Statistisk analyse af perkolatdata

Metodik til stedsspecifik risikovurdering ved deponering af affald

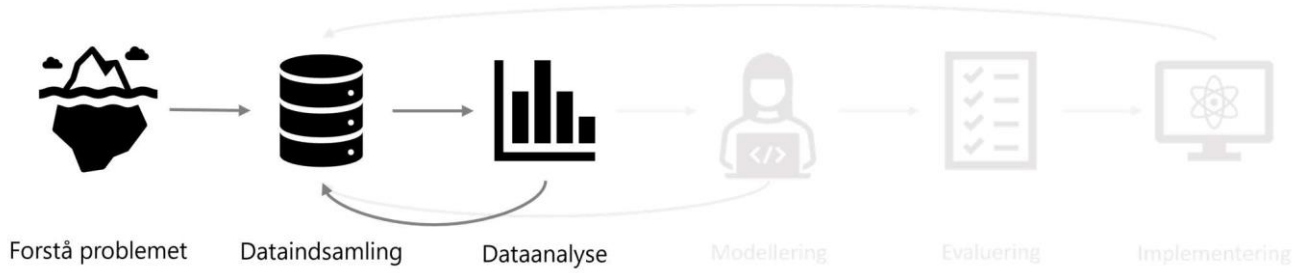
DepoNet

Dato: 10. december 2024

Indhold

1. Projektets omfang og resultat

Dette projekt er udarbejdet af NIRAS med sparring fra Danish Waste Solutions (DanWS). Projektet er udarbejdet for DepoNet med det primære formål at undersøge om man, ved brug af avanceret statistisk analyse, kan udnytte data på tværs af danske deponeringsanlæg. Dette notat omhandler punkterne 1-3 i projektbeskrivelsen, der ses illustreret i Figur 1.1 herunder.



Figur 1.1: Projektets fremgangsmåde

Fremgangsmåden i dette projekt er inddelt i tre trin: *Forstå problemet, Dataindsamling og Dataanalyse*. Projektet har fulgt en iterativ proces, hvor alle tre trin er besøgt af flere omgange. Arbejdet med data giver løbende ny indsigt, hvilket kan føre til behov for yderligere viden, nye data eller justering af analysens fokus.

Det har været afgørende at inddrage Ole Hjelmar fra DanWS og Lotte Fjeldsted fra NIRAS i arbejdet med at forstå både problemstillingen og de tilhørende data. Domæneekspertise er essentielt, når man indsamler, behandler og analyserer data, især i dette projekt, hvor dataindsamling og opbygning af en database udgjorde en væsentlig del af projektets samlede tidsramme. Den eksisterende datastruktur varierer markant på tværs af de forskellige deponeringsanlæg, hvorfor betydelige ressourcer gik til at strukturere data. Det udførte struktureringsarbejde skaber en væsentlig fordel for det videre arbejde på hvert enkelt anlæg og på tværs af anlæg. Med oprettelsen af en samlet deponidatabase er data nu flyttet fra spredte Excel-filer til en samlet og standardiseret database.

1.1 Resumé

Projektet udsprang fra en hypotese om, at det er muligt at sammenstille data fra danske deponeringsanlæg ved hjælp af avancerede statistiske metoder. Dette er værdifuldt, da det antages at en tidsserie på tværs af flere anlæg, betyder et mere robust estimat af efterbehandlingstiden og dermed et mere robust estimat af de økonomiske omkostninger relateret hertil.

Projektet vurderer, at data på tværs af anlæg kan udnyttes til at estimere stofkoncentrationer over tid og estimere, hvornår koncentrationerne vil falde under grænseværdien. Dette er muliggjort af, at data er blevet struktureret og samlet i en Deponi Database, som muliggør hurtige og standardiserede dataudtræk til analyser på tværs af deponier. Databasen udgør således en væsentligt fordel i arbejdet med at estimere kildestyrken. Derudover indikerer projektet, at nedbørsmængder kan anvendes som erstatning for perkolatmængder (jf. afsnit 3.3), hvilket er et betydeligt resultat, da datakvaliteten af perkolatmængder varierer mellem anlæg. Ved at bruge nedbørsmængden kan man inkludere flere deponeringsanlæg i analysen og dermed øge datagrundlaget for modellen. Projektet demonstrerer, at anvendelse af Fixed Effects (jf. afsnit 3.4) muliggør, at man kan sammenstille og udnytte data fra deponeringsanlæg med forskelle i f.eks. beliggenhed, størrelse og affaldssammensætning.

Projektets vurdering fremlægges dog med visse forbehold. Projektet adresserer ikke alle aspekter i en fuld analyse, da projektet har taget udgangspunkt i udvalgte scenarier. Projektet har afdækket et idealscenarie (best case scenarie), der sætter en øvre grænse for den datakvalitet, man kan forvente. Datakvaliteten af de øvrige tidsserier er ikke vurderet i dybden, hvilket medfører usikkerhed i forhold til deres indflydelse på analysen. En væsentlig udfordring er tidsseriernes længde, hvor de fleste parametre har kvartalsvis eller årlig datafrekvens. Selv ved en sammenlægning af data på tværs af anlæg, er datagrundlaget relativt begrænset, hvilket medfører en sensitiv model. Dette understreger behovet for løbende opdatering af databasen, så ny data kan styrke modellen og berige analysen på tværs af deponier. Oprettelsen af Deponi Databasen tydeliggjorde væsentlige forskelle i dataindsamlingsmetoderne på tværs af anlæggene (jf. afsnit 2). Det anbefales derfor, at der laves en fælles procedure for dataindsamling, der kan understøtte analysen på tværs.

