Norges miljø- og biovitenskapelige universitet

Master's Thesis 2024 30 ECTS
Faculty of Science and Technology

Land-atmosphere Interaction During Convective Events in Ås, Norway

Randi Maud Mork
Environmental Physics and Renewable Energy

Preface

This master's thesis marks the end of my studies in Environmental physics and renewable energies and my time as a student at the Norwegian University of Life Sciences. My time as a student has given me many good memories, friends, and experiences that I will bring with me from Ås.

First I would like to thank my supervisors Mareile Wolff and Laura Ehrnsperger for all of the support and help through the process of writing this thesis. Thank you for all of your guidance through the world of precipitation and convection.

I would also like to thank my boyfriend for always supporting me through my academic journey with all its challenges and wins. Thank you to my family for supporting me, and my friends for always having someone to discuss everything with.

Randi Maud Mork

Ås, May 2024

Abstract

Land-atmosphere interactions describe the relationship between the surface conditions and the atmosphere. One such interaction that has been studied recently is the interaction between soil moisture and convection. Convective events are still challenging to forecast but the development of higher-resolution numerical models allows for a better description of the physics leading to heavy rain more accurately. This leads to a need for more and better knowledge about the small-scale processes causing the evolution of convective events to be better equipped in a world with more and heavy precipitation due to climate change.

More knowledge about convection in higher latitudes and how convection behaves in the north is therefore being explored in this thesis. This has been done by analysing radiosonde data and other measurements such as precipitation, soil moisture and Eddy Covariance from Søråsjordet field station is Ås. To study the possible coupling between soil moisture and convection, the convective triggering potential (CTP) and low-level humidity index (HI_{low}) framework with sounding-derived parameters has been implemented for both the field station measurements and to reanalysis data from the ERA5 model.

To examine the interactions between the surface conditions and convection, daily values for the sounding-derived parameters were computed together with a classification of the different days as convective or non-convective. These values were then combined with daily aggregated values for morning and evening precipitation, soil moisture and Bowen ratio. A GLM was then fitted to the combined dataset for both the observed data and the reanalysis data.

The GLM found the CTP and HI_{low} parameters to be significant explanation variables for a 5 % significance level in the observed dataset. In the reanalysis dataset, the only explanation variable that was significant was the morning soil moisture level. In the reanalysis data for July and August 2023, a splitting of datapoints was seen in CTP, HI_{low} and soil moisture indicating that a land-atmosphere interaction might be found in Norway during convective events. The precipitative events in the observed data were placed correctly according to the CTP-HI_{low} framework. Further, almost no degree of correlation was found between the reanalysis and the observed dataset variables. The lack of correlation and the low representation of different cases in the observed dataset displays the need for a longer time series of observed data which is better balanced and includes more diversity of events than what was available at this time.

Samandrag

Samspelet mellom jord og atmosfære kan påverke ulike tilstandar i atmosfæren utifrå kva tilstandar som er i jorda. Eit slik samspel som har vore meir forska på dei siste åra er korleis ulike jordtilstandar kan påverke utviklinga av konveksjon i atmosfæren. Konveksjon er framleis vanskeleg å varsle men utviklinga av finare maskar i dei numeriske vêrvarsla let oss skildre fysikken rundt kraftig nedbør betre. Dette gjer at meir kunnskap om utviklinga av konveksjon og kva som påverkar den vil vere nødvendig for å vere betre rusta mot meir nedbør og lage betre vêrvarsel i ei verd med meir nedbør.

Meir kunnskap om konveksjon i høge breiddegrader og om korleis konveksjon oppfører seg i nord skal difor utforskast i denne oppgåva, det skal og undersøkast kva jordfukt og Eddy Kovarians målingar viser i dei ulike tilfella. Dette har blitt gjort gjennom analyse av data frå radiosondar saman med observert data frå feltstasjonen på Søråsjordet i Ås. Eit rammeverk med parametrane convective triggering potential(CTP) og low-level humidity index (HI_{low}) som er berekna frå radiosondedata har blitt teke i bruk for å undersøke samspelet mellom jord og atmosfære under konveksjon nærare. Det er i tillegg gjort ei liknande analyse av reanalysedata frå ERA5 modellen frå ECMWF.

For å undersøke samanhengane mellom jordtilstanden og konveksjon vart det berekna daglege verdiar for radiosonde parametrane og ei klassifisering av dei ulike dagane om dei inneheldt konveksjon eller ikkje vart gjort. Desse verdiane vart sett saman med dagleg aggregerte verdiar for jordfukt, Eddy Kovarians og nedbørdata delt for morgon og kveld av dagen, før ein generalisert lineær modell vart tilpassa med dei ulike aggregerte verdiane for både observert data og reanalyse data.

Den generaliserte lineære modellen fann at CTP og HI_{low} var signifikante forklaringsvariablar for det observerte datasettet med eit signifikansnivå på 5 %. For reanalyse datasettet var det verdien for jordfukt om morgonen som vart funne som signifikant forklaringsvariabel ved 5 % signifikansnivå. For reanalyse datsettet vart det og funne ei oppdeling av dei ulike konveksjonshendingane med bakgrunn i CTP, HI_{low} og jordfukt, som viser at dette rammeverket kan vere verdt å undersøke nærare. I den observerte dataen vart nedbørshendingane plassert korrekt i forhold til CTP-HI_{low} rammeverket. Elles vart det funne lite samsvar mellom dei ulike variablane frå det observerte og det modellerte datasettet. Det dårlege samsvaret mellom datasetta og den låge representasjonen av ulike tilfelle viser behovet for ei lenger tidsserie med observert data som er betre balansert og inkluderer fleire tilfelle enn det som var tilgjengeleg på dette tidspunktet.

Contents

List of Figures

List of Tables

List of abbreviations

AI	Artificial intelligence
BL	Boundary layer
CAPE	Convective available potential energy
CIN	Convective inhibition
CTP	Convective triggering potential
ECMWF	European Centre for Medium-Range Weather Forecasts
EL	Equilibrium level
$\mathbf{HI}_{low}$	Low-level humidity index
IPCC	Intergovernmental Panel on Climate Change
LCL	Lifting condensation level
LFC	Level of free convection
NWP	Numerical weather prediction

1 Introduction

The last report from the Intergovernmental Panel on Climate Change(IPCC), governed by the UN, states that climate change is happening and is man-made (Masson-Delmotte et al., 2021). One of the consequences of climate change and global warming includes more extreme weather and more extreme precipitation. Fischer and Knutti (2016) found and used the correlation between higher temperatures and wetter weather that was predicted already in the 1800s by the Clausius-Clapeyron equation which states that the vapour pressure increases with temperature. The correlation was earlier used in climate modelling and can now be observed in the longer time series of meteorological parameters available (Fischer & Knutti, 2016). The intensification of heavy rainfall suggests that the intensity and severity of convective precipitation can also emerge in the higher latitudes. Convective precipitation happens when saturated air lifts up to build heavy clouds broken down by heavy rainfall (Fischer & Knutti, 2016).

More knowledge about convection and how it evolves can be obtained by investigating the land-atmosphere coupling during convective events. Such investigations have been done on convection and how different surface conditions can affect the intensity of convection. One study that connected soil moisture content and instability parameters is Findell and Eltahir (2003a). They developed a new set of sounding-derived parameters to investigate the feedback between convection and soil moisture content. The parameters was then implemented for categorising areas in the USA as positive, negative, or no feedback zones (Findell & Eltahir, 2003b). These sounding-derived parameters have later been used by others such as Cioni and Hohenegger (2017) that used the same radiosonde data as Findell and Eltahir (2003b), but with a different weather model to model different soil moisture levels and convection. Other studies have also been done using the framework developed by Findell and Eltahir (2003a) in Europe. Jach et al. (2020) examined the land-atmosphere coupling strength for different vegetation cases and found that there were coupling hot spots in Scandinavia and Eastern Europe. The land-atmosphere sensitivity was further examined in Jach et al. (2022).

In a world with a changing climate (Masson-Delmotte et al., 2021) and developing numerical weather predictions(NWP) with smaller and smaller grids used in the calculations (Yano et al., 2018). The knowledge and observations of convective events become more important as the grid size used in the NWPs allow for convective events to be resolved. The intensity of precipitation has and will continue to increase

with increased global warming (Fischer & Knutti, 2016; Westra et al., 2013). Today's prediction of convective events is not exact enough because the grid sizes are too large to fit the local differences, and convection warnings usually cover several kilometres (Yano et al., 2018). When the NWPs evolve with even smaller grids more convection regimes will be resolved in the models but not enough knowledge or observations of such events are available for modelling today (Yano et al., 2018).

1.1 Scope of the thesis

Despite the efforts of better understanding of land-atmosphere coupling during convective events, more analysis and investigation of how convection evolves in higher latitudes is needed. The earlier analyses have also been done on modelled data or a mix of radiosonde and modelled data and a further investigation on observed data is needed.

This thesis aims to investigate land-atmosphere interactions during convective events in Norway and if the framework of the sounding-derived parameters CTP and HI_{low} proposed by Findell and Eltahir (2003a) describes the Norwegian conditions.

This will be done through three research questions that will be answered through the thesis

1. **Is precipitation occurring when the CTP and Hl_{low} indices predict it?**
2. **What do the soil moisture and Eddy Covariance measurements show in these cases?**
3. **Do the boundaries for the indices apply to Ås and subpolar and polar regions in general?**

2 Theory

2.1 Atmosphere and boundary layer

The atmosphere is traditionally divided into different layers by their different characteristics such as temperature gradient, as shown in Figure 1. The lowest layer stretching from the ground up to approximately 10 km is the troposphere. The temperature normally decreases with height in the troposphere, and 80 % of the atmosphere's mass is here. On the upper limit of the troposphere where the temperature stops declining is the tropopause, which separates the troposphere from the stratosphere. The stratosphere stretches from 10 km up to approximately 50 km above ground. The temperature in the stratosphere is no longer decreasing but increasing with height inhibiting rapid mixing and rising of air. The stratosphere is ended by the heat maximum which marks the stratopause, separating the stratosphere from the mesosphere. The mesosphere, which is the third layer, is characterised by a temperature decrease with height to a minimum temperature of the atmosphere. The mesopause is placed at this minimum temperature and marks the separation of the mesosphere and the thermosphere. The thermosphere continues until there is no air left and is the last part of the atmosphere. In the thermosphere, the temperature is again rising with height (Wallace & Hobbs, 2006).

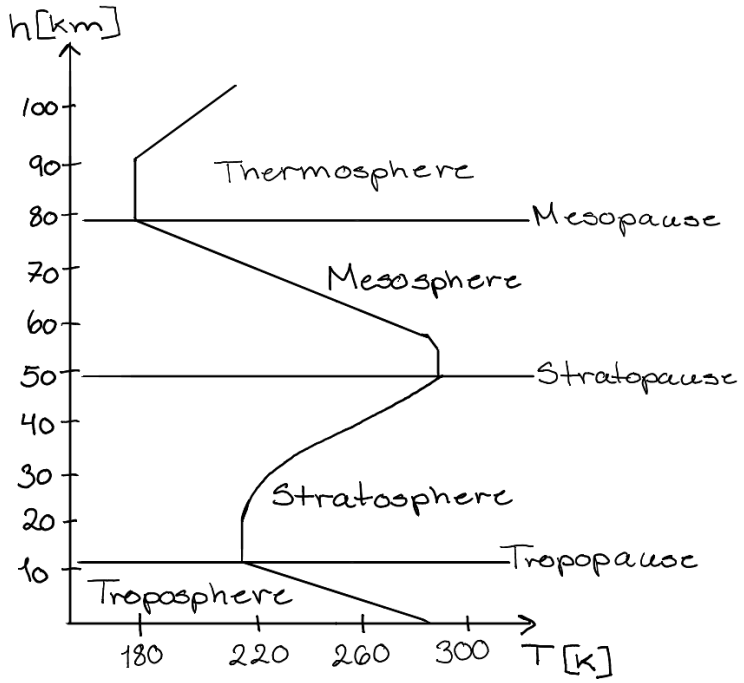


Figure 1: Temperature profile for the atmosphere and its individual layers. Figure inspired by Wallace and Hobbs (2006).

The boundary layer (BL), is a layer of the atmosphere that includes the air closest to the surface of the earth within the troposphere. The BL normally stretches up to 2 km and varies in depth according to the diurnal cycle of incoming solar radiation, and the severity of turbulence. The majority of the weather is formed in the BL (Stull, 2017).

The BL goes through a diurnal cycle of heating and cooling in which the air in the BL becomes more or less turbulent as shown in Figure 2. At night the surface gets cooler as most of the energy is transferred via long-wave radiative loss through the atmosphere. The surface is cooling and therefore the air above the surface gets cooler as there is no incoming heat flux from the sun at night, this leads to the BL collapsing. Only a residual layer and a stable boundary layer of air are left as shown in the 03:00 point in Figure 2. When the sun rises the surface and air above it gets warmed up by the incoming radiation and heat from the sun. The warmer air near the surface becomes more buoyant than the cooler air in the higher layers and begins to rise, slowly making the nocturnal stable layer more and more unstable and turbulent as more air gets entrained into the convective air from below as shown by the 15:00 point in Figure 2 (Stull, 2017).

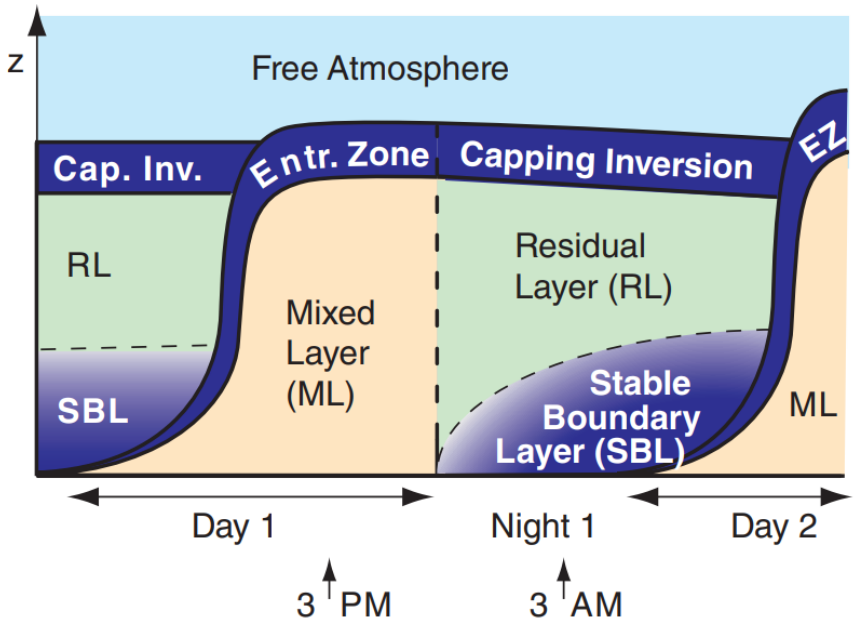


Figure 2: The diurnal cycle of the BL and its different parts. The tan sections indicate the mixed layer with high turbulence. Blue sections indicate stable layers and green sections are the residual layers. Figure reproduced from Stull (2017).

The air parcel, warmed at the surface, undergoes adiabatic changes as it rises. These changes can be described by different lapse rates that explain temperature changes

based on height or pressure. Two special case lapse rates are the dry and moist adiabatic lapse rates. These describe the change of temperature a parcel of air would experience as it rises through the atmosphere with a relative humidity of 0 % and 100 % respectively. The dry adiabatic lapse rate(Γ_D) is defined as

$$\Gamma_D = \frac{g}{C_p}, \tag{1}$$

where g is the gravitational acceleration with a value of 9.81 m/s^2, and C_p is the specific heat capacity for dry air at constant pressure with a value of 1,005 kJ/kgK (Stull, 2017).

The moist adiabatic lapse rate describes the change of temperature with the height of air parcels with a relative humidity equal to 100 %. The expression for the moist adiabatic lapse rate(Γ_s) is given as

$$\Gamma_s = \Gamma_d \cdot \frac{1 + \frac{r_s \cdot L_V}{R_d \cdot T}}{1 + \frac{L_v^2 \cdot r_s}{C_p \cdot R_v \cdot T^2}}, \tag{2}$$

where r_s is the saturation mixing ratio given in g/kg, L_v is the latent heat of vaporisation with a value of $2.5{\cdot}10^6$ J/kg and R_d is the gas constant for dry air with a value of 287 J/kgK. R_v is the gas constant for water vapour with a value of 461 J/kgK and T is the temperature of the air parcel given in K (Stull, 2017).

2.2 The hydrologic cycle

The hydrologic cycle describes the movement of water in the earth-atmosphere system. The water moves between different reservoirs, such as the atmosphere, the cryosphere which contains the ice caps and other frozen forms of water, lakes and rivers, the crust and mantle, and the oceans, together known as the hydrosphere (Wallace & Hobbs, 2006). The different reservoirs visualised in Figure 3, contain water at different lengths of time, this phenomenon is known as residence time. Some reservoirs usually contain a normal water molecule for days while others like the mantle and ice sheets have a residence time of several thousands of years. The atmosphere is one of the reservoirs with the shortest residence times which means that the atmosphere exchanges water the most with the other reservoirs. This, together with a high latent heat of vaporisation for water, makes evaporation a very effective energy transfer between the surface and the atmosphere (Wallace & Hobbs, 2006).

The different reservoirs are of different sizes and contain both saline and fresh water. The biggest reservoir is the oceans which hold 96.5 % of the total water estimated on earth. Out of the total water content, only 2.5 % is fresh water and 68.7 % of that is in glaciers and ice caps as part of the cryosphere. The groundwater accounts for about 30 % of the freshwater and only 1.2 % of the total freshwater is part of the surface freshwater. Of the surface freshwater almost 70 % is contained in ground ice and permafrost, and 20.9 % is contained in lakes. The atmosphere and soil moisture contain both about 3 % each of the surface freshwater (Gleick, 1993).

Processes that are involved in the hydrologic cycle include precipitation, evaporation, transpiration and runoff. As shown in Figure 3 water from the oceans evaporates into the atmosphere where clouds can be formed, point 1 and 2. The clouds can then become precipitative and the water will reach the oceans or land as precipitation, point 6, and become surface freshwater. The surface freshwater can then travel as runoff into lakes and rivers or downward into the soil, crust or mantle, points 3 and 4. Some water is also kept in the plants and trees and is transported into the atmosphere again by transpiration, point 5 (Wallace & Hobbs, 2006).

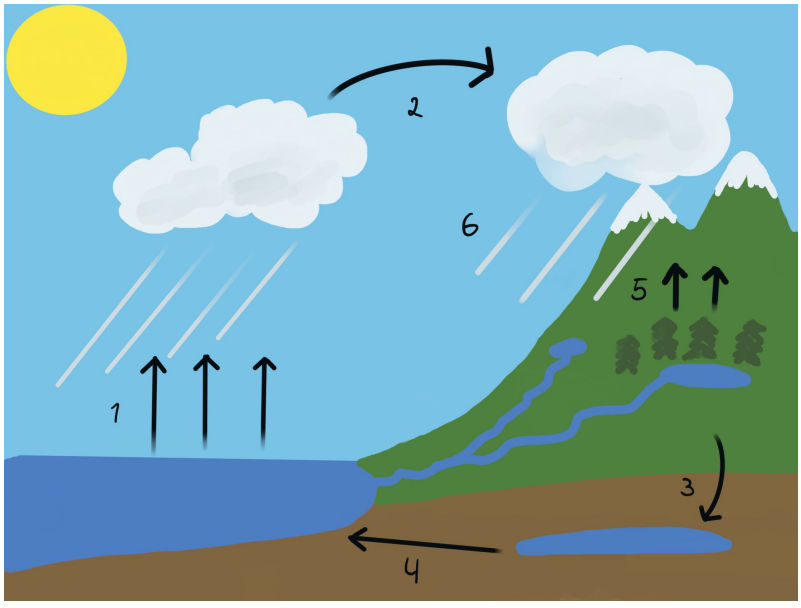


Figure 3: The different reservoirs and exchanges of water in the hydrologic cycle. 1 describes the evaporation of watermolecules from the oceans and lakes into the atmosphere, 2 is the transfer of moisture in the atmosphere, 3 is the runoff water from the soil to the mantle, 4 is the transfer of moisture from mantle into oceans. 5 represents the evapotranspiration from the surface into the atmosphere and 6 represents the precipitation.

2.3 Convection and precipitation

Precipitation is known as rain, snow or hail falling from the clouds to the surface of the earth. There are mainly two kinds of precipitation; stratiform precipitation and convective precipitation. Stratiform precipitation comes from stratiform clouds as rain showers covering large areas and long periods. Convective precipitation falls from cumulus clouds, resulting in more local and heavy rainfall over a shorter time (Barry & Blanken, 2016). There are also several sub-classes of precipitation within these two but these are not of interest for this thesis.

Convection is the upward movement of air where the warmer air near the surface rises to the colder parts of the BL, due to its lower density, and reaches its dew point temperature, at which the parcel will stop rising because of equilibrium with the surroundings. When the parcel reaches its dew point temperature cloud droplets will form. Convection may lead to precipitation if the convection is deep enough creating heavy droplets that can later fall as rain. The clouds that typically form with convection are high cumulus clouds stopping at the equilibrium level (Wallace & Hobbs, 2006).

