

Metodutveckling

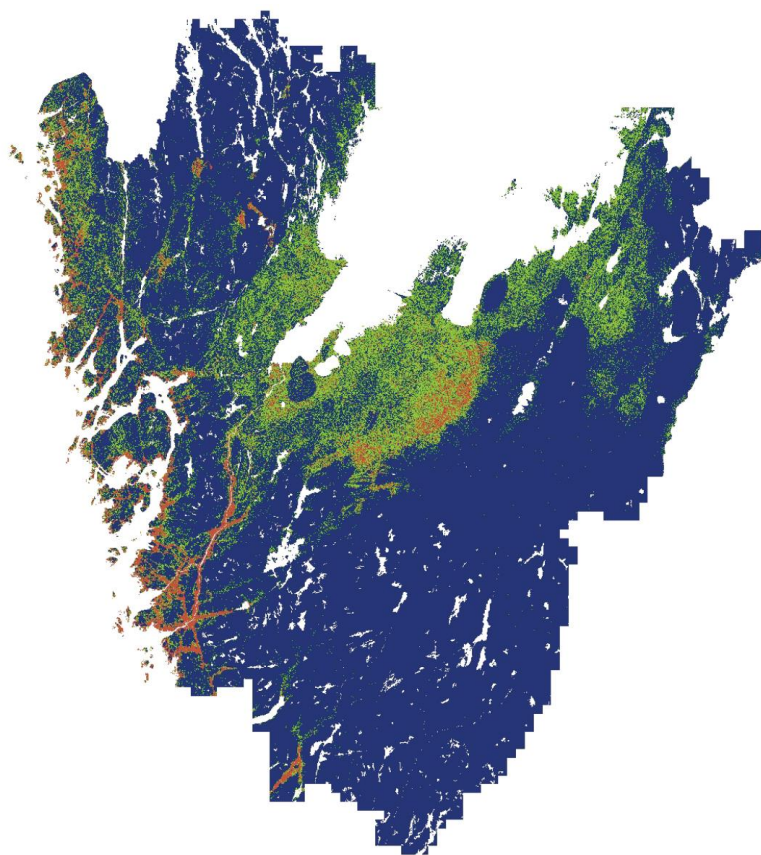
Kloridhalter i grundvatten

Modellering med hjälp av maskininlärning

Maria Åkesson, Paulina Bastviken & Marina Becher

februari 2021

SGU-rapport 2021:06



Omslagsbild: Modellering gjord via maskininlärning. Kartan visar sannolikheten för förekomst av kloridhalter >100 mg/l i grundvatten i berg i Västra Götalands län.

Författare: Maria Åkesson, Paulina Bastviken och Marina Becher

Granskad av: Lena Maxe

Ansvarig enhetschef: Mats Wallin

Redaktör: Lina Rönnåsen

Sveriges geologiska undersökning

Box 670, 751 28 Uppsala

tel: 018-17 90 00

e-post: sgu@sgu.se

www.sgu.se

INNEHÅLL

Inledning.....	4
Bakgrund.....	4
Klorid i grundvatten – orsaker och problematik.....	4
Metoder för att uppskatta kloridhalter i grundvatten.....	6
Maskininläring på SGU.....	7
Metod	8
Modelleringsprocess	8
Observationer.....	10
Miljövariabler.....	11
Resultat.....	12
Modellprestation	12
Styrande miljövariabler.....	12
Predikterade kloridhalter.....	14
Predikterad sannolikhet.....	16
Diskussion	17
Representativitet och upplösning	17
Förbättringsförslag	17
Ytterligare tillämpningsområden kopplat till SGUs grundvattenverksamhet	18
Slutsats	19
Referenser.....	19

INLEDNING

I spåren av de torra somrarna år 2016 och 2017 fick SGU i uppdrag av regeringen att under åren 2018–2020 förfinas och utöka kartläggning och karaktärisering av grundvattenresurser i särskilt utsatta områden. Särskilt utsatta områden är, enligt SGUs definition, områden med förhöjd risk för att tillgången till grundvatten av god kvalitet inte uppfyller de behov som grundvattenanvändningen medför (Hjerne 2018), och benämns framöver som bristområden. I regeringsuppdraget ingick bland annat utvecklade mätningar beträffande grundvattnets kvantitet, kvalitet och prognostisering.

Inom ramen för regeringsuppdraget har SGU drivit ett projekt i syfte att tydliggöra och utveckla myndighetens arbete med grundvattenkvalitet. I de interna direktiven för detta projekt för år 2020, uttrycktes behov av fördjupade grundvattenkemiska analyser såväl nationellt som inom utpekade bristområden. Mot denna bakgrund föddes en idé om att låta vidareutveckla metodik för analys och kartläggning av grundvattenkvalitet med hjälp av maskininlärning, baserat på verktyg och rutiner utvecklade på enheten Samhällsplanering Jord. Som pilotprojekt valdes modellering av salt grundvatten, dvs. grundvatten med förhöjda kloridhalter, vilket utgör ett stort problem i flera utpekade bristområden.

BAKGRUND

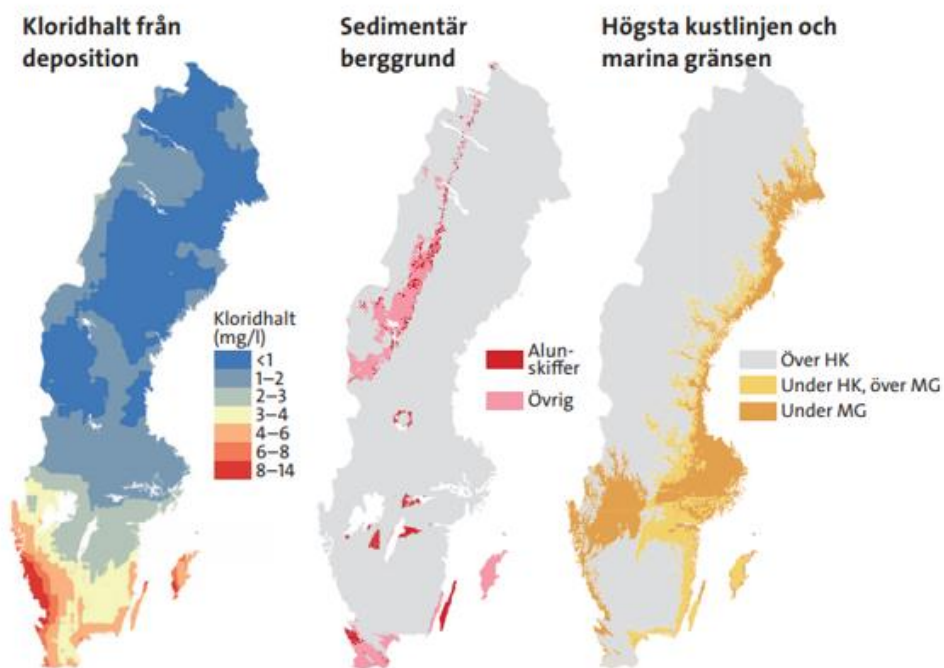
Klorid i grundvatten – orsaker och problematik

Klorid kan tillföras grundvatten såväl genom naturliga källor som av mänsklig aktivitet. Den naturliga tillförseln av klorid utgörs främst av havssalter. Depositionsbetingade kloridhalter varierar över landet, och är generellt högst i områden utsatta för västliga havsvindar. Naturligt förhöjda halter av klorid förekommer även i områden som varit täckta av hav efter den senaste istiden, p.g.a. påverkan av relik saltvatten. I strandnära områden kan påverkan av nutida havsvatten förekomma. Vidare kan förhöjda halter av klorid förekomma i grundvatten i sedimentär berggrund av marint ursprung. Figur 1 (SGU 2013a) illustrerar förutsättningarna avseende nämnda naturliga källor till klorid.

