

# Øving 6: Gruppeprosjekt del 1

## Læringsmål

Dere skal lære å jobbe med en programmeringsoppgave i grupper. Dere skal lære enkel data-analyse og plotting. Dere skal lære grunnleggende bruk av Git og Github.

## Om CSV filer

For denne øvingen vil det bli delt ut to .csv filer. De vil ha etternavn .csv.txt slik at dere får lastet dem ned riktig. Erfaringen min er at hos en del av dere vil ei fil med etternavn .csv bli åpnet i Excel og så lagret som Excel-filer (etternavn .xlsx) i stedet for .csv filer lokalt på deres egne datamaskiner. Excel-filer er binærfiler. For å lese ei Excel-fil må man bruke eksterne pakker som for eksempel Pandas, som ikke er pensum i DAT120.

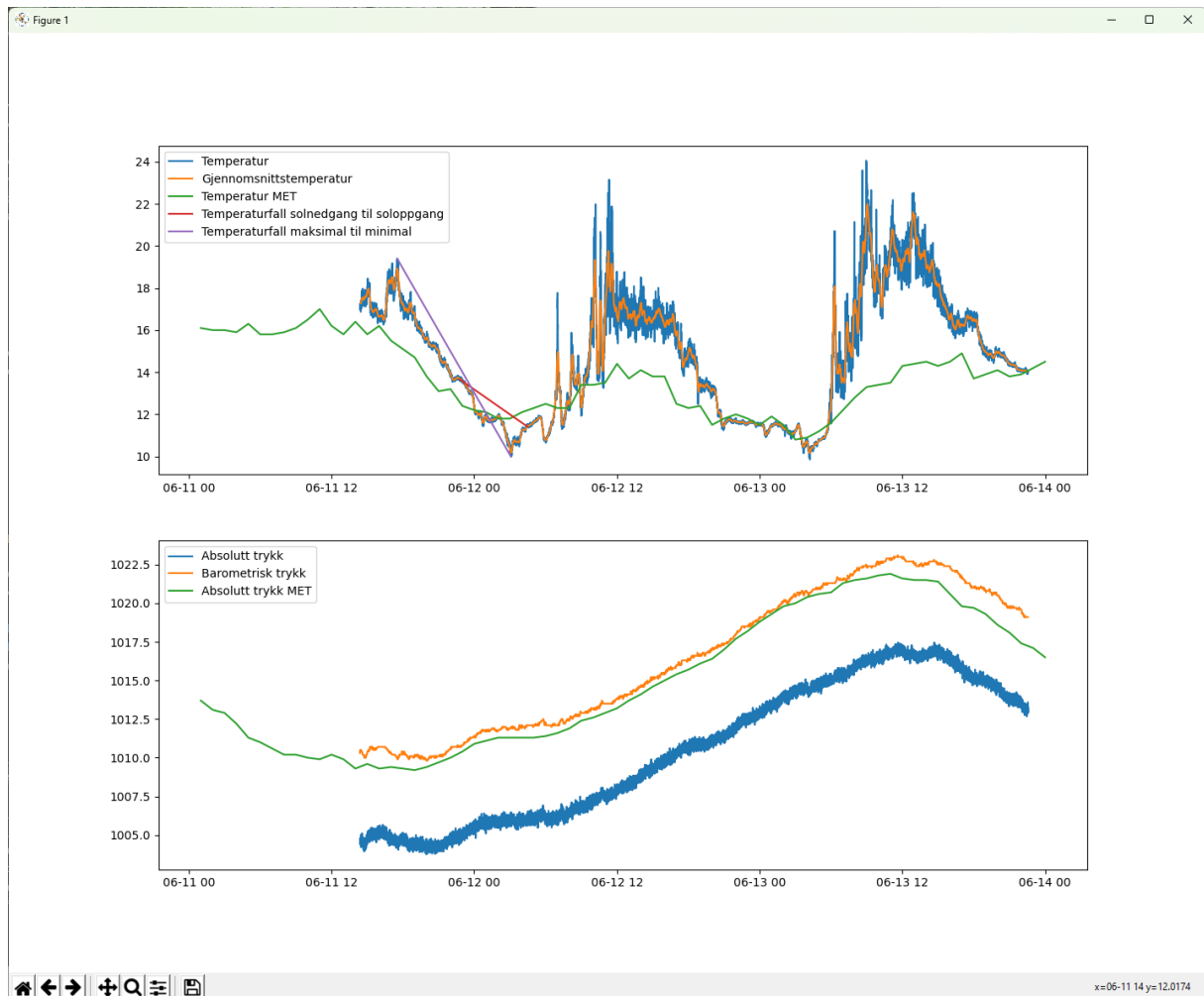
## Overordnet oppgavebeskrivelse: meteorologiske data

Dere skal lese inn to separate .csv filer som inneholder sammenliknbare værddata. Den ene fila er fra en lokal værstasjon her på UiS og kommer fra professor Rune Wiggo Time. Den andre fila er fra Sola værstasjon og er lastet ned fra Meteorologisk institutt. Denne siste fila er under en Creative Commons lisens som lar oss bruke dataene så lenge vi sier hvor vi fikk dataene fra. Begge filene dekker perioden 11. til 13. juni 2021. Begge filene inneholder data om temperatur og trykk, men med litt ulik oppløsning. Den første fila har ei linje for hver 10. sekund. Den andre fila har ei linje for hver time.