Convective precipitation presents as heavy rainfall over a short period, and cloud forming is happening on the scale of hours compared to days from the mesoscale weather systems of high and low pressure and cold- and warm fronts. Convective precipitation is therefore challenging to forecast in today's numerical weather predictions(NWPs). Convective precipitation often develops on the scale of 100 m-1 km as opposed to the grid size used in NWPs in Norway which is 2.5 km for the short-term forecasts (Norewgian meteorological institute, n.d.). Convective precipitation is a local event and may therefore not be predicted by the NWPs as the grid size of the predictions may be too big to include the local differences in weather (Šaur, 2015).

2.4 Land-atmosphere coupling

In addition to the atmospheric conditions contributing to if the convection will grow deep enough, the surface conditions may also have an impact on the degree of convection. This relationship is called a land-atmosphere coupling and represents the interactions between the underlying surface and the atmosphere and how they may accelerate or decelerate processes within the land-atmosphere boundary. Some surface conditions that affect the atmosphere include the amount and type of vegetation or sealed surfaces, the presence of water bodies, or land underneath the

atmosphere, this affects the heat fluxes and the energy exchange between the land and the atmosphere (Findell et al., 2024).

This thesis will focus on the influence of soil moisture on the convective regimes in Ås, Norway. The soil moisture can influence the BL by a difference in heat flux partitioning. The partitioning between latent and sensible heat flux will vary with soil moisture. The latent heat flux is the heat flux caused by the evapotranspiration or condensation of water at the surface. The sensible heat flux describes the heat flux by the surface turbulent heat (Barry & Blanken, 2016).

2.5 Surface heat fluxes and soil moisture

The soil's moisture content can be measured by gravimetry or by electromagnetic sensors utilising the change of electromagnetic properties in the soil with different moisture levels. Wet and dry soil act differently changing the way it affects the surroundings. Wetter soils have among others a lower albedo, and a different heat flux partitioning from dry soils (Barry & Blanken, 2016). These properties are the basis of a hypothesis stating that the soil moisture content may influence convection and precipitation or drought patterns.

The heat fluxes from the surface are split into several turbulent fluxes, two of which are latent heat flux and sensible heat flux. These fluxes among others, make up the surface energy balance where there will always be a balance between the different fluxes. A simple model of the surface energy balance is made up of the incoming radiation, conduction of heat in the ground, and the latent and sensible heat fluxes as seen in Figure 4.

The sensible heat flux and the latent heat flux are related through a heat budget where the net sum of fluxes is equal to zero. The heat budget can be represented by this equation

$$F + F_H + F_E - F_G = 0. \tag{3}$$

Where F is the net radiation between the surface and the atmosphere, F_H is the sensible heat flux, F_E is the latent heat flux, and F_G is the molecular heat conduction in between the surface and deeper levels of the soils. All heat fluxes are given in $\mathrm{W/m^2}$ (Stull, 2017). The different heat fluxes are visualised for a day with moist soil in Figure 4.

Equation (3) is a simplified equation of the surface energies. A problem associated

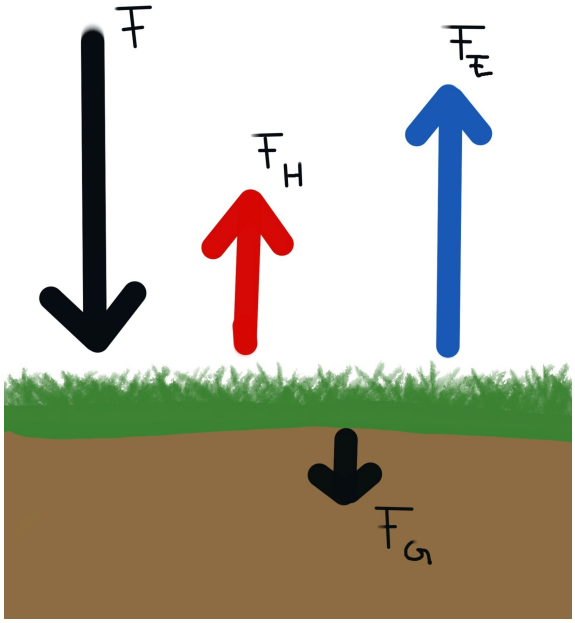


Figure 4: The different heat fluxes in the heat flux budget. The total heat flux between the surface and atmosphere is the long black arrow, F, the red arrow is the sensible heat flux, F_H, the blue arrow is the latent heat flux, F_E and the small black arrow is the heat flux in the ground F_G.

with this equation is the energy balance closure. The energy balance closure problem addresses that the sum of the heat fluxes on the surface does not necessarily match the total incoming radiation and an imbalance of heat fluxes is present. This imbalance is thought to come from the incorrect implementation of instrumental and flux correlations, the storage of energy in soil and other processes such as photosynthesis and time aggregation where the aggregated values could be a source of missing information of heat fluxes (Masseroni, 2014).

The sensible heat flux is the warming of air close to the surface by molecular diffusion. The heat will then be transferred to the overlying atmosphere by convection through the BL (Barry & Blanken, 2016). The latent heat flux describes the transfer of latent heat between the surface and the atmosphere. It involves both evapotranspiration from the moisture on the surface and condensation of moisture in the atmosphere. Evapotranspiration includes both evaporation of water from different surfaces such as soil and foliage, and transpiration from plants, as visualised by point 5 in Figure 3. This is because it is difficult to separate how much vapour comes from the different processes. The latent heat flux is therefore also known as a heat flux originating from evapotranspiration. The latent heat flux is a heat flux removing moisture from the surface and is dependent on an energy source, a moisture source, and air motion to be initiated. The latent heat flux is therefore usually higher over

wet than dry surfaces since there is more moisture available for evaporation (Barry & Blanken, 2016).

The surface fluxes can be evaluated through the Eddy Covariance method, which uses measurements of the horizontal and vertical wind profiles to calculate the different heat fluxes (Barry & Blanken, 2016). To calculate the latent heat flux, measurements of moisture are combined with the wind measurements, and for the sensible heat flux temperature measurements are used for the calculations.

One way to observe how the relationship between the latent and sensible heat fluxes is through the Bowen ratio. The Bowen ratio explains the ratio between the sensible and latent heat fluxes and is defined as

$$B = \frac{F_H}{F_E}. \tag{4}$$

The Bowen ratio will be sensitive to the water content of the soil and the amount of evapotranspiration in the area with lower values over wetter soils with more evapotranspiration and higher values for dryer soils (Wallace & Hobbs, 2006). Values higher than 5 are commonly found over arid conditions, values between 5 and 0.5 are found in the transition between dry and moist conditions, and values below 0.5 are usually found over irrigated farmland and wetter soils (Stull, 2017).

2.6 Radiosondes

Observations of the vertical composition of the atmosphere are needed to make an adequate representation of its layers and different properties. These observations are traditionally made via radiosonde. The radiosonde consists of a balloon with a sonde containing different measuring instruments attached. The balloon is typically filled with hydrogen or helium gas so that it rises through the atmosphere due to the buoyancy of the gas inside the balloon. The radiosonde is released into the atmosphere from near the ground and reaches heights up to 35,000 m before the balloon bursts (National Oceanic and Atmospheric Administration, n.d.-b).

The different measurements made by the radiosonde include among others latitude and longitude, air temperature and pressure, as well as altitude measurements and relative humidity. Observations from the radiosondes can be used to describe the stability of the BL and if there are inversions present (Stull, 2017). The measurements are also used as input for NWP models and for other purposes such as weather and climate research and as a ground truth for satellite data

(National Oceanic and Atmospheric Administration, n.d.-a). The radiosonde often makes measurements both on its ascent and descent. This gives a vertical profile of the environment in the atmosphere (Stull, 2017). The vertical profile however, will not always be the same as the radiosonde can drift up to 300 km from the release point and can therefore give the vertical profile for a new route for every sounding (National Oceanic and Atmospheric Administration, n.d.-a). There is a global system of radiosonde measurements where a radiosonde is released at 00:00 and 12:00 UTC (National Oceanic and Atmospheric Administration, n.d.-b), which is equal to 01:00 and 13:00 in Ås (UTC + 1) during the winter months and 02:00 and 14:00 during the summer months due to daylight saving time in Europe.

The radiosonde measures the temperature, pressure and the relative humidity. Other parameters are calculated, often by software included in the radiosonde itself or using the measured data from the radiosonde afterwards. One parameter that is being calculated is the dew point temperature. The dew point temperature(T_d) is calculated by a Magnus formula defined as

$$T_d = T\left[1 - \frac{Tln(\frac{RH}{100})}{L_v/R_v}\right]^{-1}, \tag{5}$$

where T is the environmental temperature from the sounding, RH is the relative humidity from the sounding, L_v is the heat of vaporisation with a value of $2.5 \cdot 10^6$ J/kg and R_v is the gas constant for water vapour with a value of 461.5 J/kgK (Lawrence, 2005).

2.7 CAPE and CIN framework

Traditionally frameworks such as the combination of convective available potential energy(CAPE) and convective inhibition(CIN) or the lifted index are used to indicate if a convective event will take place. These indices only include information about the atmosphere and do not take soil moisture or land-atmosphere coupling into account in their indications of convection. These indices are calculated from radiosonde measurements.

CAPE is defined as the integral of the upward buoyancy force of the air parcel divided by the density of the air parcel. The upward buoyancy force is measured per unit volume and is caused by the temperature difference between the parcel and its environment. CAPE could be visualised as the area between the level of free convection(LFC) and the equilibrium level(EL) in a skew-T ln(p) plot as shown in

Figure 5. The values of CAPE can vary much as shown in Table 1, and some convection could normally be anticipated with values between 1000J/kg and 2500J/kg (Wallace & Hobbs, 2006).

Table 1: The different levels of CAPE values and their generally associated convection levels.

CAPE value [J/kg]	**Convection**
0 - 1000	No convection likely
1000 - 2500	Shallow convection likely
2500 - 4000	Strong convection likely
4000 -	Extreme convective events

The other dimension of this framework is the CIN which is an abbreviation for convective inhibition. The CIN describes the energy needed to lift the air parcel to its LFC. For convection to be possible the CIN must be nonzero but not too large so that it inhibits convection altogether. A value of CIN larger than 100J/kg would mean deep convection is unlikely as this is too much energy needed for the air parcel to reach its LFC. CIN can also be visualized in a skew-T ln(p) plot as the integral between the vertical profiles of the moist adiabat and the environmental temperature from the sounding between the LCL and the LFC as shown in Figure 5 (Wallace & Hobbs, 2006).

The amount of energy that is needed for convection to be initiated varies with the different systems. In high-energy systems, more energy might be needed to lift the air parcel and in lower-energy systems, less energy might be needed. This is shown in a few studies from the higher latitudes of continental Europe. One study looked at sounding-derived parameters, such as CAPE, near hail- and thunderstorms in the Netherlands (Groenemeijer & van Delden, 2007). They found that thunder occurred for a mean CAPE value of 198 J/kg. Another study conducted in Russia on convective events proposes a threshold in CAPE of 150J/kg for severe convective events (Chernokulsky et al., 2022). This leads us to believe that the amount of energy needed to lift the air parcel in Norway also would be situated around a lower level of CAPE than presented in Table 1.

Since CAPE and CIN are dependent on the temperature and humidity of the BL big variations in time and space have been observed (Ye et al., 1998). The temporal variability is mostly connected to the diurnal cycle of available energy for convection where the amount of CAPE usually peaks in the afternoon. The spatial variability is more dependent on the local variability of temperature and humidity.

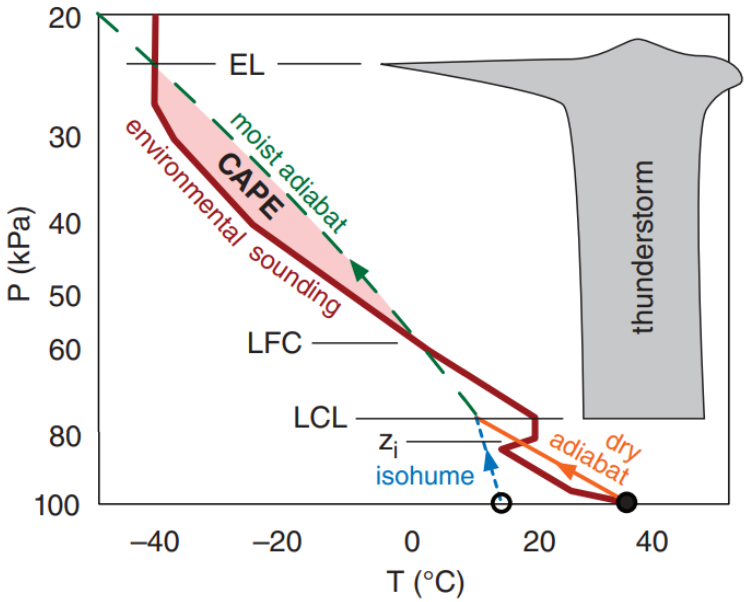


Figure 5: The CAPE is shown as the light red area between the vertical profiles of the moist adiabat and the environmental temperature from the sounding between the LFC and the EL. The CIN would be the area between the two vertical profiles between the LCL and the LFC. Figure reproduced from Stull (2017).

2.8 CTP-HI$_{low}$ framework

To better understand the land-atmosphere interaction during convective events Findell and Eltahir (2003a) proposed the combination of convective triggering potential (CTP) and a low-level humidity index (HI$_{low}$) to indicate if convection is probable during that day. Data from radiosondes were used to compute the two parameters and originally early morning soundings initiated at dawn were used (Findell & Eltahir, 2003b).

The CTP is computed by integrating the area between a moist adiabatic lapse rate starting at the air temperature 100 hPa above ground level and ending at 300 hPa above ground level, and the air temperature from the early morning radiosonde as shown in Figure 6. The region between 900 hPa and 700 hPa is defined as the critical region where the BL develops through the day (Findell & Eltahir, 2003a).

The HI$_{low}$ is calculated as the sum of the dew point depressions at the pressure levels 950 hPa and 850 hPa

$$HI_{low} = (T_{a,950} - T_{d,950}) + (T_{a,850} - T_{d,850}), \tag{6}$$

where $T_{a,950}$ and $T_{d,950}$ are respectively the air temperature and dew point tempera-

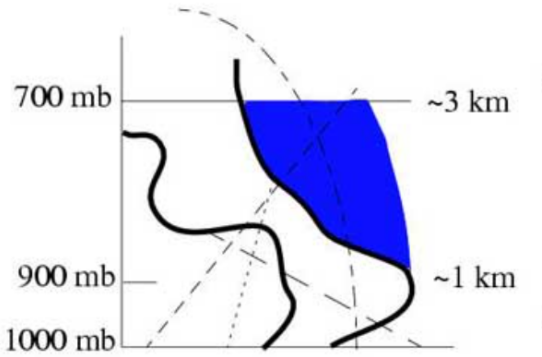


Figure 6: The CTP is highlighted in blue as the area between the moist adiabat originating at the environmental temperature at 900 hPa, and the vertical environmental temperature from the sounding between the pressure levels 900 hPa and 700 hPa. Figure reproduced from Findell and Eltahir (2003a).

ture at 950 hPa pressure level, and $T_{a,850}$ and $T_{d,850}$ are respectively the air and dew point temperatures at 850 hPa pressure level.

CTP and HI_{low} can then be used to split the cases into either atmospherically controlled or not atmospherically controlled as shown in Figure 7. In the atmospherically controlled group, the atmosphere could be either too wet making it likely to rain over both dry and wet soils, too dry making it not likely to rain over any soils or the CTP could be negative indicating a too-stable environment for convection to be triggered. In the non-atmospherically controlled group, there are combinations of CTP and HI_{low} that favour deep convection over dry soils and combinations that favour deep convection over wet soils (Findell & Eltahir, 2003a).

The combinations found to trigger deep convection more frequently over dry soils by Findell and Eltahir (2003b), were HI_{low} between 10 K and 15 K and a high CTP corresponding to high instability, CTP $>$ 200 J/kg. The combinations found to trigger deep convection over wet soils was a lower HI_{low} between 5 K and 10 K with lower positive CTP values. These limits were used in an analysis of regional climate in Europe by Jach et al. (2020).

The wet soil advantage favouring deep convection over wet soils occurs when the atmosphere is sufficiently wet, and when there are unstable conditions in the atmosphere. The dry soil advantage, favouring deep convection over dry soils, occurs with a dryer atmosphere than the wet advantage and with a very unstable atmosphere. The classification also included a transition zone where the HI_{low} is higher, 10 K to 15 K, and the CTP is intermediate with values between 50 J/kg and 200 J/kg (Findell & Eltahir, 2003a). The different zones are visualised in Figure 7.

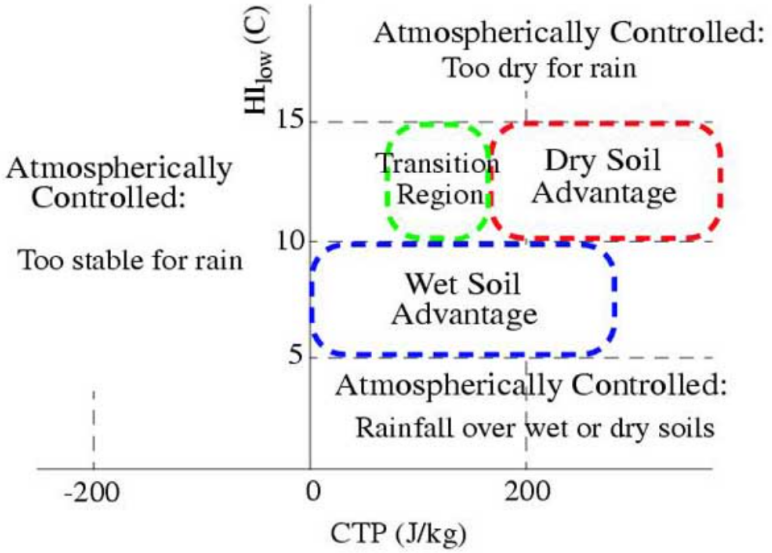


Figure 7: The figure shows the different zones which are atmospherically controlled and not atmospherically controlled. The wet soil and dry soil advantage zones found by Findell and Eltahir (2003a) are shown. Figure reproduced from Findell and Eltahir (2003a).

Findell and Eltahir (2003b) used the framework for the identification of coupling strength in the USA. They found that some areas had more dry advantage days and others had more wet advantage days, some areas closer to the oceans had most atmospherically controlled days and land-atmosphere coupling strength was low in these areas. Cioni and Hohenegger (2017) further explored the same data as Findell and Eltahir (2003a) and implemented it in a new model that allowed for more interactions between the surface and the atmosphere. They found convection to be sensitive to soil moisture and the latent heat flux, where wetter soils gave heavier precipitation, strengthening the usage of the framework.

The CTP-HI$_{low}$ framework is a framework that can be used for identification or observation of land-atmosphere coupling for an area. One study that used the framework for the identification of coupling strength is Jach et al. (2020). They implemented the framework in Europe for different surface conditions and used modelled data to identify the coupling strength between the surface and the atmosphere. They found that the coupling strength varied with location and was sensitive to what land cover was dominant in the different areas. Jach et al. (2022) did an analysis of Europe which included Norway, and Scandinavia. The analysis identified mainly atmospherically controlled days in Norway. They found little interaction between the surface conditions and the convection or precipitation. There was a little area around the Oslofjord that showed some land-atmosphere interaction.

2.9 Sub-arctic climate

Norway is a part of the sub-arctic climate zone characterised by cold or very cold winters and warm or cool summers. The Köppen Climate Classification often classifies most southern parts of Norway including Ås as Dfc; a sub-arctic climate with cool summers which is wet all year. This class is characterised by a low average temperature of -3° C or lower in the coldest month, and a higher average temperature of 10° C or higher in the warmest two or three months, and precipitation is mostly equally distributed throughout the year (Pidwirny, 2021).

Norway has a long coastline and stretches from 58°N to 71°N leading to Norway having 4 different climatic zones in addition to Dfc, according to the Köppen Climate Classification (Ketzler et al., 2021). Southern parts of Norway can be split into western and eastern, based on the mountainous region that separates the two zones. On the western part, fjords and coast climate are dominant and high levels of precipitation are measured, with a mean annual amount of 2250 mm precipitation in Bergen. As a comparison, the mean annual precipitation at Gardermoen in the eastern, more inland part of southern Norway is 862 mm (Ketzler et al., 2021). At Gardermoen, the lowest mean monthly precipitation occur in February and April, and the maximum occurs in September and October. The distribution of precipitation is however relatively even throughout the year. Drier seasons in the eastern part of Norway is expected in the spring with more stable and sunny weather being expected (Ketzler et al., 2021).

The sub-arctic region will get warmer due to climate change. With a changing climate, the already wet environment will most likely become even wetter with more intense precipitation. The intensification of heavy rainfall events was observed when data from 1981 to 2013 was compared to the last normal from 1951-1980 (Fischer & Knutti, 2016). They found that heavy rainfall that would occur 1-in-1000 days in 1951-1980 occurred about 45% more often in the later period of 1981-2013. Hennessy et al. (1997) found that in their models the high latitudes got an increase in the number of wet days and more moderate and heavy precipitation events with a doubling of the CO_2 content of the atmosphere. The precipitation in Norway is expected to increase and heavy rain events are expected to occur more often (Ketzler et al., 2021).