Projektet fokuserer på sammenlægning af data fra lukkede enheder. Det indsamlede datasæt indeholder imidlertid et begrænset antal lukkede enheder. For at kunne estimere kildestyrken for åbne enheder ville det kræve en mere omfattende og dybdegående analyse samt en forbedring af kvaliteten af affaldsdata med hensyn til datafrekvens, typeangivelse og lokationsangivelse.

I de følgende afsnit gennemgås arbejdet trin for trin, så fremgangsmåden og resultaterne tydeliggøres.

2. Dataindsamling

Dataindsamlingen der muliggør en dataanalyse er delt ind i flere trin, som illustreret i Figur 2.1.

Figur 2.1: Overblik over processen med at indsamle, strukturere, rense og opbevare data fra de danske deponier

I projektets indledende fase udarbejdede DanWS en oversigt over det nødvendige data i prioriteret rækkefølge (se bilag 1). Denne oversigt blev delt med alle medlemmer i DepoNet, og NIRAS modtog herefter data fra de seks deponeringsanlæg, der er vist til venstre i Figur 2.1. Hvert af disse anlæg leverede primært data som en samling af forskellige Excel-filer, der krævede oprydning og systematisering. Dataene blev løbende gennemgået og kontrolleret med henblik på at foretage en foreløbig vurdering af datakvaliteten, som er sammenfattet i Tabel 2.1.

NIRAS og DanWS samarbejdede om udvælgelsen og kvalitetssikringen af de indsamlede data. Derudover bidrog DanWS med input og vurderinger vedrørende deponiernes omgivende forhold samt hvilke parametre, der skulle tages i betragtning.

	Reno Djurs	Deponi syd: Grindsted	Deponi Syd: Måde	Odense Renovation	AV Miljø	Argo: Køge	Argo: Hedeland	Argo: Audebo	Nordværk
Grundlæggende informationer									
Antal etaper	2	1			10				-
Antal enheder	23	2	2	3	23	2	1	5	-
Angivelse på kort	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja
Start- og slutdato	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Nej
Areal, bundmembran	Nej	Nej	Ja*	Ja	Nej	-	-	-	Nej
Areal, overflade	Nej	Nej	Nej	Nej	Nej	-	-	-	Nej
Gns. Fyldhøjde	Nej	Nej	Nej	Ja	Nej	-	-	-	Nej
Type overdækning	Nej	Nej	Nej	Ja	Nej	-	-	-	Nej
Stofkoncentrationer									
Antal måleparametre	146	34	72	15	41	-	-	-	115
Enhedsniveau (Ja/nej)	Ja	Ja*	Ja	Ja	Ja	Ja	Ja	Ja	Ja
Affald									
Datafrekvens	Årlig	Årlig	Årlig	Årlig	Årlig	Årlig	Årlig	Årlig	Ikke modtaget
Enhed	kg	m^3	m^3, tons	kg	kg	-	-	-	
Typeangivelse	Ja	Nej	Nej	Nej	Ja	-	Ja	Nej	
Enhedsniveau	Nej	Nej	Ja	Ja	Ja	-	Ja	Ja	
Perkolat									
Datafrekvens	Månedlig	Månedlig	Månedlig	Årlig / daglig	Årlig	-	Daglig	-	Ikke modtaget
Enhedsniveau (Ja/nej)	Ja	Ja*	Ja	Ja	Nej	Nej	Ja	Estimat	
Vejrdata									
Modtaget	Ja	Ja	Ja	Ja	Nej	Nej	Nej	Nej	Nej
Andre faktorer									
Recirkulering	Ja	Nej	Nej	Nej	Nej	Nej	Nej	Nej	Nej
Flowmålinger	Ja	Nej	Nej	Ja	Nej	-	Ja	Vingeflowmåler	-
Antal lukkede enheder i modtaget data	3	0	0	2	8	2	1	3	-
Andet					Lokation ved havet	1 af enhederne har ikke bundmembran.			Perkolatdata ligger i papirform.

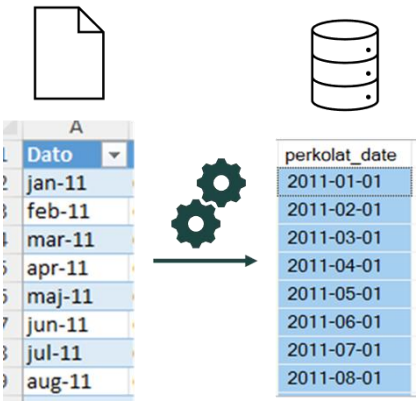
Tabel 2.1: Dataoverblik. Felter markeret med '-' indikerer, at informationen er ukendt.
** Angiver at data er modtaget for én enhed*

I Tabel 2.1 fremgår det, at der generelt mangler data om grundlæggende forhold som enhedernes størrelse og overdækning. Tabellen viser også, om data for stofkoncentrationer, affald og perkolat er angivet på enhedsniveau. Stofkoncentrationer er oplyst på enhedsniveau for alle anlæg, hvilket er nødvendigt for analysens anvendelse. Datafrekvensen for affald er generelt årlig og af varierende kvalitet. Perkolatmængder findes derimod typisk på enhedsniveau med en månedlig eller daglig frekvens. Perkolatmængderne varierer dog i kvalitet, da faktorer som placering og type af flowmåler påvirker målingerne. For eksempel er der på Audebo deponi installeret en vinge-flowmåler, som ikke kan registrere små vandmængder, hvilket reducerer datakvaliteten.