Merk: I fila for den lokale værstasjonen så skifter formatet for datoer og tidspunkt omtrent 2/3 ut i fila. Det er obligatorisk å håndtere formatet som er i første del av fila og så ignorere linjer med ugyldig datoformat. Formatet i første del av fila er på formen MM.DD.ÅÅÅÅ TT:MM. Dette betyr at det er to siffer for måned først, deretter et punktum, deretter to siffer for dag, deretter et punktum, deretter fire siffer for år, deretter et åpenrom, deretter to siffer for timer, deretter et kolon og til slutt to siffer for minutt. Dette er et normalt amerikansk datoformat. Sekundene ligger ikke her, men ligger i stedet i neste kolonne, sekunder siden start. Timer er oppgitt i 24-timersformat slik at 14 betyr to om ettermiddagen.

I fila for meteorologisk institutt er dato og tidspunkt oppgitt på formatet DD.MM.ÅÅÅÅ TT:MM.

Dere skal lage et plott som likner det på følgende figur:



## Deloppgaver

- En av dere lager et Github repository for denne øvingen. De andre inviteres med i dette repoet.
- Hver student i gruppa skal klonе dette Github repositoryet til sin lokale datamaskin og jobbe i denne mappa med øvingen. Koden dere skriver i denne øvingen skal lastes opp til Github repoet.
- Tenk over hvordan dere skal fordele oppgavene mellom gruppemedlemmene og hva som bør gjøres felles. Dere kan se på hvilke funksjoner dere skal definere i de ulike deloppgavene og så tenke over hvilke av disse funksjonene som kan skrives uavhengig av de andre.
- Analyser de to filene og last inn dataene fra dem i lister. Bruk ei liste for hver kolonne i hver fil, og sørg for at indeksen inn i disse listene passer sammen. Så for eksempel temperatur[2] inneholder temperaturen i tidspunkt[2].
- Finn ut hvordan dere håndterer datoer og tidspunkt, enten datetime-objekter eller separat lagring av komponentene. Merk at bare dag, timer og minutter strengt tale er nødvendig her. Merk at dere trenger en enkeltverdi for x-aksen så dere bør her enten bruke et datetime objekt eller kombinere dag, timer og minutter i en verdi slik at det sorteres og plottes riktig. Husk at de to filene har ulike formater for datoer og tidspunkter, men dere skal plote data fra

begge filene i samme plott, så dataene dere skal plote må ha datoer og tidspunkter i samme format.

- f) Plott temperatur fra begge filene i samme plott. Merk at du kan bruke datetime-objekter som koordinater for x-aksen i et matplotlib plott. Dette gir de blå og grønne linjene i øverste plott i figuren.
- g) I den første fila er det en god del svingninger som representerer støy i temperaturmålingene. For å redusere dette så regner man ofte ut et gjennomsnitt av flere temperaturer og plotter det. Lag en funksjon som tar inn et liste med tider, ei liste med temperaturer, og et tall n som sier hvor mange verdier den skal regne ut gjennomsnittet for. For hvert gyldig tidspunkt skal den regne ut snittet av de n forrige målingene, den nåværende målingen, og de n neste målingene. Gyldige tidspunkter er alle tidspunkter hvor du ikke havner utenfor lista ved å gå n hakk bakover eller forover. Så skal funksjonen returnere lister med gyldige tidspunkter og gjennomsnittsverdier. Plott gjennomsnittsverdiene for  $n=30$  i samme plott som for temperatur. Dette gir den oransje linja i øverste plott.
- h) Finn og plott temperaturfall kveld og natt fra den 11. til den 12. juni for den første fila. For den obligatoriske delen av øvingen skal dere gjøre dette for tidspunktene 11. juni 2021 klokka 17:31 til 12. juni 2021 klokka 03:05. Dette tilsvarer den lilla linja i figur 1. Dette krever at dere finner hvilke indekser i lista over tidspunkter som har disse tidspunktene.
- i) Plott atmosfærisk trykk fra begge filene sammen med barometrisk trykk. Dette blir det nedre plottet i figuren.
- j) **Frivillig:** Håndter begge datoformatene i fila fra den lokale værstasjonen.
- k) **Frivillig:** I stedet for de faste tidspunktene i oppgave d), finn tidspunktet med høyest temperatur den 11. juni og lavest tidspunkt den 12. juni og bruk de tidspunktene i stedet for de som er oppgitt.
- l) **Frivillig:** Finn tidspunkter for solnedgang og soloppgang for de to datoene og bruk dette i stedet for tidspunktene som er oppgitt i oppgave d). Dere kan finne hvordan å hente dette fra met.no med litt søking på nettet.