3 Materials and methods

In this analysis, two different datasets were used, one with observed data from the field station at NMBU in Ås, and one from the ERA5 reanalysis model extracted for the area over Ås municipality. The details around the two datasets are covered in section 3.1. Methods for preprocessing and data wrangling are described for both datasets in section 3.2. Followed by the implementation of the CTP-HI$_{ow}$ framework in section 3.2.1, classification of the data in section 3.2.2 and the implementation of the generalised linear model in section 3.2.3. All of the scripts used in this thesis are found in the GitHub repository found at: https://github.com/Randi-Maud-Mork/Master_Environmental_physics_and_renewable_energy_2024/tree/db5ea7d81b8b850145621594001d760108654fb7

3.1 Materials

3.1.1 Søråsjordet field station

The data used in this analysis was collected at the field station for bio-climatic studies at NMBU situated on Søråsjordet in Ås, Akershus (N59° 39'37", E10° 46'54"). The field station is situated 93.3 meters above sea level and is placed in the middle of a field with 200m to the closest housing and forest, and has Station ID: SN17850. The data was collected during the summers of 2021, 2022, and 2023, and consists of radiosonde data and data from stationary sensors at the field station. The data from the stationary sensors includes precipitation, soil moisture, radiation, surface fluxes and air temperature among others (Norges miljø- og biovitenskapelige universitet, n.d.-b).

3.1.2 Meteorological description

The summer of 2021 started with a May that was wetter than normal and had a slightly lower than normal temperature with a deviance of 0.8 °C. June, July and August were all warmer than normal. June and August were dry months where only 4 mm of precipitation was registered in August. July registered an amount of precipitation that was over the normal (Norges miljø- og biovitenskapelige universitet, n.d.-c).

The summer of 2022 started with a little lower than normal precipitation in May of 50.6 mm against the normal of 62.1 mm. The temperature was closer to normal with a deviance of 0.2 °C from the normal. June had a temperature that was 0.2 °C colder than normal while July and August had a temperature that was 1.3 °C and

0.8 °C higher than the respective normals. The precipitation in June and July was slightly under and over the normal, and in August only a third of the normal amount of precipitation was measured (Wolff, 2023).

The summer of 2023 started with a dry May with only 14 mm of precipitation for the whole month. June continued the dry spell and only 31 mm of precipitation was registered. The weather switched in July and August; these two months were cold and wet with 140 mm precipitation in July and 173 mm in August (Wolff, 2024).

3.1.3 Datasets

Radiosonde data

The radiosonde data is from radiosondes released from the field station in Ås and contains information about the atmosphere above Ås. The radiosonde data mostly consists of weekly measurements initiated at 13:05 local time and includes some additional measurements between the dates of 04.07.2021 and 14.07.2023. Some weeks do not have the weekly measurements. The data coverage therefore varies for each month and each year.

Field station data

The precipitation measurements were collected at Søråsjordet in Ås by a Geonor T-200B precipitation meter. The Geonor meter was installed in 2017 and uses frequency to measure how much precipitation has been accumulated (*Bruksanvisning nedbørmåler T-200B*, 1996).

The radiation measurements are also collected at Søråsjordet in Ås. These measurements are collected by a Kipp & Zonen CMA 11 Pyranometer installed in 2014. The Pyranometer uses Thermopile detectors and glass domes to measure the global irradiation onto the surface and the reflected radiation from the ground (*Instruction Manual - CMP/CMA series*, 2013).

The soil moisture data was collected at different sites at Søråsjordet and with various sensors. One is a SoilVUE10 sensor which uses time-domain reflectometry to measure both volumetric water content and the electrical conductivity of the soil (*Product Manual: SoilVUE 10, Complete soil profiler*, n.d.). The version used at Søråsjordet was 1m with 9 different measurement depths. In addition to the SoilVUE10 sensor, a GroPoint profile sensor was used. This sensor uses time-domain transmissometry to measure the average volumetric soil moisture content and soil temperature. The GroPoint used at Søråsjordet has 2 segments and provides mea-

surements at 2 depths, 5cm and 25cm (*GroPointTM Profile Multi-Segment Soil Moisture & Temperature Profiling Probe*, 2020).

The Eddy Covariance measurements are taken by a SmartFlux system that uses EddyPro software to compute the Eddy Covariances (*Using the LI-7500DS and the SmartFlux 3 System*, 2021). The original measurement interval is 10-minute and the EddyPro software includes a 30-minute average resulting in an output file with 30-minute interval computations (*EddyPro Software Instruction Manual*, 2021). The SmartFlux system uses the WindMasterPRO 3-axis ultrasonic anemometer for the wind measurements (*WindMasterPRO 3-Axis Ultrasonic Anemometer*, 2019).

Precipitation data from Statens Vegvesen

In addition to the precipitation data from Søråsjordet, some precipitation data from Statens vegvesens stations were acquired through the Norwegian Meteorological Institute from the Frost API. This data was used to check for rainfall in the areas around Ås. The stations are of lower quality and the measurements are only used as indicators of rain or no rain.

Reanalysis data

The reanalysis data was downloaded from the ERA5 model which is the fifth generation of reanalyses from the European Centre for Medium-Range Weather Forecasts(ECMWF) (Hersbach et al., 2023a, 2023b). The analysis contains data from 1940 onwards and is made up of model data and observations from across the world. The reanalysis from ERA5 provides hourly estimates for different atmospheric and land-surface variables for a horizontal resolution of 31 km (Hersbach et al., 2020). The data extracted from ERA5 for this analysis consisted of atmospheric variables such as air temperature and pressure, geopotential height and different wind components to mimic soundings. These variables were extracted for 09:00 every day in the extended summer, May - September (Hersbach et al., 2023a). In addition to the reanalysis data for vertical profiles, other parameters such as CAPE and CIN, precipitation, soil moisture, and surface heat fluxes were extracted for the whole time series of May through September for the three years(Hersbach et al., 2023b). This was extracted for the coordinates 59.6°N - 59.7°N and 10.7°E - 10.8°E

3.2 Methods

All preprocessing and data wrangling was done in RStudio using R version 4.3.3. The exception was the retrieval of data from the Frost API which was done using Python version 3.11. The data wrangling in R was done using the *dplyr* and *tidyr* packages

to handle data frame and vector functions. Handling of date and time formats was done by the *lubridate* package. All package versions and fields of application are found in Table 2.

Table 2: Package names, versions and Fields of applications for packages used for data wrangling and analysis in RStudio.

Package	**Version**	**Field of application**
dplyr	1.1.3	Easier data manipulation and wrangling
tidyr	1.3.1	Easier data frame manipulation and setup
lubridate	1.9.3	Handling of datetime formats
ncdf4	1.22	Open, read, and write NetCDF files in R
writexl	1.5.0	Write CSV and xlsx files from data frames in R
padr	0.6.2	Preparation of time series data with missing values
imputeTS	3.3	Visualisation and imputation of missing values in time series data
naniar	1.1.0	Imputation of time series data with missing values
factoextra	1.0.7	Extract and visualise results from data analyses
glmnet	4.1-8	Fitting of linear, logistic and multinomial regression with regularisation

Radiosonde data

The radiosonde data was saved as monthly NetCDF files which were read into RStudio and saved into data frames using *ncdf4*. To make better use of the date dimension of the NetCDF files the date dimension was converted to a POSIXct object using *lubridate*. The radiosonde data was then saved in yearly CSV files with the filenames: *sounding_21*, *sounding_22*, and *sounding_23* using the package *writexl*.

Field station data

The data from the field station at NMBU was read in and after that, *padr*, was used to impute the missing dates and hours to ensure the whole time series was included in the dataset. The distribution of missing values was then visualised using *imputeTS* and the missing values were found to be most single or small groups of missing data. This missing data got imputed using last observation carried forward from

the *naniar* package as the variables would most likely not change much from the one timestep to the next. The datasets from the three summers were then saved in three csv files and further used together with the parameters calculated from the sounding data.

The soil moisture data from the two different sensors contained some missing values. The soil moisture from the SoilVUE sensor contained only missing values in May and June 2021 as well as random missing values throughout the rest of the summer of 2021, 2022 and 2023. There were irregularities of timesteps in the measurements and an aggregation by mean was done to have hourly values ranging from May through September for the three summers. The random missing values were imputed by last observation carried forward from *naniar*.

The soil moisture data from the GroPoint sensor contained missing values from May to August in 2021, the whole summer of 2022 was missing and in 2023 some single missing values and some small groups of missing values was identified. The single missing values in 2021 and 2023 got imputed by last observation carried forward from *naniar* and 2022 was discarded as there was no information in the data.

The absolute values of the measurements from the two soil moisture sensors differ and the GroPoint sensor has been found to be the most accurate on the realistic level of soil moisture in the soil at Søråsjordet (Naalsund, 2022). The SoilVUE sensor does however capture the variations in soil moisture and the patterns of high and low soil moisture as well as the GroPoint sensor. The SoilVUE sensor is therefore included in the dataset for further analysis.

The Eddy Covariance data was saved in monthly NetCDF files which were opened and made into csv files using *ncdf4* for opening and *lubridate* for handling date and time dimensions. The data was saved in yearly csv files. The Eddy Covariance data was originally on a 30-minute resolution and therefore an aggregation by mean was done to create an hourly resolution of the data. The data was also padded to ensure the whole time series was contained in the dataset and thereafter any missing values were imputed by last observation carried forward.

Combined dataset and classification

The yearly and hourly datasets of field station data were combined into yearly datasets containing the variables precipitation, air temperature, soil moisture from the two sensors, and latent and sensible heat fluxes. The combined datasets were then used for classifications.

Soil moisture content was made into a categorical variable with two categories, high and low, where the SoilVUE data had a limit of soil moisture content higher than 20 % would be categorised as high, and lower than 20 % would be low. The GroPoint sensor had a limit at 40 % where a soil moisture content higher than 40 % would be considered high and lower would be considered low. This categorisation was done for the hours between 00:00 and 12:00 where a mean of the soil moisture content would be calculated and used for the categorisation.

The precipitation was split into morning and evening precipitation. The morning precipitation was classified as precipitation before noon and the evening was after 14:00 and later since most of the soundings were initiated at 13:05. The morning and evening precipitation was made up of a sum of the total precipitation during the respective hours. For a precipitation event to be classified as evening precipitation a total sum of 1 mm was needed as we are looking for convective precipitation that often presents as heavy rain.

From the sensible and latent heat fluxes, a Bowen ratio was calculated using equation (4). The Bowen ratio was then made into a categorical variable for the morning where the maximum Bowen ratio for the hours from 00:00 until 12:00 was used as the morning Bowen ratio and a higher Bowen ratio than 0.5 was categorised as high and below 0.5 was categorised as low.

Precipitation data from Statens Vegvesen

The Precipitation data from other stations around Ås was imported to RStudio and one station was filtered out due to unrealistic values. The remaining data was filtered for the dates with interesting CTP-HI$_{low}$ combinations to check for precipitation outside of Ås.

Reanalysis data

The extraction of reanalysis data resulted in 3 datasets for the period 2021 to 2023. The datasets were in a NetCDF format and were read and accessed using the *ncdf4* package. The data was then saved in data frames and yearly CSV files with the filenames *reanalysis_df_21.csv*, *reanalysis_df_22.csv*, and *reanalysis_df_23.csv* using the package *writexl*.

3.2.1 CTP-HI$_{low}$ framework

The CTP-HI$_{low}$ framework as introduced in section 2.8 was implemented in R with user-defined functions to calculate the HI$_{low}$ as in equation (6), the moist adiabatic

lapse rate as in equation (2) and further the CTP using the numerical integration method using the area of a trapezoid to find the integral between the moist adiabatic lapse rate and the air temperature.

The yearly datasets were transformed into a list of split datasets with a user-defined function named *split_dataset_by_DateTime*. This was done to easier make computations for each sounding. The splitting of the dataset ensured that each dataset in the list contained the data from one sounding. The user-defined function used the package *data.table* in the formatting and splitting.

The CTP values for each sounding were calculated by filtering out the critical region between 900 hPa and 700 hPa of the sounding. The data was then used to calculate the temperatures of the moist adiabatic lapse rate starting at the air temperature at 900 hPa and then changing moist adiabatically, this was done through a set of user-defined functions implementing equation (2) for each pressure level. The CTP was then implemented by a user-defined function named *calculate_CTP*, using the numerical integration to calculate the CTP as presented in section 2.8.

The HI_{low} parameter was then calculated by extracting the values closest to a pressure of 950 hPa and 850 hPa to calculate the low-level dew point depressions and the HI_{low} as explained in section 2.8 and equation (6). The CTP and HI_{low} values were then saved in a data frame together with information about the date and time and used in further analysis.

To pick out dates that may have had a convective precipitation event, the CTP and HI_{low} values were filtered. All positive CTP values were included and HI_{low} values between 5 and 20 were included after extending the limits from Findell and Eltahir (2003a) to see if there would be a change of limits due to coast climate and higher latitudes.

3.2.2 Classification

Further, each entry in the observed data were classified as convective or non-convective by implementing the CAPE values from the ERA5 model. The CAPE values were used for a binary classification where a CAPE above 300 J/kg would correspond to a convective event and a CAPE with a value below 300 J/kg would be classified as non-convective.

The classified dataset with daily values for CTP, HI_{low}, Bowen ratio, soil moisture, and precipitation was then used to fit a statistical model and clustering.

The reanalysis data contained variables describing the accumulated precipitation from convection. This variable was used for classification of the reanalysis data. If there was more than 1 mm accumulated convective rain on a day the day was classified as convective. If the accumulated rain did not exceed 1 mm the day was classified as non-convective.

The classified dataset with daily values for CTP, HI_{low}, soil moisture and precipitation was then used to fit a statistical model and clustering.

3.2.3 Generalised linear model

For a response variable that has two classes and explanation variables that are continuous a logistic regression could be a fitting way to analyse the effects of the different explanation variables. The logistic regression is fitted through a generalised linear model(GLM) that uses a link function to find the effects from each variable. For logistic regression a logit link function, $g(p)$ is used. The logit function is defined as

$$g(p) = log\left(\frac{p}{1-p}\right), \tag{7}$$

where $\frac{p}{1-p}$ is the odds ratio for the target, and p is the probability of the target class. The odds ratio can be used for describing whether the odds for the target will increase or decrease with an increase in an explanation variable. If β_1 is negative, an increase in x_1 would give a decrease in odds for the target class, and opposite if the coefficient is positive (Crawley, 2013).

The model equation for a GLM is given as

$$g(p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e, \tag{8}$$

where β_0 is the intercept, β_1 is the coefficient associated with variable x_1, β_2 the coefficient associated with x_2 and further if there are more variables. e is the residual from the model fitting and describes how well a model is fitted (Crawley, 2013).

The model assumptions for a GLM include a constant variance, normal distribution of residuals and a linear relationship between the variables.

Transformations of the dataset can be done for the data to better fulfil the model assumptions. Some widely used transformations are normalisation, standardisation and logarithmic transformation. Normalisation is done by scaling the variables to a range of [0, 1]. Standardization is done by transforming the variables so that every variable has a mean equal to zero and a standard deviation of 1. The logarithmic transformation is done by taking the logarithm of all the values in the variables. This

is done as some models can be scale sensitive and some patterns or irregularities in the data can be smoothed out and easier to handle after transformation (Raschka & Mirjalili, 2019).

3.2.4 Use of artificial intelligence

In the process of writing this thesis, some AI tools have been used with usage according to the guidelines of the Norwegian University of Life Sciences, REALTEK faculty (Norges miljø- og biovitenskapelige universitet, n.d.-a). The use is primarily for spell-checking and as an addition to a literary review. For spell-checking the generative AI Grammarly, has only been used as a spell-checking and small-scale text enhancement, no text has been generated in Grammarly (Grammarly, 2023). As an addition to the literary review, a search assistant Elicit was used for finding additional journals and articles (Elicit, 2023). The articles were then downloaded from the original source as a quality control before usage.

4 Results

4.1 Observed data

The CTP and HI_{low} parameters were calculated through the methods presented in 3.2.1 and were combined in a data frame together with aggregated values which contain morning and evening precipitation sums, as well as measurements of the morning soil moisture and Bowen ratios. From this dataset five days stood out with precipitation only occurring in the afternoon and evening hours. The data from these five days are presented in Table 3. The days with the highest and lowest CAPE values were selected for further investigation. The highest CAPE value (923 J/kg) was registered 15.08.2022, and the lowest CAPE value (33 J/kg) was registered 26.06.2023. Precipitatin amounts of 3.7 mm and 7.8 mm were registered respectively.

Table 3: Showing the CTP and HI_{low} values as well as the soil moisture in the morning, Bowen ratio(BR) in the morning and the total evening precipitation on the five days where convection resulted in precipitation. SM is the soil moisture from the SoilVUE sensor and SM IoT is the soil moisture from the GroPoint sensor.

Date	**CTP** [J/kg]	**HI_{low}** [K]	**SM** [%]	**SM IoT** [%]	**BR**	**Precipitation** [mm]	**CAPE** [J/kg]
04.07.2021	1393	8.1	-	-	0.02	0.3	591
26.07.2021	1375	11.7	6	-	-1.04	3.6	304
15.08.2022	1373	10.6	3	-	-0.45	3.7	923
21.06.2023	1063	4.5	5	39	0.31	4.9	255
26.06.2023	1181	5.5	2	38	-0.50	7.8	33

Case study 15.08.2022

This date was further explored due to its comparably high CAPE value. A plot of the global radiation on 15.08.2022 is shown in Figure 8. It shows a cloud-free morning until 11:00 when clouds started forming.

There was also recorded precipitation on this date, presented in Figure 9. Some rain showers are seen in the afternoon from 16:00-20:00 followed by more rain at night. The soil moisture of this morning is believed to be correctly categorised as low as this day was the end of a dry period of about 10 days since the last significant rainfall as seen in the time series of the soil moisture in Figure 31 in Appendix 2.

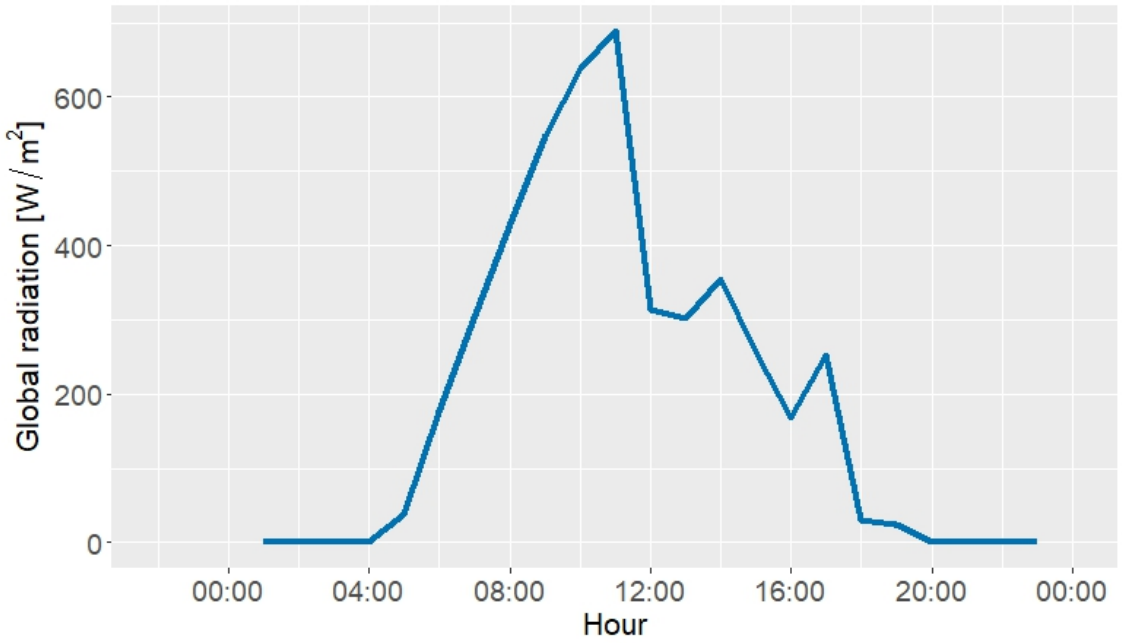


Figure 8: Diurnal course of global radiation on 15.08.2022. The steep increase in the morning indicates cloud-free conditions reaching a maximum at 11:00. The following decrease in the afternoon is typical for the formation of clouds.

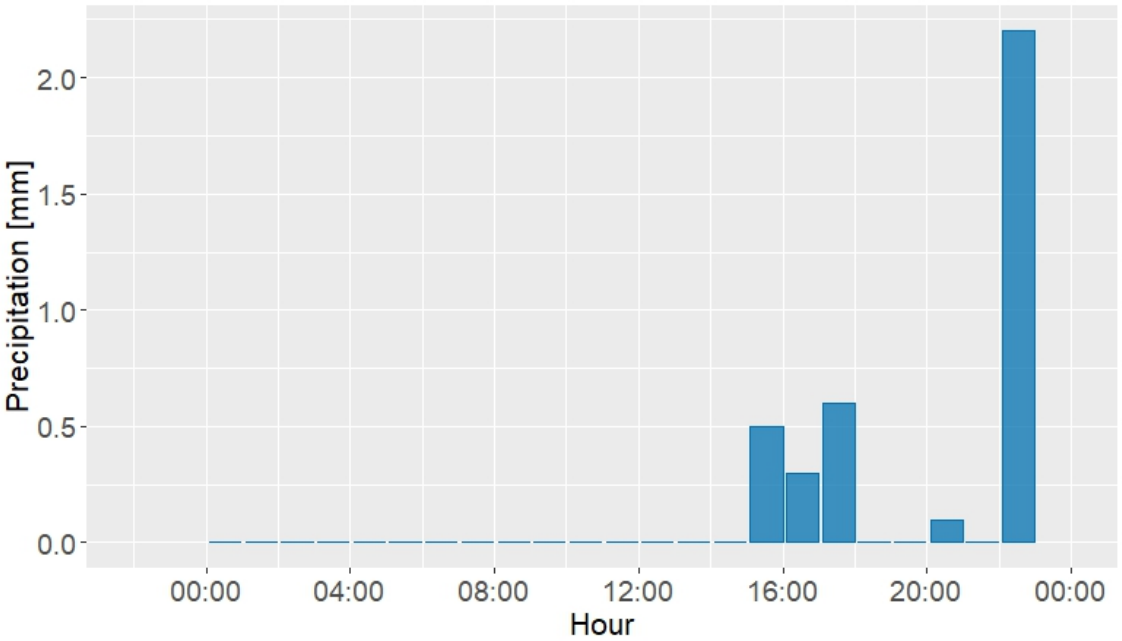


Figure 9: Distribution of hourly precipitation on 15.08.2022. There is not recorded any precipitation in the morning of the day and some precipitation is observed at 16:00 and outward before a more intense precipitative event before midnight.

