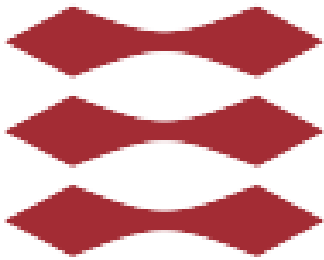


DTU



Notat om V1/V2 rådata

Jordforureningers påvirkning af overfladevand

Oliver B. Lund

29. Juni 2024

DTU Miljø

Danmarks Tekniske Universitet

Indholdsfortegnelse

1. [Introduktion](#)
2. [Datagrundlag](#)
 - [Shape-filer](#)
 - [CSV-filer](#)
 - [Hvordan laves CSV-filerne?](#)
3. [Risikovurdering](#)
 - [Trin 1: Optælling af Grundvandsforekomster](#)
 - [Trin 2: Grundvandsforekomster med Vandløbskontakt](#)
 - [Trin 3: V1/V2-lokaliteter med Aktive Forureninger i GVFK med Vandløbskontakt](#)
 - [Trin 4: Afstandsanalyse til Vandløb](#)
 - [Trin 5: Tærskel-vurdering og Kategorisering](#)
4. [Tilstandsvurdering](#)
5. [Samlet Overblik](#)
6. [Konklusion](#)

Introduktion

Projektbaggrund

Denne metodebeskrivelse dokumenterer **risikovurderingsdelen** af projektet "Beslutningstræ for grundvandsforekomsters påvirkning af overfladevand" - et samarbejde mellem DTU Miljø, GEUS og Miljøstyrelsen. Projektet udvikler en systematisk metodik til vurdering af, hvorvidt forurenede lokaliteter i grundvandsforekomster udgør en risiko for påvirkning af overfladevand.

Formål og Vision

Projektets overordnede mål er at etablere et **automatiseret beslutningstræ** til risikovurdering og tilstandsvurdering af forureningsforekomster i grundvandsforekomster og deres potentielle påvirkning af vandløb. Metoden skal kunne:

- **Screene alle danske grundvandsforekomster** systematisk for risiko
- **Identificere højrisiko-lokaliteter** der kræver nærmere undersøgelse
- **Automatiseres** så den kan håndtere landsdækkende data effektivt
- **Integreres med eksisterende modeller** (DK-model, DK-jord data)

Metodisk Tilgang

Tilgangen følger de etablerede principper for screening af jordforurening mod overfladevand fra Miljøstyrelsen, men anvender en struktureret beslutningstræ-approach med to hovedfaser:

Risikovurdering (trin 1-5):

1. **Baseline-etablering:** Optælling af det totale antal grundvandsforekomster (GVFK) i Danmark
2. **Kontakt-identifikation:** Identifikation af GVFK med kontakt til vandløbssegmenter
3. **Kildelokalisation:** Identifikation af GVFK der indeholder V1/V2-lokaliteter med aktive forureninger
4. **Afstandsanalyse:** Beregning af afstande fra V1/V2-lokaliteter til vandløbssegmenter inden for samme GVFK
5. **Tærskel-vurdering:** Kategorisering af lokaliteter baseret på afstandstærskler og stofspecifikke spredningsafstande

Tilstandsvurdering (**fremtidigt arbejde**):

- Kvantitativ fluxberegning og koncentrationsvurdering i vandløb
- Sammenligning med miljøkvalitetskrav
- Prioritering af indsatsområder

Denne Rapports Fokus

Nærværende metodebeskrivelse dokumenterer **den komplette risikovurdering (trin 1-5)** af beslutningstræet. Dette udgør den systematiske identifikation og karakterisering af alle potentielt problematiske lokaliteter baseret på afstand og forureningstyper.

Tilstandsvurderingen - den kvantitative vurdering af faktiske koncentrationer og overskridelser i vandløb - vil blive gennemført som næste projektfase efter finalisering af risikovurderingsmetodikken.

Metoden præsenteret her identificerer **2.013 højrisiko-lokaliteter** gennem stofspecifik vurdering (ud af 35.728 analyserede lokaliteter), som danner grundlag for den kommende tilstandsvurdering og prioritering af miljøindsats.