Data modtaget fra Reno Djurs, Deponi Syd, Odense Renovation og AV Miljø blev generelt vurderet som tilstrækkeligt omfattende. Derimod var det ikke muligt at indsamle nok data fra ARGO og Nordværk inden for projektets tidsramme. Dette skyldtes blandt andet, at en del af Nordværks data kun foreligger i papirform, og at ARGO's data på visse områder ikke var tilfredsstillende. Det blev derfor vurderet, at det ikke var muligt at behandle disse data inden for projektets tidsplan, og derfor er data fra disse deponeringsanlæg ikke inkluderet i databasen.

I projektoplægget nævnes det, at "Danske deponeringsanlæg er konstrueret eller drevet på lige så mange forskellige måder, som der er anlæg". Under dataindsamlingen blev det tydeligt, at det samme gør sig gældende for formateringen og strukturen af modelanlæggenes data. Det første skridt mod opsætningen af en database var derfor at ensrette de indsamlede data ved hjælp af Power Query. Med Power Query kan man definere en "opskrift" for databehandlingen af en given datafil, så samme metode kan anvendes på fremtidige data i samme format. For hver modtaget Excel-fil blev der opsat en specifik opskrift, som kan genbruges, hvis laboratoriet fastholder formatet for fremtidige analyseresultater. Dette vil muliggøre hurtig integration af ny data i databasen. Antallet af opskrifter er dog højt, og en strømlining af formaterne vil derfor øge effektiviteten.

Når data er ensrettet via en opskrift og har fået den ønskede struktur, importeres den til Python, hvor den endelige databehandling og upload til databasen udføres. Et eksempel på en behandlingsprocedure er standardisering af datoformatet, så alle datoer i databasen har samme format. Figur 2.2 illustrerer en dato, der er ændret fra formatet *jan-11* til *2011-01-01*. Dette enkle eksempel viser, hvordan data er blevet opryddet og strømlinet.



Figur 2.2: Standardisering af dato formater

2.1 Databasestruktur

Databasen er udarbejdet i databaseformatet SQLite, som er et opensource databaseformat, der er filbaseret, og derfor let at udveksle mellem forskellige parter. Databasen er struktureret som en relationel database med fakta-tabeller (fact) og dimensionstabeller (dim), der er relateret til hinanden via primary (PK) og foreign keys (FK).

Figur 2.3 illustrerer enhedsrelationsdiagram (ERD) for den opstillede Deponi Database. Et ERD viser alle tabellerne i databasen med angivelse af hvilke informationer, som hver tabel indeholder og en angivelse af relationerne mellem tabellerne.

Figur 2.3: Entity Relationship Diagram (ERD) for Deponi Databasen[1]

[1] NIRAS blev den 21/11/2024 gjort opmærksom på den gældende definition af enhed og celle i Deponeringsbekendtgørelsen. Denne definition er indarbejdet i rapportens indhold og vurderinger, men afspejles ikke i databasens navngivning af visse kolonner, hvor *unit* er anvendt som etape og *cell* som enhed. Ved opstart af en eventuel anden fase af projektet vil dette blive justeret, så det stemmer overens med Deponeringsbekendtgørelsens definition.

Faktatabellerne indeholder numeriske data, mens dimensionstabellerne indeholder kontekst og baggrundsinformation. I Figur 2.3 fremgår det, at databasen er opstillet med to dimensionstabeller, dimDeponi og dimEnhed. DimDeponi indeholder en række for hvert deponi med et unikt DeponiID, mens dimEnhed har en række for hver enhed. Kombinationen af deponi, etape og enhed udgør en unik kombination, som får tildelt et EnhedID. DeponiID og EnhedID fungerer som referencer i faktatabellerne. Der er indsamlet data om perkolatmængder, analysedata, vejrdata og affaldsdata, og for hver af disse er der oprettet en faktatabel med reference til enten DeponiID, EnhedID eller begge. Vejrdata er indsamlet per deponi og ikke på enhedsniveau, da enhederne normalt ligger inden for en radius, hvor de samme vejrforhold antages at gælde. Derfor er factWeather kun knyttet til DeponiID.

De øvrige faktatabeller er relateret til både DeponiID og EnhedID. Ideelt set burde perkolatmængder, analysedata og affaldsdata være tilgængelige på enhedsniveau, men da dette ikke er tilfældet for alle deponier, er det nødvendigt også at knytte DeponiID til tabellen.

Alle tabellerne er oprettet med et TimeStamp, som angiver, hvornår data blev lagt ind i databasen. Dette er en standardpraksis, når man opsætter databaser for at sikre mulighed for versionsstyring af data.

3. Dataanalyse

Med oprettelsen af Deponi Databasen er det muligt at lave hurtige og effektive dataudtræk til dataanalysen. Formålet med projektet er, som nævnt, at vurdere, om avancerede statistiske metoder kan anvendes til at estimere kildestyrken for et givent stof, dvs. stofkoncentrationen pr. tidsenhed. Under ideelle forhold haves data af høj kvalitet for bl.a. perkolatmængde, stofkoncentrationer, nedbør og affaldsmængder. Problemet opstår, da de fleste anlæg ikke har tilstrækkeligt lange tidsserier til, at en kildestyrkemodel kan udarbejdes uden betydelige antagelser. Endvidere er data for perkolatmængder af varierende kvalitet på tværs af de forskellige anlæg og enheder, hvorfor man ikke altid kan bruge dette data til at forudsige kildestyrken.