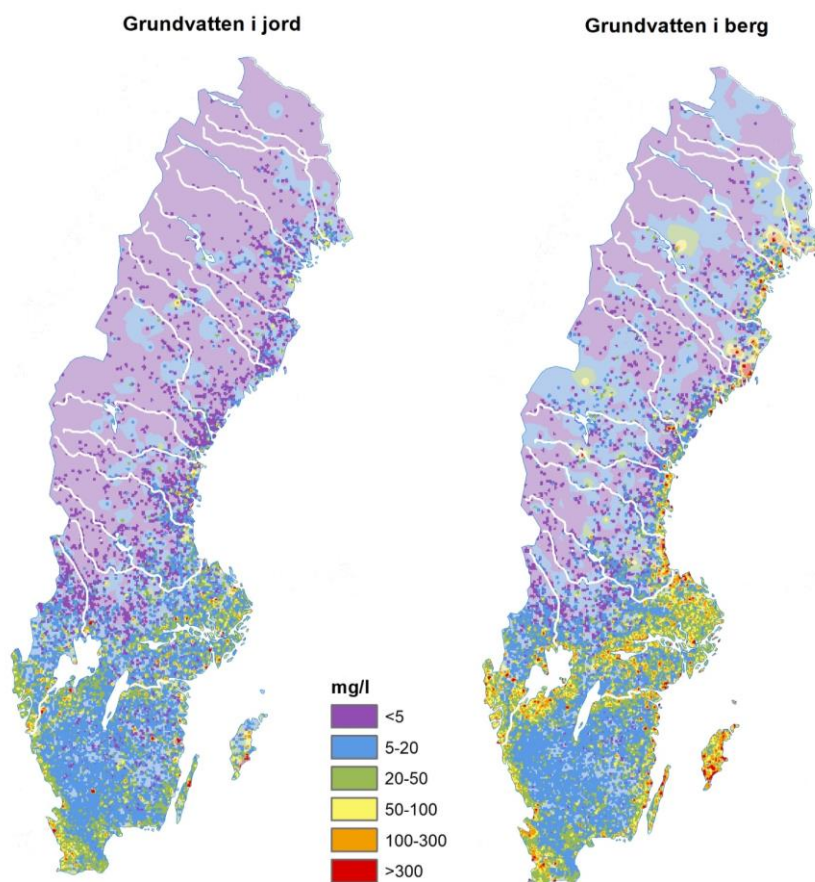
Mänsklig tillförsel av klorid till grundvatten sker via exempelvis vägsaltning, avlopp och lakvatten från deponier. Det är vidare vanligt att kloridhalten i en given punkt ökar vid uttag av grundvatten, till följd av förändrade och mer koncentrerade flödesmönster.

I figur 2 ges en översiktlig bild av kloridhalter i grundvatten i Sverige. Materialet är framtaget inom ramen för pågående uppdateringar av SGUs Bedömningsgrunder för grundvatten (2013a), och bygger på interpolering av uppmätta halter från olika databaser vid SGU.

Höga kloridhalter i grundvatten kan vara problematiskt på flera sätt. Ett alltför salt vatten är inte gott att dricka och utgör en onödig exponering sett till den enskildes totala saltintag. Höga kloridhalter kan också innebära begränsningar i användning i livsmedelsproduktion, exempelvis i förhållande till bevattning. Kloridhaltigt vatten kan vidare föranleda stora problem och kostnader genom korrosion av exempelvis ledningsnät och armeringsjärn i betongkonstruktioner. Livsmedelsverkets gränsvärde för tjänligt dricksvatten med anmärkning är 100 mg/l och satt mot bakgrund av tekniska risker (SLV 2017). Halter högre än 300 mg/l innebär risk för smakförändringar.



Figur 1. Viktiga naturliga källor till klorid i grundvatten, från SGU (2013a).



Figur 2. Uppmätta och interpolerade kloridhalter i grundvatten i jord och berg. Kartorna utgör icke offentliggjort arbetsmaterial.

Vid höga koncentrationer finns även risk för ekologisk påverkan i fall av utströmning i ytvatten, inte så mycket av kloridjonerna i sig som av den ökade surhetsgrad och de förhöjda metallhalter som kan följa på resulterande jonbytesprocesser.

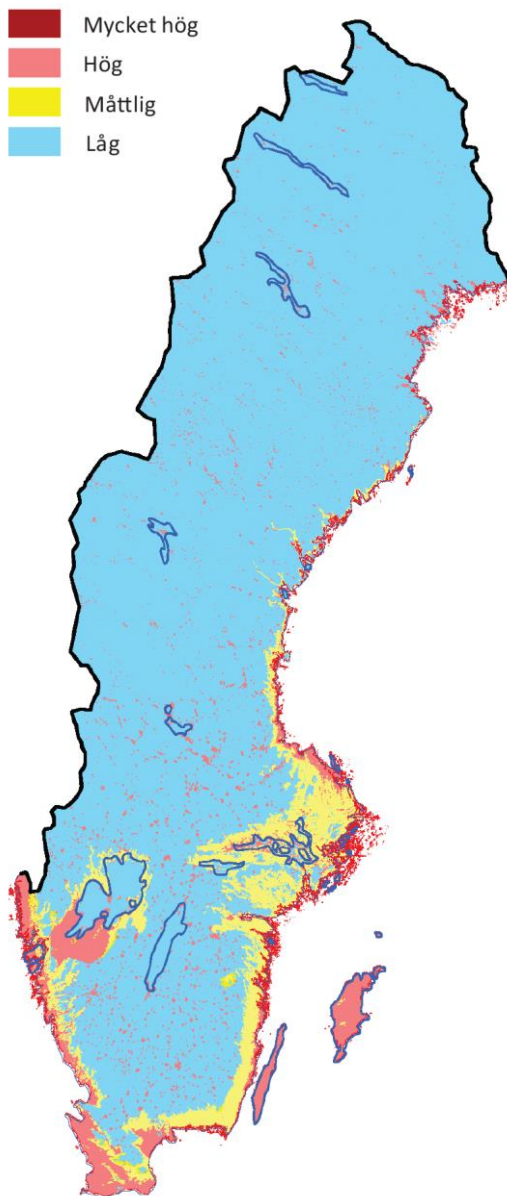
Kommande klimatförändringar bedöms innebära minskad grundvattenbildning i stora delar av södra Sverige. Minskad grundvattenbildning i kombination med stigande havsnivå och oförändrade alternativt ökande uttag, riskerar att öka problematiken med förhöjda kloridhalter i grundvatten i dessa områden, till följd av minskad potential för utspädning, ökade uttag från större djup, samt ökande saltvatteninträngning i kustområden.

Metoder för att uppskatta kloridhalter i grundvatten

En av de vanligaste metoderna för att uppskatta förekomst av klorid i grundvatten är den s.k. GALDIT-metoden (Lobo-Ferreira m.fl. 2007), som bygger på en sammanvägd bedömning av miljövariablerna akviferstyp, genomsläpplighet, grundvattennivå i förhållande till havsnivå, avstånd från kust, rådande saltvattenpåverkan och akvifersmäktighet. GALDIT-metoden kan sägas vara en variant av DRASTIC-metoden, som tillämpas mer generellt för föroreningar i grundvatten och utvecklats av den amerikanska motsvarigheten till Naturvårdsverket (Aller m. fl. 1987).

I Sverige har vid upprepade tillfällen varianter av den så kallade Risk Variabel (RV)-metoden tillämpats för att uppskatta kloridhalter i grundvatten (t.ex. av Bucht och Larsson 1995, Lindberg m. fl. 1996, Ebert m. fl. 2016). Vilka miljövariabler som beaktas, för vilka en osäkerhetsbedömning ingår, avgörs utifrån statistiska analyser. I Lindberg m.fl. (1996) beaktas följande miljövariabler: nederbörd, marklutning, jordtäckte, jorddjup, regionala sprickzoner, lokal sprickfrekvens, brunnsdjup, markens exploateringsgrad, uttagsbrunnars lokalisering, sanitär standard samt avstånd till kust, sjö, våtmark, saltad väg och andra saltkällor (ex. deponi).

