænsning af de danske grundvandsforekomster: Ny afgrænsning og delkarakterisering samt fagligt grundlag for udpegning af drikkevandsforekomster. (Danmarks og Grønlands Geologiske Undersøgelse Rapport; Bind 2020, Nr. 1). GEUS. <https://doi.org/10.22008/gpub/32641>

## VIDEREUDVIKLING OG KLARGØRING AF METALSTAT FOR LANDSDÆKKENDE BEREGNINGER

Denne rapport indgår i et større arbejde med udvikling af statistiske modeller til estimering af koncentrationer af miljøfarlige stoffer i overfladevand for Miljøstyrelsen ved at videreudvikle og anvendeliggøre en pilotudgave af en statistisk model (MetalStat) for forekomsten af 5 metaller som opløst koncentration i vandløbsvand. Rapporten tager således udgangspunkt i afrapportering af MetalStats principper givet i Sørensen et al., 2022 og 2023. Modellen er sat op, så metalkoncentrationen i vandløbet regnes for at være bestemt af det ID15 opland, som målestationen er placeret i, hvor et ID15 opland har en varierende udstrækning på typisk 10-15 km<sup>2</sup>. Den primære opgave for 2023 er at få pilotudgaven af MetalStat gjort brugbar med levering af resultater, der kan indgå i Miljøstyrelsens forvaltning