Datagrundlag

Analysen er baseret på følgende datafiler:

Shape-filer

- **VP3Genbesøg_grundvand_geometri.shp**: Grundvandsforekomster (GVFK) - 2.043 unikke forekomster
- **Rivers_gvf_rev20230825_kontakt.shp**: Vandløbsstrækninger med tilknyttet GVFK og kontaktflag (14.454 segmenter, 7.496 med kontakt)
- **V1FLADER.shp**: V1-lokaliteter som polygoner (28.717 polygoner, 23.209 unikke lokaliteter)
- **V2FLADER.shp**: V2-lokaliteter som polygoner (33.040 polygoner, 21.269 unikke lokaliteter)

CSV-filer

Følgende CSV-filer er genereret via **V1_V2.py** scriptet og "Fremgangsmåde til klassifikationer af forurenede grunde.docx" notatet (lavet af: Luc Taliesin Eisenbrückner, september 2024), som behandler og kombinerer lokalitetsdata med grundvandsforekomster:

- **v1_gvfk_forurening.csv**: V1-lokaliteter med GVFK-tilknytning, forureningsdata og brancheoplysninger (84,601 rækker, 23,209 unikke lokaliteter)
- **v2_gvfk_forurening.csv**: V2-lokaliteter med GVFK-tilknytning, forureningsdata og brancheoplysninger (134,636 rækker, 21,269 unikke lokaliteter)

Hvordan laves CSV-filerne?

- **DK-jord udtræk (27-09-2024)**: Fra Danmarks Miljøportal med .shp filer (alle V1 og V2 kortlagte grunde) og .csv filer med lokation, forurening, branche, aktivitet og forureningsstatus
- **Geometri-forbehandling**: ArcGIS Dissolve værktøj anvendt på V1 og V2 .shp filer med Lokalitetsnummer som dissolve field og "create multipart feature" aktiveret, så hver unik lokalitet blev til én multipart feature i stedet for opdelte polygoner
- **Kodeliste-join**: Simpelt join mellem forskellige koder fra DK-jord data

Kobling til grundvandsforekomster (ArcGIS spatial analyse)

- **Overlapsanalyse**: Spatial join mellem de grundvandstruende V1 og V2 lokaliteter og .shp fil med 2,050 grundvandsforekomster (VP3)
- **Join-operation**: "One to many" join med "keep all target features" aktiveret og match option sat til "intersect"
- **Resultat**: Hver V1/V2 lokalitet får tilknyttet de grundvandsforekomster, den geografisk overlapper med

Databehandlingsproces (**V1_V2.py**)

1. Indlæsning af `dkjord-View_Lokaliteter` med lokation, forurening, branche og aktivitetsdata
2. Ekspansion af stoffdata (opdeling af `Lokalitetensstoffer` ved semikolon til separate rækker)
3. Fjernelse af dubletter baseret på alle kolonner
4. Join med de ArcGIS-forbehandlede V1/V2 GVFK-data på lokalitetsnummer
5. Fjernelse af GIS-relaterede kolonner og oprydning af datasæt

Overlap mellem datasæt

3,608 lokaliteter findes i både V1 og V2 data.

Vigtige kolonner til risikovurdering

- `Lokalitetsbranche`: Branche/industri-information
- `Lokalitetsaktivitet`: Aktivitetstype
- `Lokalitetsstoffer`: Forureningsstoffer (kun lokaliteter med data medtages)
- `Navn`: GVFK tilknytning fra ArcGIS spatial join

Risikovurdering

Trin 1: Optælling af Grundvandsforekomster

Formål: Etablere baseline for det totale antal unikke grundvandsforekomster (GVFK) i Danmark til sammenligning med filtrerede undersæt.

Input Data (fra Datagrundlag):

- **Fil:** `VP3Genbesøg_grundvand_geometri.shp`
- **Anvendte kolonner:**
 - `Navn`: Unik tekstidentifikator for hver grundvandsforekomst (primær nøgle)
- **Datatype:** GeoDataFrame med GVFK-polygoner

Proceslogik (`step1_all_gvfk.py`):

1. **Indlæsning:** Læser shapefil med `geopandas.read_file()`
2. **Validering:** Kontrollerer eksistens af `Navn`-kolonne
3. **Optælling:** Beregner antal unikke værdier med `Navn.nunique()`
4. **Lagring:** Gemmer hele GeoDataFrame i hukommelsen til videre brug

Output:

- **Ingen filer gemt** (kun i hukommelsen)
- **Returnerer:** (GeoDataFrame, antal_unikke_GVFK)

Aktuelle Resultater:

- **2.043 unikke grundvandsforekomster** identificeret
- Data videreføres til Trin 2 for filtrering

Trin 2: Grundvandsforekomster med Vandløbskontakt

Formål: Identificere det kritiske undersæt af grundvandsforekomster hvor grundvand-overfladevand interaktion forekommer, hvilket er afgørende for forureningsspredning til vandløb.