Dataanalysen i projektet fokuserer således på følgende to spørgsmål:

1. Kan nedbørsmængder anvendes som erstatning for perkolatmængder?

2. Kan data udnyttes på tværs af deponier til at estimere stofkoncentrationer over tid og forudsige, hvornår de falder under grænseværdien?

En dybdegående analyse af alle deponier og enheder ligger uden for projektets tidsramme, hvorfor analysen udføres på et begrænset datasæt. For at vurdere, om nedbør kan anvendes som erstatning for perkolatmængden, er det afgørende at have en enhed med ideelle eller næsten ideelle forhold. Det er nødvendigt at have en lukket enhed med historiske data af høj kvalitet for både perkolatmængder og nedbørsmængder. I samarbejde med DanWS og med styregruppens godkendelse blev enhed 1B ved Odense Deponi udvalgt. Idealscenariet består dermed af data fra enhed 1B i den lukkede periode fra år 2013 og frem med fokus på Ammonium-N (NH4-N).

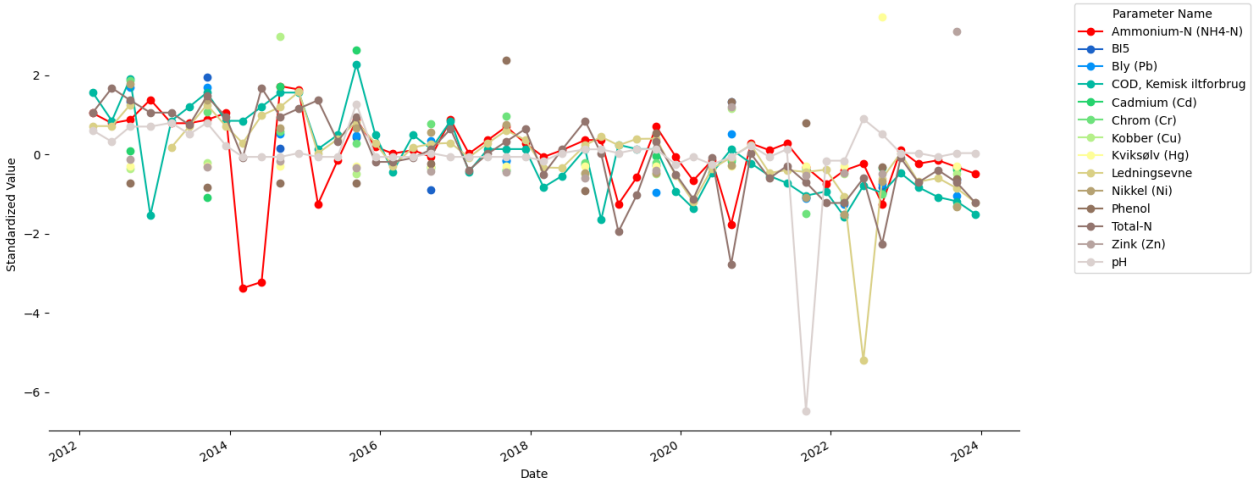
For at illustrere hvordan data kan udnyttes på tværs af anlæg, er det relevant at sammenlægge to af de bedste enheder og analysere effekten på modellens robusthed. Et udvidet scenarie opstilles således ved at tilføje data fra Odense Deponi, enhed 1A til idealscenariet for at illustrere effekten af et større, men mere varieret datasæt.

I denne dataanalyse er der foretaget en grundig analyse af idealscenariet samt det udvidede scenarie. En tilsvarende analyse bør udføres ved inddragelse af flere enheder, for at sikre det bedst mulige datagrundlag uden støj. Dette er et nødvendigt og vigtigt trin i databehandlingsprocessen, da et solidt datagrundlag er fundamentet for en pålidelig model.

3.1 Datakvalitet

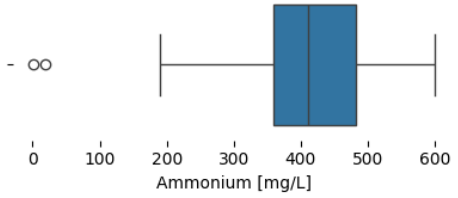
Indledningsvist undersøges datakvaliteten for idealscenariet, da det vil give en indikation af den maksimale datakvalitet, der kan forventes. Datasættet for enhed 1B dækker perioden fra nedlukningen i 2012 og frem. I denne periode er målt for 14 forskellige parametre i forbindelse med kontrol af enhedens perkolat. Parametrene analyseres typisk fire gange årligt, undtagen metaller, som analyseres én gang årligt. Data viser, at der kun er få manglende værdier for enkelte parametre, og for Ammonium-N (NH4-N) er der ingen manglende værdier. Figur 3.1 præsenterer de standardiserede analyseresultater, hvilket gør det muligt at sammenligne parametre på tværs af skalaer og enheder.

Figuren viser, at flere parametre har værdier, der skiller sig markant ud fra alle andre observation (outliers), som fremgår af markante udsving i tidsserierne. Ammonium-N (NH4-N), markeret med rød, viser især et markant udsving omkring år 2014-2015 og da den videre dataanalyse fokuserer på denne parameter, undersøges de nærmere. Den følgende analyse af outliers illustrerer, hvordan data behandles inden modellering.



