

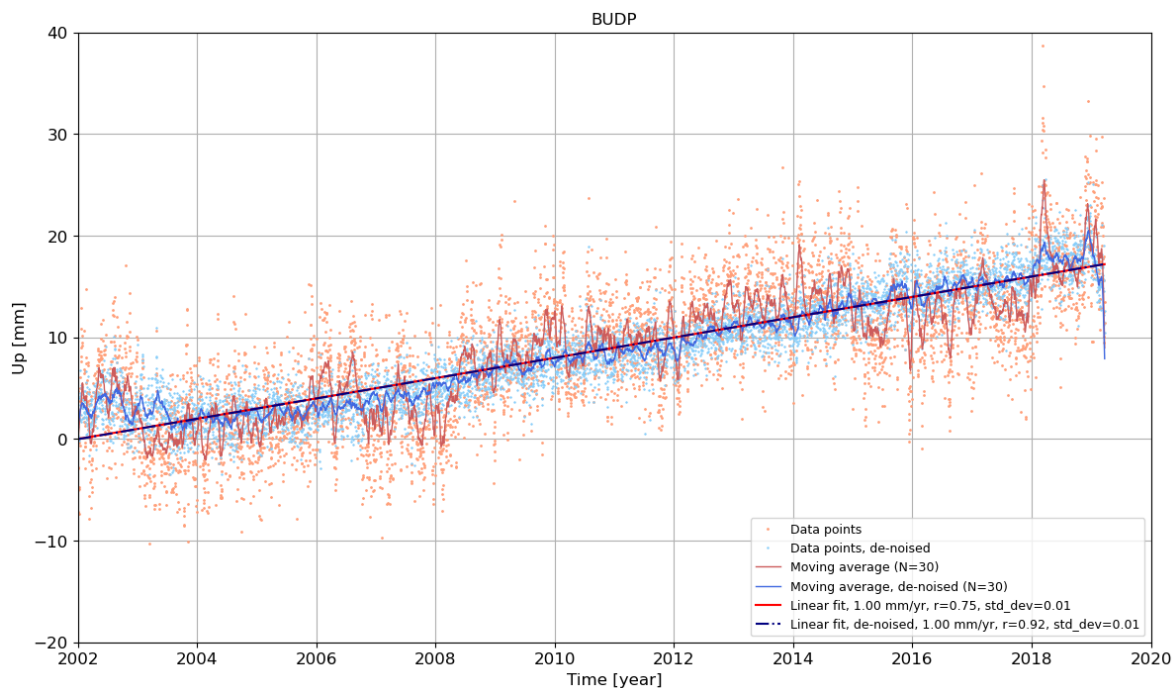
1 Introduktion

Denne rapport beskriver python programmet gnssts, som er skrevet i 2022. Formålet med programmet er at undersøge hvordan antallet af målepunkter i en tidsserie påvirker, om det data man har er retvisende. Målingerne der undersøges her er fra permanente GNSS stationer, som sub-samples med forskellige antal punkter. I denne rapport vises kun resultater fra BUDP, men resultaterne er konsistente med de andre stationer i undersøgelsen. Koden kan findes på [github](#), hvorfra resultaterne kan genskabes. Resultaterne kan også (midlertidigt) findes på F-drevet i mappen F:\GRF\Medarbejdere\MJENS\gnssts\out.

2 GNSS Time Series

2.1 Præsentation af data

Data i denne rapport er fra stationen BUDP, og er taget fra 2002, og frem til og med 2019. Figur 1 viser data fra stationen, samt tendenslinje og rullende gennemsnit for disse. Både tendenslinje og rullende gns. er lavet for de rå data såvel som for de-noised data. De-noising er udført ved at detrende data fra samtlige stationer, og derefter finde gennemsnittet af disse, for at få et udtryk for den generelle støj.



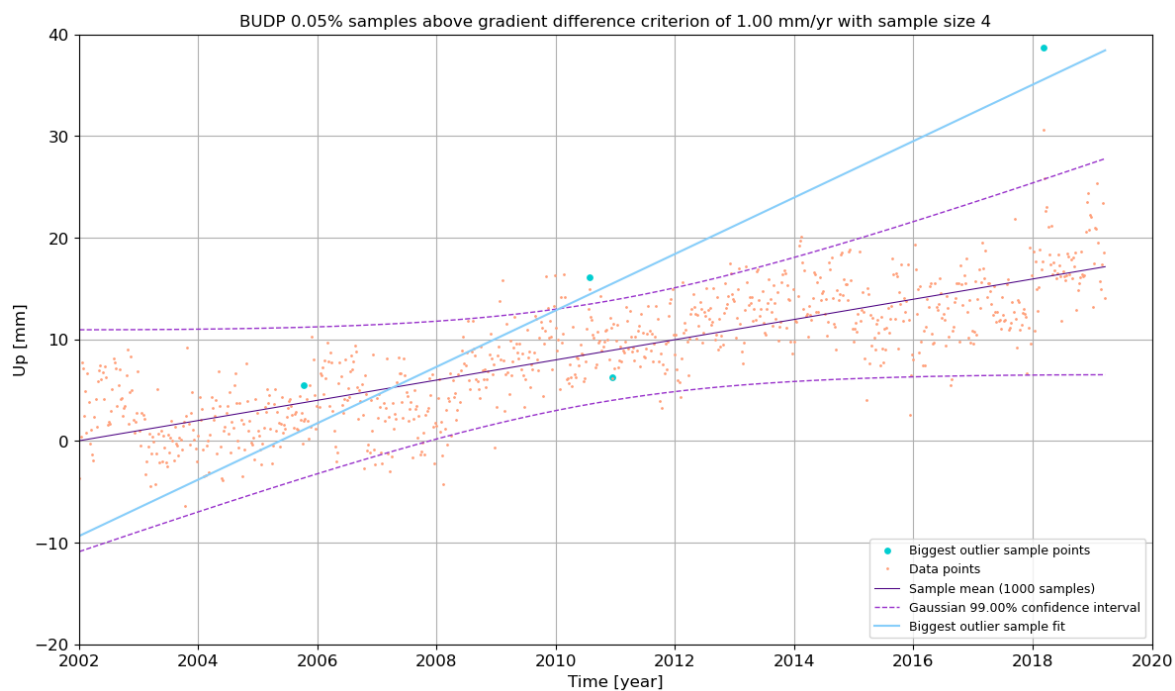
Figur 1: Data fra BUDP med rullende gns. samt tendenslinje. Begge dele er lavet både for de rå data og de-noised data.