SGU har vid ett flertal tillfällen bedömt sannolikhet för förekomst av salt grundvatten, på olika skalor och för olika områden. Samtliga bygger på statistisk analys och prediktion. Ett tidigt exempel är den så kallade saltriskkarta som togs fram för Upplands Bro kommun, som bygger på analys av kloridhalter i förhållande till läge och marknivå (Anderberg 1997). Ett senare exempel är den nationella karta över Sannolikhet för salt grundvatten i bergborrade brunnar som togs fram av SGU 2014 (fig. 3). Kartan bygger på en regressionsanalys av uppmätta kloridhalter i berg i förhållande till geografiskt läge, läge i förhållande till marina gränsen samt region enligt SGUs Bedömningsgrunder för grundvatten (SGU 2013a). Fasta buffertområden sattes i förhållande till avstånd till vägar och kust. Kartan var tänkt att läggas upp som en egen Kartvisare på SGUs webbplats men produkten fastnade i en betaversion med uppdateringsbehov och färdigställdes aldrig helt. Kartan var tänkt för bedömningar på nationell och regional nivå.



Figur 3. Sannolikhet för salt grundvatten i bergborrade brunnar, enligt SGU 2014.

Maskininlärning på SGU

På enheten Samhällsplanering Jord på SGU har nyligen en metod för att modellera utbredningen av sura sulfatjordar utvecklats (Becher m. fl. 2019). Sur sulfatjord är problematiskt, då fluktuation av grundvattennivån i sådan jord kan leda till försurning och mobilisering av skadliga metaller till närliggande vattenmiljöer.

Principen för metoden är att sambandet mellan punktoobservationer av sur sulfatjord och bedömt styrande miljövariabler kan utnyttjas för att träna upp en modell, vilken sedan kan användas för att prediktera sannolikheten för förekomst av sur sulfatjord på platser där observationer av sur sulfatjord saknas men information om miljövariablerna finns. Under utvecklingen av metoden testades flera olika typer av matematiska modeller och den som visade sig ge bäst resultat var Random Forest (Breiman 2001), vilken sammanlänkar data via ett stort antal klassificeringsträd.

Resultaten från modelleringen av sura sulfatjordar redovisas i SGUs Kartvisare, där för närvarande (år 2020) täckningen omfattar områden längs Västernorrlands, Västerbottens och Norrbottens kust (<https://www.sgu.se/produkter/kartor/kartvisaren/jordkartvisare/sur-sulfatjord/>).

Under 2020 har metoden vidareutvecklats i syfte att ta fram underlag till jordartskarteringen. Punktoobservationer av jordarter och miljövariabler över hela Sverige används för att skapa en modell som används för att prediktera jordart i ett område. Den matematiska modellen som användes i detta fall var Extreme Gradient Boosting (Friedman 2001).

METOD

Modelleringsprocess

Arbets sättet som utvecklats för att modellera jordarter på SGU utgör grundstommen för den metodik som använts för att modellera kloridhalter i grundvatten i denna studie. I figur 4 visas en konceptuell bild över processen.

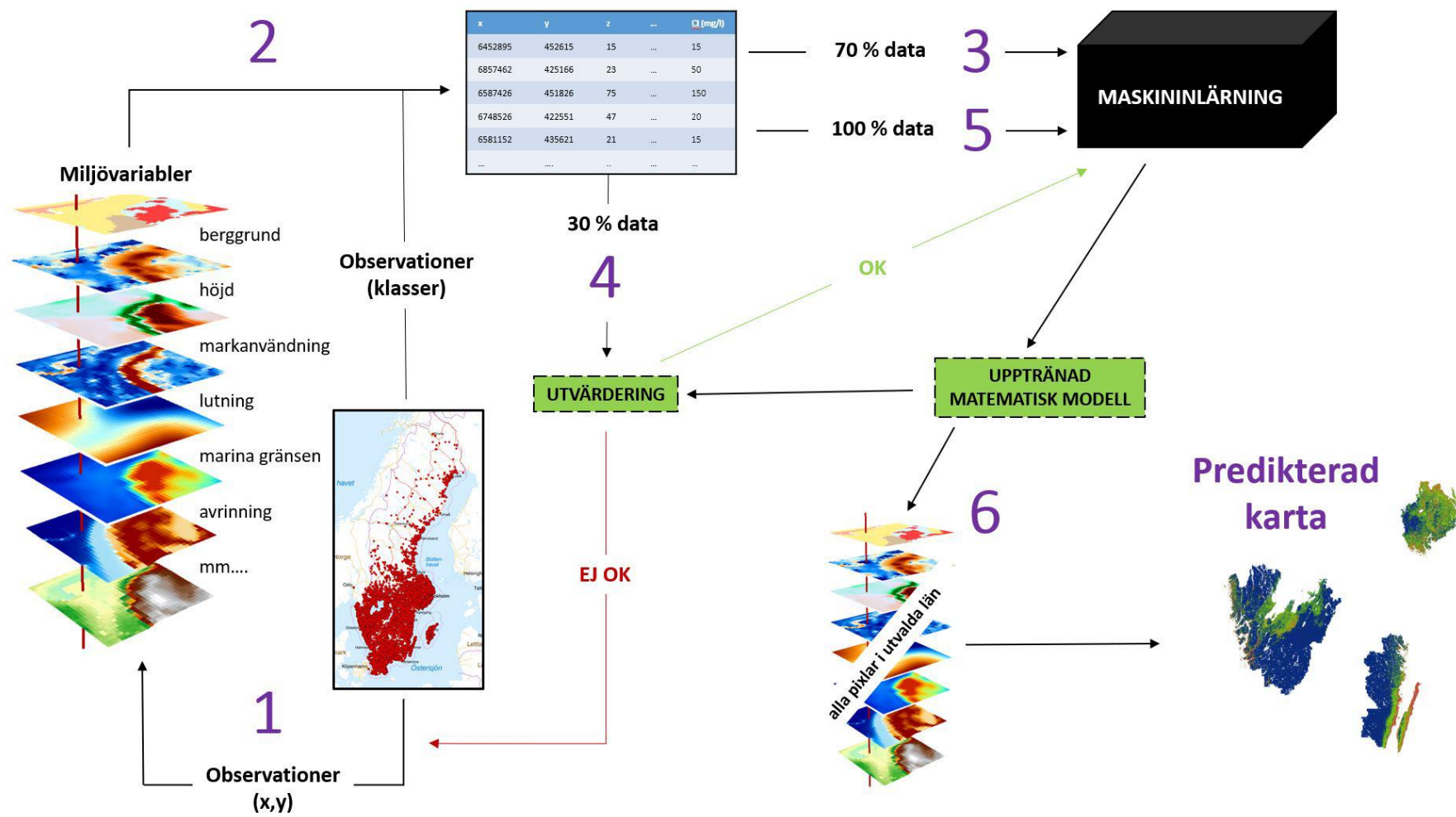
I ett första steg används koordinater för tillgängliga observationer av klorid i grundvatten för att extrahera värden av miljövariabler i samma punkter (1). Därefter sammanställs miljövariabler och kloridhalter, klassificerade i haltintervall, till ett gemensamt dataset (2). I nästa steg används 70 % av datasetet till maskinlärning, dvs till att upprätta och träna en matematisk modell som förklarar sambandet mellan kloridobservationerna och miljövariablerna (3). Modellen utvärderas sedan med hjälp av resterande 30% av datasetet (4). Om utfallet inte är tillfredställande får datasetet (observationer och/eller miljövariabler) revideras och processen återgå till första steget (1) Om utfallet efter steg (4) däremot är tillfredställande, används 100 % av punkterna i datasetet för att generera en upptränad matematisk modell (5) som används för att prediktera kloridklasser för hela den yta över vilken information om miljövariablerna finns tillgängliga (6).

Utvärderingen av modellen enligt steg (4) görs genom att beräkna *overall accuracy* och det så kallade *kappa-värdet*. Overall accuracy beskriver den totala observerade modellprestationen och utgör alltså ett mått på modellens förmåga att prediktera de faktiska observationerna. Kappa-värdet är ett mått på prediktionens prestation jämfört med slumpen, dvs. hur pass bra modellen presterar jämfört med utfallet av en helt slumpmässig prediktion.