Figur 3.1: Standardiserede koncentrationer af måleparametre for Odense Deponi, enhed 1B i den lukkede periode

Et boksplot er et velkendt værktøj til at identificere outliers. Boksplottet i Figur 3.2 viser data for Ammonium-N (NH4-N) og to markante outliers, indikeret med cirkler, ses omkring værdierne 0-50 mg/L. Disse to værdier antages at være et resultat af fejlbehæftet data og erstattes med lineært interpolerede værdier.



Figur 3.2: Boksplot for koncentrationen af Ammonium-N (NH4-N) ved Odense Deponi, enhed 1B i den lukkede periode

En tilsvarende analyse af outliers udføres for de resterende datapunkter for enheden. Efter enhedens nedlukning i 2012 findes der flowmålinger af perkolat med start fra den 28-12-2011 og daglig datafrekvens uden manglende værdier. En outlier-analyse viser dog en betydelig outlier på 48176 m^3, hvilket må antages at skylde en fejl i data, da den gennemsnitlige værdi uden denne ligger på 17,6 m^3. Denne outlier erstattes derfor af en lineær interpolation. Ligeledes analyseres nedbørsdata leveret af Odense Renovation, der indeholder daglige målinger uden manglende værdier. En analyse af outliers i nedbørsdata viser en lang hale af outliers, som gør dem plausible, og de fjernes derfor ikke fra datasættet.

En mere dybdegående analyse kan potentielt afdække hvad de identificerede outliers er et resultat af. Dette er dog udenfor projektets tidsramme og relevans, da det ikke antages at have en effekt på den overordnede analyse i projektet.

3.2 Feature Engineering

Når kvaliteten af enhedens data er sikret og valideret, bør en transformation af data overvejes. Feature engineering er en proces, hvor man skaber eller transformerer data for at fremhæve mønstre, som kan forbedre modellens præcision. Det kan betragtes som en måde at "klargøre" data på, så modellen lettere kan forstå og forudsige de ønskede resultater. Features er som modellens byggesten, og kvaliteten af disse er afgørende for modellens evne til at opdage mønstre og lave præcise forudsigelser.

De målte værdier for perkolat og nedbør på analysedagen kan ikke bruges, da de ikke afspejler, hvordan vand har bevæget sig gennem enheden over tid. Koncentrationerne bliver typisk analyseret kvartalsvis eller årligt, så det er mere meningsfuldt at se på mængden af vand, der er sivet igennem enheden over en længere periode, som for eksempel de seneste seks måneder. Således er data blevet aggregeret ved at beregne gennemsnit for både perkolatmængden og nedbørsmængden over den seneste halvårsperiode på hver analysedato.

De nye aggregerede variable hjælper modellen med at fange tendenser, som de enkeltstående daglige værdier ikke kan fange. En drøftelse i styregruppen afdækkede, at der kan gå op til et år eller mere før nedbør er sivet igennem de store enheder. I fremtidige analyser kan det således være relevant at undersøge, om der er et andet aggregeringsniveau, der bedre understøtter modellens præcision. Ved fremtidige analyser kunne man desuden inkludere andre statistiske variable såsom standardafvigelse, minimum og maksimum for den forudgående periode. Eksempelvis kan et skybrud i perioden skabe udsving, der ikke afspejles i gennemsnittet, men som fremgår af maksimumværdien og kan forklare et pludseligt spring i koncentrationerne.

3.3 Modellering

Efter en grundig databehandling påbegyndes modelleringen. I dette projekt er fokus på at afprøve metoden og redegøre for hvordan en Machine Learning model kan opstilles. For at illustrere dette, anvendes en af de mere simple Machine Learning regressions modeller, kaldet Lasso. Denne model giver et enkelt og fortolkbart resultat, samtidig med at den kun fokuserer på det vigtigste i datagrundlaget. Dette er med til at gøre modellen robust ved små datasæt. De følgende to afsnit, beskriver resultatet af en Lasso-model trænet på hhv. én enhed og to enheder, for at illustrere hvordan data fra flere enheder kan sammenlægges.

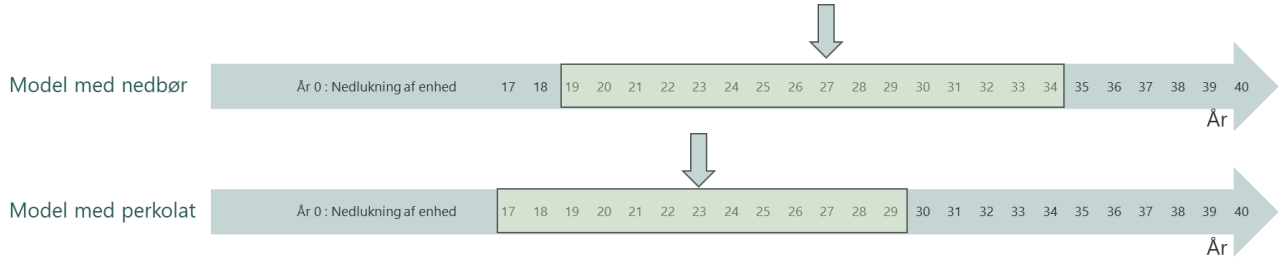
3.3.1 Model med én enhed

Den første model opstilles med data fra idealscenariet (Odense Deponi, enhed 1B) for at have et sammenligningsgrundlag for den udvidede model med to enheder. Endvidere trænes modellen på data for hhv. perkolatmængden og nedbørsmængden, for at vurdere betydningen af hver af disse faktorer i kildestyrkemodellen.