Input Data (fra Datagrundlag):

- **Fil 1:** `Rivers_gvf_rev20230825_kontakt.shp` (14.454 vandløbssegmenter)
 - **Anvendte kolonner:**
 - `GVForekom`: GVFK-navn tilknyttet hvert vandløbssegment
 - `Kontakt`: Numerisk flag (1 = har kontakt, 0 = ingen kontakt)
 - **Datatype:** GeoDataFrame med vandløbslinjer
- **Fil 2:** `VP3Genbesøg_grundvand_geometri.shp` (genbrugt fra Trin 1)
 - **Anvendte kolonner:**
 - `Navn`: GVFK-identifikator for matching med vandløbsdata
 - **Datatype:** GeoDataFrame med GVFK-polygoner

Proceslogik (`step2_river_contact.py`):

1. Vandløbsfiltrering:

- Indlæser vandløbsdata med `geopandas.read_file()`
- Filtrerer til kun segmenter hvor `Kontakt == 1` (7.496 af 14.454)
- **Årsag:** Kun segmenter med aktuel grund-/overfladevand interaktion er relevante

2. GVFK-ekstraktion:

- Udtræk unikke værdier fra `GVForekom`-kolonnen
- Fjernelse af None-værdier og ikke-teksttyper
- Oprettelse af liste med 593 GVFK-navne

3. Geometri-kobling:

- Indlæser GVFK-geometri fra Trin 1
- Filtrerer hvor `Navn` findes i vandløbskontakt-listen
- **Årsag:** Bevar kun GVFK-geometrier med dokumenteret vandløbskontakt

4. Output-lagring:

- Gemmer filtrerede GVFK-geometrier til `step2_river_gvfk.shp`

Output:

- **Fil:** `step2_river_gvfk.shp` (588 GVFK-geometrier)
- **Returnerer:** (liste_med_593_GVFK_navne, antal_unikke_GVFK, GeoDataFrame)

Aktuelle Resultater:

- **593 GVFK har kontakt med vandløb** (29,0% af alle GVFK)
- **588 GVFK-geometrier gemt** med vandløbskontakt
- **Forskel mellem 593 og 588 skyldes:**

- Fejl i navne mellem `VP3Genbesøg_grundvand_geometri.shp` filen og `Rivers_gvf_rev20230825_kontakt.shp`
- 5 GVFK-navne i vandløbsdata findes ikke i geometrifilen
- **Videre dataflow:** Dette undersæt danner grundlag for analyse af V1/V2-lokaliteter

Trin 3: V1/V2-lokaliteter med Aktive Forureninger i GVFK med Vandløbskontakt

Formål: Identificere V1/V2-lokaliteter med aktive forureninger i grundvandsforekomster med vandløbskontakt. Bevare en-til-mange lokalitet-GVFK relationer som er kritiske for korrekte afstandsberegninger i Trin 4.

Eksempel på en-til-mange relation: Lokalitet "12345" kan overlape flere GVFK-polygoner ("GVFK_A" og "GVFK_B"), hvilket resulterer i to kombinationer: (12345, GVFK_A) og (12345, GVFK_B). Hver kombination kræver separate afstandsberegninger.

Input Data (fra Datagrundlag):

1. CSV-filer (pre-processeret via ArcGIS spatial join):

- `v1_gvfk_forurening.csv` (84.601 rækker)
 - **Anvendte kolonner:**
 - `Lokalitetsnr`: Lokalitetsidentifikator
 - `Navn`: GVFK-navn (fra ArcGIS spatial join)
 - `Lokalitetensstoffer`: Forureningsstoffer (kritisk filterkolonne)
 - `Lokalitetensbranche`, `Lokalitetensaktivitet`: Metadata for Trin 5
 - **Datatype:** DataFrame med lokalitet-GVFK relationer
- `v2_gvfk_forurening.csv` (134.636 rækker)
 - **Samme kolonnestruktur som V1**

2. Shapefiler (geometrisk data):

- `V1FLADER.shp` (28.717 polygoner -> 23.209 unikke lokaliteter)
 - **Anvendte kolonner:**
 - `geometry`: Polygongeometrier for forurenede lokaliteter
 - `Lokalitet_`: Lokalitetsidentifikator (matcher Lokalitetsnr fra CSV)
 - **Datatype:** GeoDataFrame med lokalitetspolygoner
- `V2FLADER.shp` (33.040 polygoner -> 21.269 unikke lokaliteter)
 - **Samme struktur som V1**

3. Fra Trin 2: Liste med 593 GVFK-navne med vandløbskontakt

Proceslogik (`step3_v1v2_sites.py`):

1. Aktiv forureningsfiltrering (kritisk kvalitetskontrol):

- Filtrer hvor `Lokalitetensstoffer` ELLER `Lokalitetensbranche` ikke er null/tom

- **V1:** 84.601 -> 84.401 rækker (99.8% retained)
- **V2:** 134.636 -> 134.491 rækker (99.9% retained)
- **Årsag:** Inkluderer lokaliteter med dokumenterede aktive forureninger ELLER brancheoplysninger, da branch-information kan indikere potentielle forureningsrisici også uden dokumenterede stoffer

2. Geometri-processering:

- Indlæs shapefiles med `geopandas.read_file()`
- Dissolve geometrier efter `Lokalitet_` for at håndtere multipart polygoner
- **Årsag:** Enkelte lokaliteter kan bestå af multiple separate polygoner