Som figuren viser er der ændringer i det rullende gennemsnit, men stort set ikke i tendenslinjen. Det samme gælder for de andre stationer. Derfor bruges blot den rå data i den videre sampling.

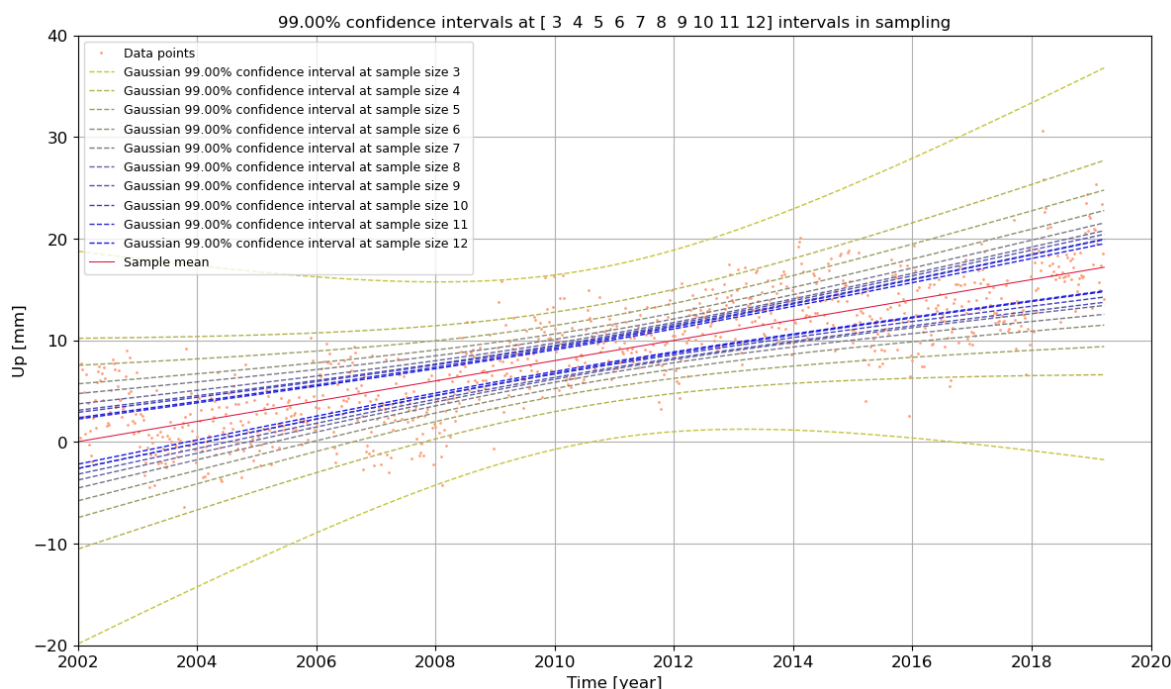
2.2 Sampling

Metoden til sampling er som følgende: Først splittes det pågældende datasæt op i N lige store dele, hvorefter et enkelt punkt udtages fra hver del. I figur 2 er dette koncept vist med en sample størrelse på 4 og 1000 samples. Ud fra disse

er der konstrueret et 99% konfidensinterval for at undersøge spredningen af samples. Det er dog svært at undersøge spredning ud fra denne sampling alene.



Figur 2: Resultat af sampling med BU DP data. I dette tilfælde er det lavet med 1000 samples af 4 punkter. Der vises et 99% konfidensinterval i lilla, samt det største outlier-sample i blå, med tendenslinje for denne.



Figur 3: Konfidensintervaller for flere sample størrelser fra BUDP data. Her er samlet med sample størrelse 3-12. Den mindste sample størrelse giver det største konfidensinterval og vice versa.

Derfor viser figur 3 99% konfidensintervaller for sampling med flere sample størrelser. Dette er for at give en idé om, hvornår der er nok samples til, at man kan stole på, at data er retvisende for den reelle tendens. Jo mindre sample størrelsen er, jo bredere er konfidensintervallet, hvilket ses på de mere gule intervaller i figuren. Modsat giver en større sample størrelse et smallere konfidensinterval hvilket ses på de mere blå intervaller. Det kan observeres i figuren at omkring en sample størrelse på 7-8, går det væsentligt langsommere med at konfidensintervallet bliver smallere, ved en yderligere stigning i antal samples. Denne tendens er gennemgående for alle stationer. Laves i stedet et 95% konfidensinterval, ses denne tendens allerede ved 5-7 samples alt efter station.

3 Konklusion

Da et 95% konfidensinterval viser at der skal 5-7 samples til før en relativt stabilitet opnås, og et 99% konfidensinterval viser 7-8 samples, anbefales det at der som absolut minimum anvendes 5 målepunkter, men helst 7 eller flere.