I denne analyse bruges tiden *t* (målt i antal måneder siden enhedens nedlukning) samt de gennemsnitlige værdier for nedbørsmængden og perkolatmængden. I en fremtidig model kunne man inkludere flere forklarende variable,

men i denne analyse fokuseres der udelukkende på effekten af en model trænet på nedbørsdata versus en model trænet på perkolatdata. Figur 3.3 viser resultaterne for hver model, hvor den estimerede efterbehandlingstid er markeret med en pil, og usikkerhedsintervaller med en lysegrøn boks. Efterbehandlingstiden (antal år siden nedlukning) er defineret ud fra hvornår stof-koncentrationen når grænseværdien.

Modellen baseret på nedbørsmængden estimerer en efterbehandlingstid for Ammonium-N (NH4-N) på 27 år, mens modellen baseret på perkolatmængden estimerer 23 år. Usikkerhedsintervallerne bør også vurderes, da modellen med nedbør har en variation på 15 år (19-34 år), mens modellen med perkolat har en mindre variation på 12 år (17-29 år). Dette indikerer en let fordel ved perkolatmodellen, dog stadig med betydelig usikkerhed. Samlet set er fordelen ved perkolatmodellen ikke stor nok til at udelukke nedbørsmængder som erstatning, selvom det medfører en lidt højere usikkerhed.



Figur 3.3: Estimeret efterbehandlingstid af Ammonium-N (NH4-N) og usikkerhedsinterval for en model trænet på hhv. nedbørsmængder og perkolatmængder. Data kommer fra Odense Deponi, enhed 1B.

Datagrundlaget er generelt begrænset til korte tidsserier, især for parametre med årlige målinger. Derfor er det relevant at undersøge, om datakombinationer på tværs af deponeringsanlæg kan øge præcisionen i estimaterne og reducere usikkerhedsintervallerne. En mulig tilgang er at øge datasættets størrelse ved at sammenlægge data fra flere enheder, forudsat at de samme variable af høj kvalitet kan indsamles. Hvis data for perkolatmængder ikke kan indhentes for alle enheder, må man enten helt udelade den pågældende enhed eller finde alternative variable, som er tilgængelige for alle enheder. Her kan nedbørsdata være et oplagt alternativ til perkolatmængder, da de kan indhentes fra kilder som DMI, hvis anlægget ikke selv har udført disse målinger.

3.3.2 Udvidet model med to enheder

I det udvidede scenarie kombineres data fra to enheder, hhv. enhed 1B (idealscenariet) og enhed 1A fra Odense Deponi. Formålet med denne model er at illustrere, hvordan man kan kombinere data på tværs af deponier/enheder for at opnå et mere præcist estimat af efterbehandlingstiden. Modellen benytter nedbørsmængder som beskrivende variabel.

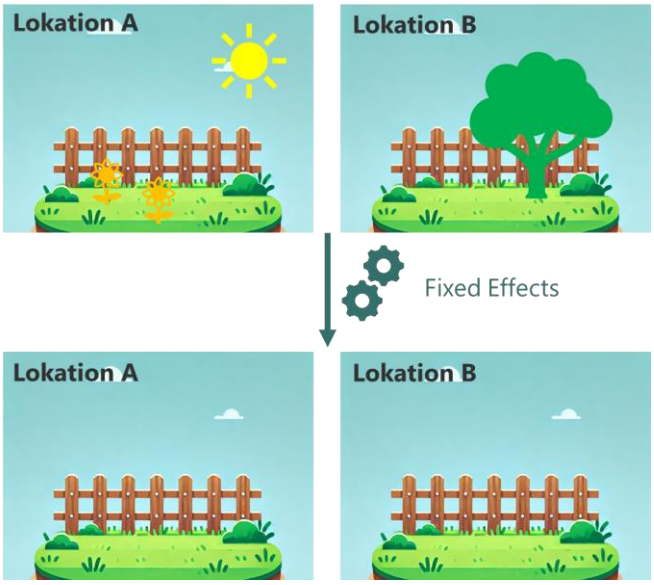
Når man arbejder med data fra forskellige kilder - her to enheder med forskellige karakteristika - kan en dummyvariabel anvendes til at tage højde for forskellene. En dummyvariabel er en binær variabel, der fungerer som en markør i datasættet, og som angiver, hvilken enhed en given observation kommer fra. I dette tilfælde tildeles værdien 1 til observationer fra den ene enhed og værdien 0 til observationer fra den anden enhed. På den måde kan modellen skelne mellem de to enheders data og tilpasse sin beregning, så de unikke forhold for hver enhed medtages. Ved sammenlægning af flere end to enheder, kan metoden udvides, ved at inkludere flere dummyvariable i modellen. Anvendelsen af dummy-variable kan benyttes, fordi der er tale om enhedsspecifikke effekter (Fixed Effects), der er tidsinvariante (konstante).