Den matematiska modell som slutligt använts för att modellera kloridhalter är tidigare nämnda *Extreme Gradient Boosting*, som precis som *Random Forest* bygger på beslutsträd. Till skillnad från *Random Forest*, som bygger träden oberoende av varandra, bygger *Extreme Gradient Boosting* ett träd i taget där det senare trädet lär sig av det föregående trädet. På det sättet kombinerar *Extreme Gradient Boosting* resultatet av träden till skillnad från *Random Forest*, som kombinerar resultatet i slutet genom att ta ett medelvärde av resultaten av träden. Fördelen med *Extreme Gradient Boosting* är att den ger ett bättre resultat när datat är obalanserat och att den går snabbare att köra än *Random Forest*. I steg 3 ovan, testades båda modellerna varvid *Extreme Gradient Boosting* predikterade träningsdatat bättre än *Random Forest*.

Maskininlärning och prediktering har utförts med en upplösning (cellstorlek) av 2×2 m.

På grund av tidsbegränsningar kunde inte predikteringssteget (6) utföras för hela Sverige, utan avgränsades till Kalmar, Uppsala och Västra Götalands län. Dessa län valdes givet förekomst av bedömda bristområden samt variabilitet avseende beaktade miljövariabler.



Figur 4. Konceptuell bild över modelleringsprocessen.

Observationer

Som kända punktoobservationer av klorid i grundvatten valdes kloridhalter från analyser utförda på vatten från enskilda brunnar, vilka insamlats till SGUs databas i två skeden. Den äldre delen av datamängden härstammar från SGUs kemiarkiv medan prover insamlade år 2007 och senare kommer från ett projekt benämnt just "Enskilda brunnar" där data samlas in via kommersiella laboratorier.

Analyser från brunnar med osäkra koordinater, där uppgift om akvifärstyp (berg/jord) saknades eller rening av vattnet uttryckligen förekommit filtrerades bort. För koordinater där fler än en analys fanns tillgänglig valdes den högsta uppmätta halten. Efter denna bearbetning återstod ett dataset med 28 539 observationer av klorid från brunnar i berg, och 11 930 observationer av klorid från brunnar i jord, se tabell 1.

Observationerna fördelades i fyra olika klasser. Som lägsta gräns valdes 15 mg/L, motsvarande en uppskattad halt under vilken naturliga bakgrundsvärden av klorid normalt kan anses motsvara ett opåverkat vatten. Tre ytterligare klasser, med stigande grad kloridpåverkan, valdes sedan i förhållande till de två fastställda miljökvalitetsnormer som utgör utgångspunkten för att vända trend (50 mg/L) och riktvärdet (100 mg/L) för klorid på nationell nivå i SGUs Föreskrift om miljökvalitetsnormer och statusklassificering för grundvatten (SGU-FS:2013:2).

Tabell 1. Sammanställning över använda observationer av kloridhalt i grundvatten.

Klass	Kloridhalt (mg/l)	Antal observationer inom respektive område							
		Sverige		Uppsala län		Kalmar län		Västra Götaland län	
		Berg	Jord	Berg	Jord	Berg	Jord	Berg	Jord
Naturlig bakgrundshalt	<15	12 771	6 969	560	231	457	324	1 343	915
Låg påverkan	≤15 och <50	10 183	3 798	434	90	303	192	2 030	920
Lika med eller över utgångspunkt för att vända trend	≥50 och <100	2 667	707	126	29	97	33	461	166
Lika med eller över riktvärde	≥100	2 918	456	163	22	130	30	466	94
Summa		28 539	11 930	1 283	372	987	579	4 300	2 095

Miljövariabler

För maskininläringen kompletterades den uppsättning miljövariabler som använts för prediktion av jordarter på SGU, med tre variabler som bedömdes vara speciellt relevanta i fråga om förekomst av kloridhaltigt grundvatten. Samtliga beaktade miljövariabler framgår av tabell 2.

De miljövariabler som lades till var:

- Läge i förhållande till **marina gränsen**. Sannolikheten för höga kloridhalter i grundvatten är högre i områden under marina gränsen än i områden över.
- **Bergart**. Sannolikheten för förekomst av salt grundvatten i berg kan påverkas av bergartstyp. Utöver att bergets kemiska uppbyggnad kan spela roll, kan även den dominerande typen av grundvattenflöde (i porsystem eller i sprickor) ha betydelse. För maskininläringen gjordes en förenklad indelning av Sveriges berggrund i neoproterozoiska–fanerozoiska plattformstäcket respektive ”övrig berggrund” i syfte att, om än grovt, spegla vissa fundamentala skillnader i dessa avseenden.
- Det neoproterozoiska-fanerozoiska plattformstäcket motsvarar ungefär sedimentära bergarter med hög potentiell kloridhalt där grundvattenflöde kan ske både i porsystem och i sprickor, medan ”övrig berggrund” i princip motsvarar kristallina prekambrika bergarter med ringa tillskott från bergmassan och ett helt sprickdominerat grundvattenflöde.
- **Nettonederbörd**. Sannolikheten för höga kloridhalter i grundvatten påverkas av grundvattnets potential för nybildning och omsättning, då detta avgör hur klorid transporteras och späds ut i vattenkolumnen. Som komplement till redan ingående miljövariabler med inverkan på denna potential, valdes nettonederbörd per avrinningsområde enligt SMHI:s S-HYPE-modell (normalperiod 1981–2010). Nettonederbörden, eller avrinningen, utgör den maximalt möjliga grundvattenbildningen i ett givet område.

Tabell 2. Beaktade miljövariabler.

Miljövariabel	Beskrivning	Källa/ursprung
DEM	Höjd över havet	Lantmäteriet
EAS 1ha-	Hur långt det är från en pixel till närmaste vatten från 1 ha vattendragsnätverk	Beräknat från DEM
EAS 10ha-	Hur långt det är från en pixel till närmaste vatten från 10 ha vattendragsnätverk	Beräknat från DEM
DI2m	Sluttningsindex	Beräknat från DEM
CVA	Lutningsorientering	Beräknat från DEM
SDFS	Standardavvikelsen av lutningen	Beräknat från DEM
DFME	Avvikelser från medelhöjd i varje pixel	Beräknat från DEM
Rugged	Topografisk heterogenitet	Beräknat från DEM
MSRM	Terräng ojämlikhet på olika skalor	Beräknat från DEM
HKDepth	Djup under högsta vattenytan sedan deglaciationen	SGU
LandAge	Deglaciationstiden i varje pixel	Interpolerad från Stroeve et al. 2016.
Marina gränsen	se text	SGU
Bergart	se text	SGU
Nettonederbörd	se text	SMHI