3. Vandløbskontakt-filtrering:

- Filtrer CSV-data hvor `Navn` findes i Trin 2's GVFK-liste
- **Årsag:** Bevarer kun lokaliteter i GVFK med dokumenteret vandløbskontakt

4. Data-kobling:

- Standardiser kolonnenavne (`Lokalitetsnr` -> `Lokalitet_`)
- Join CSV-attributter med dissolved geometrier via `Lokalitet_`
- **Resultat:** Komplet spatial+attribut datasæt

5. Deduplikering og datakonsolidering (to-trins proces):

- **Trin 1:** Aggreger lokalitet-GVFK kombinationer inden for V1/V2
 - **Stoffer:** Sammensæt alle unikke stoffer med semikolon-adskillelse
 - **Andre felter:** Bevar første værdi (identisk indenfor samme kombination)
 - **Årsag:** Bevarer ALLE forureningsstoffer per lokalitet-GVFK kombination
- **Trin 2:** Håndter lokaliteter i både V1 og V2 (marker som "V1 og V2")
 - **Stoffer:** Sammensæt stoffer fra både V1 og V2 registreringer
 - **Årsag:** Sikrer komplet stoffortegnelse for lokaliteter med dobbelt klassificering

Hvorfor en-til-mange relationer bevares:

- En enkelt forurennet lokalitet kan overlappe flere GVFK-polygoner
- Trin 4 kræver alle kombinationer for at finde nærmeste vandløb inden for HVER GVFK
- Essentielt for korrekte afstandsberegninger per GVFK-lokalitet par

Output:

- **Filer:**
 - `step3_v1v2_sites.shp`: Alle lokalitet-GVFK kombinationer med geometri
 - `step3_gvfk_with_v1v2.shp`: GVFK-polygoner med V1/V2-lokaliteter
 - `step3_site_gvfk_relationships.csv`: Detaljerede relationsdata
- **Returnerer:** (sæt_med_GVFK_navne, v1v2_kombineret_GeoDataFrame)

Aktuelle Resultater:

- **35.728 unikke V1/V2-lokaliteter** med aktive forureninger eller brancheoplysninger
- **69.627 totale lokalitet-GVFK kombinationer** efter deduplikering

- **491 GVFK har V1/V2-lokaliteter** (82.8% af vandløbs-GVFK fra Trin 2)
- Gennemsnitligt 1,9 GVFK per lokalitet

Lokalitet-fordeling efter type:

- **V2:** 15.610 lokaliteter (43.7%)
- **V1 og V2:** 3.099 lokaliteter (8.7%)
- **V1:** 17.019 lokaliteter (47.6%)

Videre dataflow: Lokalitet-GVFK kombinationer fra dette trin danner grundlag for afstandsberegninger i Trin 4

Trin 4: Afstandsanalyse til Vandløb

Formål: Beregne minimumsafstanden fra hver V1/V2-lokalitet til vandløbssegmenter med grundvandskontakt inden for samme GVFK. Dette kvantificerer forureningsspredningsrisikoen baseret på fysisk afstand.

Inddata:

- **Fra Trin 3:** 35,728 unikke V1/V2-lokaliteter med 69,627 lokalitet-GVFK kombinationer
- **Fra datagrundlag:** Vandløbsstrækninger med kontakt til grundvand ([Rivers_gvf_rev20230825_kontakt.shp](#))

Proceslogik ([step4_distances.py](#)):

Trin 4 håndterer én-til-mange relationer mellem lokaliteter og GVFK ved at beregne afstande for hver kombination separat:

1. Afstandsberegning per lokalitet-GVFK kombination: For hver af de 69,627 kombinationer:

- Hent lokalitetens geometri og tilknyttet GVFK-navn
- Find matchende vandløbssegmenter hvor **GVForekom** = lokalitetens GVFK OG **Kontakt** = 1
- Beregn minimumsafstand mellem lokalitetspolygon og alle matchende vandløbssegmenter
- Gem resultatet for denne specifikke kombination

2. Identifikation af final minimumsafstand per lokalitet:

- Gruppér resultater efter lokalitets-ID
- Find den GVFK-kombination med absolut korteste afstand til vandløb
- Markér denne som lokalitetens primære risikosti (**Is_Min_Distance** = True)

Attributmatchingslogik: Korrekt afstandsberegning kræver præcis matching mellem lokalitets-GVFK tilknytninger og vandløbssegmenter:

- Lokalitet har præ-defineret GVFK-tilknytninger fra CSV-filer (kolonne **Navn**) - disse blev skabt ved tidligere spatial analyse
- Trin 3 tilføjer kun geometrier til disse eksisterende lokalitet-GVFK relationer
- Vandløbssegment har præ-defineret GVFK-tilknytning i kolonne **GVForekom** (udført af Lars Trolborg/DKModel)
- Kun når **Navn** = **GVForekom** AND **Kontakt** = 1 beregnes afstand
- Dette sikrer at forurening kun kan nå vandløb gennem faktisk grundvand-vandløb kontakt

Koordinatsystem og afstandsmåling:

- Alle beregninger i UTM32/EUREF89 koordinatsystem (meter-baseret)
- Afstande beregnet med `geometry.distance()` - minimum euklidisk afstand mellem geometrier
- Beregning: Korteste afstand mellem ethvert punkt på lokalitetspolygonen og ethvert punkt på vandløbslinjen

Aktuelle Resultater: Algoritmen behandlede 69,627 lokalitet-GVFK kombinationer med følgende resultater:

- **35,728 unikke lokaliteter** har alle afstande til vandløb (100% success rate)
- **Gennemsnitlig afstand:** 3,116 meter til nærmeste vandløb
- **Median afstand:** 1,368 meter til nærmeste vandløb
- **Minimum afstand:** 0,0 meter (lokaliteter direkte ved vandløb)
- **Maksimum afstand:** Varierer afhængigt af GVFK størrelse

Output-filer:

1. **step4_final_distances_for_risk_assessment.csv:** Én række per lokalitet med minimum afstand
 - Kolonner: `Lokalitet_ID`, `Final_Distance_m`, `Closest_GVFK`, samt forureningsmetadata
 - Bruges direkte af Trin 5 til risikovurdering
2. **step4_valid_distances.csv:** Alle lokalitet-GVFK kombinationer med gyldige afstande
 - Bruges til visualiseringer og detaljeret analyse
3. **unique_lokalitet_distances.shp:** Shapefil med lokalitetsgeometri og minimum afstande
 - Bruges til GIS-baserede visualiseringer

Særlige overvejelser:

- Lokaliteter uden matchende vandløbssegmenter i deres GVFK filtreres fra
- Step bevarer alle forureningsattributter (branche, aktivitet, stoffer) til videre analyse
- GVFK-information bevares for sporbarhed af kritiske forureningsstier

Eksempel: Lokalitet 12345 med Multiple GVFK:

Inddata fra Trin 3:

- Lokalitet 12345 findes i 3 lokalitet-GVFK kombinationer:
 - Lokalitet 12345 -> GVFK_A (Navn = "DK_GVF_001")
 - Lokalitet 12345 -> GVFK_B (Navn = "DK_GVF_002")
 - Lokalitet 12345 -> GVFK_C (Navn = "DK_GVF_003")

Niveau 1: Beregning per kombination:

- **Kombination 1:** Find vandløbssegmenter hvor `GVForekom` = "DK_GVF_001" AND `Kontakt` = 1
 - Findes: 3 matchende vandløbssegmenter
 - Afstande: 450m, 720m, 890m -> **Minimum: 450m**
- **Kombination 2:** Find vandløbssegmenter hvor `GVForekom` = "DK_GVF_002" AND `Kontakt` = 1
 - Findes: 2 matchende vandløbssegmenter

- Afstande: 320m, 580m -> **Minimum: 320m**
- **Kombination 3:** Find vandløbssegmenter hvor `GVForekom = "DK_GVF_003"` AND `Kontakt = 1`
 - Findes: 1 matchende vandløbssegment
 - Afstand: 1200m -> **Minimum: 1200m**

Niveau 2: Final minimum per lokalitet:

- Sammenlign: 450m (GVFK_A), 320m (GVFK_B), 1200m (GVFK_C)
- **Final resultat:** Lokalitet 12345 = 320m afstand via GVFK_B ("DK_GVF_002")

Output: Én række i `step4_final_distances_for_risk_assessment.csv`:

```
Lokalitet_ID: 12345
Final_Distance_m: 320
Closest_GVFK: DK_GVF_002
[+ metadata kolonner]
```

Trin 5: Tærskel-vurdering og Kategorisering

Formål: Identificere højrisiko V1/V2-lokaliteter baseret på afstand til vandløb og stofspecifikke mobilitetsegenskaber. Implementerer to-lags risikovurdering med både generelle og stofspecifikke tærskler.

Inddata:

- **Fra Trin 4:** `step4_final_distances_for_risk_assessment.csv` med 35.728 lokaliteter og deres minimumsafstande
- **Excel-baseret kategorisering:** `compound_categorization_review.xlsx` med litteraturbaserede faner.