Case study 26.06.2023

The day with the lowest CAPE and highest precipitation also showed some charac-

teristics that can be associated with convection. In the global radiation plot, Figure 10, it is observed a clear sky in the morning with more clouds from 11:00 and outward with a small breakage of cloud cover at 14:00. In the plot of the precipitation, Figure 11, during that day it is shown that there was some precipitation around 13:00 followed by over 2 mm rain in the hour between 15:00 and 16:00, and some more rain the rest of the evening. This kind of precipitation can match the intensity expected from a convective event.

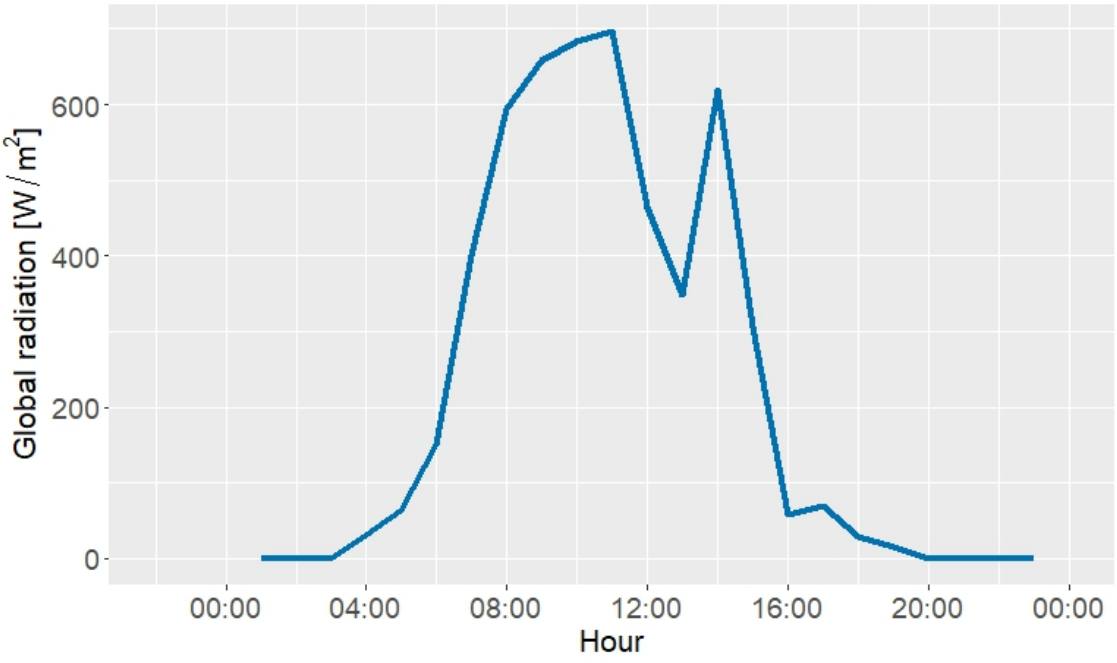


Figure 10: Diurnal course of global radiation on 26.06.2023. The steep increase in the morning indicates cloud-free conditions reaching a maximum at 11:00. The following stepwise decrease in the afternoon is typical for the formation of clouds.

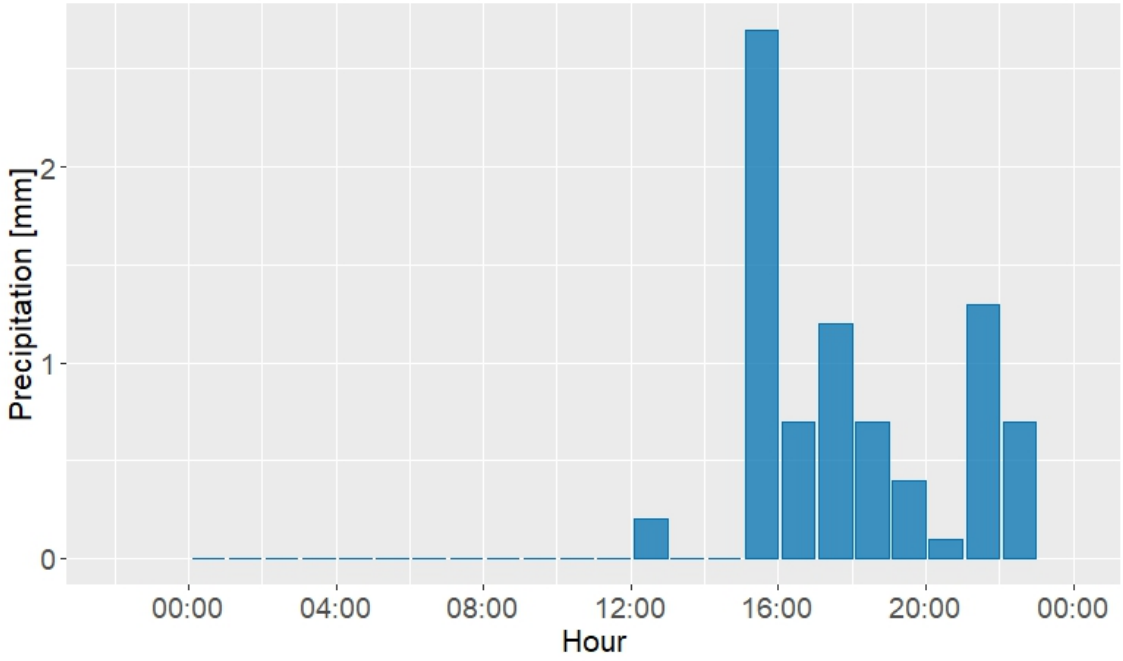


Figure 11: Distribution of hourly precipitation on 26.06.2023. There was no recorded precipitation in the morning of the day and some precipitation was observed after 12:00. An intense precipitation event with over 2 mm precipitation at 16:00 was recorded. Precipitation was then recorded for the rest of the day.

Analysis of the full dataset

The correlations between the different variables used in the dataset are shown in Figure 12. The correlations are generally low as they all are at or under 0.4. The highest positive correlation is between CAPE and morning precipitation meaning a higher CAPE is often seen with more morning precipitation. The highest negative correlation is between the HI_{low} and the morning Bowen ratio meaning a higher HI_{low} is correlated with a lower Bowen ratio. Generally, there are some negative correlations between the two precipitation variables $NB_{Morning}$ and $NB_{Evening}$ with the HI_{low} indicating a lower chance of precipitation with a higher HI_{low}.

To see the distributions of CTP and HI_{low} through the three summers of sounding data, CTP was plotted against HI_{low} in Figure 13, where different variables are made visible through shape, colour, and size. There are mostly observations of days without any rain(orange crosses), two days with rain the whole day (blue dots), eight days with rain only in the evening (orange dots) and seven days with rain only in the morning (blue crosses). Most days with morning precipitation are found beneath the 5 K limit of HI_{low} and most days without rain are found above the limit. The points that match the description of a possible convective event are the orange

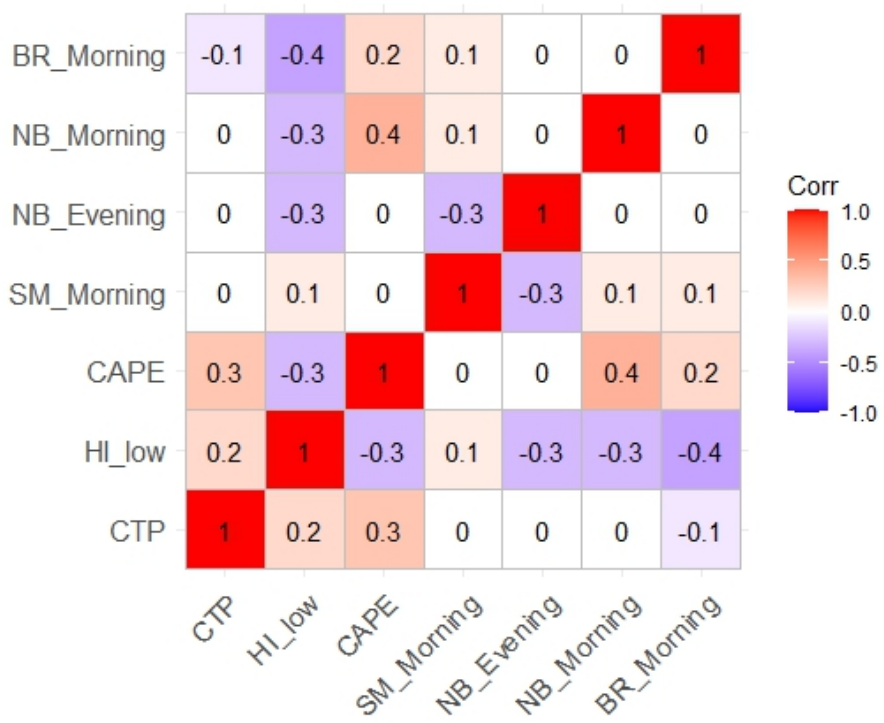


Figure 12: Correlation matrix of the observed variables. All correlations are at or below an absolute value of 0.4. The highest negative correlation is between HI_{low} and the morning Bowen ratio. The highest positive correlation is between the morning precipitation and the CAPE.

dots that had only rain in the evening. Most of these points are located between HI_{low} of 4 K and 20 K.

The CTP values distribution shown in Figure 26 in Appendix 1 is almost normal but with a left tail, and has a mean value of 1056 J/kg. The distribution of the HI_{low} is shown in Figure 27 in Appendix 1, the distribution is closer to a poisson distribution with most points in the lower values.

The sounding dates were classified as either convective or non-convective by their highest CAPE value during the day. When a day achieved a CAPE value above 300 J/kg it was classified as convective and if the CAPE did not reach 300 J/kg throughout the day the day was classified as non-convective. The classification is presented in Figure 14 where the colours and shapes show the same as in Figure 13 but the size shows whether it was classified as a convective event or not. The classification led to three of the dates in Table 3 to be classified as convective. The

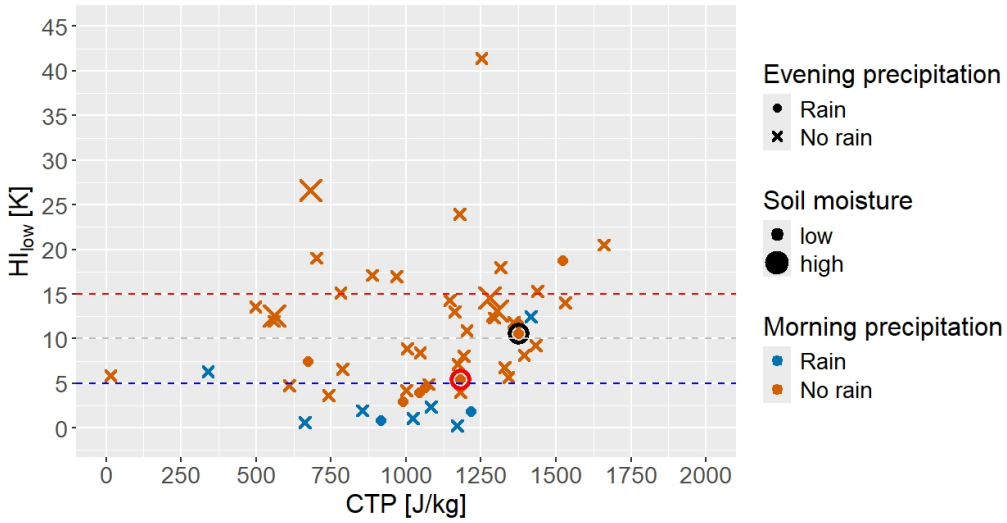


Figure 13: Plot of the observed data where blue colour indicates morning precipitation orange colour no morning precipitation, dots represent evening precipitation, and crosses represent no evening precipitation. The black circle marks the case study of 15.08.2022, and the red circle marks the case study of 26.06.2023. The interesting points in this case are the orange dots, as they did not have any rain in the morning and did get rain in the evening. The size of the points tells us if the soil moisture was categorized as high or low in the hours from 00:00 to 12:00.

points that are classified as convective are placed on a diagonal line in Figure 14.

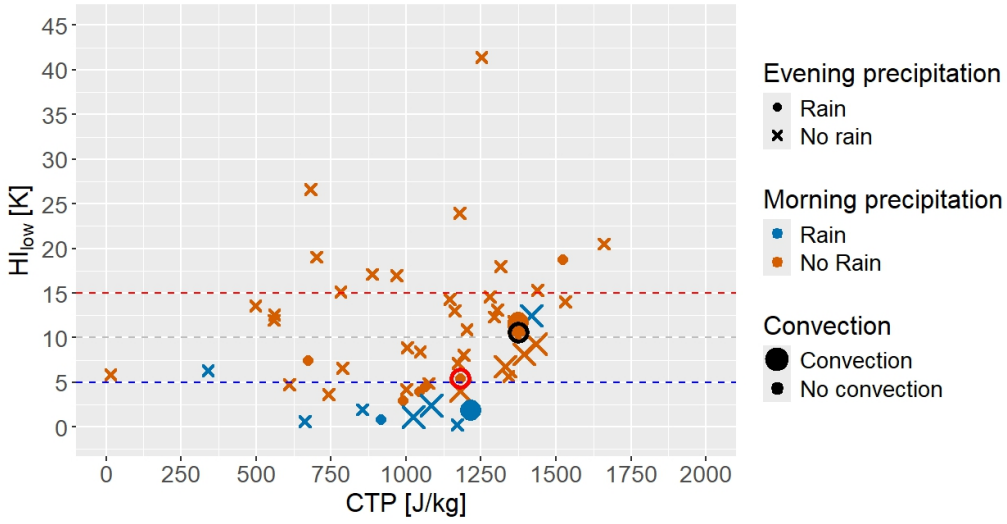


Figure 14: Plot of the sounding data with CTP and HI_{low} as x- and y-axis. The colour differentiates between rain(blue) and no rain(orange) in the morning, and the shape differentiates between rain(dots) and no rain(crosses) in the evening. The size of the point indicates whether it is considered a convective event or not due to its CAPE value extracted from reanalysis data. The black circle marks the case study of 15.08.2022, and the red circle marks the case study of 26.06.2023. The convective points are on a diagonal line in the plot.

Bowen ratio

To see if there is a correlation between convection and the surface heat fluxes the CTP and HI_{low} were plotted with the size showing if the Bowen ratio was considered high or low. This is presented in Figure 15. Most of the cases with HI_{low} below 5 K also have a low Bowen ratio. For the points above 5 K a higher Bowen ratio is mostly observed.

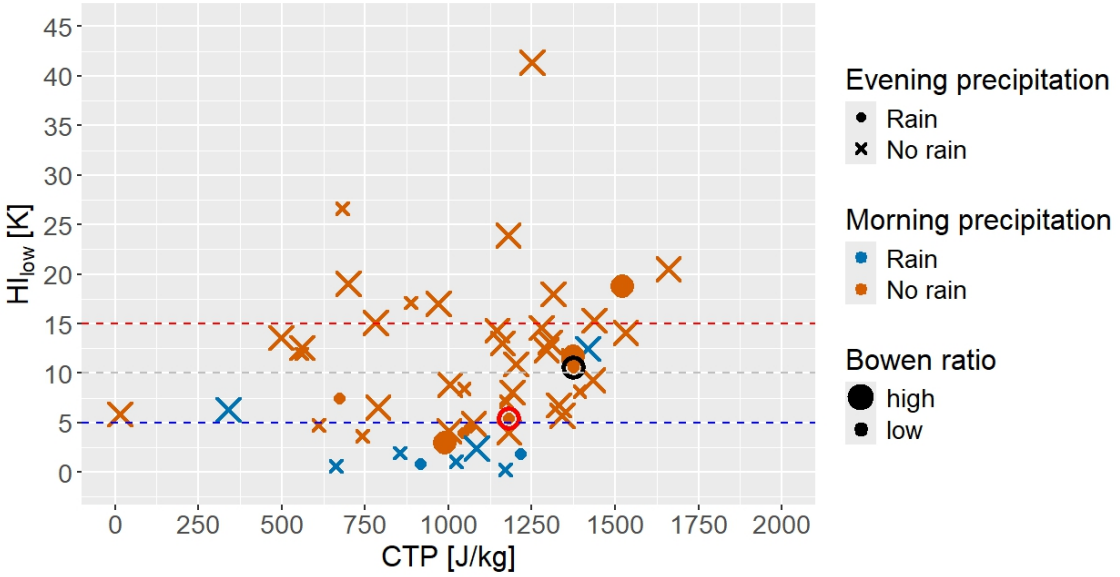


Figure 15: CTP plotted against HI_{low} with the Bowen ratio as size. The bigger points are the ones with a higher Bowen ratio and drier soils. A Bowen ratio with an absolute value above 0.5 was considered high and lower than 0.5 was considered low.

4.1.1 Statistic analysis

A generalised linear model(GLM) is used for the model fitting and as the target is a two-class problem the link function used is the logit function with a binomial distribution. The continuous explanation variables and the factorised target fit with this model. The dataset was standardised before fitting the model equation to achieve a better fit. A reduced model equation was chosen including CTP, HI_{low} and soil moisture in a standardised form. The model equation representing the data is

$$g(p) = b_0 + b_1 x_{1,i} + b_2 x_{2,i} + b_3 x_{3,i},$$

where $g(p)$ represents the link function for the probability of the target convection, b_0 represents the intercept coefficient, b_1 represents the coefficient of CTP, b_2 the

coefficient of HI_{low}, and b_3 the coefficient of soil moisture. The x variables represent the different values of the variables CTP, HI_{low} and soil moisture respectively.

The coefficients obtained in the model fitting are represented in Table 4. In the GLM, CTP and HI_{low} were identified as significant effects with a 5 % significance level. The soil moisture was not found to have a significant effect on convection in the observed data and can therefore be taken out of the model equation without much loss of accuracy.

Table 4: Statistics table of a GLM fitted to a convection target variable with CTP, HI_{low} and morning soil moisture as predictor variables.

Coefficient	**Estimate**	**Std. error**	**z-value**	**P-value**
Intercept	-4.9025	1.5434	-3.177	0.00149
CTP	4.8963	1.7371	2.819	0.00482
HI_{low}	-3.9729	1.4096	-2.819	0.00482
$\mathrm{SM}_{Morning}$	-0.2341	0.5871	-0.399	0.69015

4.2 Reanalysis data

The reanalysis data was used to calculate CTP and HI_{low} values as presented in section 3.2.1 and a dataset including CTP, HI_{low}, CAPE, soil moisture, Bowen ratio and different precipitation variables was created. In the analysis of the reanalysis data, only variables drawn from the ERA5 model are being used, including the vertical temperature profiles, CAPE, precipitation variables, soil moisture and latent and sensible heat fluxes. The correlations between the different variables are shown in Figure 16.

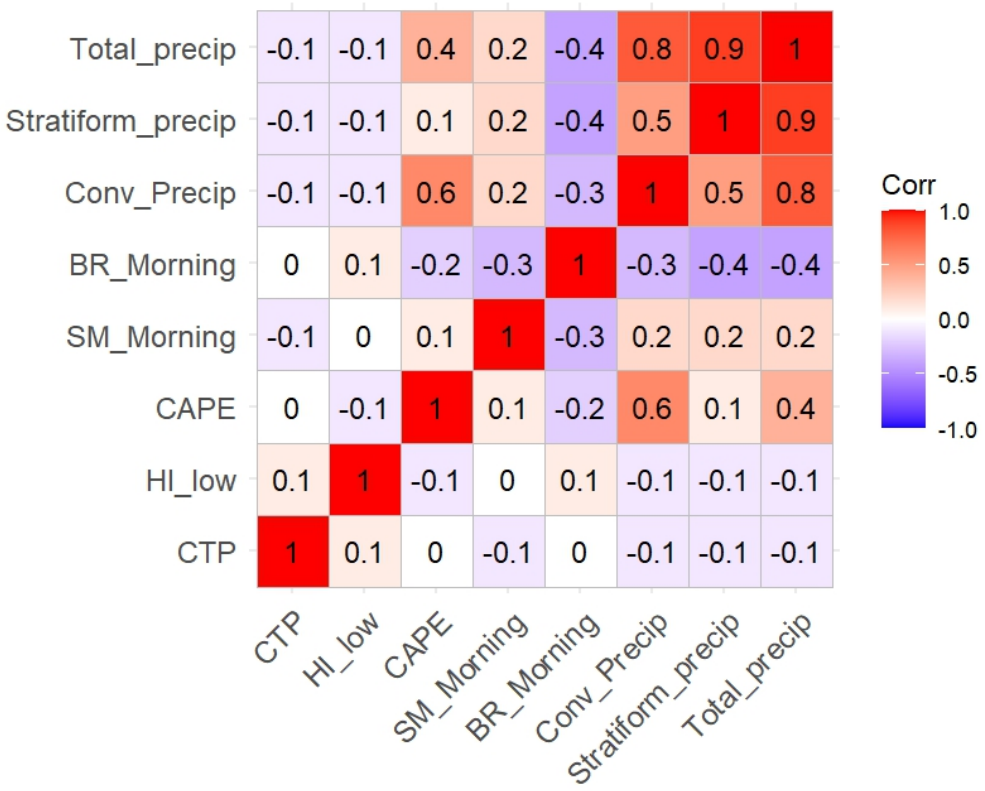


Figure 16: Correlation matrix of the different variables in the reanalysis dataset. The different precipitation variables are highly correlated with values of 0.9, 0.8, and 0.5, where the lowest correlation is between the amounts of convective and stratiform precipitation. There is also a correlation of 0.6 between the amount of convective precipitation and the CAPE. The rest of the variables do not have a high correlation and most correlations are not higher than 0.2.