RESULTAT

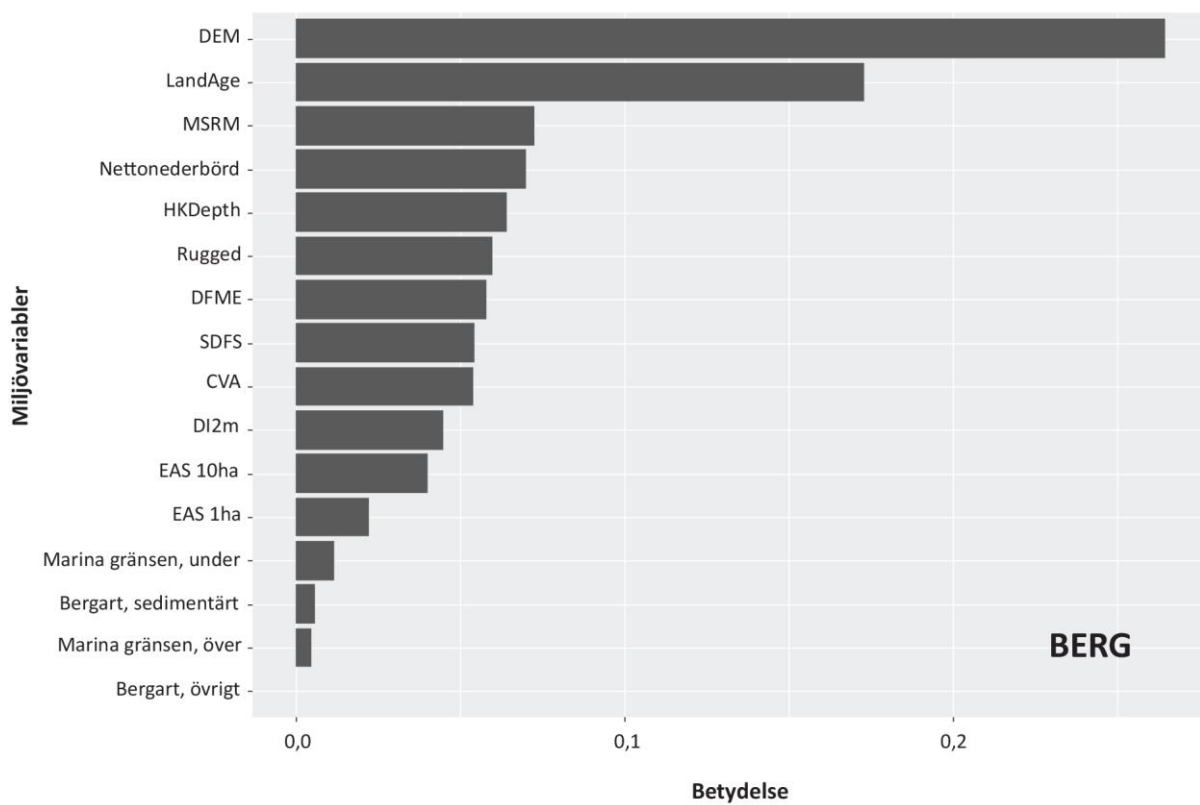
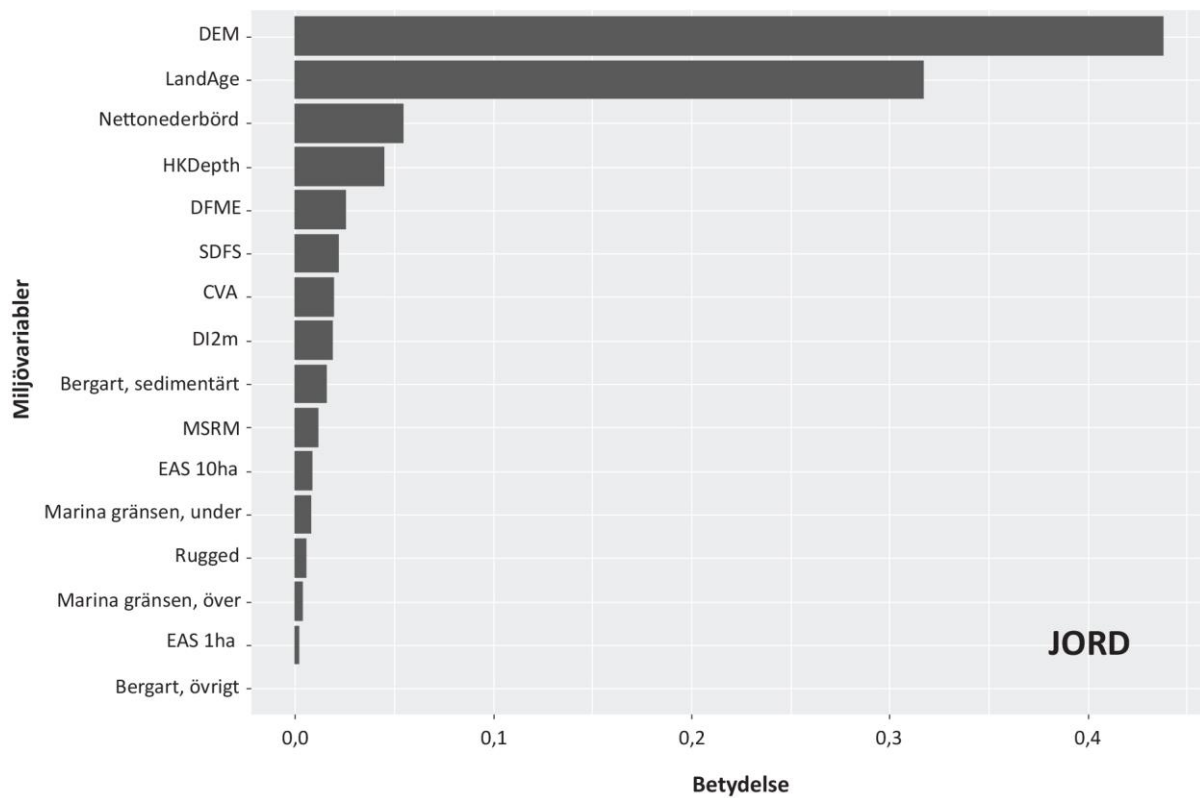
Modellprestation

Den upptränade matematiska modellen för kloridhalter i grundvatten i jord har, enligt valideringssteget, en overall accuracy på 0,67 och ett kappa-värde på 0,30. Motsvarande värden för den upptränade matematiska modellen för kloridhalter i grundvatten i berg är 0,55 respektive 0,25. Båda modellerna predikterar således rätt haltklass för mer än hälften av observationerna. Kappa-värdena faller båda inom ett intervall som brukar kategoriseras som fair agreement. Den matematiska modellen för prediktering i jord presterar något bättre än motsvarande modell i berg.

Eftersom den slutliga prediktionen bygger på maskininlärning med samtliga kloridobservationer är den slutliga modellens prestation sannolikt något bättre än vad som framgår av valideringen, som utvärderar den modell som presterar bäst i förhållande till enbart 70 % av observationerna. Denna skillnad är dock troligtvis förhållandevis liten.

Styrande miljövariabler

De miljövariabler som har störst betydelse för prediktionen av kloridhalter i grundvatten i jord respektive berg framgår av figur 5. Av störst vikt i båda fallen är (1) den absoluta höjden i läget för observationen i marknivå (DEM), och därefter (2) den tid som förflutit sedan den senaste deglaciationen i läget för observationen (LandAge). Även storleken på variabeln Nettonederbörd, observationens nivå i förhållande till Högsta Kustlinjen (HKDepth) samt ett antal av de variabler som beräknats från Lantmäteriets DEM, är klart betydelsefulla. Betydelsen av tillagda variabler för Marina gränsen samt Bergart är högst begränsad.



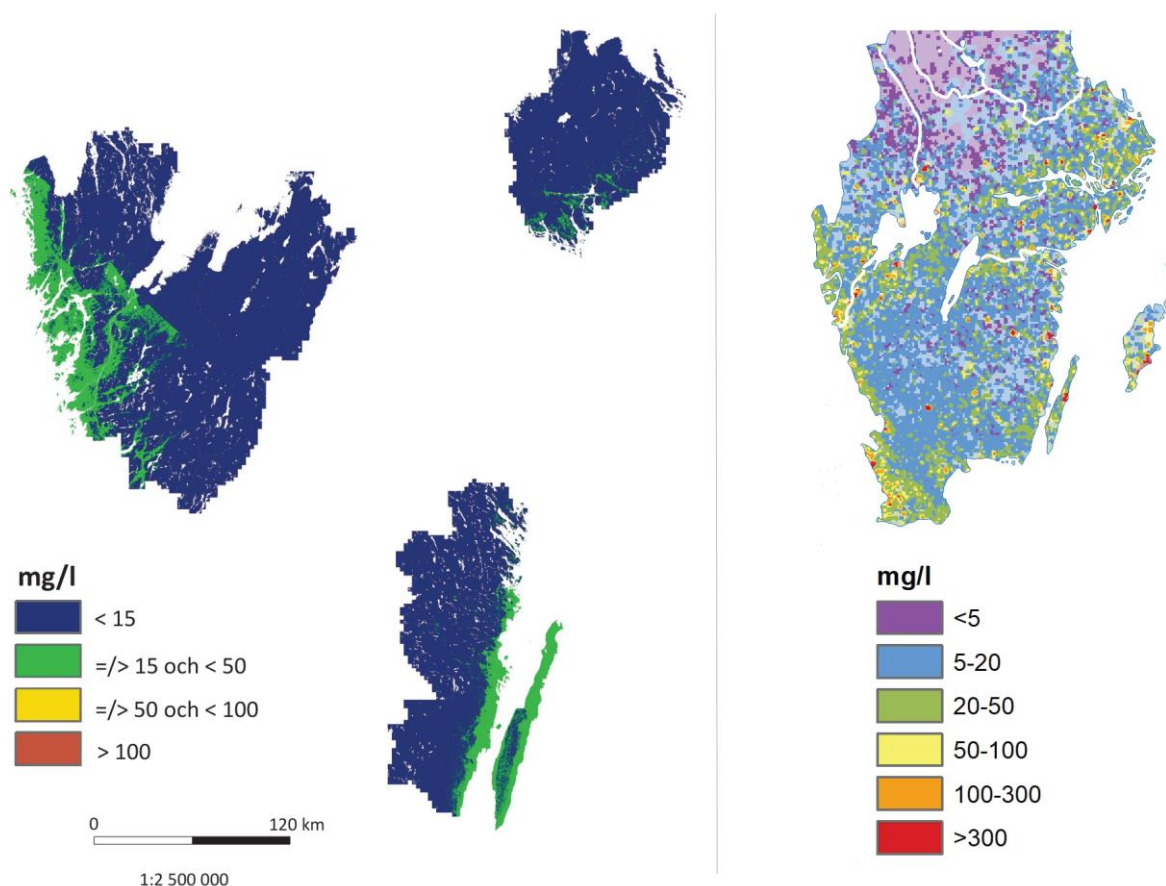
Figur 5. Beaktade miljövariablers betydelse i slutlig modell.