0. For-kvalificering (data qualification):

- Trin 5 modtager alle lokaliteter fra trin 4 (minimumsafstand per lokalitet).
- Funktionen `separate_sites_by_substance_data` opdeler datasættet i to grupper:
 - *Kvalificerende sites:* Mindst én registreret substans eller losseplads nøgleord i branche/aktivitet. Disse fortsætter til de efterfølgende vurderinger.
 - *Parkede sites:* hverken substansdata eller relevante losseplads nøgleord. De gemmes i `step5_unknown_substance_sites.csv` til dokumentation og manuel opfølgning.
- De kvalificerende sites udgør grundlaget for både den generelle 500 m screening og den stofspecifikke analyse.

Proceslogik (`step5_risk_assessment.py`):

1. Generel risikovurdering (500 m universal tærskel)

- Filtrer lokaliteter hvor `Final_Distance_m ≤ 500 m`
- Konservativ screening uafhængig af forureningstype
- Output: `step5_high_risk_sites.csv` og GVFK-shapefiler

2. Stofspecifik risikovurdering med losseplads-override (to-fase tilgang)

Fase 1 – Kategorisering og initial screening

Kilde: Jordforureningens påvirkning af overfladevand, delprojekt 2: Afstandskriterier og fanebredder. Miljøprojekt nr. 1565, 2014. Tabel 2 og 3.

- Parse semikolon-separerede stoffer per lokalitet og hent kategori samt stofspecifik tærskel fra Excel-mappingen
- Kategorierne spejler de aktive mobilitetsklasser (fx **PAH_FORBINDELSER** 30 m, **BTXER** 50 m, **PHENOLER** 100 m, **UORGANISKE_FORBINDELSER** 150 m, **POLARE_FORBINDELSER** 300 m, **KLOREREDE_OPLØSNINGSMIDLER** 500 m, **PESTICIDER** 500 m, **LOSSEPLADS** 500 m, **ANDRE** 500 m) og respekterer stofspecifikke overrides som Benzen 200 m
- Registrer allerede her om lokaliteten har losseplads-karakteristika via **Lokalitetensbranche** eller **Lokalitetensaktivitet**
- Hvis lokaliteten matcher en kategori i **LANDFILL_THRESHOLDS**, sættes den effektive tærskel til $\max(\text{kategori-tærskel}, \text{losseplads-tærskel})$ for ikke at frasortere kombinationer der kun består på grund af en lempelig lossepladsgrænse; mere restriktive losseplads-tærskler håndteres efterfølgende
- Lokalteter uden stofdata men med losseplads-brancher/aktiviteter klassificeres som **LOSSEPLADS** med en 100 m screenings-tærskel via **categorize_by_branch_activity**
- Evaluer hver stof-lokalitet (eller branch-only) kombination mod den effektive tærskel og gem kvalificerende rækker

Fase 2 – Losseplads-override (post-processering)

Kilde: Risikovurdering af lossepladsers påvirkning af overfladevand. Bjerg, Poul Løgstrup; Sonne, Anne Thobo; Tuxen, Nina; Skov Nielsen, Sanne; Roost, Sandra. 2014. Tabel 7.1.

- Arbejder kun videre med kombinationer fra Fase 1 og fokuserer på lokaliteter med losseplads-karakteristika
- For kategorier der findes i **LANDFILL_THRESHOLDS** anvendes nu den faktiske losseplads-tærskel; kombinationer over grænsen fjernes, mens de resterende reklassificeres
- Reklassificering sætter **Qualifying_Category** = "LOSSEPLADS", udfylder **Losseplads_Subcategory** = "LOSSEPLADS_<originalkategori>", markerer **Landfill_Override_Applied** = True og logger oprindelsen i **Qualifying_Substance** = "Landfill Override: <originalkategori>"
- Kombinationsrækker uden losseplads-flag eller uden relevant tærskel bevares uændret

Eksempler

- **BTXER ved 60 m med losseplads-flag**: passerer Fase 1 med 70 m ($\max(50,70)$) og forbliver i datasættet efter override, nu mærket som **LOSSEPLADS_BTXER**
- **Phenoler ved 80 m med losseplads-flag**: passerer Fase 1 (100 m), men fjernes i Fase 2, fordi losseplads-tærsklen er 35 m
- **Branch-only losseplads ved 120 m**: falder allerede i Fase 1, fordi den branch-baserede 100 m tærskel overskrides

Tærskel-sammenligning (Normal vs. Losseplads-specifik)

Kategori	Normal Tærskel	Losseplads Tærskel	Repræsentativ Forbindelse	Override Status
BTXER	50 m	70 m	Benzen	✓ Override (lempligere)
KLOREREDE_OPLØSNINGSMIDLER	500 m	100 m	Tetrachlorethylen (TCE)	✓ Override (strengere)
PHENOLER	100 m	35 m	Phenol	✓ Override (strengere)
PESTICIDER	500 m	180 m	MCPP	✓ Override (strengere)
UORGANISKE_FORBINDELSER	150 m	50 m	Arsen	✓ Override (strengere)
PAH_FORBINDELSER	30 m	-	-	Ingen override
POLARE_FORBINDELSER	300 m	-	-	Ingen override
LOSSEPLADS	100 m	-	-	Ingen override (allerede losseplads)
ANDRE	500 m	-	-	Ingen override

