

Noteringar till databearbetning och beräkningar av trender i grundvattennivåer

Claudia von Brömssen och Erik Olsson

2024-10-23

Nedladdning av data

Data och stationsinformation laddas ner via SGUs API enligt en stationslista som tillhandahålls av länsstyrelsen. 694 stationer ingick i urvalet. De flesta stationerna hade både lodade och loggade mätningar under olika tidsperioder. Observationer som är kodade som "misstänkt felvärde" eller "påverkat av testet" tas bort.

Beräkning av variabler

Värden från 2024 utesluts för att kunna göra beräkningar för årsdata med enbart kompletta år. Även första mätåret i varje serie tas bort, dels för att undvika data som kan vara osäkra i uppstarten av loggern eller mätningen och dels för att ha fullständiga år för att beräkna årsmedianer, samt årsminimum och -maximum.

Årsmedian

För loggade data fylls först saknade värden genom linjär interpolation. Detta görs genom att först beräkna median per station och dag och sen interpolera dagar som inte har något värde. Om glappet är större än 40 dagar så fylls det inte. Årsmedian beräknas som median av dyngsmedianerna.

För lodade data extraheras mätningen i mitten av månaden (mellan dag 8 och dag 22). Om flera mätningar görs inom denna tidsperiod så tas medelvärde av dessa.

Loggade och lodade data kombineras till en gemensam serie, där loggade data används för de år de är tillgängliga och lodade data för tidigare år. Årsmedianer beräknas enbart för lodade data som har observationer för alla 12 månader för ett specifikt år.

Årsminimum och årsmaximum

Årsminimum och årsmaximum beräknas från dagsmedianer av loggade data som 5% och 95% percentil. För lodade data beräknas årsminimum och årsmaximum enbart för år som har minst en observation varje månad och då igen som 5% och 95% percentiler. Att använda faktiska minimum och maximumvärden ökar risken att få med felaktiga värden, därav valet av percentiler. Om detta är rätt sätt att hantera data behöver kanske utvärderas när verktyget har testats.

Månadsvärden

Månadsmedianer beräknas för loggade serier och kombineras med månadsvärden för lodade serier. Data delas upp i de enskilda månaderna för att skapa serier på årsbasis för varje månad.

Enheter och värden under detektionsgräns

Grundvattennivåer anges i meter under markytan. Detektionsgränser i form av värden mätt under eller över mätgräns finns i data och tas med som det rapporterade värdet. Värdena antas ha liten effekt för de beräknade medianerna, medan medianerna skulle kunna över- eller underskattas om dessa värden tas bort från analyserna.

För att trenderna ska vara intuitiva så kodas variabeln till negativa värden. Lägre negativa värden är lägre grundvattennivå. Minskade/ökande statistiska trender anger minskade/ökande grundvattennivåer.

Urval av mätvärden och stationer

Trender beräknas enbart för stationer där det finns minst 6 år med (icke-saknade) observationer. Om serien uppvisar en längre period (mer än 10 år) utan mätdata mellan perioden det har mätts med klucklod och perioden där det har mätts med logger, så presenteras enbart data från loggern.

Beräkning av Generaliserade additiva modeller (GAM)

Modellerna som beräknade utjämnade trendkurvor baseras på årsdata. För att få rimliga utjämnningar för alla olika variabler och stationer används en styrning av parametern k (gam i mgcv paketet), som väljs som halva antalet observerade årsmedelvärden. Detta leder till att modellenpassningen startar med en rimlig komplexitet och anpassningen alltid kan genomföras. För infärgning av kurvan (mörkblått indikera ökande och ljusblått indikerar minskande trender) används 95% konfidensintervall för derivatan av kurvan.

Beräkning av Mann-Kendall tester

Mann-Kendall tester beräknas på årsvisa data, dvs årsmedian, årsminimum och årsmaximum, samt månadsmedianer som årsserier (ett januarivärde per år). Mann-Kendall tester beräknas för de senaste sex åren (om det finns data för varje år), för de senaste 12 åren (om det finns minst 6 observationer under dessa 12 år), samt för hela tillgängliga serien.

Filer på Github

Alla filer för beräkningar av tidstrender, samt input- och outputfiler finns på Github (<https://github.com/claudiavonbromssen/Trendverktyg-grundvatten>).

Indata:

Nedladdningar via API (nedladdning gjord den 21 oktober 2024)

Filer med stationsinformation

Beräkningar i R:

Inläsningsfilen _grundvattennivåer.Rmd: Inläsning via API, data sparas som csv-filer.

Inläsningsfilen _stationsinformation.Rmd: Inläsning via API, data sparas som csv-fil.

Databearbetning _grundvattennivåer.Rmd: Beräkning av variabler och trendanalyser

Export grundvatten.Rmd: döpa om variabler och fixa outputformat

Outputfiler:

Tidstrender grundvatten – Data.csv: Serier för variablerna inklusive GAM-trender

Tidstrender grundvatten – Mann-Kendall.csv: Mann-Kendall testresultat

Tidstrender grundvatten – Metadata: Metadata som beskriver stationerna

Tidstrender grundvatten – Mätvariabler: Beskriver variablerna, inklusive enhet