Predikterade kloridhalter

Predikterade kloridhalter, i haltintervall, i grundvatten i jord respektive berg ges av figur 6. I samma figur redovisas ett urklipp av den interpolerade haltkartan från arbetet med de nya bedömningsgrunderna (se figur 2).

Generellt predikteras relativt begränsade kloridhalter i grundvatten i jord. I Uppsala län dominerar det lägsta haltintervallet (<15 mg/l, ”opåverkat”) helt. I Västra Götaland och Kalmar län indikeras förhöjda halter över något större, framförallt kustnära områden. Modellen predikterar till synes inga halter ≥ 50 mg/l. I jämförelse med den interpolerade kartan över kloridhalter tycks utfallet av modelleringen överlag indikera något lägre halter. Interpoleringen indikerar större område med förhöjda kloridhalter, samt högre halter utmed framförallt kustlinjer men även fläckvis i vissa områden inom samtliga län.

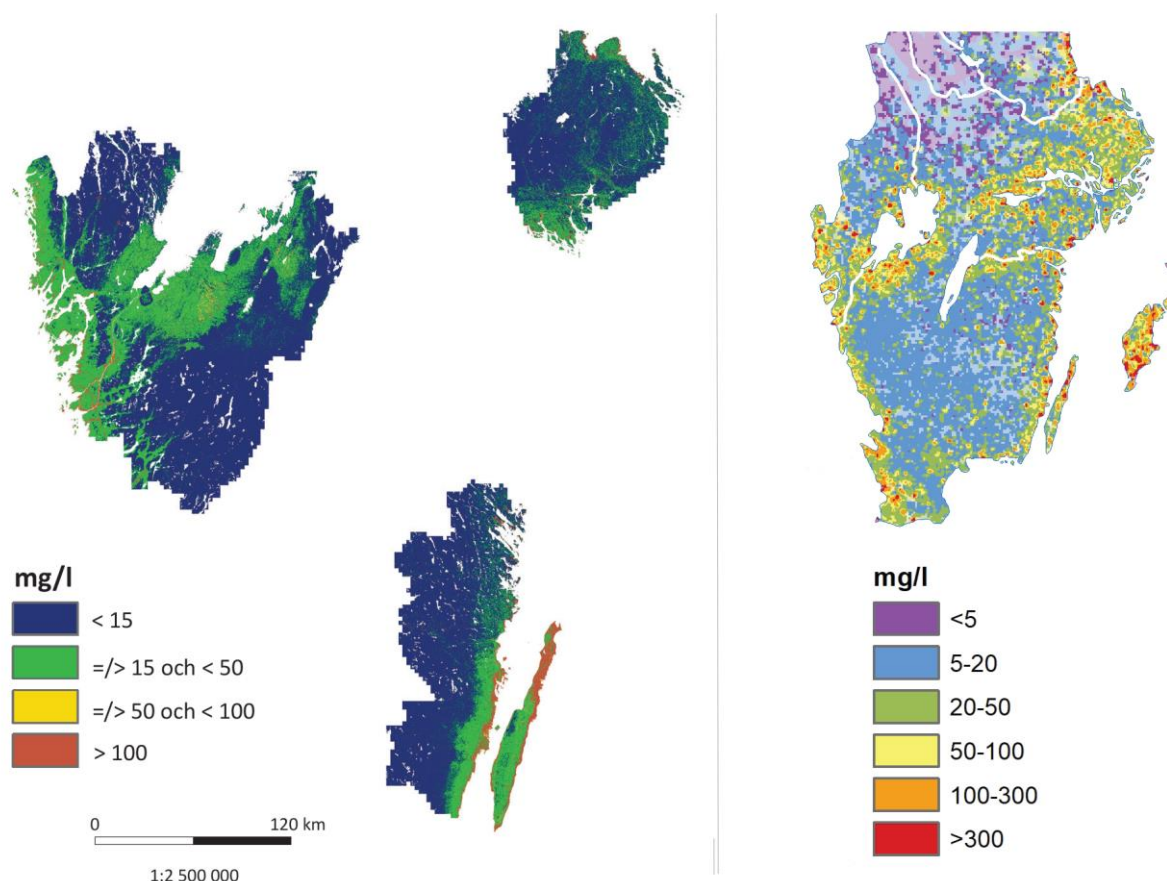
En struktur syns SV om Vänern, där modellen indikerar en skarp, rak gräns mellan opåverkat och påverkat (>15 mg/l) grundvatten. Mönstret bedöms inte spegla reella förhållanden utan snarare vara ett resultat av någon modellteknisk företeelse, alternativt någon enskild miljövariabel.



Figur 6. Predikterade kloridhalter, i haltintervall (mg/l), i grundvatten i jord i Uppsala, Västra Götaland och Kalmar län (vänster). Till höger ett urklipp från kartan i figur 2, som illustrerar uppmätta och interpolerade halter i södra Sverige. Notera att de olika kartorna är framtagna med olika metoder och datamängder samt att haltklasser och färgskalor skiljer sig åt.

För grundvatten i berg (figur 7) predikteras överlag en högre grad av kloridpåverkan, vilket också är i enlighet med vad som kan förväntas. Det lägsta haltintervallet bland de så kallat påverkade klasserna, dvs. 15–50 mg/l, dominerar helt. Så gott som inga prediktioner syns inom haltintervallet 50–100 mg/l, däremot syns en del prediktioner av halter i det högsta haltintervallet dvs. >100 mg/l. Denna typ av speciellt höga halter predikteras framför allt närmast kusten och utmed strandlinjer (samtliga län), fläckvis i ett område söder om Vänern samt mer genomgående över norra Öland.

Som för modellresultaten avseende kloridhalter i grundvatten i jord, tycks motsvarande resultat för grundvatten i berg generellt prediktera något lägre halter än vad som ges av den interpolerade kartan. Dock är överensstämmelsen avseende opåverkade områden och områden med speciellt höga halter (>100 mg/l) överlag relativt god, med undantag för Uppsala län.



Figur 7. Predikterade kloridhalter, i haltintervall (mg/l), i grundvatten i berg i Uppsala, Västra Götaland och Kalmar län (vänster). Till höger ett urklipp från kartan i figur 2, som illustrerar uppmätta och interpolerade halter i södra Sverige. Notera att de olika kartorna är framtagna med olika metoder och datamängder samt att haltklasser och färgskalor skiljer sig åt.