In the correlation matrix, Figure 16, a high correlation can be seen between total precipitation and the two other precipitation variables convective and stratiform precipitation. This is because they are dependent of each other and the stratiform precipitation variable is calculated from the total and convective precipitation. Further the correlation between CAPE and the convective precipitation, with a value of 0.6, is to be noticed as this is the highest correlation if the correlations between

the precipitation variables are disregarded.

Some of the information contained in the dataset is visualised in Figure 17. The CTP values calculated in the reanalysis data have a wider distribution, lower values, and more cases with negative CTP values than for the observed data, resulting in a mean value of 376.2 J/kg as seen in Figure 28 in Appendix 1. The HI_{low} values span about the same area and are more similarly distributed as the observed data as seen in Figure 29 in Appendix 1. The HI_{low} is close to a poisson distribution and has more points in the lower values.

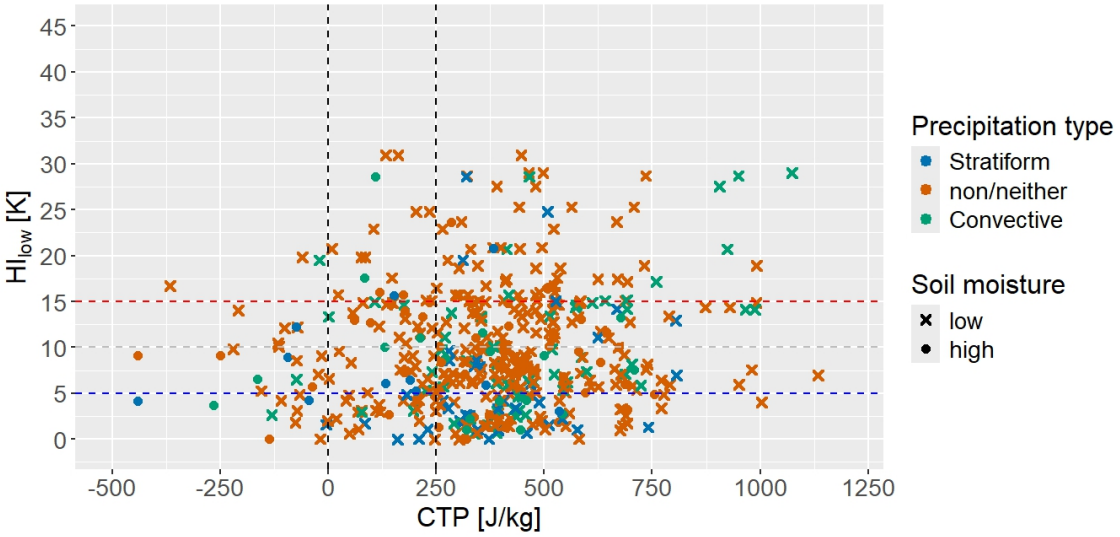


Figure 17: Plot of the reanalysis data from the three summers 2021, 2022, and 2023. The data is coloured by precipitation type where the most dominant precipitation type is represented as blue points for stratiform dominant and green points for convective dominant. The orange points have either no precipitation or equal amounts of stratiform and convective precipitation. The shape of the points indicates soil moisture where the crosses have low soil moisture and the dots have higher soil moisture.

The five days with precipitation from the days with soundings are used for comparison of reanalysis and observed data. The calculated values of the five days are presented in Table 5. The CTP values from the reanalysis data are all lower than the CTP values for the observed data, while the HI_{low} values are both higher and lower. The amount of convective precipitation does not seem to correlate with the CTP values for the reanalysis data. A more distinct correlation is seen between the CAPE and the amount of convective precipitation where the days that achieved a CAPE value greater than 300 J/kg had all above 10 mm precipitaiton and the two days with lower CAPES had below 5 mm precipitation.

Table 5: CTP, HI_{low}, CAPE, soil moisture, and total and convective precipitation on the days where convection was most likely in the observed data. The three days with a CAPE above 300 J/kg all have cumulative amounts of above 10 mm of convective precipitation and the ones with a CAPE below 300 J/kg do not have an accumulative amount higher than 5 mm.

Date	**CTP**	$\mathbf{HI}_{low}$	$\mathbf{SM}_{morning}$	**Precipitation**		**CAPE**
	[J/kg]	[K]	[%]	**Total** [mm]	**Convective** [mm]	[J/kg]
04.07.2021	422.14	5.89	0	10	10	590.53
26.07.2021	354.87	13.30	0	20	17	303.59
15.08.2022	1.17	13.34	0	11	11	922.93
21.06.2023	161.16	0.00	7	14	4	255.46
26.06.2023	280.55	8.58	2	8	2	32.93

Case study July and August 2023

July and August 2023 were rain-heavy months with more rain than average and typical summer convective events were expected to be found. The dates from July and August 2023 were filtered out and plotted by themselves in Figure 18 to look further into the data.

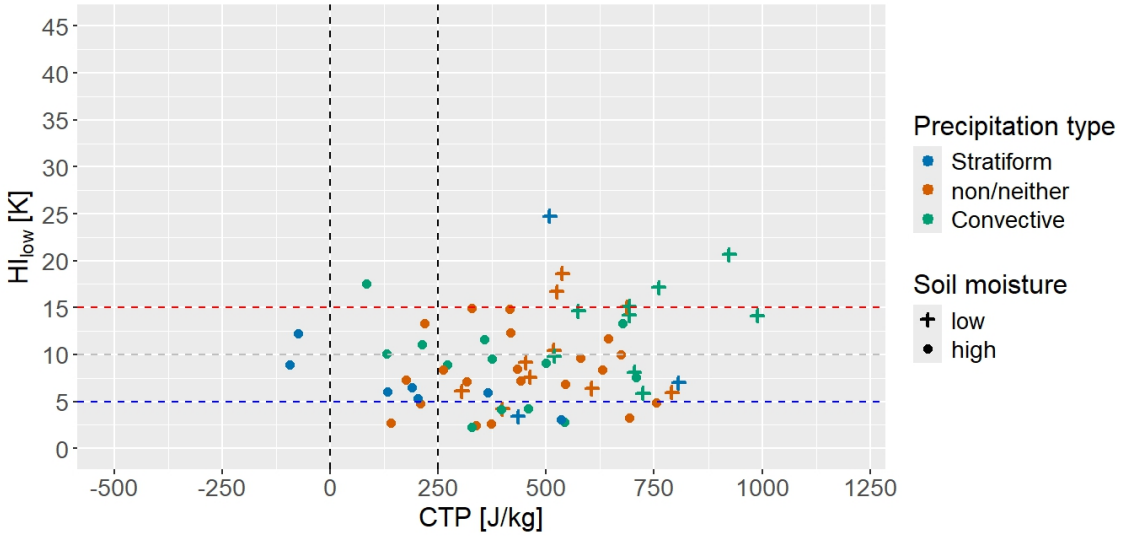


Figure 18: CTP vs HI_{low} for reanalysis data 2023 July and August. The crosses indicate low soil moisture levels, the dots indicate high soil moisture levels. The blue points are dominantly stratiform rain, and the green points are dominantly convective rain. The orange points contain no or equal amounts of stratiform and convective rain.

It is observed that most of the stratiform dominant (blue) points are around the 5 K level of HI_{low} and most of the convective (green) points are between the 5 K and 15 K levels of HI_{low}. It can also be observed that all of the convective dominant

points with a CTP below 500 J/kg have a soil moisture classified as high and the convective points with a CTP above 750 J/kg have a soil moisture classified as low.

Only convective cases

Another filtering is done on the full dataset where only the points with convective precipitation above 1 mm are included. The result of this filtering is presented in Figure 19. It can be observed here that most of the stratiform dominant events are located beneath or around the 5 K level and many of the convective dominant events are located under the 15 K level. It can also be observed that all of the convective events with a CTP higher than 750 J/kg have soil moisture classified as low and most of the high soil moisture convective events have a CTP under 500 J/kg.

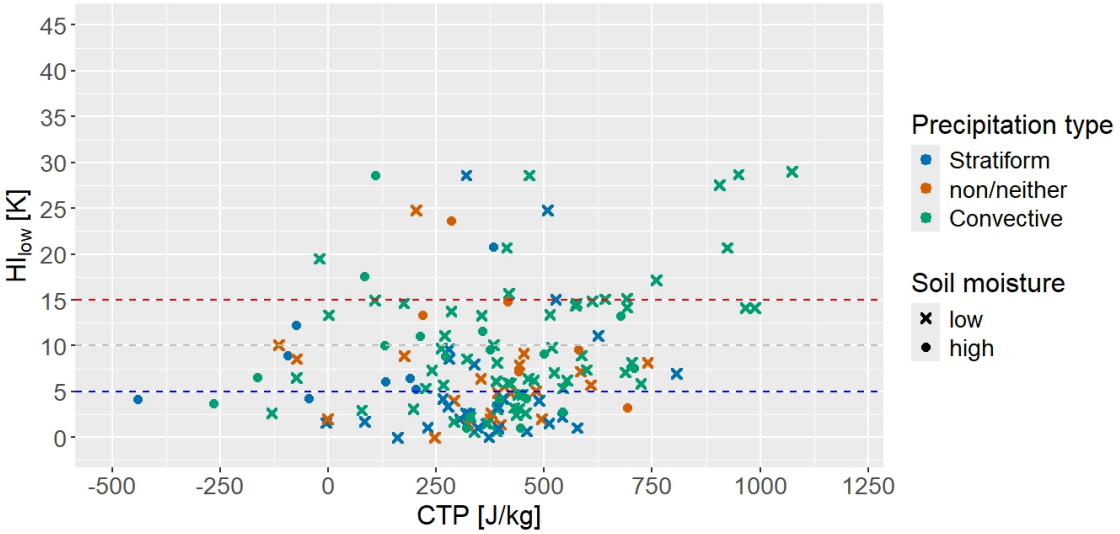


Figure 19: CTP and HI_{low} values for only positive convection cases. The crosses indicate low soil moisture levels, the dots indicate high soil moisture levels. The blue points are dominantly stratiform rain, and the green points are dominantly convective rain. The orange points contain no or equal amounts of stratiform and convective rain..

Dates with a corresponding observed sounding

For a better comparison between the observed data and the reanalysis data filtering was done on the reanalysis data to only include the dates where a corresponding sounding exists. The result from this filtering is shown in Figure 20. It is observed lower values than in the observed data for the CTP where most points are situated between 200 J/kg and 600 J/kg for the reanalysis. The blue points with morning precipitation are mostly situated below the 5 K limit of HI_{low}. Another observation is the orange crosses that had no rain all day and are placed well under the 5 K limit.

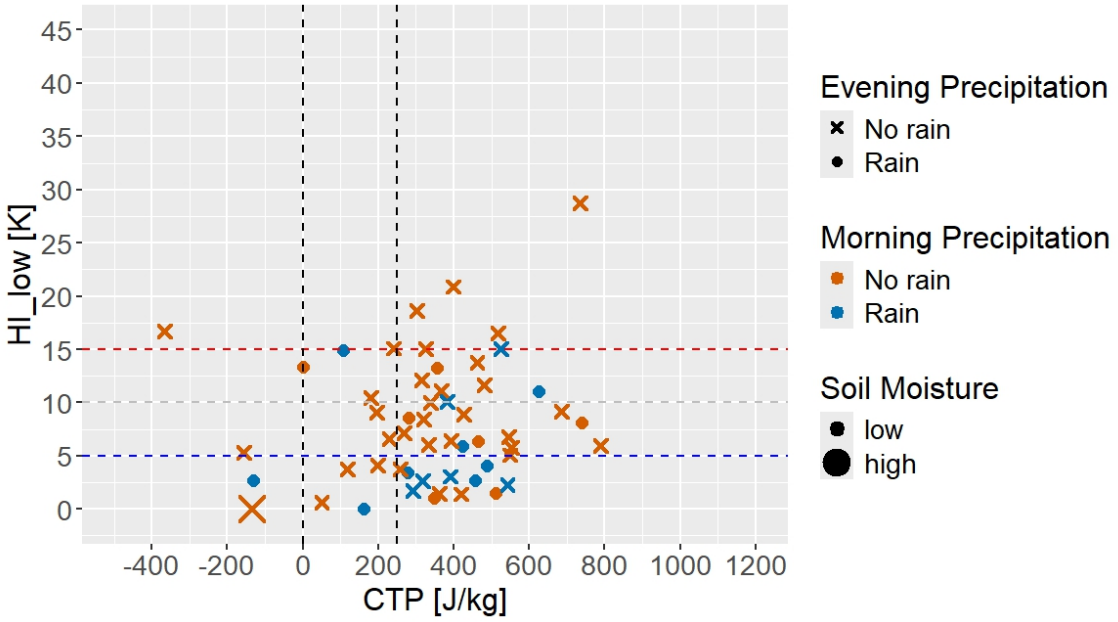


Figure 20: Plot of reanalysis data where morning and evening precipitation is portrayed. Orange dots had no rain in the morning but rain in the evening, orange crosses had no rain in the morning or evening, blue dots had rain the whole day and blue crosses had only rain in the morning. The interesting points are the orange dots with no morning rain and more evening rain. The size of the points indicates high or low soil moisture.

4.2.1 Statistic analysis

A GLM was fitted to the reanalysis data and the model followed this equation

$$g(p) = b_0 + b_1 x_{1,i} + b_2 x_{2,i} + b_3 x_{3,i},$$

where g(p) is the link function connected to the probability of convection, b_0 is the intercept, b_1 to b_3 are the coefficients fitted to respectively CTP, HI_{low}, and soil moisture, and the x variables represent the input values of the variables respectively CTP, HI_{low}, and soil moisture.

The results from the fitting of the model to the reanalysis data are presented in Table 6. In the fitting of the model, the soil moisture is the only effect that has a significant effect with a 5 % significance level.

Table 6: Statistics table of a GLM fitted to a convection factor as response and CTP, HI_{low} and morning soil moisture as predictor variables for the reanalysis data. The morning soil moisture was the only variable with a significant effect with a significance level of 5 %.

Coefficient	Estimate	Std. error	z-value	P-value
Intercept	-1.64701	1.10801	-1.486	0.13716
CTP	0.26623	0.18554	1.435	0.15133
HI_{low}	-0.20187	0.10923	-1.848	0.06458
Soil moisture	0.13537	0.05049	2.681	0.00733

4.3 Comparison of observed and reanalysis data

The results from the reanalysis and sounding data are different and therefore a comparison of different parameters used in the calculations is done.

4.3.1 Correlations of air temperature

The air temperature is used directly in the calculation of CTP and HI_{low} and is therefore interesting to compare. The vertical air temperatures are plotted against each other in Figure 21 for the same pressure levels and are filtered to only contain the pressures from 1000 hPa to 700 hPa. In Figure 21 it can be observed that an overestimation is done in September and August and an underestimation is done in May and June. The correlation seems to follow a linear trend besides the over- and underestimations.

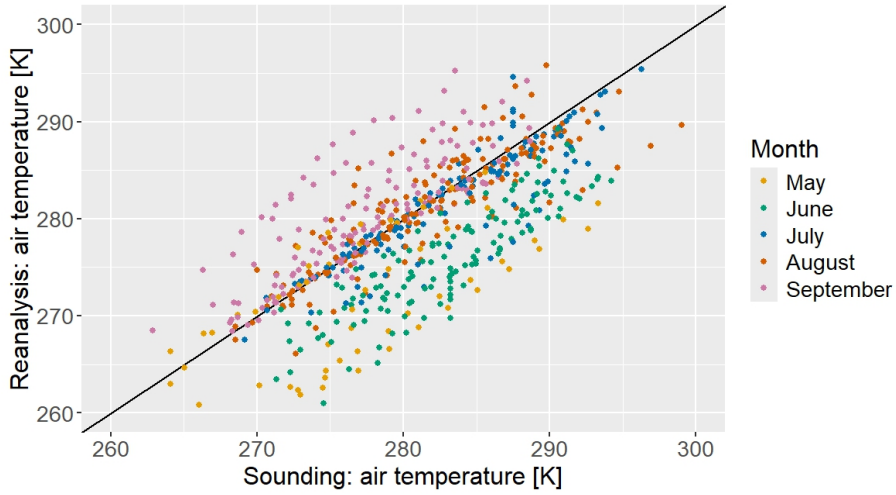


Figure 21: Correlation of vertical profile of air temperature between the reanalysis on the y-axis and the radiosonde on the x-axis. The colour represents which month the data is representing. An overestimation is seen in September and an underestimation is seen in May and June. The black line represents a perfect correlation between the two parameters.

4.3.2 Correlations of relative humidity

The relative humidity is used for calculations of the dew point temperatures used in HI_{low}. The correlations are plotted in Figure 22. The relative humidities are plotted for the same pressure levels and only for the levels between 1000 hPa to 700 hPa. The correlation plot shows no correlation and the plot looks like random noise.

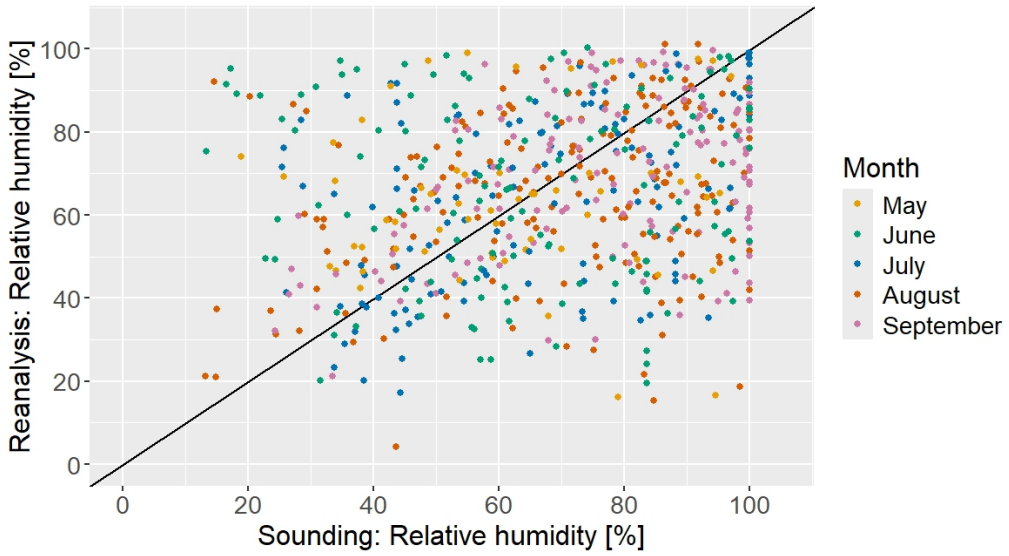


Figure 22: Correlations between the relative humidity in the reanalysis data on the y-axis and the sounding data on the x-axis. The colour indicates which month the data represents. The data is from the same dates but not necessarily from the same hour. The black line represents a perfect correlation between the two parameters.

4.3.3 Correlations of dew point temperature

The dew point temperatures are used to estimate HI_{low}, and for the reanalysis data, the dew point temperatures are estimated from relative humidity and air temperature amongst others.

The correlation of the dew point temperatures is plotted in Figure 23, and the dew point temperatures are plotted for the same pressure levels and are filtered for the pressure levels from 1000 hPa to 700 hPa. There is less of a pattern in the dew point temperatures but still a similar linear trend. The lower values are less correlated than the higher values.