Litteraturbaserede stofgrupper ([refined_compound_analysis.py](#))

- [refined_compound_analysis.py](#) grupperer stoffer i 10 kategorier baseret på *Jordforureningens påvirkning af overfladevand, delprojekt 2* (Tabel 2 og 3).
- Kortlægningen danner grundlag for ovenstående tærskler og bruges både i foranalysen og i Trin 5.
- Kategorierne og deres standardtærskler er:
 - **BTXER** – 50 m (BTEX + oliefraktioner; tabel 2).
 - **KLOREREDE_OPLØSNINGSMIDLER** – 500 m (klorerede opløsningsmidler; tabel 2).
 - **POLARE_FORBINDELSER** – 300 m (MTBE m.fl.; tabel 2).
 - **PHENOLER** – 100 m (fenoler; tabel 2).
 - **KLOREDE_KULBRINTER** – 200 m (klorerede/bromerede kulbrinter; tabel 3).
 - **KLOREREDE_PHENOLER** – 200 m (klorerede fenoler; tabel 3).
 - **PAH_FORBINDELSER** – 30 m (PAH; tabel 3).
 - **PESTICIDER** – 500 m (pesticider inkl. PFAS-stoffer; tabel 2).
 - **UORGANISKE_FORBINDELSER** – 150 m (metaller og uorganika; tabel 3).
 - **LOSSEPLADS** – 100 m (landfill-relaterede kilder).
- Kategoriseringsscriptet anvender også stofs specifikke overrides (fx Benzen 200 m, cyanid 100 m), som går forud for både normale og losseplads-specifikke kategori-tærskler.

- Stoffer der ikke matcher nogen kategori lander i **ANDRE** og får default 500 m (efter manuel opfølgning).
- **Output:** `step5_compound_detailed_combinations.csv` med alle kvalificerende kombinationer efter override
- Output: `step5_compound_detailed_combinations.csv` med alle kvalificerende kombinationer efter override

Aktuelle Resultater:

Generel vurdering (500m tærskel):

- **4.156 lokaliteter** kvalificerer som højrisiko (11,6% af alle analyserede)
- **300 unikke GVFKs** påvirket (14,7% af alle danske GVFK)
- **Top kategorier:**
 - *Brancher:* Servicestationer (657), Autoreparationsværksteder (638), Drift af affaldsbehandlingsanlæg (600)
 - *Aktiviteter:* Andet (935), Benzin og olie, salg af (666), Benzin og olie, erhvervsmæssig oplag af (461)
 - *Stoffer:* Tungmetaller (675), Bly (661), Olieprodukter (561)

Stofspecifik vurdering med losseplads-override (kategori-baserede tærskler):

- **2.013 unikke lokaliteter** kvalificerer (5,6% af alle analyserede)
- **4.235 stof-lokalitet kombinationer** (gennemsnit 2,1 stoffer per lokalitet)
- **705 kombinationer** blev reklassificeret til LOSSEPLADS via override-systemet
- **367 kombinationer** blev fjernet pga. strengere losseplads-tærskler

Losseplads-override fordeling efter kategori:

- **UORGANISKE_FORBINDELSER:** 299 kombinationer (50m tærskel)
- **BTXER:** 191 kombinationer (70m tærskel)
- **PESTICIDER:** 114 kombinationer (180m MCPP-tærskel)
- **KLOREREDE_OPLØSNINGSMIDLER:** 63 kombinationer (100m tærskel)
- **PHENOLER:** 38 kombinationer (35m tærskel)

Multi-stof distribution:

- **1.483 lokaliteter** (64%): 1 kvalificerende stof
- **322 lokaliteter** (14%): 2 kvalificerende stoffer
- **190 lokaliteter** (8%): 3 kvalificerende stoffer
- **313 lokaliteter** (14%): 4+ kvalificerende stoffer
- **Maksimum:** 32 stoffer (Lokalitet 751-00018)

Kategori-fordeling efter forekomst:

- **UORGANISKE_FORBINDELSER:** 1.055 forekomster (588 lokaliteter) - 150m tærskel
- **KLOREREDE_OPLØSNINGSMIDLER:** 926 forekomster (480 lokaliteter) - 500m tærskel
- **LOSSEPLADS:** 854 forekomster (793 lokaliteter) - 500m tærskel
- **BTXER:** 780 forekomster (433 lokaliteter) - 50m tærskel

- **PESTICIDER:** 642 forekomster (277 lokaliteter) - 500m tærskel
- **ANDRE:** 260 forekomster (238 lokaliteter) - 500m tærskel
- **PAH_FORBINDELSER:** 137 forekomster (77 lokaliteter) - 30m tærskel
- **PHENOLER:** 72 forekomster (49 lokaliteter) - 100m tærskel