Predikterad sannolikhet

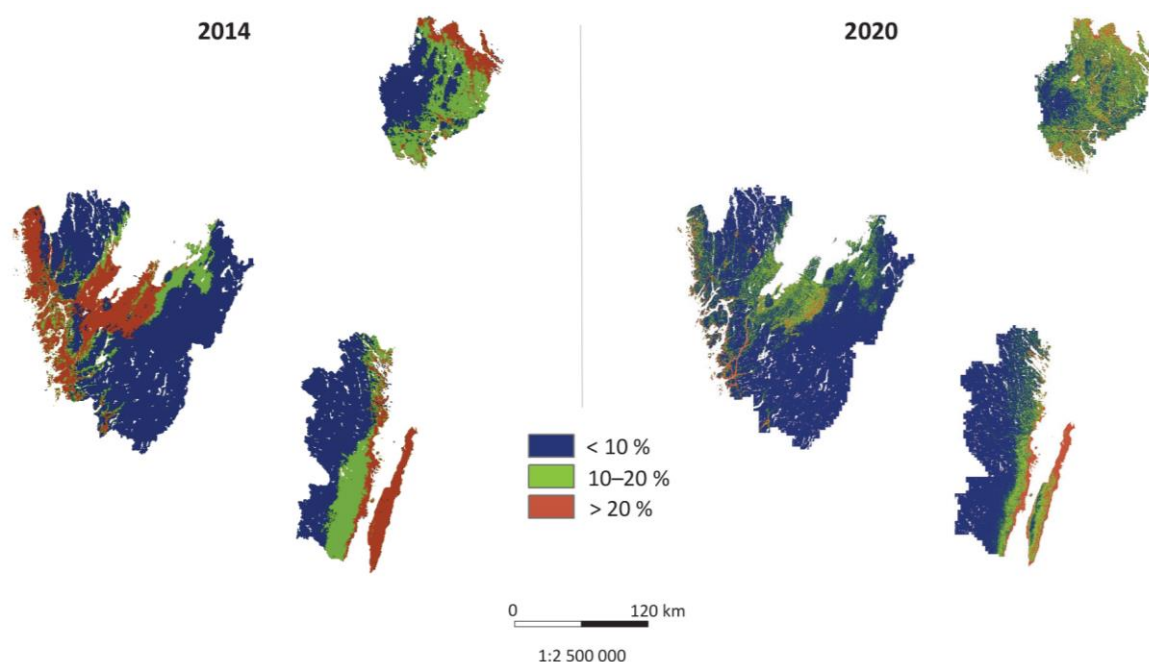
Som ett komplement till de modellerade haltkartorna, producerades även en karta som visualiserar predikterad sannolikhet för förekomst av kloridhalter högre än 100 mg/l i grundvatten i berg. En sannolikhetskarta är en typ av kartprodukt som inte kan tas fram enbart genom interpolering av uppmätta halter. Kartan bedöms speciellt intressant i och med att 100 mg/l utgör det beslutade riktvärdet inom vattenförvaltningen och Livsmedelsverkets gränsvärde för tjänligt dricksvatten med anmärkning.

Kartan togs också fram i syfte att möjliggöra en mer direkt jämförelse med den nationella sannolikhetskartan som togs fram av SGU 2014. En jämförelse mellan kartorna ges av figur 8. Den äldre sannolikhetskartan har en upplösning om 250 × 250 m.

Sett till ytor med en bedömt förhöjd sannolikhet (>10%) för förekomst av grundvatten med kloridhalter >100 mg/l, är överensstämmelsen mellan de båda produkterna relativt god. Vissa skillnader i detta avseende föreligger dock. I Kalmar län, inklusive Öland, ger resultaten från maskininläringen att sannolikheten för förhöjda kloridhalter avtar relativt närmre kusten än vad resultaten från tidigare regressionsanalys indikerar. I Uppsala län är situationen den omvända, dvs maskininläringen indikerar förhöjd sannolikhet över större ytor jämfört med regressionsanalysen.

Mer utmärkande är att regressionsanalysen överlag indikerar en högre grad av sannolikhet för förekomst av salt grundvatten än vad maskininlärningsmodelleringen ger. En del av förklaringen till detta är att 2014 års kartprodukt inkluderar en fast buffert för klass hög (>20% sannolikhet) omfattandes tätorter, vägar samt en buffertzona om 200 m till dessa och strandlinjer i salta och bräckta vatten. Detta enskilt förklarar dock inte de relativt stora skillnader i utbredning av denna typ av områden som ges av en jämförelse mellan de båda kartprodukterna.

Det bör understrykas att de olika sannolikhetsbedömningarna inte bara grundas i olika metoder utan att även dataunderlaget skiljer sig åt. Det är således inte möjligt att jämföra de två rakt av, och det är heller inte möjligt att använda det ena alternativet som ett facit i förhållande till den andra.



Figur 8. Bedömd sannolikhet för förekomst av kloridhalter >100 mg/l i grundvatten i berg enligt SGU 2014 (vänster) respektive modellering genom maskininläring (höger).

DISKUSSION

Representativitet och upplösning

I maskininlärningsmetoden tränas en matematisk modell upp, baserat på observationer från hela landet. Den upptränade modellen beskriver därmed samband mellan miljövariabler och kloridhalter på ett nationellt plan. Resultatens representativitet, det vill säga lämplig skala för redovisning och tolkning, bedöms därför i första hand vara nationell, eller möjligtvis regional. För att bedöma representativiteten mer lokalt krävs fördjupad utvärdering. I detta avseende kunde det vara intressant att låta utföra maskininlärning och modellering med en mer lokal datamängd.

En annan aspekt som kan vara relevant att testa i framtida liknande studier är att variera tillämpade klassgränser, alternativt modellera faktiska halter snarare än haltintervall. Valet föll på klasser denna gång, på grund av att SGUs modellering av jordarter genomförts på motsvarande vis. Som resultaten visade gjordes dock få prediktioner inom klassen 50–100 mg/l. Möjligen skulle det vara fördelaktigt att, i fall av fortsatt modellering av klasser/haltintervall, begränsa sig till ett mindre antal klasser såsom under/ovan ett specifikt utvalt gränsvärde. Bedömningsvis torde dock maskininlärning och prediktion av faktiska observationer (halter) snarare än klasser, till följd av större grad av nyansering och frihet, leda till en bättre modellprestation totalt sett.

Upplösningen för resultaten är som nämnt 2×2 m, vilket motsvarar den cellstorlek för vilken jordarter modellerats. Det kan diskuteras om en så pass hög upplösning är relevant och motiverad i fallet modellering av kloridhalter, åtminstone om avsikten är nationell och/eller regional redovisning och användning. Frågan borde rimligen utredas närmre för framtida analyser då den är central för modelleringens tidsåtgång.

Förbättringsförslag

Den framtagna modellen ger i stort förhållandevis rimliga resultat som är intressanta och användbara för att visualisera och diskutera mönster och samband på nationell och regional nivå. Valideringsresultaten och jämförelse med tidigare/liknande kartprodukter/analyser visar dock att det finns utrymme för förbättringar.

I första hand torde ytterligare och i vissa fall förfinade miljövariabler inkluderas i datasetet. Konkret föreslås, baserat på tidigare studier och kända samband, att jordart, jorddjup, markanvändning, geografiskt läge och grundvattenuttag beaktas. Jordart, jorddjup, markanvändning och geografiskt läge i form av X- respektive Y-koordinat var ursprungligen tänkta att ingå i tillämpad uppsättning miljövariabler, men så blev ej fallet.

Jordart och jorddjup föll bort i och med att metoden kom att byggas på den metod som tagits fram för prediktion av just jordart. I datasetet för att prediktera jordarter ingår varken jorddjup eller – av förklarliga skäl – jordart. Markanvändning fick uteslutas i och med att kloridobservationernas position inte täckte in samtliga typer av markanvändning. Detta gör att den upptränade matematiska modellen inte kan se några mönster och inte heller prediktera kloridhalter inom alla typer av markanvändning. X- och Y-koordinat uteslöts mot bakgrund av problematiska erfarenheter vid modellering av jordarter då dessa variabler beaktats. Grundvattenuttag bortprioriterades på grund av tidsbrist och då relevant underlag tenderar att vara svårt att få tag på.

I och med att varken markanvändning eller geografiskt läge ingått i metoden har inga slutsatser kunnat dras huruvida det går att se samband mellan predikterade kloridhalter och avstånd från exempelvis kust och saltade vägar. Att addera miljövariabler som representerar detta kan således potentiellt vara en utvecklingsmöjlighet värd att testa i framtida studier.