Ved at udnytte Fixed Effects kan modellen skelne mellem "inden for"-variation og "imellem"-variation. Inden for-variation dækker ændringer over tid inden for hver enhed og fokuserer på faktorer som nedbør og

perkolatmængde. Imellem-variation refererer til de tidsinvariante forskelle mellem enhederne, som f.eks. affaldsblanding og volumen. Ved at fokusere på inden for-variation og ignorere imellem-variation kan Fixed Effects-modellen identificere de faktiske sammenhænge mellem variablerne (som kildestyrken og nedbør) på tværs af tid i den enkelte enhed.

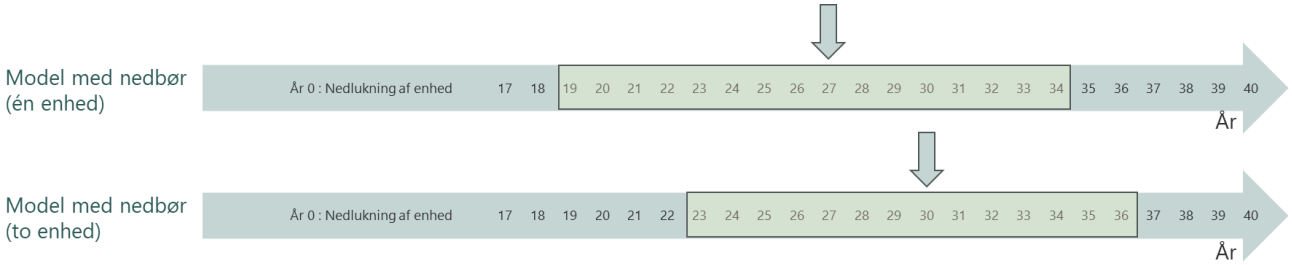
Figur 3.4 illustrerer anvendelsen af Fixed Effects-metoden med et forenklet, hypotetisk eksempel. I figuren vises to græsplæner (Lokation A og B), der er placeret to forskellige steder med varierende beplantning, der antages at være konstante, hvilket medfører forskellig eksponering for sollys. Den ene græsplæne får meget sol, mens den anden ligger i skygge på grund af et stort træ. Disse lokationsspecifikke forhold kan påvirke græssets vækst, uafhængigt af gødningens effekt. For at kunne isolere og analysere den rene effekt af gødning fjernes disse forskelle statistisk ved hjælp af Fixed Effects-metoden. Metoden justerer således for tidsinvariante effekter, hvilket muliggør en sammenligning af gødningens virkning uden indflydelse fra lokationernes unikke karakteristika.

På samme måde har enhederne 1A og 1B enhedsspecifikke effekter (Fixed Effects), som kan påvirke efterbehandlingstiden. Eksempelvis har enhed 1B et større volumen end enhed 1A, hvilket kan betyde en længere efterbehandlingstid. Da begge enheder er lukkede, forbliver deres størrelse og placering uændrede, hvorfor det er tidsinvariante variable. Ved at introducere dummyvariable kontrolleres for disse forskelle. Faktorer, der ændrer sig over tid og kan påvirke kildestyrken, bør stadig inkluderes i modellen. Af den årsag medtages stadig stedsspecifikke nedbørsmængder samt en enhedsspecifik tidsvariabel *t*, der beskriver tiden siden hver enheds nedlukning.



Figur 3.4: Illustration af hvordan man vha. Fixed Effects metoden kan "fjerne" konstante forskelle i karakteristika mellem to lokationer.

Figur 3.5 viser resultaterne for den udvidede model sammenlignet med modellen baseret på én enhed. Modellen baseret på data fra to enheder estimerer en efterbehandlingstid for på Ammonium-N (NH4-N) på 30 år med et usikkerhedsinterval på 13 år (23-36 år). Sammenlignet med modellen trænet på data fra én enhed, der havde en variation på 15 år, er modellen blevet mere præcis. Dette indikerer, at et øget datagrundlag gør modellen mere robust og pålidelig. Dette skyldes, at modellen kan lære af flere observationer og dermed bedre kan opfange generelle tendenser.



Figur 3.5 Estimeret efterbehandlingstid af Ammonium-N (NH4-N) og usikkerhedsinterval for en model trænet på nedbørsmængder og med data fra hhv. én enhed (Odense Deponi, enhed 1B) og to enheder (Odense Deponi, enhed 1A og 1B)

Det vurderes, at data fra flere enheder kan kombineres i samme model, hvilket muliggør en bedre udnyttelse af data på tværs af enheder. Dette er muligt på baggrund af teorien om Fixed Effects og brugen af dummyvariable. Analysen indikerer en positiv effekt af, at sammenlægge data fra flere enheder. Det er dog et stærkt stiliseret eksempel, der kun inkluderer to enheder og få variable. Modellen er derfor sensitiv og usikkerhedsintervallet er fortsat stort, hvilket primært skyldes den begrænsede datamængde. Da hver enhed har korte tidsserier, kræves det at sammenlægge data fra mange enheder for at opnå en robust og pålidelig model. Dette kan være udfordrende på grund af antallet af lukkede enheder i det indsamlede datasæt samt den tilhørende datakvalitet.

4. Anbefalinger

Baseret på resultaterne af projektet fremlægger NIRAS en række anbefalinger til den videre proces. Anbefalingerne er inddelt i generelle anbefalinger samt anbefalinger til den videre proces.

4.1 Generelle anbefalinger

De generelle anbefalinger omfatter forbedringer til af dataindsamling og datastruktur for DepoNets medlemmer.