GVFK-filtreringskaskade:

- **Total GVFK Danmark:** 2.043 (100,0%)
- **Med vandløbskontakt (Trin 2):** 593 (29,0%)
- **Med V1/V2-lokaliteter (Trin 3):** 490 (24,0%)
- **Med sites ≤500m (Generel):** 300 (14,7%)
- **Med stofs specifik risiko (Trin 5):** 246 (12,0%)

Metodiske forbedringer implementeret:

- **Stofaggregering i Trin 3:** Sikrer alle forureningsstoffer bevares gennem workflow
- **Multi-stof håndtering:** Lokaliteter med både V1/V2-klassifikation får kombineret stoffortegnelse
- **Stofspecifik risikoevaluering:** Hver stof evalueres mod sin kategori-specifikke tærskel
- **GVFK-spring:** Bruger **Closest_GVFK** fra Trin 4 til at identificere primær risikosti

Tilstandsvurdering

Status: Fremtidigt arbejde - planlagt som næste projektfase efter finalisering af risikovurderingsmetodikken.

Planlagt Metodisk Tilgang

Tilstandsvurderingen vil bygge videre på de 2.013 højrisiko-lokaliteter identificeret gennem stofs specifik risikovurdering og omfatte:

Kvantitativ Fluxberegning

- Beregning af forureningsflux fra de **2.013 højrisiko-lokaliteter** identificeret i risikovurderingen
- Anvendelse af infiltrationsdata fra DK-modellen: **Flux = Areal × Koncentration × Infiltration**
- Transport af flux langs strømlinjer til relevante kontaktstrækninger inden for de **232 påvirkede GVFK**

Koncentrationsvurdering i Vandløb

- Beregning af blandingskoncentration: **C_{mix} = Forureningsflux / Vandføring**
- Sammenligning med miljøkvalitetskrav (MKK) for specifikke stoffer
- Identifikation af overskridelser på stofniveau

Prioritering og Kvantificering

- Kategorisering af overskridelser efter alvorlighedsgrad
- Vurdering af mindre overskridelser (1-10 gange MKK)
- Udarbejdelse af prioriterede indsatslister

Samarbejde med GEUS

Tilstandsvurderingen kræver tæt samarbejde med GEUS vedrørende:

- Kontaktzoner og strømningsveje fra DK-modellen
- Vandløbsstrækninger og vandføringsdata
- Automatiserede udtræk til landsdækkende anvendelse

Samlet Overblik

- **Datagrundlag: 2.043 grundvandsforekomster** i Danmark
- **Vandløbskontakt: 593 GVFK** (29,0%) har kontakt med vandløb
- **Aktiv forureningsfiltrering:**
 - V1: 84.601 -> **34.232 lokaliteter** med aktive forureninger (60% reduktion)
 - V2: 134.636 -> **121.984 lokaliteter** med aktive forureninger (9% reduktion)
 - Eliminere lokaliteter uden dokumenterede forureningsstoffer
- **Endelig analyse: 35.728 unikke lokaliteter** med både aktive forureninger og vandløbskontakt
- **Generel screening: 4.156 lokaliteter** inden for 500m (300 GVFK påvirket)
- **Stofspecifik risikovurdering: 2.013 højrisiko-lokaliteter** baseret på litteraturbaserede tærskler (232 GVFK påvirket)
- **Output:** Præcise afstande til vandløb med komplet forureningsinformation til risikovurdering

Konklusion

Denne metodebeskrivelse præsenterer en systematisk og robust tilgang til identificering af forurenede lokaliteter med potentiel risiko for påvirkning af overfladevand. Ved at kombinere spatial analyse med detaljerede forureningsdata opnås et præcist grundlag for risikovurdering og prioritering af miljøtiltag.

De udviklede metoder muliggør:

- **Effektiv risikoidentifikation:** 2.013 højrisiko-lokaliteter identificeret gennem stofspecifik vurdering (ud af 35.728 analyserede lokaliteter)
- **Prioriteret indsats:** Fokus på lokaliteter inden for stofspecifikke tærskler med dokumenteret grundvandskontakt
- **To-lags vurdering:** Både generel screening (4.156 sites) og stofspecifik analyse (2.013 sites)
- **GVFK-reduktion:** Fra 2.043 danske GVFK til 232 højrisiko GVFK (11,4%)
- **Kvantificeret risiko:** Præcise afstandsmålinger og stofspecifikke tærskler som grundlag for risikovurdering
- **Sporbarhed:** Komplet dokumentation af databehandling og analysemetoder

Metoderne danner et solidt fundament for fremtidig miljøovervågning og kan tilpasses forskellige tærskler og kriterier alt efter specifikke beslutningsbehov.