Avseende de tillagda miljövariablerna, var betydelsen som nämnt högst begränsad för såväl läge i förhållande till marina gränsen som bergartstyp. Sannolikt vore det i fråga om bergart mer relevant att beakta en eller flera mer detaljerade variabler, som sammantaget är mer och bättre beskrivande i fråga om såväl bergartskemi som grundvattenflöde. Generellt gångbara beskrivande raster avseende Sveriges berggrund (kemi, struktur) vore onekligen värdefullt för att underlätta även för framtida bredare tillämpning av maskinlärning och modellering.

Vad gäller marina gränsen så är denna gräns behäftad med osäkerhet och i behov av uppdatering i delar av Sverige. Mot bakgrund av detta, och det faktum att variabeln är binär utan någon form av nyansering, är det inte helt förvånande att betydelsen för prediktion av specifika halter är begränsad. En liknande slutsats har dragits även av Müllern & Thunholm (1998). I föreliggande fall är det tydligt så att de mer nyanserade miljövariabler som relaterar till Högsta Kustlinjen (LandAGe, HKDepth) bättre speglar variationen hos kloridobservationerna. Möjligen vore det därför, i framtida liknande studier, bättre att tillämpa en eller fler miljövariabler som mer mångsidigt speglar läge i förhållande till de olika insjöar och hav, med olika salthalt, som ryms inom Östersjöns utveckling.

En annan aspekt som skulle kunna förbättra metoden är att utveckla representationen av den variation i kloridhalt som kan förväntas med djupet. Uppdelningen av datamängden i jord respektive berg innebär i viss mån en djupindelning, men att inkludera denna tredje dimension i analysen på ett bättre och mer detaljerat sätt skulle på sikt kunna ge ett mer användbart resultat. Ett alternativ skulle kunna vara att utföra maskininlärning och prediktion separat för olika djupintervall. Detta skulle å ena sidan reducera antalet observationer per dataset, men å andra sidan innebära en lägre grad av variabilitet inom respektive dataset och därmed möjligen även bättre modellprestation. Det bör dock nämnas att information om uttagsdjup saknas för en stor del av de kloridobservationer som använts i denna studie och att ett krav om sådan information därmed skulle reducera antalet tillgängliga observationer.

Slutligen bör en utökning av antalet kloridobservationer i datasetet lyftas som en förbättringsmöjlighet. SGU har fler kloridobservationer i grundvatten i såväl jord som berg, tillgängliga i andra datamängder än ”enskilda brunnar”. I det fall fler observationer används borde modellens prestation öka. Man bör dock vara uppmärksam på att andra datamängder kan inrymma observationer från andra typer av provtagningsplatser än just enskilda brunnar. Då olika typer av provtagningsplatser kan vara representativa för exempelvis olika ytor och djup bör effekter av att blanda observationer från dessa beaktas.

Ytterligare tillämpningsområden kopplat till SGUs grundvattenverksamhet

Metodiken bedöms vara tillämplig såväl för analys och kartläggning av andra substanser förekommande i grundvatten, dvs. kvalitativa frågeställningar, som för kvantitativa studier. Exempelvis kan man tänka sig modellering av grundvattennivåer, hydraulisk konduktivitet och uttagskapacitet.

Vidare skulle metoden gå att tillämpa för att modellera kvalitativa och kvantitativa förhållanden under förändrade omständigheter, avseende styrande miljövariabler. I fallet kloridhalter kan man exempelvis tänka sig att utvärdera effekterna (förändrade halter/förändrad sannolikhet för specifika halter) av ett förändrat klimat (förändrad nettonederbörd mm), förändrad markanvändning och förändrade grundvattenuttag.

SLUTSATS

Metodik för modellering av jordarter med hjälp av maskininlärning, har anpassats och tillämpats för modellering av kloridhalter i grundvatten i jord och berg. Resultaten är överlag rimliga, men indikerar samtidigt att uppsättningen av datamängder och dess respektive innehåll behöver utvecklas för att uppnå ett bättre och mer användbart resultat. Potentialen att kunna utnyttja metodiken för olika analyser av såväl kvalitativa som kvantitativa hydrogeologiska frågeställningar bedöms som stor.

Arbetssättet erbjuder en möjlighet att förhållandevis enkelt dra nytta av, samköra och maximera utväxlingen av de stora, och ofta på ett eller annat sätt relaterade, datamängder som finns såväl inom SGUs olika verksamheter, som hos andra myndigheter i Sverige. I och med detta kan sannolikt många samordningsvinster göras inom SGU om det framtida arbete med maskininlärning bedrivs över enhets- och avdelningsgränser, samt med god dialog med andra aktörer.

REFERENSER

- Aller, L., Bennet, T., Lehr, J. H., Petty, R. J. & Hackett, G., 1987: DRASTIC: A Standardized System for Evaluating Ground Water Pollution Potential Using Hydrogeologic Settings. *US EPA Report 600/287/035*. U.S. Environmental Protection Agency, 622 s.
- Anderberg, J., 1997: Beskrivning till kartan över Grundvattentillgångar i Upplands Bro kommun. *Sveriges geologiska undersökning An 4*, 124 s.
- Becher, M., Sohlenius, G., & Öhrling, C., 2019: Sur sulfatjord – egenskaper och utbredning. *SGU-rapport 2019:13*. Sveriges geologiska undersökning, 37 s.
- Breiman, L., 2001: Random Forests. *Machine Learning* 45, 5–32.
- Bucht, M. & Larsson, P-O., 1995: *Risikobedömning för salt grundvatten – förslag till ny metodik*. Magisteruppsats 1995:6, Kungliga Tekniska Högskolan, 106 s.
- Ebert, K., Ekstedt, K. & Jarsjö, J., 2016: GIS analysis of effects of future Baltic sea level rise on the island of Gotland, Sweden. *Nat. Hazards Earth Syst. Sci.* 16, 1571–1582.
- Friedman, J.H., 2001: Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics* 29, 1189–1232.
- Hjerne, C., 2018: Särskilt utsatta områden. *SGU-PM dnr 314-951/2018*, Sveriges geologiska undersökning, 29 s.
- Lindberg, J., Olofsson, B. & Gumbrecht, T., 1996: Risk mapping of groundwater salinization using Geographical Information Systems, 14th Salt Water Intrusion Meeting (SWIM). *Rapporter och Meddelanden* 87, Sveriges geologiska undersökning, 188–197.
- Lobo-Ferrera, J.P., Chachadi, A.G., Diamantino, C. & Henriques, M.J., 2007: Assessing aquifer vulnerability to seawater intrusion using the GALDIT method: Part 1 – application to the Portuguese Monte Gordo aquifer. I J.P. Lobo Ferreira & J.M.P. Viera (red.): *Proceedings of Water in Celtic Countries: Quantity, Quality and Climate Variability*. IAHS Publication 310, International Association of Hydrological Sciences, 161–171.
- Müllern, C-F. & Thunholm, B., 1998: Djupt, salt grundvatten i Sverige. Rapport utarbetad på uppdrag av SKB dnr 08-172/1998. Sveriges geologiska undersökning, 29 s.
- SGU, 2013a: Bedömningsgrunder för grundvatten. *SGU-rapport 2013:01*. Sveriges geologiska undersökning, 238 s.
- SGU, 2013b: Föreskrift om miljökvalitetsnormer och statusklassificering för grundvatten. *SGU-FS 2013:2*. Sveriges geologiska undersökning, 10 s.

- SLV, 2017: Livsmedelsverkets föreskrifter om ändring i Livsmedelsverkets föreskrifter (SLV-FS 2001:30) om dricksvatten. *LIVSFS 2017:2*. Livsmedelsverket, 48 s.
- Stroeven, A. P., Hättstrand, C., Kleman, J., Heyman, J., Fabel, D., Fredin, O., Goodfellow, B.W., Harbor, J.M., Jansen, J.D., Olsen, L., Caffee, M.W., Fink, D., Lundqvist, J., Rosqvist, G.C., Strömberg, Bo., & Jansson, K., 2016: Deglaciation of Fennoscandia. *Quaternary Science Reviews*. 91–121(147).