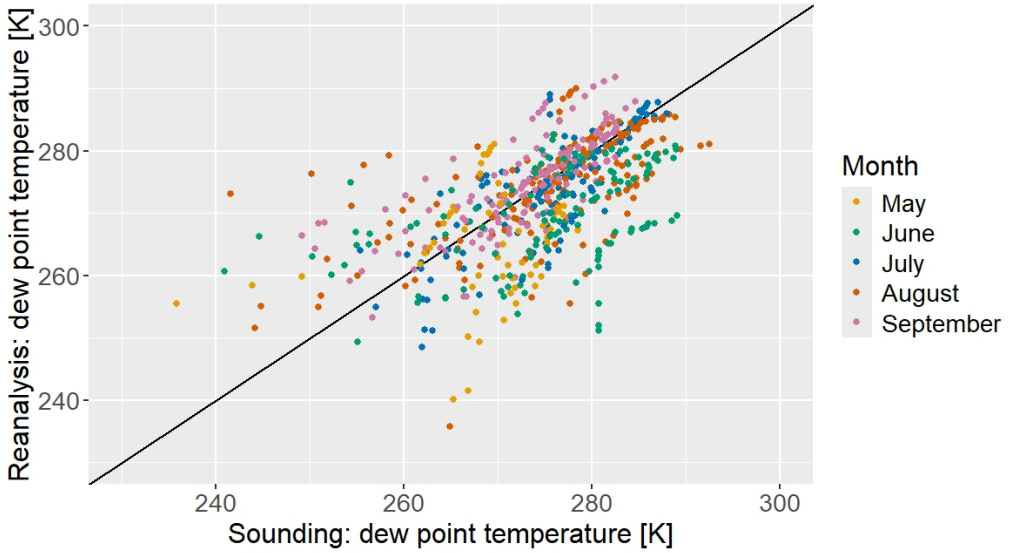
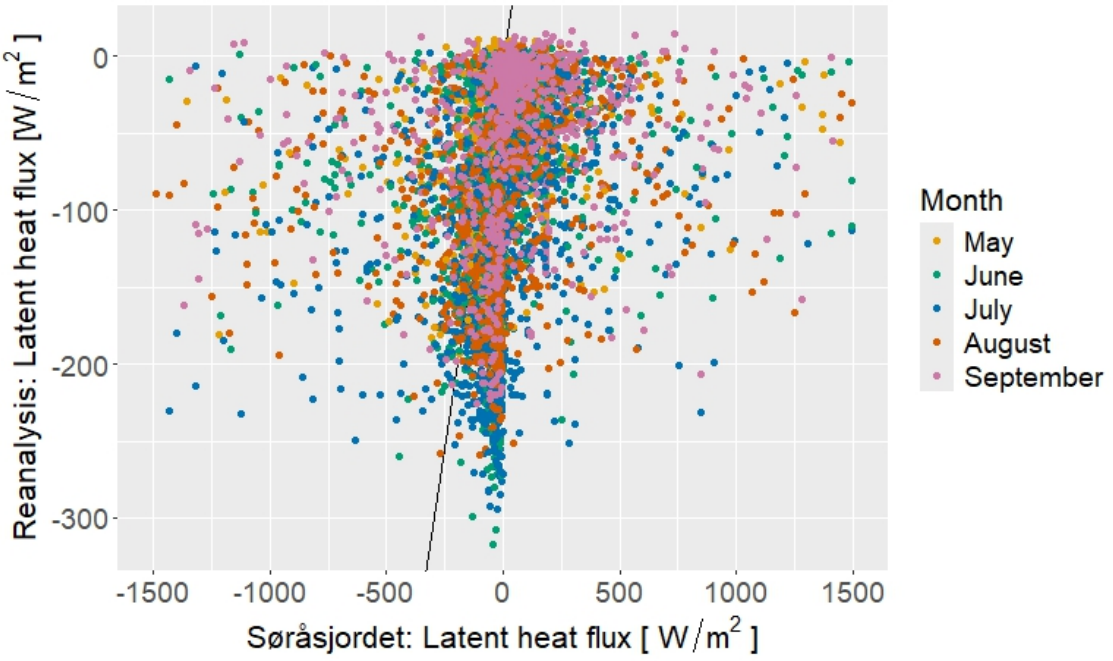


Figure 23: Correlations between the dew point temperature from the reanalysis on the y-axis and radiosonde on the x-axis. The colour represents which month the data represents. The black line represents a perfect correlation between the two parameters.

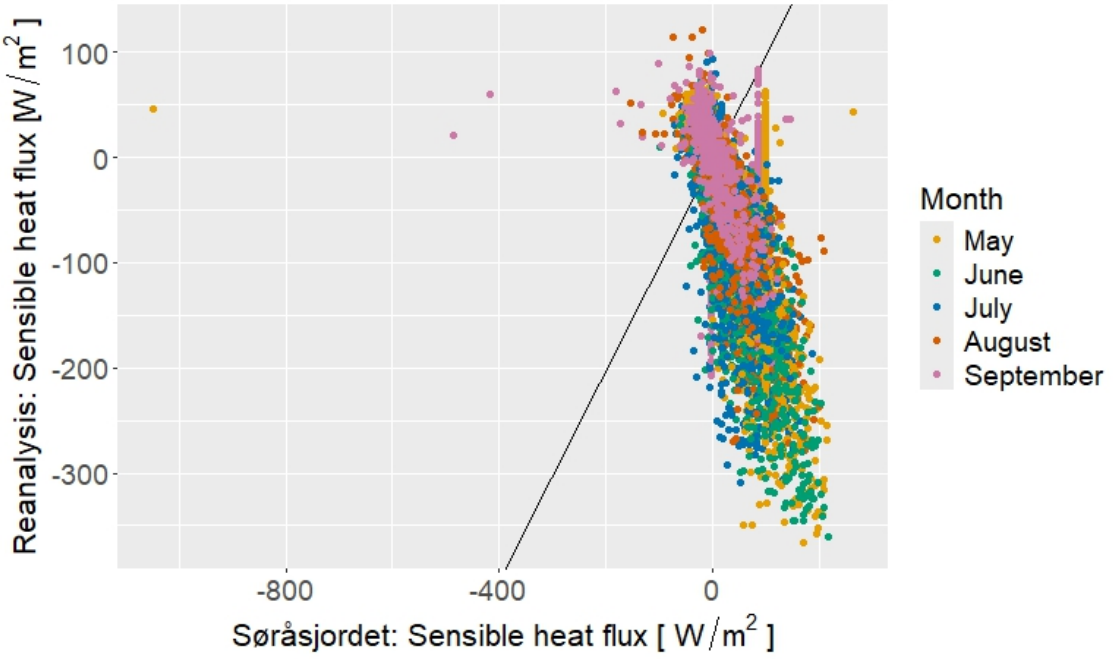
4.3.4 Correlations of heat fluxes

The correlations between the latent and sensible heat fluxes in the sounding and reanalysis data are shown in Figure 24. The latent heat flux in Figure 24a has most points near the black line representing perfect correlation. There is also a big variance in the plot with many miscalculations of the latent heat flux as soon as the flux deviates from zero.

The correlation of sensible heat between the two datasets is shown in Figure 24b. There is a negative linear correlation and an overestimation of the sensible heat flux as it gets greater than zero. The big deviances from the observed data lead to the Bowen ratio being taken out from the analysis of the reanalysis data.



(a) Correlation of reanalysis and observed latent heat flux.



(b) Correlation of reanalysis and observed sensible heat flux.

Figure 24: **(a)** Correlation between the latent heat flux in reanalysis and observed data. The correlation seems to be mostly linear as most of the points surround the black line that represents perfect correlation. **(b)** Correlation between latent heat flux from reanalysis data and observed data. There is a linear correlation between the two but it is negative and with an overestimation of the absolute values in the reanalysis data.

4.3.5 Correlations of soil moisture

The correlation in soil moisture in Figure 25 has a lower variance for values of soil moisture around 30 %. For other values of soil moisture, the variance in the data gets bigger and spans a bigger area indicating a low correlation between the observed data and the reanalysis data. The maximum value in the reanalysis data for soil moisture is around 40 % where the real data spans all the way to 68 %. The points creating lines in the plot are due to the nature of soil moisture that gets high with intense precipitation and steadily decreases as the soil dries afterwards.

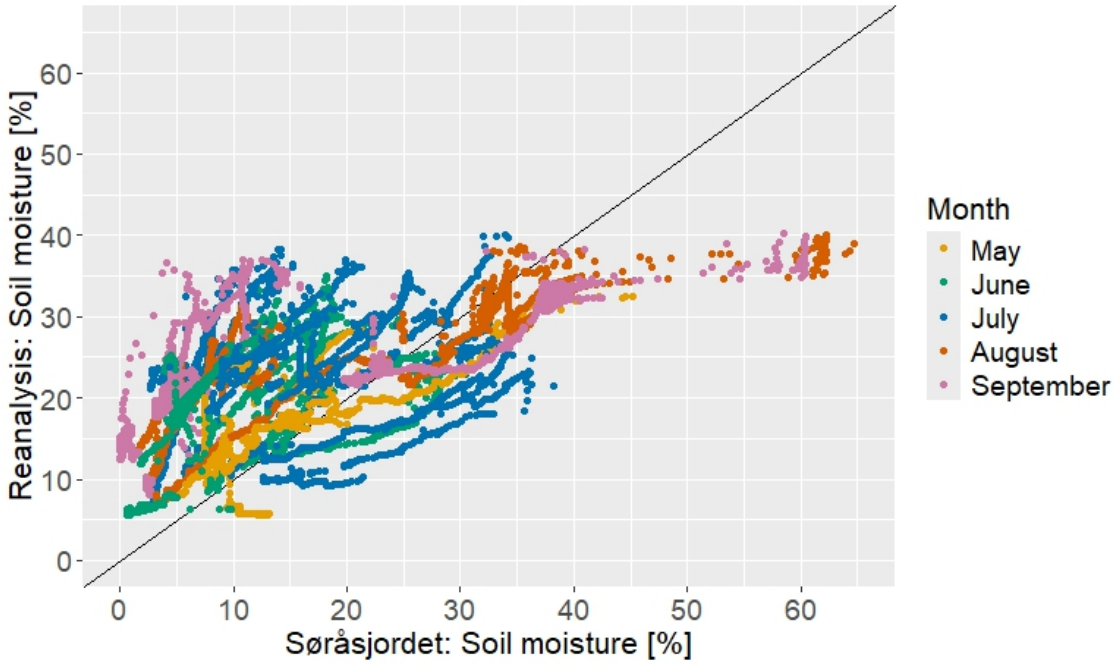


Figure 25: Correlation of soil moisture in sounding and reanalysis data. The colours indicate which month the data represents. There is a difference in the maximum values with a difference of 20 %.

5 Discussion

The results from the analysis will now be discussed in the same order as they were presented in section 4. First to be discussed is the observed data. Then the reanalysis data will be discussed in the same order as the observed data before a comparison of the reanalysis data and the observed data is done. Lastly the limitations of the analysis will be discussed.

5.1 Observed data

In this section, the results from the observed data will be discussed. First, an evaluation of the CTP-HI_{low} framework will be done, followed by an evaluation of the classification of convective events. Then the soil moisture and Bowen ratio will be discussed before an evaluation of the dataset's quality is done.

5.1.1 CTP-HI_{low} framework

The validity the CTP-HI_{low} framework for Norway would depend on how the parameters describe the actual observations and data at hand. In Figure 12 a weak positive correlation of 0.3 is found between CTP and CAPE. This can indicate that the CTP can be used as an instability measurement and that a high instability indicated by CAPE goes along with a high instability indicated by CTP as well. This is also seen in Table 3 where the higher CTP values occur with the higher CAPE values and the lower CTP values with the lower CAPE values. The high CTP values displayed here and in Figure 26 in Appendix 1 could be due to a temporal variability that is a known characteristic of CAPE. CTP will probably also have a diurnal cycle with rising values during the day with a maximum in the afternoon, such as the CAPE.

The global radiance plot for the case study of 15.08.2022 [8], matches the expected radiance for a day with a convective event where the sky is clear in the morning and gets cloudy during the day. The amount of precipitation, visualised in Figure 9, between 16:00 and 20:00 is not necessarily intense enough to state it was a convective event. Nevertheless, the high CAPE value of 923 J/kg strengthens the case for this date to have a convective event. From the CTP and HI_{low} values this day should be a dry advantage day and therefore should have given a convective event given the dry soil recorded. The HI_{low} is just over 10 K which can be one reason for the convection not resulting in heavy rain. The convection might have been deeper if the air was even drier and better placed into the dry advantage zone.

The global radiance plot, [10], and the precipitation plot, [11], for the case of 26.06.2023, have characteristics that match the expectations of a convective event. The CTP and HI_{low} values of respectively 1181 J/kg and 5.5 K from Table 3, suggest the case should have a wet advantage with more convection over wet soils. The soil moisture recorded was low which did not seem to slow the convection down. The CAPE value on the other hand with a value of 33 J/kg suggests a quite stable atmosphere where convection would be unlikely. The mismatch between the two frameworks for convection could be due to the CAPE value being taken from the ERA5 reanalysis and not being calculated from the soundings themselves. This could be due to a low correlation between the reanalysis and the observed data. Which might be caused by the coarse spatial resolution of the reanalysis data, missing the small-scale convection.

The CTP values were close to normal distributed as shown in Figure 26 in Appendix 1, which shows that the soundings that were taken could make a somewhat representative subset of the summers in Ås. The mean value of the CTP from the soundings is regarded as a high value of CTP according to Findell and Eltahir (2003a) who set medium levels of CTP to be between 0 J/kg and 200 J/kg, and above 200 J/kg as high values. All, but two, of the CTP values calculated are above 500 J/kg. This might be due to the initiation of the soundings being mostly at 13:05 as the CTP might have the same temporal variability and cycle as the CAPE. The patterns shown in Figure 7, are not seen as clearly in this data. More soundings on days that rained could give a better picture of when convective precipitation occurs and more soundings on days with higher soil moisture are needed to balance the dataset better.

The difference of wet and dry advantage with CTP and HI_{low} values are not observed in this dataset. This might be because of the few instances of wet soil and the many instances of dry soil. One reason behind this is because of many missing values in the dataset especially in 2021, where the whole of May and June had missing soil moisture values. This therefore had to be excluded from the analysis. The months covered by soundings were also drier than normal which made the dataset more biased with less rain and fewer instances of wet soils.

The HI_{low} limits of 5 K and 15 K fit nicely with the data from the radiosondes. The lower limit could have been even lower according to Figure 13 where some points with no rain are observed under the 5 K level and most of the points with morning rain are located well under the 5 K level. A proposal for a new limit on the basis of the observed data would be at 4 K as the morning rain cluster is situated beneath

4 K and some convective points are found between 4 K and 5 K. More data is needed to have more confidence in the limits of HI_{low}.

There is not enough representation in the soil moisture data of days with higher soil moisture values. Four out of 53 days were categorised as having a high soil moisture content which resulted in an unbalanced dataset. Better measurements with fewer missing values could help have a more balanced dataset. The balance in the dataset could also improve with more dates covered by the soundings and therefore more dates to include for soil moisture and the other variables.

The CTP and HI_{low} variables were significant effects with the same p-value of 0.00482, as presented in Table 4 in the GLM. The GLM was fitted with a logit link function, with convection as the target variable. This might indicate that the framework could be used as a link between the morning boundary layer and later convection on the same day. On the other hand, the model that was fitted did not fit the data well and did not fully fulfil the constraints of randomly distributed residuals or normal distributed values. Although the computation of the coefficients did converge it is not necessarily a good fit.

A density-based clustering was performed with no luck in creating any clusters or detecting any patterns in the real dataset. This might be because the dataset is too small to see tendencies and clusters and additional soundings would be needed for more identifications of clusters or systems to appear in the data.

5.1.2 Classification of convective events

The classification of whether a day with a radiosonde had a convective event or not was done by using a threshold of CAPE values from the ERA5 model. The threshold was chosen to be 300 J/kg as some studies had found thunderstorms to occur with CAPE values as low as 150 J/kg and 200 J/kg in Russia and the Netherlands respectively. A threshold that was a little higher was then chosen since this data is only from the summer months where higher energy is expected to be needed for convection. Based on the correlations between the vertical profiles of radiosonde data and model data from ERA5 of air temperatures as shown in Figure 21 and relative humidity as shown in Figure 22 a mix of both reanalysis data and historical data might not be a good dataset to analyse. This is because the radiosonde data will describe one atmosphere as it was historically, and the ERA5 model data will describe a modelled atmosphere with other characteristics regarding relative humidity. The values from the ERA5 model would therefore not be a good approximation of the characteristics of the historical data.

The CAPE values were retrieved for the whole day and the maximum value of the day was used for checking if the CAPE grew enough through the day to achieve a potential energy of above 300 J/kg. Since the CAPE values were used to classify convective events the convective events that did not lead to rain could also be included as it was not the rain itself that was used for classification.

The use of CAPE values from ERA5 could give a wrong picture of the actual events on the dates with soundings. Instead of using ERA5 values for the CAPE, observed values of CAPE could have been made if there were more than one sounding for each day. The CAPE builds up during the day and gets released in the case of a deep convection event or rainstorm. Therefore the change of CAPE could be a good indicator for whether a convective event did happen or not. This would require several soundings on each day.

Another way to detect convection is by the vertical wind velocity. Convection itself is an updraft of wind and the vertical wind velocity would therefore be a more direct observation of convection. This would also need several measurements throughout the day and a good algorithm for calculation of vertical wind speed throughout the atmosphere from sounding data (Zhang et al., 2019).

5.1.3 Soil moisture and Eddy Covariance

Soil moisture

For all five precipitative days presented in Table 3 the soil moisture is low. With the SoilVUE sensor, the soil moisture level was measured to be between 2 % and 6 % while the GroPoint measured soil moisture of 38 % and 39 % when available. The five days had HI_{low} values spanning from 5 K to 12 K and all had a CTP above 1000 J/kg. Based on the imbalance of soil moisture distribution in the dataset with 4 days of high soil moisture, no splitting of the data due to soil moisture is seen. The soil moisture also showed no correlation with CTP and HI_{low} parameters in the correlation plot, [12], where the correlation was calculated to 0.0 and 0.1 respectively. This is also further supported by the GLM fitted to the dataset, with coefficients presented in Table 4, which found no significant effect of soil moisture on convection.

The two soil moisture sensors are measuring quite different values of soil moisture where the GroPoint has been found to give the most realistic results. The SoilVUE on the other hand had the longest continuous timeseries and is therefore the one used for classification as it is only classified as high or low soil moisture. The SoilVUE

sensor did catch the patterns of higher and lower soil moisture just as well as the GroPoint sensor and is therefore used in the classification of soil moisture.

Eddy Covariance

The latent heat flux is dependent on available moisture to start removing moisture from the surface and into the atmosphere. This moisture could come from soil moisture, plants, and foliage but could also originate from other moisture sources nearby. Exactly where the moisture is coming from is not easy to measure which is why the term evapotranspiration is used to cover some sources. A low Bowen ratio will not necessarily indicate wet soil but could generally indicate a high moisture availability.

The Bowen ratio for the five precipitative days in the sounding dataset has a range of values in the order of 10^{-2} to 1 which shows that some days have a wetter soil or more available moisture than other. When the Bowen ratio has an absolute value below 1, the latent heat flux is bigger than the sensible heat flux. This indicates that especially 04.07.2021 was a day with more available moisture than the other days. The Bowen ratio also indicates that 15.08.2022 and 26.06.2023 had similar availability of moisture. The availability of moisture can come from soil moisture. This can be confirmed by the SoilVUE sensor which has a difference of only 1% between those two cases. The Bowen ratio could therefore be a good indicator of soil moisture, where there is missing data.

The correlation between CTP and the morning Bowen ratio is small and has a value of -0.1, [12]. The lack of correlation between Bowen ratio and CTP is also seen in Figure 15, where any CTP value corresponds to both high and low Bowen ratios. A weak correlation between HI_{low} and the Bowen ratio is, however, present with a value of -0.4, [12], and splitting of Bowen ratio can be seen in Figure 15 where most of the instances with a low Bowen ratio are situated under the 5 K line of HI_{low}. This is probably because they both record the humidity at a lower level and the humidity at the 950 hPa level and the surface level could be similar.

5.1.4 Quality of the dataset

The sounding dataset contained mostly weekly soundings and some extra soundings on interesting days beginning 04.07.2021 and through the whole summer of 2022 and until 14.07.2023. Where most of the soundings were initiated at 13:05 local time. If a day had several soundings, the one that was closest to the morning was chosen to represent the current day. This resulted in 53 usable soundings. This was to better

recreate the original analysis from Findell and Eltahir (2003b).

The other measurements of precipitation, soil moisture and heat fluxes were filtered out from continuous measurements from the field station at Søråsjordet. In this dataset, there were not many missing values except for the time series of soil moisture. The soil moisture time series consists of measurements from a GroPoint sensor and a SoilVUE sensor which had different measurements. The measurements differed in the range of values but contained the same patterns. The GroPoint sensor with the better accuracy of values had a lot of missing values in 2021 and 2022 and therefore the soil moisture used in the plots is from the SoilVUE sensor. Longer continuous measurements with fewer missing values from the more accurate GroPoint soil moisture sensor would possibly make the dataset better.

5.2 Reanalysis data

In this section, the results from the reanalysis dataset will be discussed. First is an evaluation of the CTP- HI_{low} framework followed by an assessment of the classification of convective events before an analysis of the soil moisture and Bowen ratio is done. Lastly, the GLM is evaluated.

5.2.1 CTP-HI_{low} framework

In the reanalysis dataset, the distribution of CTP values is close to normal distributed as seen in Figure 28 in Appendix 1, and the mean value was closer to a medium range of CTP than the mean value for the observed dataset. The distribution of values around a mean of 376 J/kg is closer to the same distribution as seen in Findell and Eltahir (2003b). This is most likely due to the time the data was extracted for, which was 09:00. This time is closer to the sunrise and indicates that a temporal variation is seen in CTP.