Estimaterne af efterbehandlingstiden er begrænset af kvaliteten af datagrundlaget, især på grund af korte tidsserier. Det anbefales derfor, at øge datamængden kontinuerligt, hvilket kræver en systematisk tilgang til dataindsamling på tværs af deponier. NIRAS anbefaler, at DepoNet fremhæver betydningen af høj datakvalitet og tydeliggør de potentielle gevinster for de danske deponier ved en standardisering på tværs. Dette kan øge motivationen for en ensartet databehandling og styrke det samlede datagrundlag, hvilket vil forbedre betingelserne for fremtidige analyser.

En systematisk dataindsamling, kan danne grundlag for etablering af et Power BI-dashboard til løbende analyse og overvågning. Dette vil understøtte kvalitetssikringen af data samt driften af deponier, ved at muliggøre alarmer, som aktiveres, hvis kritiske grænseværdier nærmes eller hvis der opstår afvigende værdier, som kræver undersøgelse. På denne måde kan markante outliers identificeres og håndteres allerede i dataindsamlingsfasen.

4.2 Anbefalinger til den videre proces

I projektoplægget lægges der op til potentiel gennemførelse af punkt 4-6 der dækker: *Modellering, Evaluering og Implementering af model.*

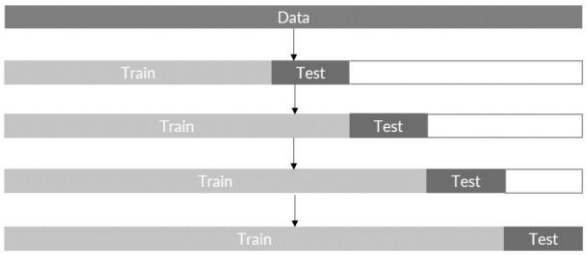
Under punkt *4. Modellering*, planlægges opsætning og træning af en AI model. NIRAS ser et potentiale for videre arbejde med modellering af det indsamlede data, der findes i Deponi Databasen. Analysen i dette projekt er baseret på et begrænset datasæt, hvilket giver høj sensitivitet, men den indikerer en positivt effekt ved at opsætte en model på tværs af deponeringsanlæg. Det vurderes relevant at eksperimentere med inddragelsen af flere

lukkede enheder, samt mere avancerede Machine Learning modeller. Dette arbejde afhænger dog af datakvaliteten for de øvrige lukkede enheder, og det anbefales at indsamle analyseresultater fra flere lukkede enheder, hvis muligt. For at styrke datagrundlaget yderligere anbefales det også, at undersøge potentialet i at inddrage den åbne periode af enhederne. I denne periode varierer affaldsmængden, hvorfor den ikke er tidsinvariant og dermed ikke dækkes af Fixed Effects. For at udnytte den åbne periode er det nødvendigt at inddrage affaldsdata i modellen. Kvaliteten af affaldsdata kan dog blive en udfordring, da data typisk er årligt aggregeret og i nogle tilfælde mangler specifikationer for lokation og type. Alligevel anbefales det, at mulighederne undersøges nærmere.

Den strukturerede data kan også udforskes gennem Machine Learning for at afdække nye, potentielt uventede indsigter om danske deponier. For eksempel kan perioder med høj nedbørsmængde være forbundet med en øget udvaskningsgrad. Ved at analysere sæsonudsving kan sådanne tendenser integreres i modellen, hvilket potentielt kan styrke modellens robusthed og bidrage til en bedre forståelse af f.eks. fordelene ved recirkulering.

I et videre arbejde anbefales det at sammenligne resultaterne fra Machine Learning modeller med kildestyrkemodellen, for at vurdere om den øgede kompleksitet i Machine Learning modellerne kan forsvares af forbedret performance. Dette falder under punkt *5. Evaluering* i projektoplægget, hvor en AI-model testes og evalueres.

For at evaluere Machine Learning modellerne bør Time Series Cross Validation bruges (se Figur 3.3). Metoden opdeler data i trænings- og testsegmenter, hvor testdata altid ligger senere end træningsdata. Evalueringen gentages flere gange med skiftende tidsvinduer, hvor trænings- og testperioder løbende rykkes fremad i tid. Denne fremgangsmåde sikrer en robust vurdering af modellens evne til at forudsige fremtidige hændelser baseret på historiske mønstre, og reducerer risikoen for overtilpasning til specifikke datasegmenter.



Figur 4.1: Time Series Cross Validation

For punkt *6. Implementering af model,* ønskes en løbende registrering af data. NIRAS anbefaler at etablere en struktur for løbende opdatering af databasen. Dette kunne eksempelvis være en fælles rapporteringsplatform, hvor deponeringsanlæg løbende indsender data i et fastlagt format, som derefter lagres i databasen og anvendes til rapportering. Da de fleste parametre rapporteres kvartalsvis, kunne en kvartalsvis dataindsamling styrke datagrundlaget over tid og give mulighed for løbende justering af den estimerede efterbehandlingstid.

På nuværende tidspunkt kan NIRAS kan ikke fastslå, om projektets anden fase vil resultere i en AI-model, der kan implementeres og driftes kontinuerligt, da projektets resultater ikke entydigt kan bekræfte robustheden af en sådan model på det indsamlede data. NIRAS anbefaler dog, at der fortsat er fokus på forbedring af estimaterne for efterbehandlingstiden, da det vil understøtte økonomisk planlægning ved at sikre, at omkostningerne til behandling dækkes af deponeringsafgifterne. Med de historiske data tilgængelige vurderes det oplagt at fortsætte en sådan analyse.