In the correlation matrix, [16], no correlation is found between the different forms of precipitation and the CTP and HI_{low} as they all have a correlation with a value of -0.1. This is also seen in Table 5 where it seems random whether a point has high or low CAPE and CTP values which also has a correlation of 0 in the correlation matrix, [16].

A correlation that is seen in both the correlation matrix, [16], and Table 5 is the correlation between the highest achieved CAPE and the amount of convective precipitation. The three days with the highest CAPE also have the most convective precipitation of the five. This indicates that a CAPE derived from the actual ver-

tical temperature profile could be a good classifier of convective precipitation. The CAPE and the amount of convective precipitation also have a higher correlation of 0.6 in Figure 16.

From Table 5 one can also see that the date with a CAPE of 255.46 J/kg has little convective precipitation compared to the total precipitation and the date with a CAPE value of 303.59 J/kg has almost only convective precipitation. This may suggest that a limit of 300 J/kg for the likelihood of deeper convection would be a good estimate for the data from Ås. This strengthens the hypothesis that less energy is needed for deeper convection in Norway and the higher latitudes.

In the data from July and August of 2023, [18], most of the stratiform dominant rain is located around the 5 K level of HI_{low} with some points having higher HI_{low} values of 12 K and 25 K. All of the convection points with a CTP under 500 J/kg had a soil moisture content classified as high and the convection points between 500 J/kg and 750 J/kg were classified as both high and low soil moisture content. The convective cases with a CTP higher than 750 J/kg are all classified as having low soil moisture which might indicate that a better limit for medium-range CTP values in July and August might be situated between 500 J/kg and 750 J/kg for Ås. A pattern like this is however not observed in the full dataset, [5] and a proper suggestion of where the limits between wet and dry advantage could be for Ås can not be done.

In the subset of only positive convective events, [17], there are also only points with low soil moisture for CTP values above 750 J/kg. This strengthens the presence of an upper limit of the medium-range to be situated here. A lower limit for medium-range CTP values is however not found in the plot of positive convective events. As there are both high and low soil moisture convective events for the lower CTP values.

Most of the stratiform dominant precipitation days are situated under the 5 K limit and fit with the atmospherically controlled limit for humid air. This strengthens the placement of the 5 K limit in the HI_{low} as the limit for where it is expected to become rain independently of the value of the CTP. Most of the convective dominant cases are between the 5 K and the 15 K lines which strengthens the placement of these limits. However, the limits could have been tweaked to fit the Norwegian climate better. From Figure 17, it seems as the lower limit of originally 5 K could have been moved further down to higher humidity and the upper limit moved to lower humidity. This would include more of the datapoints that are convective but more data would be needed for this to be done. The clustering in the original framework

was distinct and did not float as much into each other as these do and one may therefore argue that the land-atmosphere coupling in Norway is not as strong as in Illinois where the framework was first developed.

In the GLM for the reanalysis data, only the soil moisture was identified as a significant effect with a significance level of 5 % and the CTP and HI_{low} are therefore not identified as significant effects for the occurrence of convection. With a significance level of 10 % the HI_{low} is also a significant effect. This might indicate that the humidity of the atmosphere might have an effect on the occurrence of convection this would need to be further investigated to have anything more sure to say about this.

Clustering was tried on the reanalysis data as well without a clear outcome, there were no apparent clusters to be found in the dataset and therefore also no connection between the soil moisture, CTP and HI_{low} values. A tendency of most stratiform precipitation to have a HI_{low} of around 5 K and lower could be observed. This could indicate that a lower HI_{low} could lead to rain regardless of wet or dry soil advantage as stated in the original framework.

5.2.2 Classification of convective events

In the reanalysis data, the classification of convective events was done using the parameter convective precipitation. This was one of the parameters that was available for single pressure levels in the ERA5 model and gives an estimate for how much precipitation came from convection on an hourly basis. Two different classifications were done using this parameter. One binary classification where an event was classified as convective if the total amount of convective rain during the day was higher than 1 mm and one multiclass classification where the dominant precipitation type was classified. The dominant precipitation type was set as convective if there was over 1 mm more convective precipitation that day, stratiform if there was recorded over 1 mm more stratiform rain than convective rain and as non/neither if there were no or equal amounts of convective and stratiform rain during that day. The binary classification was used for the GLM and the multiclass classification was better for the visualisation of atmospherically controlled and not atmospherically controlled days.

To better compare the reanalysis data and the sounding data the CAPE should maybe have been used for classification. This would lead to the same basis of parameters being used in the classification and maybe some convective events that did not lead to precipitation could have been included in the convective class. On the

other hand, the same data would be used for classifying the two different datasets and that would not be the best outcome either. To use CAPE as a classifier for both datasets, the CAPE belonging to the observed data should have been computed from the soundings and the reanalysis CAPE used on the reanalysis dataset.

The reanalysis data also contained information about the vertical wind velocity in the atmosphere. Classification could therefore also have been done through evaluation of the vertical wind velocity for the reanalysis data. This would not have given us any better basis of comparison as the observed dataset did not contain this information, and there was not enough time to implement a computation of the vertical velocity. For future evaluations and analysis of convection, it is suggested to compute CAPE values from real soundings or the vertical wind velocity for better comparison with reanalysis data.

5.2.3 Soil moisture and Eddy Covariance

The observation of a division of datapoints by soil moisture in the plot of July and August 2023, [18] might indicate a soil moisture dependency. This is although not seen as clearly in the full dataset, [17]. The full dataset shows a more random distribution of convection and soil moisture. This might be because of the imbalance in the dataset in regards to the soil moisture. In 2021 and 2022 there were no instances of soil moisture above 25 % and in 2023 there were some instances. This leads to a dataset with almost only low soil moisture values and these three years are therefore not suited to analyse an interaction between surface conditions and convection.

5.2.4 Generalised linear model

In the GLM the soil moisture was the only variable that was found to have a significant effect on the convection for a significance level of 5 %, as seen in Table 6. The CTP and HI_{low} was not found to be significant for a 5 % significance level. The outcome could be because most of the convective events occurred with the presence of low soil moisture contents and the soil moisture data is unbalanced. The fit of the model to the data is therefore questionable and more balanced data could be of help in fitting a better model.

5.3 Comparison of observed data and reanalysis data

In this section, a comparison of the observed data and the reanalysis data is done. The results from each CTP-HI_{low} implementation is compared and the different

correlation plots are further discussed.

5.3.1 CTP-HI$_{low}$ framework

In the radiosonde data the mean value of the CTP at 1057 J/kg are higher than in the reanalysis data where the mean was 376 J/kg. This could be due to the time of day that the radiosondes were initiated. In the original framework, the radiosondes should be initiated at sunrise and capture a predisposition of the atmosphere for convection or not. Most of the soundings in Ås were initiated at 13:05 local time and since CTP is defined similarly to CAPE but with other limits it is logical to assume that the CTP also will change with more energy in the system as the day evolves. More accurate CTP values from the soundings could therefore be obtained with soundings that are initiated near sunrise which would be around 04:00 - 05:00 in the summer in Ås. The CTP values from the reanalysis data is computed on soundings initiated at 09:00 and is therefore closer to the same distribution as in Findell and Eltahir (2003b).

The CTP and HI$_{low}$ was only found to be significant variables in the model fitted to the observed data and not the reanalysis data. This is most likely due to the more linear placement of the points classified as convective in the observed data than the reanalysis data. Since the parameters were identified as significant a further analysis with a better data foundation is needed.

Many of the other analyses that have been done on land-atmosphere interactions during convective events, with CTP-HI$_{low}$, have been done using a combination of real soundings and models to model different outcomes based on different surface conditions. The earlier studies have also been done on bigger areas on the size of continents and not on the size of one municipality. If a similar data foundation should exist with only observed data a continuous time series of 10 years would be a start.

5.3.2 Correlations between the vertical profiles

The reanalysis dataset did not have a lot of missing values with an exception from the CIN which had more missing values than actual values and was therefore disregarded. The assessment of the use of the reanalysis dataset therefore concerns mostly the quality of the data and the correlations with reality. Therefore correlation plots of the different parameters that should contain the same information were made to assess the quality of the dataset and the analysis.

The vertical profiles had a linear correlation with the air temperature even though some over- and underestimations were observed in the fall and spring months, [21]. The soundings were mostly launched at 13:05 and the reanalysis data of the vertical profiles was retrieved at 09:00. This could have been one of the reasons for the correlation not being a perfect fit as the temperature usually varies through the day.

The vertical profile of the relative humidity in Figure 22 was not correlated with the sounding data and is more questionable to use as a substitute for real data. One reason could be that the grid used for computing the data is too big and that local differences in relative humidity and other variables that affect the relative humidity are too big to be handled well by the grid used in the computations.

The dew point temperature is computed using the relative humidity and the air temperature. The correlation between the dew point temperatures of the sounding and reanalysis data in Figure 23 is a sum of both. Most of the datapoints with a dew point temperature above 270 K seems linearly correlated but with some over and underestimation. The lower has a bigger variance and is more randomly distributed. This is most likely because of the calculations that include both relative humidity and air temperature and that the air temperature and relative humidity have different effects for different values.

5.3.3 Correlations between surface conditions

The correlations between the sounding and reanalysis data regarding the heat fluxes in Figures 24a and 24b, indicate a lack of correlation between the real data from Søråsjordet field station and the reanalysis data. This could also be because of the big local differences in terrain and ground coverage within the grid used in the computations. Søråsjordet measures the Eddy fluxes in the middle of a field with grass. In comparison, the reanalysis data has to consider the paved areas, buildings, lakes, and the Oslofjord that might be within its grid.

The correlation in soil moisture between reanalysis and real data from Søråsjordet in Figure 25 shows a better correlation for values around30 %. The variance increases for values above and below 30 % soil moisture content. The highest values are also badly correlated where the reanalysis data does not have any points above 40 % soil moisture while the highest soil moisture content measured at Søråsjordet is closer to 65 %. This shows again that the local differences might be too big for the grid of the reanalysis data to capture. It is also to be noticed that an overestimation of the lower soil moisture values is also done by the reanalysis data.

The reanalysis data might not be a good fit for analysis of a local event like convection. The grid seems to be too big to capture local changes and details such as humidity and soil moisture well. The reanalysis data might be better fitted to analyses on a greater scale and observations of mesoscale events rather than local phenomena such as convection.

5.4 Limitations

5.4.1 Observed data

The limitations of this analysis are mostly related to data quality and missing data. The 53 radiosondes which captured only five days with significant precipitation did not make a good basis for an analysis of the interactions between the surface conditions. The lack of diversity in surface conditions with only four out of 53 days with a high soil moisture content did not strengthen the foundation of the analysis and a more balanced dataset would be beneficial for further analysis.

The two sensors used for measurements of soil moisture content, GroPoint and SoilVUE, had many missing values and missing time series. The GroPoint sensor had been found to be the most accurate and was also the sensor with the shortest continuous time series with measurements mostly from 2023. The SoilVUE sensor which had lower quality measurements had missing values in May and June 2021 but had a more continuous time series for the rest of the period used in the analysis. The use of the SoilVUE sensor instead of the GroPoint sensor in the fitting of the GLM and for other classifications of high and low soil moisture is also a limitation of the analysis.

The period of analysis itself of three months from 2021, the whole extended summer from 2022 and two and a half months from 2023 is not a long enough period to cover the actual variance of weather in the region. The months covered by soundings were dry and did not have much precipitation. More instances of wetter soils and precipitative events might be observed for longer continuous time series of soundings and soil moisture measurements.

Time was also a limitation as better variables and predictors could have been calculated or developed for classification of convective events if time would not have been limited.

5.4.2 Reanalysis data

The usage of reanalysis data with a grid size of 31 km for a local event such as convection, and surface heat fluxes is shown to have its weaknesses. The surface heat fluxes, soil moisture and relative humidities have a low correlation with the observed data and the grid size of the reanalysis data is too big for such local differences. Ås is close to the Oslofjord and the grid cell used to calculate the humidity and heat fluxes in Ås probably includes parts of the Oslofjord. There is also a lake in Ås that could have influenced the modelling of the data. The size of the grid is therefore a concern for the usage of such models when investigating local events and interactions.

The correlations of the moisture dependent variables, such as relative humidity, soil moisture, and heat fluxes, between the sounding and reanalysis data might reveal that the computation of moisture dependent variables in the modelled data is not correct. A better foundation of data from real observations is therefore preferred over the use of reanalysis data for this analysis.

6 Conclusions & further work

The HI_{low}parameter limits seem to fit with the occurrence of precipitation. In observed data, most of the morning and evening precipitation registered happened when HI_{low} predicted it. On the other hand, were the values from the CTP skewed against higher values in both observed and reanalysis data. This is mostly due to the temporal variation of CTP due to the difference in the initiation of the radiosondes between the original framework, reanalysis and the radiosondes initiated in Ås.

There was no splitting or pattern of higher or lower soil moisture in relation to convection in the observed data. This is mostly due to the low soil moisture levels these three years and very few instances of high soil moisture in the observed data. The reanalysis data for July and August 2023 displayed a splitting in soil moisture with high soil moisture for CTP values below 500 J/kg and low soil moisture for CTP values above 750 J/kg. This indicates that a splitting might exist in Norway but longer time series with higher soil moisture levels are needed to confirm.

A split of the points with high and low Bowen ratios was seen in the observed data. Most of the points with a Bowen ratio corresponding to high moisture availability had a HI_{low} below 5 K. This also strengthens the hypothesis of the presence of land-atmosphere coupling in Norway.

In total, the HI_{low} boundaries seem to fit for Ås and sub-polar regions. The CTP values found in this analysis do not fit with the boundaries, but an analysis of soundings initiated at sunrise would be better in finding more correct boundaries for Ås and sub-polar regions.

6.1 Further work

The data needed to better observe land-atmosphere interactions are longer and more continuous time series of the surface conditions and the vertical profiles. Longer continuous time series are needed from the more accurate GroPoint soil moisture sensor, and longer time series of the sensible and latent heat fluxes are needed to get a better data foundation and cover as many events as possible.

The measurements of the vertical profile have to be taken with shorter intervals and several vertical profiles a day would be beneficial for better observation of convection and the development of potential energy in the boundary layer. Hourly vertical profiles of the boundary layer would be beneficial but coarser time series would also be of great help. Findell et al. (2024) suggests vertical profiles every three hours,

which seems reasonable. Since only data from the boundary layer, which is almost always below 3 km above ground level, is needed radiosondes might not be necessary and other ways of retrieving the vertical profiles might be used such as drones or other methods.

Since 2021, 2022, and 2023 were years with mostly dry summers, the length of the time series also has to be longer and include more than almost three summers. A basis of ten years or more would be a good foundation as ten years would most probably include a greater set of combinations of the different parameters and not be as influenced by the extremes.

To have a better foundation to make decisions about limits of CTP and HI_{low} daily soundings initiated at sunrise is needed to get a better representation of the summers and how convection evolves through the day in Ås. In addition to daily soundings, some days that are of more interest could get more soundings. The most interesting days start with a blue sky and no cloud coverage, as well as a lack of mesoscale events. On these days could a combination of cloud coverage, CAPE values, and vertical wind velocity measurements throughout the day be helpful to see whether convection was initiated or not.

Better ways to classify convective events and how to classify convection that does not necessarily lead to precipitation are also needed. One way to classify convection that could not be used in this case is the vertical wind speed measurements. This could have been the next step to investigate if time would not have been a limit.

References

Barry, R. G., & Blanken, P. D. (2016). *Microclimate and local climate*. Cambridge University Press.

Bruksanvisning nedbørmåler t-200b. (1996). Geonor. Østerås, Norway.

Chernokulsky, A. V., Eliseev, A. V., Kozlov, F. A., Korshunova, N. N., Kurgansky, M. V., Mokhov, I. I., Semenov, V. A., Shvets', N. V., Shikhov, A. N., & Yarinich, Y. I. (2022). Atmospheric severe convective events in russia: Changes observed from different data. *Russian Meteorology and Hydrology*, *47*(5), 343–354. https://doi.org/10.3103/S106837392205003X

Cioni, G., & Hohenegger, C. (2017). Effect of Soil Moisture on Diurnal Convection and Precipitation in Large-Eddy Simulations [Publisher: American Meteorological Society, Section: Journal of Hydrometeorology]. *Journal of Hydrometeorology*, *18*(7), 1885–1903. https://doi.org/10.1175/JHM-D-16-0241.1

Crawley, M. J. (2013). *The r book* (Second edition). Wiley.

Eddypro software instruction manual. (2021). LI-COR. Lincoln, Nebraska, USA.

Elicit. (2023, January 24). *Elicit: The ai research assistant*. https://elicit.com

Findell, K. L., & Eltahir, E. A. B. (2003a). Atmospheric controls on soil moisture–boundary layer interactions. part i: Framework development. *Journal of Hydrometeorology*, *4*(3), 552–569. https://doi.org/10.1175/1525-7541(2003)004<0552:ACOSML>2.0.CO;2

Findell, K. L., & Eltahir, E. A. B. (2003b). Atmospheric controls on soil moisture–boundary layer interactions. part II: Feedbacks within the continental united states. *Journal of Hydrometeorology*, *4*(3), 570–583. https://doi.org/10.1175/1525-7541(2003)004<0570:ACOSML>2.0.CO;2

Findell, K. L., Yin, Z., Seo, E., Dirmeyer, P. A., Arnold, N. P., Chaney, N., Fowler, M. D., Huang, M., Lawrence, D. M., Ma, P.-L., & Santanello Jr., J. A. (2024). Accurate assessment of land–atmosphere coupling in climate models requires high-frequency data output [Publisher: Copernicus GmbH]. *Geoscientific Model Development*, *17*(4), 1869–1883. https://doi.org/10.5194/gmd-17-1869-2024

Fischer, E. M., & Knutti, R. (2016). Observed heavy precipitation increase confirms theory and early models [Publisher: Nature Publishing Group]. *Nature Climate Change*, *6*(11), 986–991. https://doi.org/10.1038/nclimate3110

Gleick, P. H. (1993). *Water in crisis: A guide to the world's fresh water resources* (Pacific Institute for Studies in Development, Environment, and Security & Stockholm Environment Institute, Eds.). Oxford University Press.

Grammarly. (2023). *Grammarly*. https://www.grammarly.com/

Groenemeijer, P. H., & van Delden, A. (2007). Sounding-derived parameters associated with large hail and tornadoes in the netherlands. *Atmospheric Research*, *83*(2), 473–487. https://doi.org/10.1016/j.atmosres.2005.08.006

Gropointtm profile multi-segment soil moisture & temperature profiling probe. (2020). GroPoint. North Saanich, BC, Canada.

Hennessy, K. J., Gregory, J. M., & Mitchell, J. F. B. (1997). Changes in daily precipitation under enhanced greenhouse conditions. *Climate Dynamics*, *13*(9), 667–680. https://doi.org/10.1007/s003820050189

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., & Thépaut, J.-N. (2023a). *Era5 hourly data on pressure levels from 1940 to present* [Accessed on 15-03-2024]. https://doi.org/10.24381/cds.bd0915c6

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., & Thépaut, J.-N. (2023b). *Era5 hourly data on single levels from 1940 to present.* [Accessed on 15-03-2024]. https://doi.org/10.24381/cds.adbb2d47

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., ... Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Instruction manual - cmp/cma series. (2013). Kipp & Zonen. Delft, The Netherlands.

Jach, L., Schwitalla, T., Branch, O., Warrach-Sagi, K., & Wulfmeyer, V. (2022). Sensitivity of land–atmosphere coupling strength to changing atmospheric temperature and moisture over Europe [Publisher: Copernicus GmbH]. *Earth System Dynamics*, *13*(1), 109–132. https://doi.org/10.5194/esd-13-109-2022

Jach, L., Warrach-Sagi, K., Ingwersen, J., Kaas, E., & Wulfmeyer, V. (2020). Land Cover Impacts on Land-Atmosphere Coupling Strength in Climate Simulations With WRF Over Europe. *Journal of Geophysical Research: Atmospheres*, *125*(18), e2019JD031989. https://doi.org/10.1029/2019JD031989

Ketzler, G., Römer, W., & Beylich, A. A. (2021). The climate of norway. In A. A. Beylich (Ed.), *Landscapes and landforms of norway* (pp. 7–29). Springer International Publishing. https://doi.org/10.1007/978-3-030-52563-7_2

Lawrence, M. G. (2005). The relationship between relative humidity and the dew-point temperature in moist air: A simple conversion and applications. *Bulletin of the American Meteorological Society*, *86*(2), 225–234. https://doi.org/10.1175/BAMS-86-2-225

Masseroni, D. (2014). Limitations and improvements of the energy balance closure with reference to experimental data measured over a maize field. *Atmósfera*, *27*(4), 335–352. https://doi.org/10.1016/S0187-6236(14)70033-5

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J., Maycock, T., Waterfield, T., Yelekçi, O., Yu, R., & (eds.), B. Z. (2021). Ipcc, summary for policymakers. in: Climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change, 3–32. https://doi.org/10.1017/9781009157896.001.

Naalsund, I. L. (2022). *Investigation of spatiotemporal variations in soil moisture in søråsfeltet, ås, applying different measurement techniques* [Master thesis]. Retrieved January 31, 2024, from https://nmbu.brage.unit.no/nmbu-xmlui/handle/11250/3011249

National Oceanic and Atmospheric Administration. (n.d.-a). *Radiosonde observation*. Retrieved April 25, 2024, from https://www.weather.gov/upperair/factsheet

National Oceanic and Atmospheric Administration. (n.d.-b). *Radiosondes*. Retrieved February 24, 2024, from https://www.noaa.gov/jetstream/upperair/radiosondes

Norewgian meteorological institute. (n.d.). *Weather forecasts on yr - how are they made?* Retrieved May 12, 2024, from https://hjelp.yr.no/hc/en-us/articles/360004008874-Weather-forecasts-on-Yr-how-are-they-made

Norges miljø- og biovitenskapelige universitet. (n.d.-a). *Kunstig intelligens ved REALTEK*. Retrieved May 14, 2024, from https://www.nmbu.no/fakulteter/fakultet-realfag-og-teknologi/kunstig-intelligens-ved-realtek

Norges miljø- og biovitenskapelige universitet. (n.d.-b). *Meteorologiske data | NMBU*. Retrieved March 18, 2024, from https://www.nmbu.no/forskning/grupper/meteorologiske-data

Norges miljø- og biovitenskapelige universitet. (n.d.-c). *Værstatistikk - kommentarer for 2021*. Retrieved May 10, 2024, from https://www.nmbu.no/fakulteter/fakultet-realfag-og-teknologi/vaerstatistikk-kommentarer-2021

Pidwirny, M. (2021). *Physical geography lab manual: The atmosphere and biosphere*. University of British Columbia. https://pressbooks.bccampus.ca/physgeoglabmanual1/back-matter/appendix-2-koppen-climate-classification-system/

Product manual: Soilvue 10, complete soil profiler. (n.d.). Campbell scientific. Logan, UT USA.

Raschka, S., & Mirjalili, V. (2019). *Python machine learning* (Third edition). Packt Publishing.

Šaur, D. (2015). Evaluation of the accuracy of numerical weather prediction models. In R. Silhavy, R. Senkerik, Z. K. Oplatkova, Z. Prokopova, & P. Silhavy (Eds.), *Artificial intelligence perspectives and applications* (pp. 181–191). Springer International Publishing. https://doi.org/10.1007/978-3-319-18476-0_19

Stull, R. (2017). *Practical meteorology: An algebra-based survey of atmospheric science.* University of British Columbia. Retrieved January 25, 2024, from https://www.eoas.ubc.ca/books/Practical_Meteorology/prmet102/Practical_Meteorology-v1.02b-WholeBookColor.pdf

Using the li-7500ds and the smartflux 3 system. (2021). LI-COR. Lincoln, Nebraska, USA.

Wallace, J. M., & Hobbs, P. V. (2006). *Atmospheric science: An introductory survey* (2nd ed). Elsevier Academic Press.

Westra, S., Alexander, L. V., & Zwiers, F. W. (2013). Global increasing trends in annual maximum daily precipitation [Publisher: American Meteorological Society, Section: Journal of Climate]. *Journal of Climate*, *26*(11), 3904–3918. https://doi.org/10.1175/JCLI-D-12-00502.1

Windmasterpro 3-axis ultrasonic anemometer. (2019). Gill Instruments Limited. Hampshire, United Kingdom.

Wolff, M. A. (2023, July 13). *Været 2022 på ås*. Retrieved May 10, 2024, from https://www.nmbu.no/fakulteter/fakultet-realfag-og-teknologi/vaeraret-2022-pa

Wolff, M. A. (2024, January 19). *Været i ås 2023*. Retrieved May 10, 2024, from https://www.nmbu.no/fakulteter/fakultet-realfag-og-teknologi/vaeret-i-2023

Yano, J.-I., Ziemiański, M. Z., Cullen, M., Termonia, P., Onvlee, J., Bengtsson, L., Carrassi, A., Davy, R., Deluca, A., Gray, S. L., Homar, V., Köhler, M., Krichak, S., Michaelides, S., Phillips, V. T. J., Soares, P. M. M., & Wyszogrodzki, A. A. (2018). Scientific challenges of convective-scale numerical weather prediction [Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society]. *Bulletin of the American Me-*

teorological Society, *99*(4), 699–710. https://doi.org/10.1175/BAMS-D-17-0125.1

Ye, B., Genio, A. D. D., & Lo, K. K.-W. (1998). CAPE variations in the current climate and in a climate change. *Journal of Climate*, *11*(8), 1997–2015. https://doi.org/10.1175/1520-0442(1998)011<1997:CVITHCC>2.0.CO;2

Zhang, J., Chen, H., Zhu, Y., Shi, H., Zheng, Y., Xia, X., Teng, Y., Wang, F., Han, X., Li, J., & Xuan, Y. (2019). A novel method for estimating the vertical velocity of air with a descending radiosonde system. *Remote Sensing*, *11*(13), 1538. https://doi.org/10.3390/rs11131538

A Appendix

A.1 Appendix 1 - Distribution plots

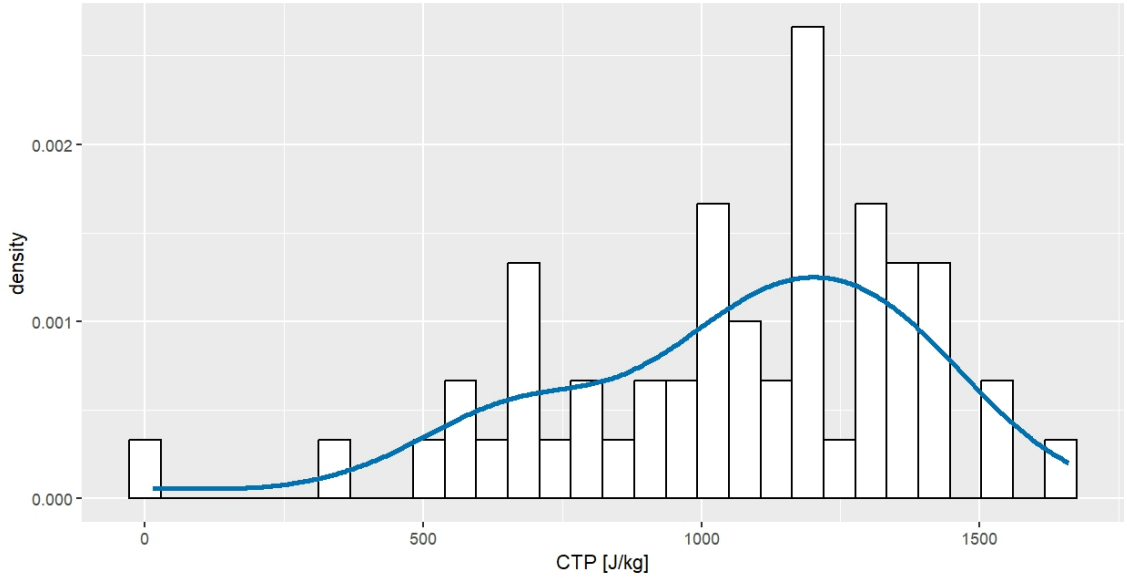


Figure 26: Distribution plot of the CTP values in the radiosonde data. CTP is close to normal distributed but has a tail to the lower values and deviates therefore some from a perfect normal distribution.

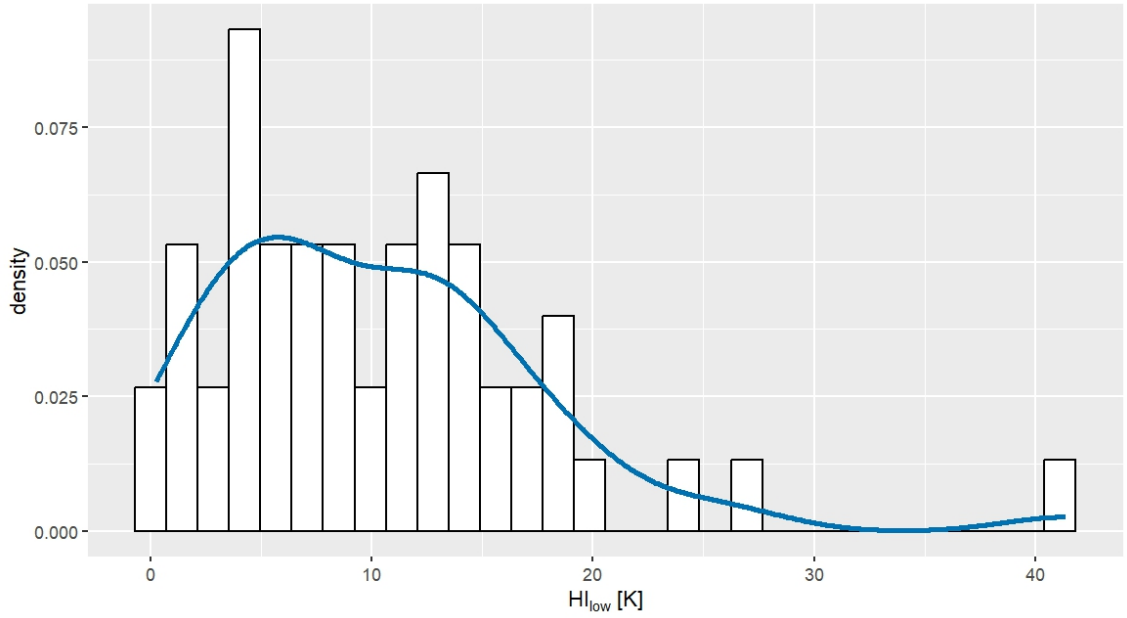


Figure 27: Distribution plot of the HI_{low} in the radiosonde data. The distribution is close to a gamma distribution with the high occurrence of low values.

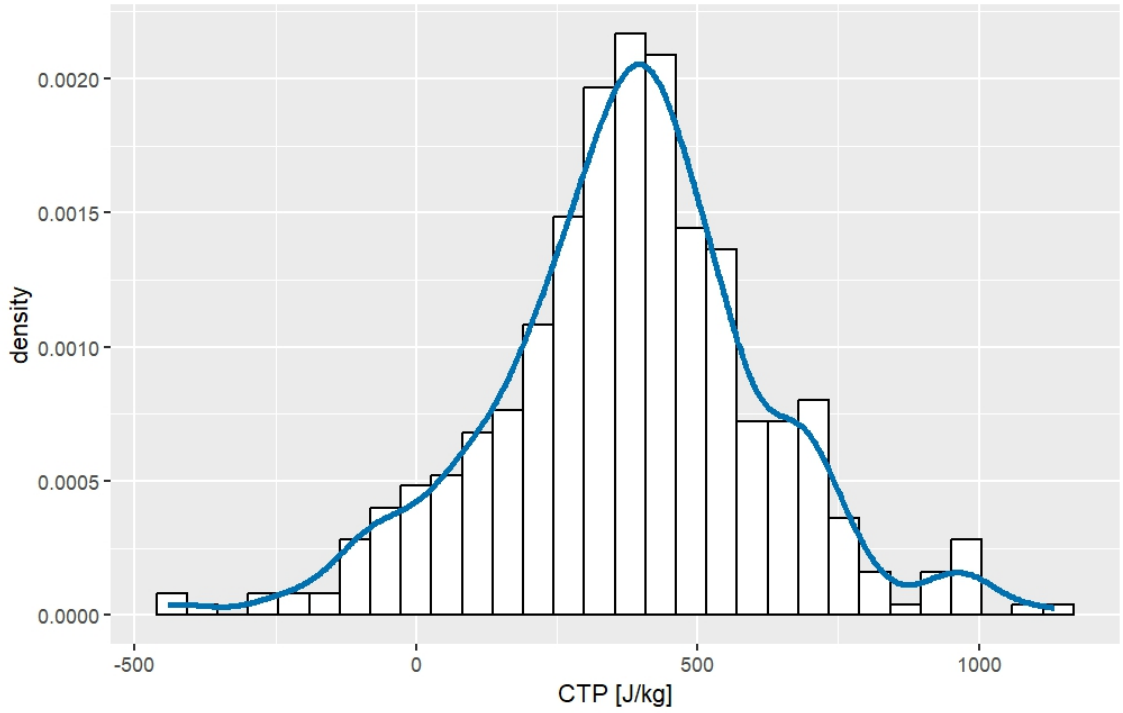


Figure 28: Distribution plot of CTP from reanalysis data. the distribution of values is close to normal distributed.

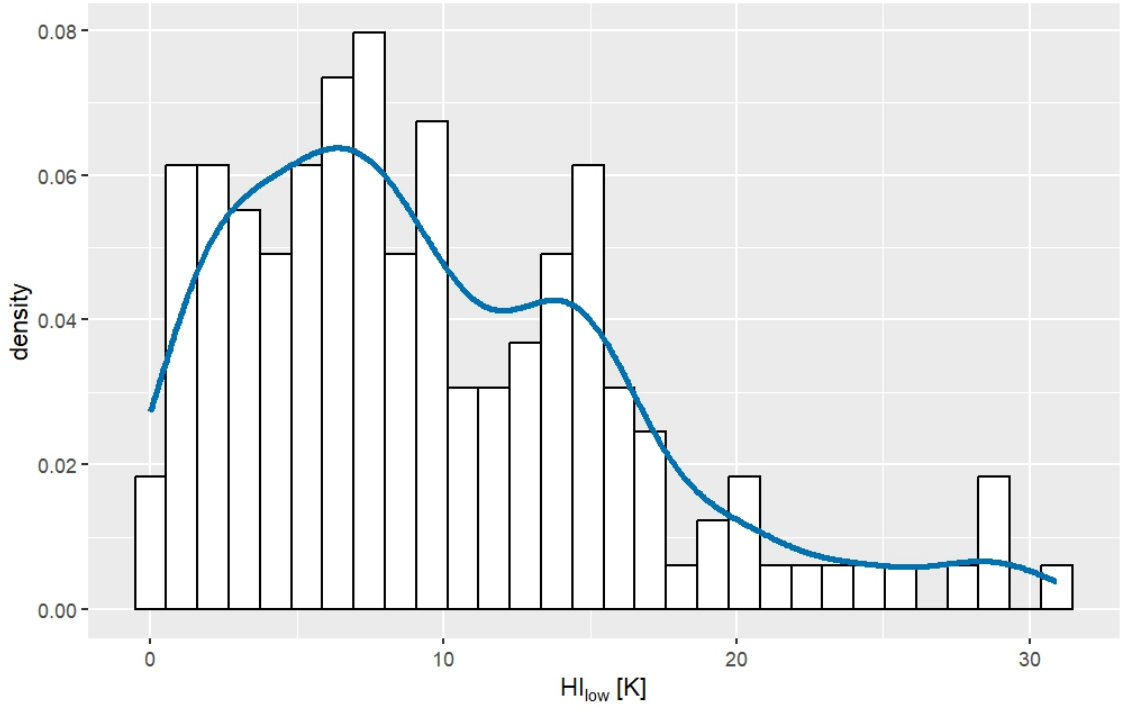


Figure 29: Distribution plot of HI_{ow}. The distribution is close to a version of the gamma distribution with a higher occurrence of the lower values.

A.2 Appendix 2 - Soil moisture

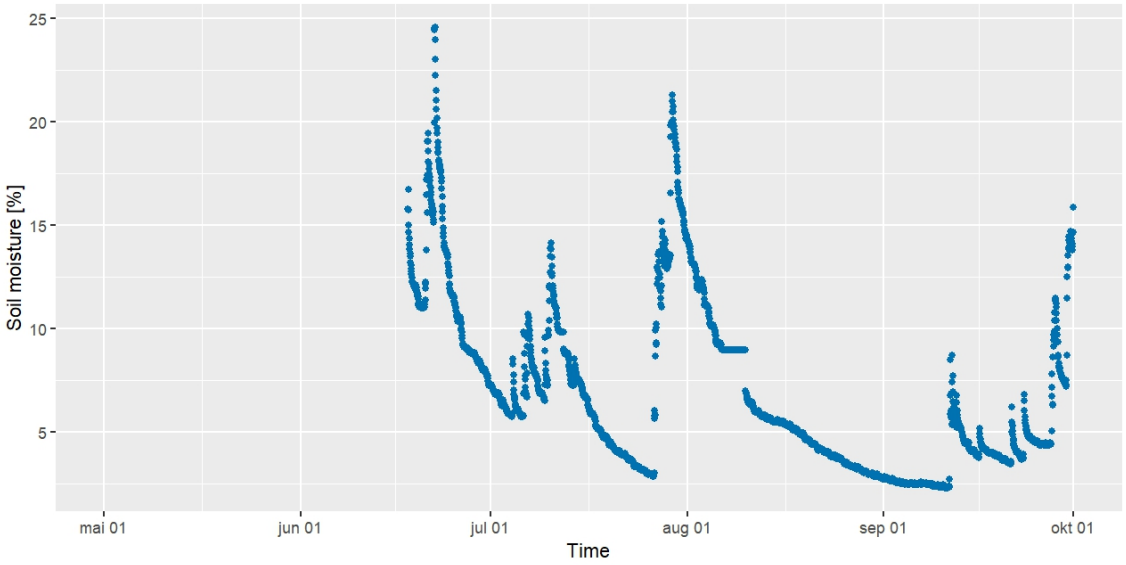


Figure 30: Time series of soil moisture in the summer of 2021 from the SoilVUE sensor.

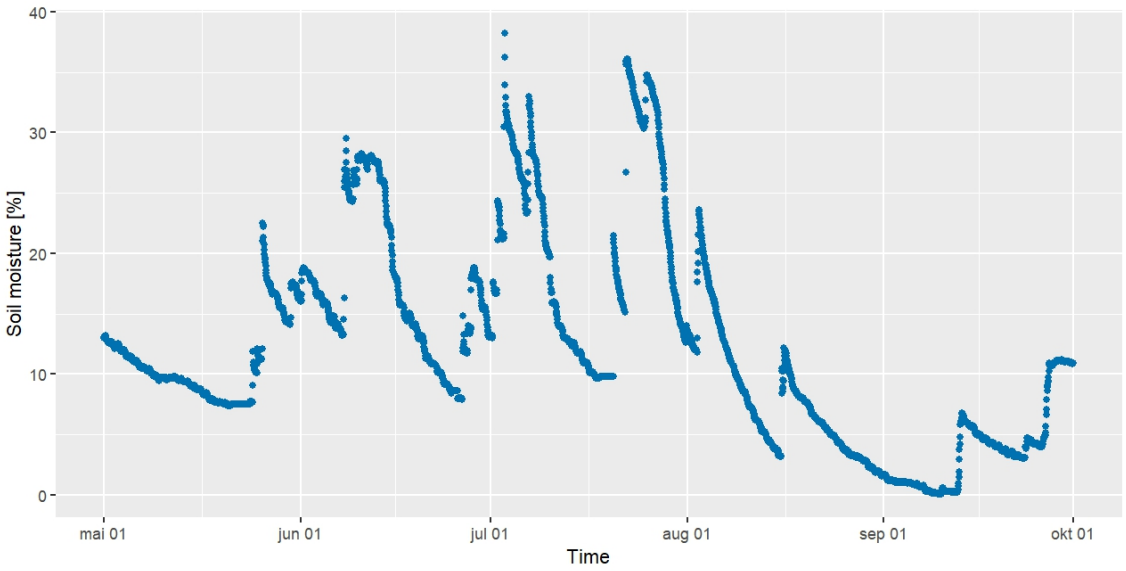


Figure 31: Time series of soil moisture in the summer of 2022 from the SoilVUE sensor.

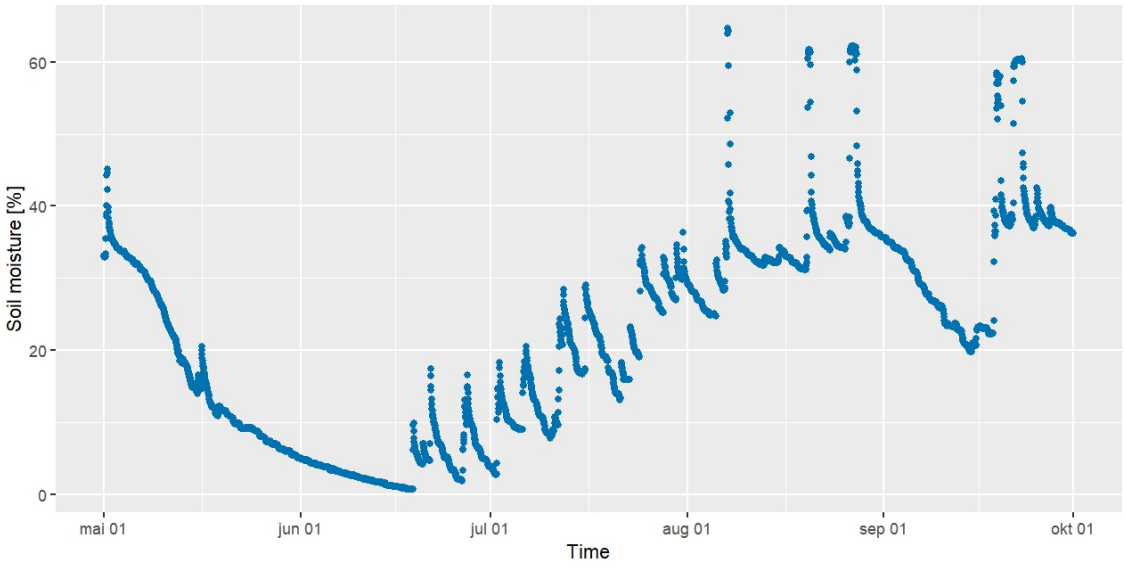
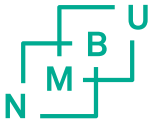


Figure 32: Time series of the soil moisture in the summer of 2023 from the SoilVUE sensor.



Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway