Applying artificial intelligence to perform climate predictions

Data driven approach for parametrizing cloud cover

Hanna Svennevik

Thesis submitted for the degree of

Master of science in Computational Science

60 Credits

Department of Geoscience

Faculty of Mathematics and Natural Sciences

UNIVERSITY OF OSLO

January 14, 2020

© 2020 Hanna Svennevik

Applying artificial intelligence to perform climate predictions

Data driven approach for parametrizing cloud cover

This work is published digitally through DUO – Digitale Utgivelser ved UiO

http://www.duo.uio.no/Printed: Reprosentralen, University of Oslo

All rights reserved. No part of this publication may be reproduced or transmitted, in any form or by any means, without permission.

Abstract

Ikke skriv masteroppgaven som om LSTM kommer til å funke. Skriv generelt at i denne oppgaven skal vi test to metoder. Hugo. Skriv også om alle valg som er tatt i hensyn til når du lager datasettet. Dette er viktigere enn sykt mye teoridel.

Introduce topic and why its important

Over the last decades researchers have been working on determining the climate sensitivity. The fifth assessment report result in a range of 1.5K to 4.7K. Here 4.7K allows for a doubling of the CO_2 emissions compared to 1.5K. Most of this uncertainty is attributed to how clouds are resolved in climate models. (Her har jeg vell sagt motsatt ting.) We known clouds are important in today's climate. Since then reflect 50% of the solar radiation back into space.

Introduce a challenge or unresolved issue that you will try and solve.

Large uncertainties associated with the effect of clouds in future climates. Its unclear to which level of sophistication these sub-grid scale parametrizations need to be in order model their effect on climate.

What have you done to try and solving this.

In this thesis we have tried a new approx using supervised learning to regress clouds amount to cloud macroscopic properties in time and space.

Improve parametrization of clouds using machine learning. Uncertain to what level of sophistication is necessary to model cloud amount.

Main result - Include the numerical result of your best model.

The implications in the context of 1+2.

Acknowledgement

This thesis is a joint work between UiO and SimulaMET. First and foremost I would like to thank my supervisors Trude, Hugo and Micheal for giving so much freedom to make my own choices regarding the data. **Setning om Trude**. **Setning om Michael** og **Setning om Hugo** .

I would like to thank my fellow master students Ingvild, Marit and Johanne for adventures, the support and encouragement during this thesis. You have been a X.

I would also like to thank the **navn** at the Helpdesk in EUMETSAT. **Navn** at Copernicus Data Senter. Their expertise have been invaluable when generating the data set. The IT-support on the GEO-department your technical support regarding storage space and other stuff related to this.

I would also like to thank Raymond and Ina for input and proof reading this thesis. Your contribution have been valuable for my work.

A big thanks goes to my sister Sara for always being there, rooting for me. You are good to have and your company on long days at Blindern have been much appreciated.

Contents

List of Figures

List of Tables

Introduction

Clouds play a important role in the climate system. Both affecting the radiative budget and the hydrological cycle. *Consist/Composed of liquid droplets, ice crystal or both.* Understanding how clouds form in the complex system of the atmosphere involves both knowledge about the large scale influence by the circulation and the small scale influenced by aerosols. To this day the micro-physics of all phases are not fully understood. Here mixed phase clouds, clouds consisting of both liquid and ice, *proofs* to be the most difficult.

Climate models are the most useful tool for studying past, present and future climate change. Clouds and aerosols are acknowledged as the factor contributing with the largest uncertainty to the equilibrium climate sensitivity, ECS . Also know as global mean temperature increase due to a doubling of the pre-industrial levels of CO_2 (280 ppm). *It remains unclear to which level of sophistication is adaquate to model their effect om climate.* **Siter ch7 AR5**.

Make sure you include everything that's related to parametrized processes. It is understood that cloud formation requires suitable aerosol and sufficient supersaturation. Aerosols include both gases and solid particles suspended in air. They interact with the clouds by serving as particles which vapour and ice can condensate or deposit upon. The different phases require different properties and the nuclei's are called CCN for liquid droplets and INP for ice crystals. Saturation is usually archived by a temperature decrease in rising air masses. Thus the stability of the atmosphere affect plays a key role for convective motions. **Legg inn bilde a skyer en i is fase og en i liquids. Skriv noe som "the sharp outoline suggest that the cloud is consisting of liquid droplets, even at temperagtures below 0."**

Growth processes are phase dependant. Liquid droplet grow by diffusion and later by collision and coalescence. At temperatures -38/-40 degrees **kilde** they will spontaneously freeze and could play the role as INP. When both phases are present in a cloud, the saturation vapour pressure over ice is higher than over liquid. This may cause the droplets to evaporated and depositing on to the ice crystals. This is called the Wegeron-Bergeron-Findeisen process. Clouds consisting sole of ice crystals first grow by deposition of vapour then by aggregation.

Due to the complex nature of clouds. Lots of different processes occurring simultaneously on different scales. Incorporating all these interactions into a model framework has proven to be difficult. **Read ch. 2 at Statkraft i helga.** Explain convection and fronts. Mention cumulus, status and cirrus clouds? Include observational changes in the hydrological balance.

Clouds in the current climate

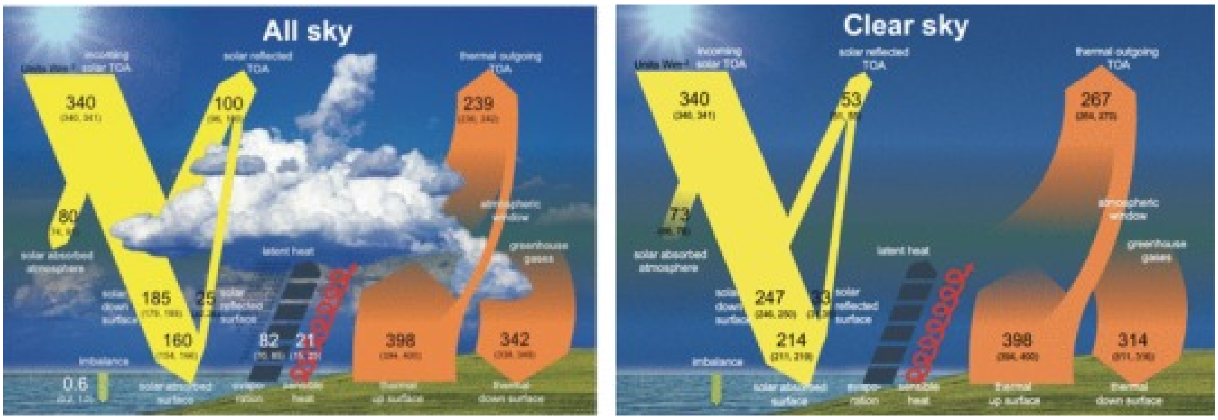

Figure 1.1: *The all-sky and the clear sky. Figure 14 Wild 2019. Ikke i bruk vennligst kommenter om denne er bedre enn den andre som viser differences mellom disse subplottene.*

Based on satellite and ground based measurements Wild et. al. 2019 **siter** have quantified the contribution of elements in the radiative budget. Subtracting the clear-sky from the all-sky climatology to compute the cloud radiative effect, CRE . This is shown in equations (1.1) and (1.2). Wild et. al. 2019 **siter** concludes with a reduction in shortwave radiation of $-47Wm^{-2}$ by clouds. In other words clouds reflect approximately 50% of the incoming solar radiation. Longwave component is $28Wm^{-2}$. This give a net CRE of $-19Wm^{-2}$. Proving that the net effects of clouds on the radiative budget is negative.The altitude along with the composition determines the radiative properties of the cloud. **Relate this to the black body properties of clouds..? LES Artikkel fra Jonah**

$$CRE_{sw} = SW\uparrow_{clear-sky} - SW\uparrow_{all-sky} \tag{1.1}$$

$$CRE_{lw} = LW\uparrow_{clear-sky} - LW\uparrow_{all-sky} \tag{1.2}$$

The physical properties causing the interaction with radiation is described below. Dense low level clouds reflect solar radiation. This is called the albedo effect. Albedo being the ra-

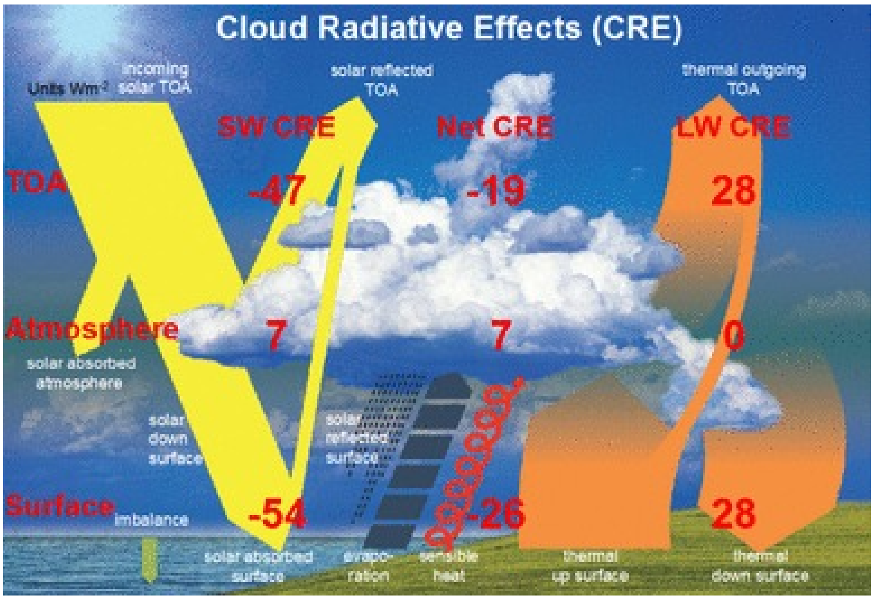

Figure 1.2: *Cloud radiative effect, CRE is the differece between the radiative components of the Clear sky radiative and the all sky. Cite this figure as fig 15 in Wild 2019*

tio between reflected to incoming radiation. The higher number concentrations of droplets in a cloud the higher the total surface area of droplets. The more radiation gets reflected back into space. Clouds absorb longwave radiation and re-emits it. The absorbed radiation originates from the surface and is given by Stefan-Boltzmann forth-power law, see equation (1.3). The emmisivity, ε depends on the (composition, compactness and surface roughness) of the medium. Water, snow and ice have different spectral emmisivity. Huang et. al., 2018. Different parts of the globe are covered by different surfaces and Huang et al 2016 proved that assuming a constant suface emmisivity effects effects the TOA polar enegy buget. High clouds have low temperatures and since the re-emitted flux is a function of the cloud temperature. The greenhouse effect increases with the cloud altitude. **Fokuser på at uavhengig av emmisiviteten til et medium er deet en funksjon av the fourth power of** T_{skin}

$$F = \sigma \varepsilon T^4 \tag{1.3}$$

Clouds in future climates

Wild et. al. 2019 **siter** finds a imbalance of $0.6 W m^{-2}$. This heat gets trapped in the earth system, forcing the surface temperature to increase in order to close the radiative budget. The imbalance in the radiative budget at TOA is the radiative forcing. Climate drivers include

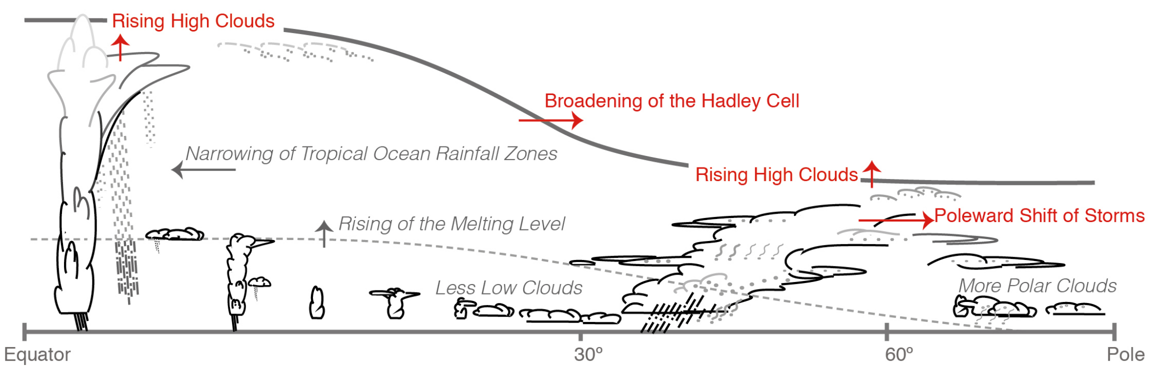

Figure 1.3: *Cloud climatology in future climate. Developed based feedbacks in climate models, the different adjustments have different sikkerhet. Cite the fifth assessment report IPCC report.*

both natural and antropogentic forcings. This can be everything from natural variability in the solar energy output, vulcanic eruptions and green house gas emisions. The overall goal is to compute the climate sensitivity/equilibrium temperature as a function of forcing. For different representative concentration pathways, RCP scenarios one get a different temperature increase.

The global temperature will keep rising until we have reached the equilibrium climate temperature. This temperature increase induces climate changes. The IPCC suggest the following shift in cloud schemes (see figure 1.3). Figure 1.3 shows a summary of the most likely cloud feedbacks. First, a broadening of the Hadley cell causes a poleward shift of storms. This dries up the subtropics and moistens the higher latitudes. The clouds move further into the polar night, decreasing the albedo effect. The greenhouse effect of clouds still persist without sunlight leading to a net heating in the Arctic. Second, rising higher clouds causing a stronger greenhouse effect. Third, less low level clouds, this is assumed to be partly offset by a increase in the melting layer, leading to more opaque clouds. Rising of the meltlayer cause ice crystals to melt resulting in more opaque clouds. These have a higher albedo and reflect more sunlight.

Aerosols can alter the cloud micro-physics and in terms alter the radiative properties of the cloud. A polluted cloud gets extra CCN, this results in more smaller droplets, as they share the available liquid. This increases the total surface area of the droplets. Which again reflect more radiation and could led to a enhanced lifetime, since it takes longer for the droplets to reach precipitation size. When clouds persipitate they clean the air by removing particles.

Cloud micro-physical processes are not yet fully understood. Along with the fact that clouds are formed a smaller scale then can be resolved in your average climate models. Parametrizations are used to include the contribution from the subgridscale processes to the mesoscale proceses (weather phenomenons) in climate models and in weather predictions in general. Over the last years this has gained more attention since its acknowleged as the largest contributor to the uncertainty in climate models. I'll get back to this in chapter 2.

Deep Learning

Må begynne noen mer gripende eksempler på hva machine learning kan gjøre. Artificial intelligence dates back to the fifties when pioneers start talking about *automating task normally performed by humans*. Machine learning is a means to achieve artificial intelligence. The progress in the field follows a sigmoid curve. Slow increase at first, then really steep, before it slows down again (see figure **ref sigmoid activation func**). Over the years there have been several discoveries kick-starting the development in machine learning. The very first algorithm's include probabilistic modelling. Using the principles of statistics to analyse data. This includes Naive Bayes classifiers and logistic regression. Two algorithm's which predates computers and are still useful today. Three main bottlenecks of advances in AI is hardware, data and algorithm's. Internet continuous to provide large amounts of data from Wikipedia, Flicker (tagged images) and YouTube. Advances in computational powers, such as graphical processing units, GPU's *provide a environment/platform to learn in/on*. These where originally develop for the gaming industry, but in 2007 they realised a interface called CUDA (2007) which allows for computing **find a up to date cost and flops (floating point operations per second)**. **siter Chollet bok**

For clarity, the deep in deep learning refer to the number of layers. Moving from shallow networks to deeper ones (more than 10) algorithmic advances in gradient propagation was needed. This includes activation functions, weight initialisation and optimisation schemes. I'll get back to that in section 2.3. Other advances like batch normalisation, depth wise separable convolutions attributed to the revolution of AI. Earth system monitoring provides a global view of variables across meteorological systems. **Some thing about satellite era**.

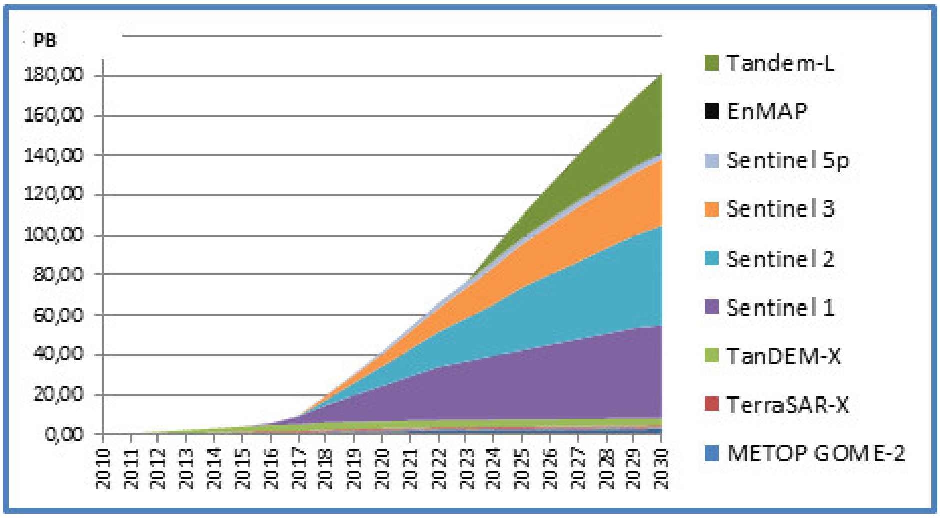

Figure 1.4: *Data volume. By the continuous earth system monitoring, meteorology/ climate science have progressed toward becoming a big data science. Observations are most used for verifying climate models and quantifying the current state of climate. https://www.dlr.de/eoc/en/desktopdefault.aspx/tabid-12632/22039_read-51751.*

These large amounts of data and the flexible nature of the neural network makes is a suitable method also in geosciences. **With enough data neural networks can serve as a universal function approximate given a suitable hyper parameter tuning and input data.** The last couple of years reaserchers have been attempting to use this for wide range of problems like rainfall runoff modeling (krazerts), high-resolution weather forecasting (Rodrigues), Air quality forecasting (sun and liu), precipitaiton nowcastng (Shi et al) and **kanskje: LES** *deep neural network based feature representation for weather data.* **lui et al** . **noe med forskjellig hell.** Another more comprehensive machine learning project is lead by Tapio Schneider at Caltech. Along with his team og technologist they have ambitions to create a earth system model using machine learning. With his team of from MIT and former employees of Microsoft and Google they hope to create a platform which can resolve clouds and hopefully reduce the spread in climate sensitivity. **cite Science**

Theoretical Background

In this chapter I will present motivation, necessary theoretical background and practical implications to understand thee need for new predictions of cloud cover/cloud amount. Investigating a data driven approach for parameterization of clouds. Developing new tools for approximating the complex process of cloud formation. Hopefully this will contribute to the understanding of the physical process which is cloud formation. The data used in this project is a mixture of satellite retrievals and reanalysis data. No ancillary information is provided. See sections 2.2.1 to 2.2.2 for more detailed descriptions.

The numerical methods used in this thesis are described in section **??**. All the code is available on GitHub in the project repository named MS on https://github.com/hannasv/MS. At this repository you will find everything you need to perform this experiment yourself. Descriptions for downloading data and retrieving the correct licences.

There is project environment ready for installation, called *sciclouds*. This is a conda envionment, the yaml-file lists the python packages and their versions used for running this code. Notebooks for conducting the experiments. Supplementary material for remapping satellite data and land-sea masks are available in a supplementary repository called *MS-suppl* https://github.com/hannasv/MS-suppl.

Parameterizations of clouds

Parametrizations are a tool used in climate models to include the effect of subgridscale prosesses. This is done for several processes **give examples**. This thesis is only conserned with parametrizaions of cloud cover. The simplest form of cloud scheme is binary. Either the eintire pixel is covered by clouds or there is no clouds present. This is implemented as follows, if $RH > 100 \rightarrow CLA = 1$ else $CLA = 0$. Sub gridscale variability in humidity is necesary to achive fractional cloud cover. This can be combined with a sub gridscale of temperature. Over the years reaserchers have tried to draw the distrubutions of these variables from observations and implement them into models. Vitualy all probability density functions, PDF's have been used to model either cloud cover or its dependant variables humidity, temperature and so on. They have not been sucesfull in finding a adaquate repreentation of cloud cover using this approach. **(siter Tomkins summary)**

Cloud cover is usually a combination of several parametrizations. Its common to have seperate schemes for ice-, liquid clouds and convections. **Read more Tomkins**

Parametrizations of clouds - Climate models

Climate models are an important tool for studying the effects of emissions/forcing on future climates. The intergovernmental panel of climate change, IPCC provide assessments report every 10th year or so, providing a state of the art status update on the current knowledge of climate change. Since the previos assesment report there has been three spesial report A, B and C. **Les special report.** The previous report published was Assesment report 5, AR5 in 2013 and the next report is scheduled to be published in 2021. The ensamble of climate models included in AR5 is the coupled model intercomparison project phase 5, CMIP5 . CMIP6 are now being evaluated, thus there is less published literature. Even though its a bit old, we will mostly focus on the results in AR5. **Plus the findings in the special reports.**

In resent years a lot of effort have been invested in improving the parametrizations of subgrid-scale processes. Among these clouds contribute with the largest uncertainty, approximately three times as large as other process i.e. relative humidity-lapse rate feedback (these processes shoukd not be viewed in isolation). The contributions of the clouds to the short wave component in the radiative budget is the main contributor to the uncertainty. Short wave cloud feedback. *To this day neither observastions of global climate models, GCM's provide clear eveidence or contradict the low level clouds feedback.* There is no accepted basis to refuse a GSM *a priori* **this increases the muli-model mean spead in climate sensitivity**. Missing representations of clouds microphysical processes related to opacity or cirrus (high altitude, composed of ice) clouds.

The computational cost of generating these large ensambles are limiting factor. Simplyfied models in terms of resolution and/ or complexity is common/often necesarry.

Using idealised experiments they give model spread in equilibrium climate sensitivity, ECS . This describes the *equilibrium change in global and annual mean surface temperature after doubling the CO_2 conditions from preindustrial times.* For CMIP5 the ECS is 2.1^oC to 4.7^oC.The is *very high confidence* that clouds are the primary factor attributing to the wide range. This is not a very big improvement from Hansen et. al. 1984 first estimate of climate sensitivity which was the range 2.0^oC to 5.0^oC. **explain idealised experiments.** Hansen et. al. ran their experiments using a coarse resolution of $8^o \times 10^o$ grid box (lat $\times$ lon) and a doubling of the CO_2 concentrations from 315ppm to 630ppm. The CMIP5 **What more where you thinking here..?**

ERA5

ERA5 is produced used IFS cycle 4lr2. This has a new cloud scheme or hydrological cycle. **Artikkelen forteller om oppdateringer fra sist era-interim produksjon ikke alt som finnes. Bedre å skrive denne etter du har skrevet i datasettet hvor** . Husk Tomkins jobber for ECMWF.

Practical implications

When conducting large machine learning project, such as this thesis its good to have a understanding of needs of the end product. What should it be used for? And how does it need to be implemented to be usefull in such a way. **Explain the strength and weaknesses of this approach for future implementations. Hva er tatt høyde for ved valg av data og lignende. Kan være nyttig for om folk reproduserer det og skal jobbe videre...** One of the obvious downsides of using this data driven approach is the rigid resolution. It needs to be retrained in another resolution for it to be usefull in climate research. Climate models provide data in a wide range of different of spatiotemporal resolutions. Before implementing this model it would need to be retrained on the resolution of the climate model under development. Which includes both remapping the data set and retraining the model. This is a time consuming process involving finding a new set of hyperparameters suitable for the new resolution. Once trained machine learning models provide fast results even for complex parametrizations which is what makes them suitable. For global climate models you need to have access to training data all over the globe otherwise there is no point. This thesis is concerend with a region, in order to compute a proof of consept. Antoher issue is that most machine learning packages are in python programming languange while climate models are in **fortran 90/95 - dobbelsjekk**. **how to solve this**

Data set and Methods

ERA5

ERA5 is the latest in the series of reanalysis produced by European Centre for Medium-Range Weather Forecasts, ECMWF . Re-analysis is as close to observations as one can get coherent in space and time. Sometimes people forget that it is assimilated against observations not observations. Data assimilation takes in observations and tries to make a accurate estimate of the state of the system. This includes observations from ground based, ships, bouyes, airplanes and satellites. The analysis is produced in the operational system, making it available within five days of real time. ERA5 is based on the Integrated Forecasting System, IFS cycle 4lr2. The data is available in 0.25^o degree and hourly resolution. Its an important product in the continuous climate monitoring of the earth system. Since we are learning from data its important to mention that the all sky radiance's from meteosat second generation, MSG in the period 2003-2012 is included in the assimilation. This is the same satellite as I gather the cloud masks. From more information on how the cloud mask are computed please read section 2.2.2.

Reanalyses data is often mistakenly referred to as observations. The differences between them was the theme of a essay in Bulletin American Meteorological Society, BAMS , 2015. They conclude with that they are not to different. Both involve inference (theory based calculations) and re-analysis relies on forecast and observations does not. This is not a significant difference as long as the forecast is sufficiently accurate. Its important to be aware of that the uncertainty of the reanalysis is less well known than for observations. This makes it harder to judge appropriate use of the reanalysis.

METEOSAT Second Generation - SEVRI

The highest resolved satellite retrivals come from geostationary meteorological weather satellites. Polar orbiting satellites usually provide daily resolution. **(STUBENAU)**. There is also a wide range of viewing angles. Optical properties is always a function of viewing angle and frequency (Huang et. al. 2018).

The angle attributes to small differences in detected cloud mask. This becomes evident when the standby and operational satelite scan simultaneously. By default the stanby satelite is adjusted to fit the position of the operational. By taking the difference some small patterns becomes visible. This is not accounted for when using the data.

The second generation satellite consist of METEOSAT 8 to 11. The MSG system provides a two satellite system. The operational satellite at a nadir point of 0^o latitude.

Sometimes both satellites gather data at the same time. Then the standby-satellite grid is rectified to a a grid of the operational one (Personal correspondence with the help-desk). In these images it becomes evident that viewing angle affects the retrievals. When this duplication occurs, the operational satellite is chosen. The temporal resolution of 15min and the nadir pixel size is 3km (Taravat, 2015). The sensor SEVERI har 12 channels. One broadband visible channel, three solar channels (0.6, 0.8 and 1.6 μm) and 8 thermal infrared channels (3.9, 6.2, 7.3, 8.7, 9.7, 10.8, 12.0 and 13.4 μm). (Taravat, 2015 - should probably find the source online on EUMATSAT web pages).

More practical implications. This one has two partners covering the Indian and Pacific ocean. Together they give almost a global view, discarding the poles. This will be useful if this trail run is successful. **programming language. how to best implement them in**

EUMETSAT Cloud Mask

The EUMETSAT cloud mask, CLM relies on the fact that clouds are colder and more reflective than the surface. They also reclassify isolated pixel. The data is available on Earth Observation Portal on EUMETSATS web pages. CLM consist of four classes, zero - clear sky over ocean, one which is clear sky over land, 2 denotes cloudy and 3 is outer space/off earth disk/no data. These classes are derived from almost all channels except (X) **cite article 10 in Tavarat, 2015**.

The cloud mask is distributed in GRIB-format (no coordinates) and NC-format (coordinates). Due to spatial limitations (Since all files have the same coordinates) one NC-files is downloaded for the coordinates and the rest of the data has been downloaded in GRIB-format. Some retrievals are excluded because of high uncertainty in the measurements this creates gaps in the data set.

Thoughts when choosing the data set of satellite images

Choose the finest temporal resolution possible. The average lifetime of a cloud is an hour. Since this is a proof of study is seems reasonable to choose the finest resolution available. Here this is METEOSAT.

In future work it would be interesting to asses how data driven parametrization compare to the existing parametrizaions available in the state of the art climate models. Here both the temporal and spatial resolution is a lot coarser. Other data sets could be considered. The masks in other data sets are computed based on more channels than in METEOSAT but the temporal resolution is a lot worse.

European Cloud Cover

Add somewhere from personal correspondence the gaps are caused by: We regret to inform you that unfortunately we cannot process your orders 1368572 and 1368598. The data was corrupted prior archiving to tape and this data file cannot be recovered. For the purpose of this thesis a new data set was built. The clouds in ERA5 is models according to the descriptions in 2.1.2 so I had to look elsewhere for a suitable *true value* to regress the meteorological variables against. Keep in min the hourly temporal resolution and 0.25^o spatial resolution in ERA5. Most satellite products move in polar orbits and finer resolution than daily is rare. Vertically resolved data was also unnecessary. **cite Calipso and Modis**. This guided me in the direction of the satellites in geostationary orbit. The only geostationary satellite covering the east Atlantic is the 0 degree meteosat second generation, MSG satellite. With the high temporal resolution of 15min and a fixed grid is seemed like a reasonable choice. Preserving the resolution available from ERA5, remapping the cloud mask to cloud fractions also know as cloud amount. *ECC comprises of five variables collected from two sources; ERA5 and EUMETSAT*The final product consist of the variables temperature, pressure, cloud amount, specific and relative humidity. Hourly data on a 0.25^o resolution in the period **put in first date** to **last date**. For this project the geographical domain has been restricted latitude $\in [30, 50]$ and longitude $\in [-15, 25]$. This becomes 80x160 pixels for each time step. As always when working with observations, data is missing. Since the individual pixels are remapped to fractions by using the area weighted mean, NaN's are not a issue. For sometimes no cloud masks are available. Then the closest time step available within the previous and trailing 45 minutes are chosen. Other gaps are documented in **X**.

The original data is described in table 2.1 and the finished product is described in the **??**. The demanding process of remapping the mask to fraction is presented in section 2.2.5 Fractional cloud cover is computed from the cloud mask product retrieved by the second generations METEOSAT satellites. You can read more about this data in section 2.2.2. For simplicity we will refer to this dataset as European Cloud Cover Dataset, ECC from now on.

The mapping from the curve-linear grid of the geostationary satellite to the uniform grid of era5 is quite technical and is described in section **??**. The cloud amount of a pixel is the sum of the area weighted cloud mask contributing to a cell. **transponer tabellen!!!!**

Physical basis of variable decision

The variables have been chosen based on availability, *uncertainty/ quality* and their contribution to the physical processes. As mentioned in the introduction clouds require a aerosol and sufficient supersaturation. *However up-draft velocities and aerosols (especially type) is difficult to impossible to know?* **kilde?** I hypothesis that some of this information is present in temperature, humidifies and surface pressure in space and time and thus the cloud cover can be inferred patterns of these variables. Except for rural areas there is usually CNN and/or INP present. **lov å si kilde!**

Temperature

The temperature is the two meter temperature produced by ERA5. For simplicity this will be referred to as temperature from now on. High temperature is related to convection. The air close to the surface gets heated, this reduces the density and starts the process of rising

Source	Type	Variables	Projection	Availability	Licence
ERA5	Surface 1000 hPa	2m Temperature Surface pressure Relative Humidity Specific Humidity	Uniform grid	1979 - dd.	Need user on Copernicus Data Store. Available for everyone.
MSG	Satellite retrival	Cloud Mask	Curvelineaer grid	2004-dd.	Higher than 3hourly resolution requires a reaserchers liscence.

Table 2.1: *Data description on the data present in the dataset ECC (European Cloud Cover).* ***Add projection as a column, Availability for download and the period of data.***

air masses. As the air rises it cools as a consequence of the work that has been done to the surrounding because of expansion. The temperature could also be a seasonal proxy.

Humidity

Both specific and relative humidity is included as features. Conditions where relative humidity exceeds 100% are called supersaturated. Intuitively this should be a good predictor. However since the the clouds don't form at the surface its not clear if specific humidity is a better predictor. Specific humidity is the actual amount of vapour in the atmosphere in units of **kg/kg?**. *Trude: Can I say that the rate of evaporation at the surface is proporsjonale to the humidity?* To some extent these variables are temperature dependant. Warm air can retain more vapour than cold air. Which is the main reason from precipitation.

Let humidity be a collective term for both humidity's. The data is gathered from the model level closest to the surface. Which is at a height of 1000hPa.

In order to form a cloud you need the air to exceed a saturation with respect to water or ice. This can with the presence of cloud condensation nuclei/ ice nuclei particles start the

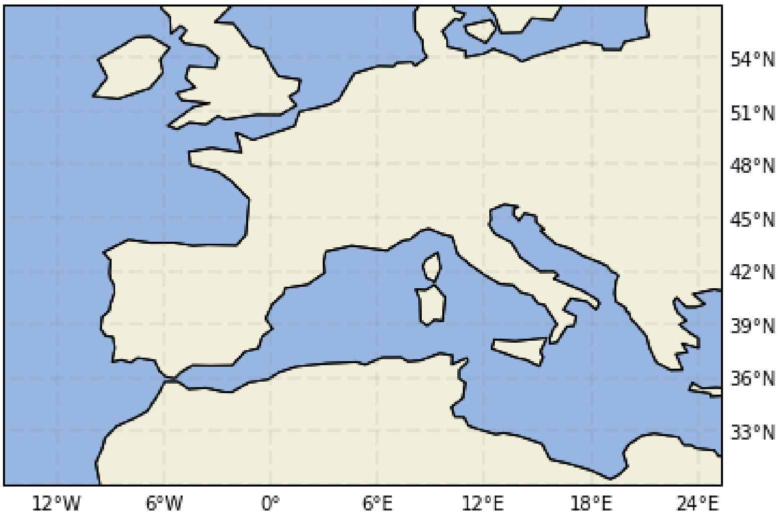

Figure 2.1: *Map showing the domain. The region cover southern Europe and northern Africa. The coordinate system is Plate Caree, which is the same as ECC. The image have been generated using the python package cartopy ref?* ***new figure with correct range and some vegetation or topography on***

cloud formation processes. On the other hand specific humidity is the actual amount of water vapour present in the atmosphere.

Surface Pressure

Due to the earth geometry and the angle of rotation there is a energy surplus at the equator. Winds transport some of this heat pole ward. Geostrophic winds are the large scale balance between the Coriolis force and the pressure gradients. This wind flows parallel to the isobars, lines of constant pressure. Any factor that generates a pressure gradient can create disruptions in the wind pattern's. Topography for instance. For convenience surface pressure will be referred to as pressure in this thesis. **Don't think including the equation from geostrophic balance improved this section.**

Computing cloud fractions

The cloud fractions in ECC is the area weighted cloud masks from MSG. The masks in MSG referred to as clear (0) or cloudy (1). If there is a NaN present, this pixel is simply excluded. Calculation wise this has the same affect as a clear pixel. **Should this be updated to remove the area of the cloudy pixel.**

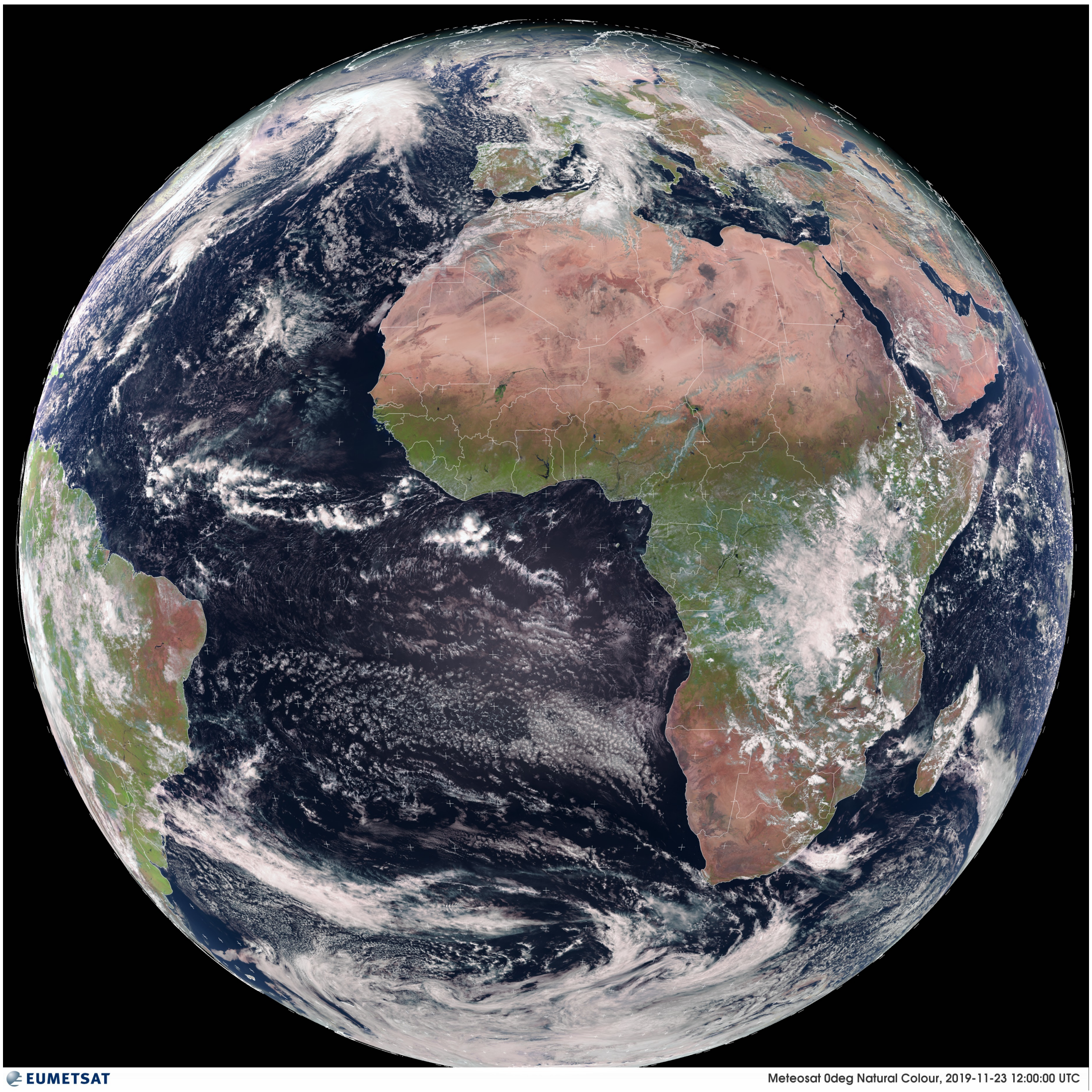

Figure 2.2: *The view of the earth from MSG. The picture is dated noon on the 11 November 2019.* ***Cite EUMETSAT****. By studying the patterns it becomes evident that clouds are influenced by the circulations.*

The METEOSAT cloud mask provided in NetCDF-format contains the coordinate information. Since the coordinate system is constant only one NetCDF files are downloaded.The surface area of a square on a sphere can be computed by solving the integral in (2.1). This is the result in (2.2). Since the netCDF-files only contain latitude and longitude informa

Before using this formula approximations of the extent of the cell needs to be made.

Calculating the area based on a curve-linear grid that only contain information about latitude and longitude involves some simplifications. From (2.2)

, which can be rewritten into equation (2.2). The latter one is used for implementations. Here

R denotes the distance to earth centre, θ the latitude and ϕ the longitude. Approximation of $d\phi$ and $d\theta$ have been done based on the two-dimensional fields of latitude and longitude values according to equations (2.3) and (2.4).

The visual comparison between raw satellite images and cloud amount seem to agree. The cloud fractional distribution also retain the same shape as ERA5 and MODIS 6.1 terra in the period from 2004 to 2018.

Update equation to use a upside down delta.

$$A = -R^2 \int_{\theta-\delta\theta}^{\theta+\delta\theta} \int_{\phi-\delta\phi}^{\phi+\delta\phi} cos\left(\theta'\right) d\phi' d\theta' \tag{2.1}$$

$$A\left(\theta, \phi, \delta\theta, \delta\phi\right) = 2R^2 \left(sin\left(\theta+\delta\theta\right) - sin\left(\theta-\delta\theta\right)\right)\delta\phi \tag{2.2}$$

The latitude and the extend of the pixel is terms in this equation. The extent of a pixel can also be interpreted as the **what.** The changes on longitude at a certain pixel is the the average distance to neighbouring points.

$$\delta\phi_{i,j} = \left|\frac{\phi_{i+1,j} - \phi_{i-1,j}}{4}\right| \tag{2.3}$$

$$\delta\theta_{i,j} = \left|\frac{\theta_{i,j+1} - \theta_{i,j-1}}{4}\right| \tag{2.4}$$

The latitude, longitude information is retrieved from the product of the satellite images. In order to keep the data storage to a minimum most files are download in grib-format. All satellite images from the 0 degree service are given with the same coordinates. This became evident from conversation with the help desk.

The cloud cover will be referred to as cloud amount, fraction or simply the clouds.

Licences and Downloading Data

Scripts for downloading the ERA5 data used in this thesis is available in the project GitHub on https://github.com/hannasv/MS/tree/metos/downloading_RA. However you will need to create a CDS-user. I suggest you follow the instructions on ECMWF homepages on *how to download ERA5*. There are no scripts available for downloading METEOSAT data this is

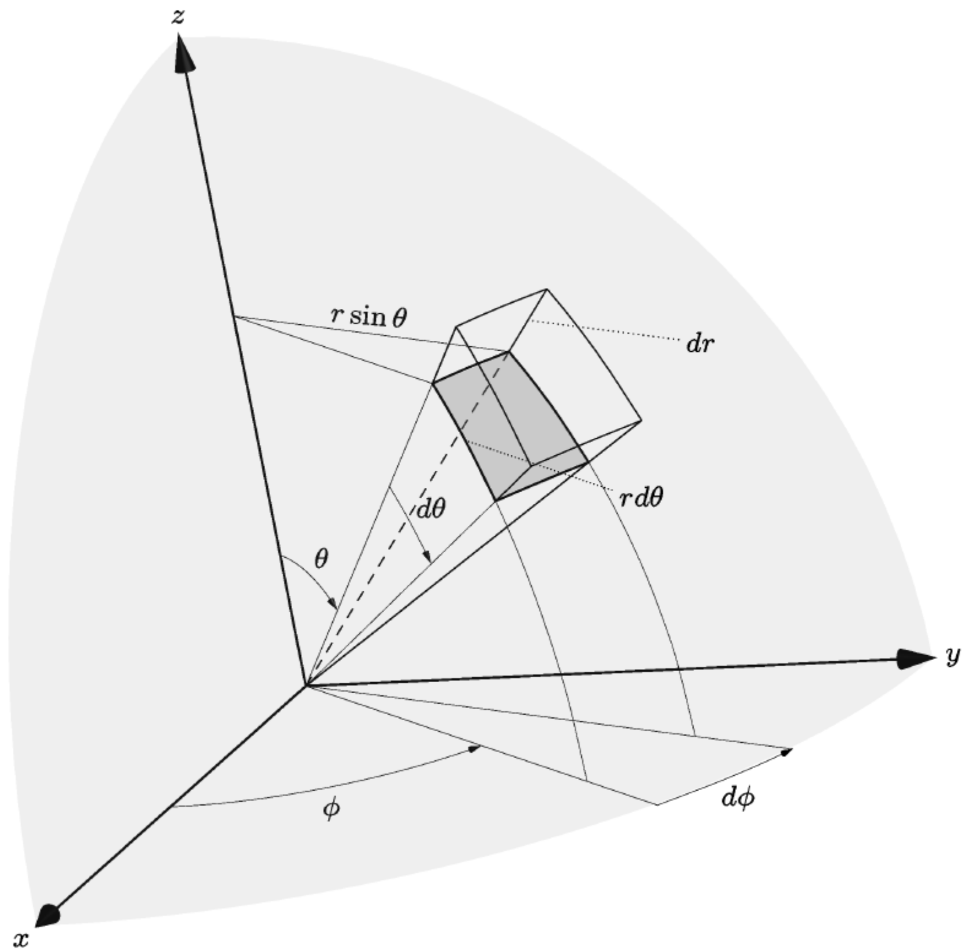

Figure 2.3: *Credit, https://tex.stackexchange.com/questions/159445/draw-in-cylindrical-and-spherical-coordinates*

done using satellite retrievals at EUMETSATs Earth Observation Portal. Its freely available in hourly resolution. Scientist can apply for increased resolution up to 15min. You choose the cloud mask product in grb-format. By running **X - legg inn filnavn** you can remap your own files.

summary

In this thesis I want to test if its possible to make a parametrization on total cloud cover based on macro-scale variables like humidity, surface temperature and pressure. Move away from the subgridscale processes, by regressing historical observations against macro physical properties which affect clouds. The question remains: Is there be enough information in humidity, temperature and surface pressure to predict clouds in a time and space.

1. How representative is the training period we choose?

2. politikk - machine learning velger personalisert reklame i forhold til valg.
3. speech recognition (google speaker)
4. spam filter 1990
5. selv kjørende buss - aker brygge
6. manipulere video - at du får politikere / andre til å si ting de aldri har sagt
7. face generation
8. testing (in-sample error) and validiation (out of sample error) - generalization error
9. pipeline - transformations of cloud cover data.
10. predicting house prices is a typical regression task. Predicting if a person will default on its loan is another one.
11. clustering and classification as examples on mnist (overused dataset)
12. Stuff on downloading data and installing the envionment can be called *setting up your workspace – > downloading data*
13. Look at the correlation in the data. Pearson correlation? Someone else correlation?
14. *This plot reveals several things. First, the correlation is very strong. Second, ...*
15. *You will often gain good insight on the problem by examining*
16. Ensamble methods in climate models and machine learning. It true that for both domains the model mean usually outperform the single model.

Machine learning

In this section I will explain the computational methods used for generating the numerical experiments conducted in this thesis. Starting with the the performance metrics used to evaluate the models. Followed by the auto regressive model and recurrent neural networks. For recurrent nets I will start by explaining the simple feed forward network building up to the more complex recurrent networks. This network is also known as convolution long short-term memory network. **Anatomy and architecture.**

Machine learning is a part of our daily life. From self driving bussed in your

is a word on a lot of peoples tounge. It becoming more and more a part of our daily life from search engines to googles speakers responding to speech. There are lot of different types of machine learning, suitable for solving different tasks. Figure 2.4 shows the types of machine learning and their subcategories. Supervised learning is the part of machine learning concerned with learning the relation between input data, x and labelled data y. Regression predict continuous values. Replicating a function. Classification is discrete, since it assigns a category to the input. Reinforcement learning is solves games, labyrinth and fluid mechanics **(Jean Rubalt)**. Unsupervised learning tries to detect patterns in unlabelled data. This includes contains clustering and dimensionality reduction. Unsupervised and reinforcement learning is out of the scope of this thesis and will not be further discusses. There is a ongoing

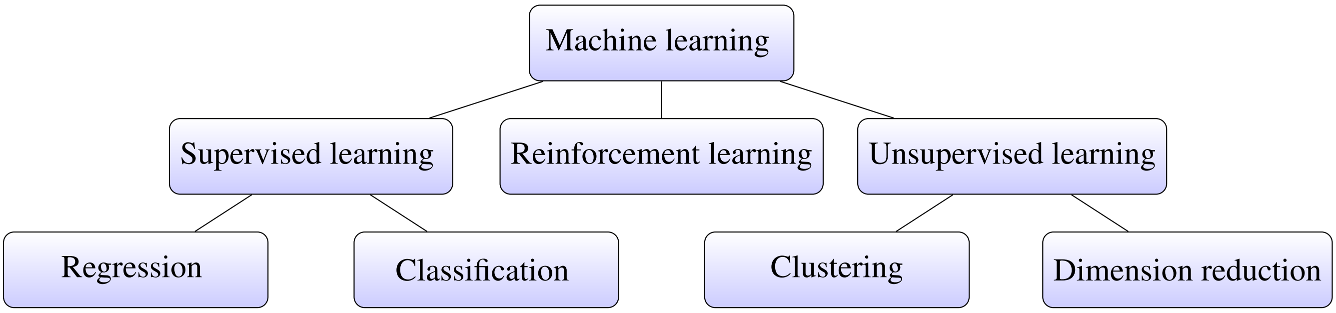

Figure 2.4: *Graph showing types of machine learning and their subcategories.*

debate on what is intelligence. Traditionally a machine would be considered intelligent if it would beat a human at a given task. This has later been abandoned. The abilities a machine need to posses in order to beat a human in chess is completely different. A human needs X while the machine need Y. **kilde Chollet google** This will not be further discussed in this thesis since this is about task specific intelligence. Is it possible to train a network to gain sufficiently intelligent at the task of prediction European cloud cover?

Supervised learning

Linear regression is the simplest form of supervised learning. Finding a suitable curve for a set of points. Working with real data, there is usually noise present in the dataset. In order to compensate for this we split the data into training and test (validation) sets. Overfitting becomes evident when you have a increase in the difference between the test- and training error. In non-mathematical terms, you have adjusted to much to the training data and where not able to find the general relation or "rules". See figure 2.5

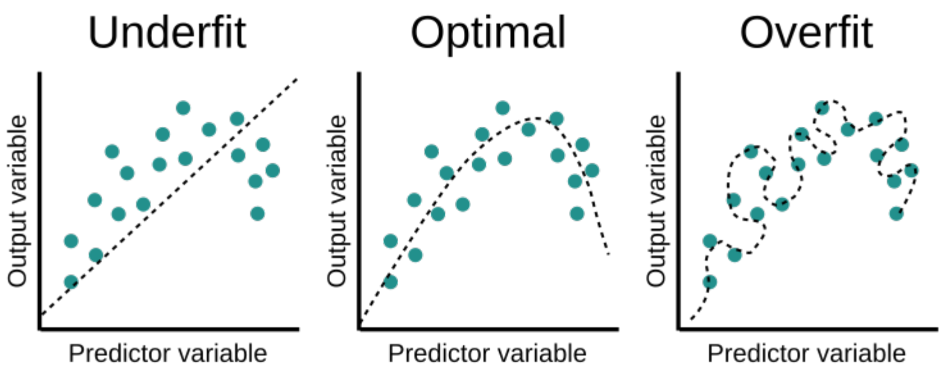

Figure 2.5: *Fitting at different levels. The optimal fit is the most general one. This is applicable to many cases.*

Autoregressive models

The autoregressive model is a form of linear regression models where you allow a certain number of timesteps to be predictor variables. Equation (2.5) describes how to make a prediction based on the optimal weighs, β. The expression for finding the optimal betas in a matrix form is given in equation (2.6). In the case for linear regression (using MSE- loss) there is only one solution to the optimal beta values. This makes it computationally very fast, as long as the matrix $X^T X$ is non-singular and thus its inverse exits. **Should i derive equation** (2.6)**?**. For more complicated loss surfaces, there is no analytical solution. Gradient descent is a common algorithmn used to work around this. **write a section on gradient descent. Illustrert med mann som går ned fjellet og**

$$\hat{Y_n} = \hat{\beta_0} + \sum_{j=1}^{p} X_j \hat{\beta_j} + \sum_{i=1}^{n_{ts}} Y_{n-i} \widehat{\beta_{p+i}} \tag{2.5}$$

$$\hat{B} = \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X} \bar{y} \tag{2.6}$$

Transforming data

Since cloud cover fractions are in the range from zero to one, a common approach can be to transform the data so it includes values from the entire real axis $(-\infty, \infty)$. This is done using the inverse of the sigmoid function (see figure 2.6). **mer?**

Neural network

Figure 2.7 shows a simple architecture of a neural network. This consist of four input nodes or neurons, one for each of the variables relevant for the problem at hand. This is a fully connected network, meaning that every input node has a connection/weight to the next layer. After the input data has been passed to the hidden layer (sum over the matrix multiplication of the data and weight) it goes though a activation. The activation function in neural networks introduce the non-linearity's. Without then this would be piece-wise linear functions**..?**. Figure 2.8 illustrates the activation in one node based on the input of variables. In a neural network the connections between the nodes and the biases of the nodes need to be trained. This means that the optimal first layer in three layer model, is not equal to the optimal first layer in a two layer model.

Three important concepts to every model. The weights, loss-function and optimiser. **neurons..?** The models is built from weights. The loss is a measure on how close you are to what you want to learn. Optimiser updates/adjust the weights in the direction of a lower loss. **Backpropagation is the learning algorithmn?**

$$tcc = f(T_{2m}, q_v, RH, sp) \tag{2.7}$$

The rules I hope to achieve in this thesis is the physical relation of total cloud cover based on temperature, pressure and humidity. Like described in equation (2.7).

Explain Difference between a neural network used for classification and regression, is choises in cost-function and activation function in the output layer (linear for regression and softmax for classification).

Recurrent networks

Introduction to recurrent networks.

Convolution

Convolution is a mathematical operation where you move a filter across you two dimensional data. Here the result in one cell is the affected by its neighbours. In a convolution neural network the trained weights are the filters.

Climate models are implementations of physical equations and parametrizations. After some

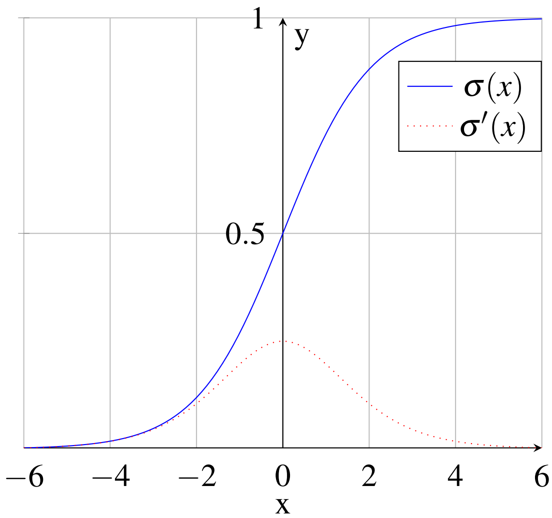

Figure 2.6: *Sigmoid function and its derivative.*

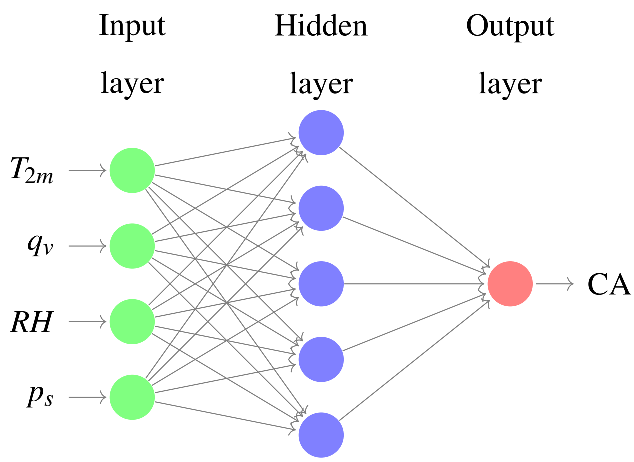

Figure 2.7: *One layer hidden neural network. Taking temperature, humidities and pressure as input and estimating a fractional cloud cover - cloud amount - CA*

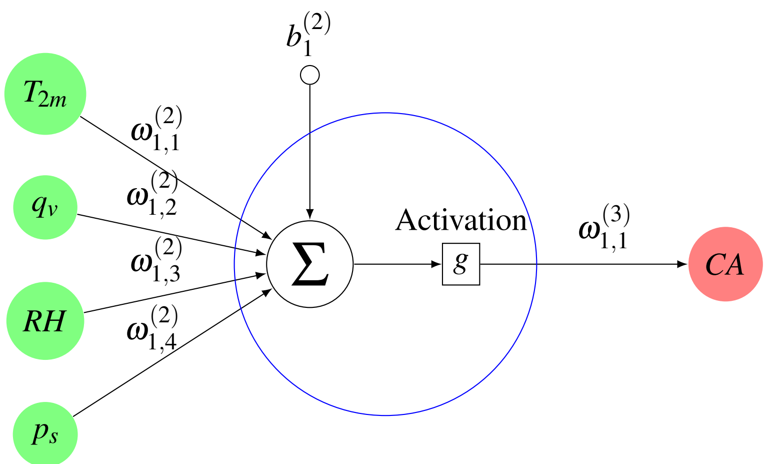

Figure 2.8: *Sketch of the activation on one of the nodes in the hidden layer. Edit figure to have one output node. Consider using another subscript than g for activation function..?*

tuning they are thought to provide the answers given a state. You can read more about climate models in section **??**. Since Hansen et. al. 1984 **siter** first attempted to estimate the equilibrium climate sensitivity, ECS , reaserchers have been working on reducing the spread. More complex models introduce other uncertainties and they have not yet been able to reduce it sufficiently **hvordan si "nok" - uten at et blir slemt**. With data available, more complex relations can be described. This requires deeper networks and the depth in deep learning refers to the number of layers. More complex models come with a higher risk of overfitting. **Kilde book Chollet**.

Metrics

In order to acquire a certain skill you need a measure determining how close you are. Use the sum of square or abosolute values in order to not penalize point on the lower side of the line. Or not having to deal with negative distances.

$$MSE(\hat{y},\hat{\tilde{y}}) = \frac{1}{n}\sum_{i=0}^{n-1}(y_i - \tilde{y}_i)^2 \tag{2.8}$$

$$ASE(\hat{y},\hat{\tilde{y}}) = \sum_{i=0}^{n-1}(y_i - \tilde{y}_i)^2 \tag{2.9}$$

$$R^2(\hat{y},\tilde{\hat{y}}) = 1 - \frac{\sum_{i=0}^{n-1}(y_i - \tilde{y}_i)^2}{\sum_{i=0}^{n-1}(y_i - \bar{y})^2} \tag{2.10}$$

where mean value of $\hat{y}$ is defined as $\bar{y} = \frac{1}{n}\sum_{i=0}^{n-1} y_i$. R^2 describes how much of the variation in the dataset you are able to capture with your model.

Hyperparameters - Automatic Optimization

Keras-tuner. **Explain all the params you tune. Might be beneficially with a figure. See Rune's MS-thesis.**

Convolutional Neural Networks

Convolutional LSTM

Key equations in convolutional LSTM is listed in equation (2.11) - (2.15). Here $\circ$ denoted the Hademand product, which is a component wise multiplication, and * is convolution.

$$i_t = \sigma\left(W_{xi} * x_t + W_{hi} * H_{t-1} + W_{ci} \circ C_{t-1} + b_i\right) \tag{2.11}$$

$$f_t = \sigma\left(W_{xf} * x_t + W_{hf} * H_{t-1} + W_{cf} \circ C_{t-1} + b_f\right) \tag{2.12}$$

$$C_t = f_t \circ C_{t-1} + i_t \circ tanh\left(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c\right) \tag{2.13}$$

$$o_t = \sigma\left(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \circ C_t + b_o\right) \tag{2.14}$$

$$H_t = o_t \circ tanh\left(C_t\right) \tag{2.15}$$

Learning algorithmns - optimizen (implementation of sequence length)

Describe the following

1. Loss, cost, performance metric? epoch batch

From Chollet book s. 11 *the fundamental trick in deep learning is to use this score (result from performance metric) as a feedback signal to adjust the value of the weights a little, in a direction that will lower the loss score for the current example.* The adjustment is the job of the optimizer, which implements backpropagation algorithmn which is the sentral learning algoritmn. **Explain this in words.**

Klipp og lim

When describing shapes, do it by keras convention. Which is channel first? **Talk about working with real data and not synthetic.** In this *Learning means finding a suitable representation of model parameters that minimize a loss function for a given set of training data samples and their corresponding targets.* **Does the setup here resemble videos mostly? Since its a sequence of frames containing meteorological data..**

Future work

1. Describe how to make parametrization to implement in CMIP6.

2. Programming language in climate models vs python.

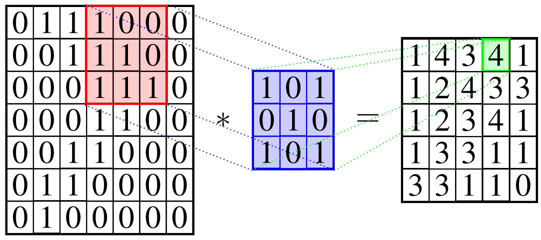

Figure 2.9: *Diagram showing a convolutional operation.*

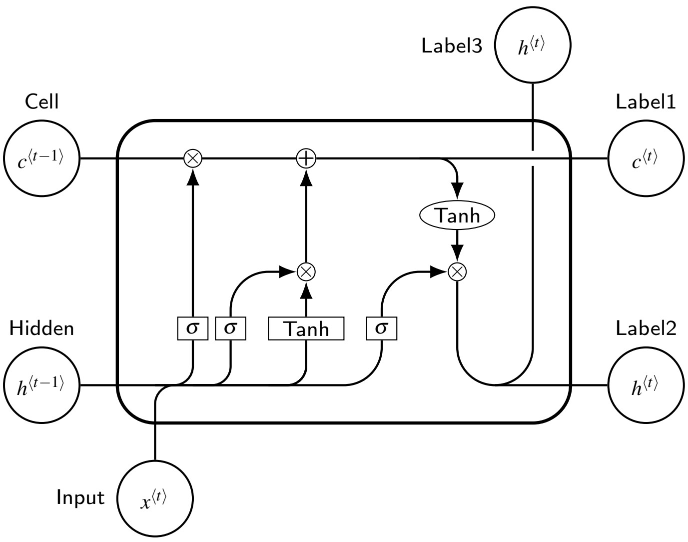

Figure 2.10: *Architecture of a LSTM cell.*

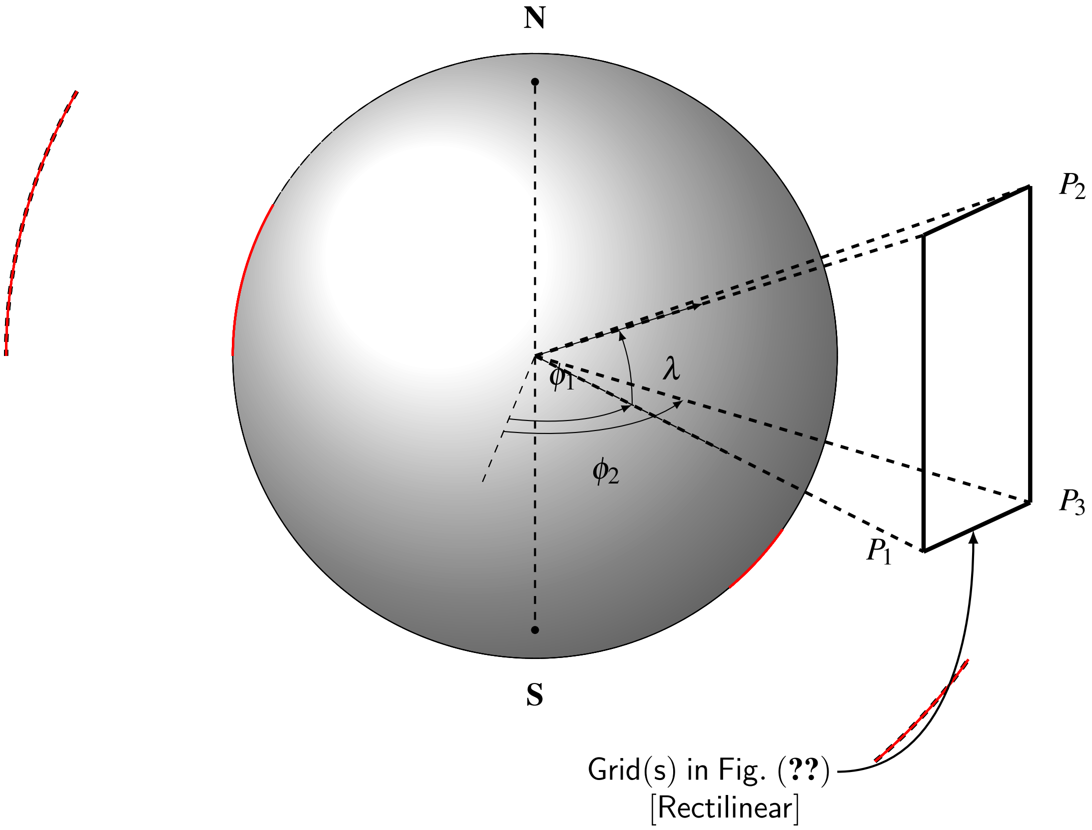

Figure 2.11: *Representation of spherical (red) and cartesian (black) co-ordinate systems. Latter gives an example of unstructured grids. Both unstructured. Conversion from former to latter involves a deformation factor which is acceptable within a given spatial limit. For my case, only unstructured flat meshes are employed (Lisboa Geodetic datum: black grid on the right). Confront above represented points (P_1, P_2, P_3) with Fig.(**??**). Mathematically frames change accordingly: see Eq.(**??**).*

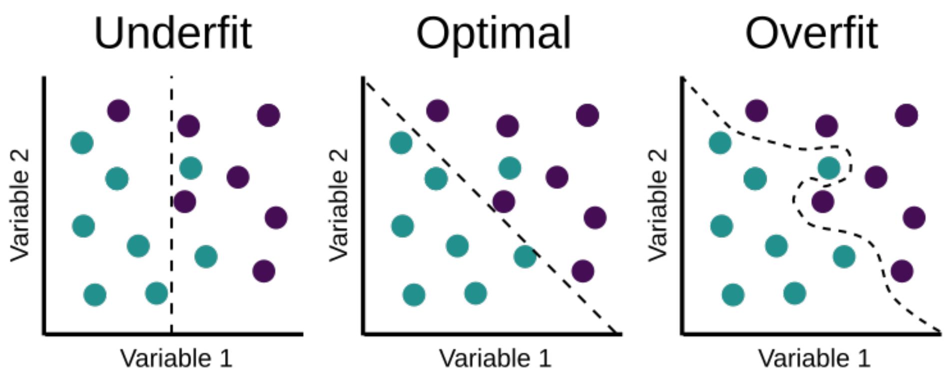

Figure 2.12: *Caption*

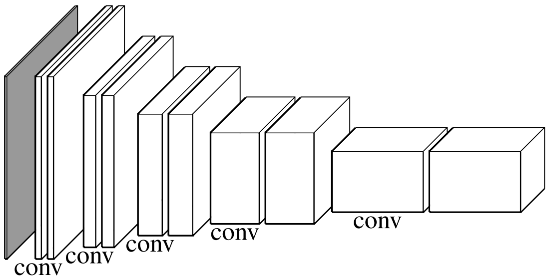

Figure 2.13: *Example CNN.*

Methods

Results

Conclusions

Discussion - currently inspo

INSPO

1. No further adjustment is needed for a proper comparison.

Siter Schubenau og si at satelitter definerer skyer basert på optical depth og at det derfor er nokså stor forskjellen på global gjennomsnittlig skydekket. I tillegg har man forskjeller på retrival algorithmns and sensors.

NOTES – to be removed

I dagen klima vet vi at skyer er vektig Wild et al 2019 radiative budget.

Usikkerhet for fremtidig klima. IPCC for klimafølsomhet og endringer i sky regimer.

Manaby and Waetherall 1967 og (75?). første klimamodeller Hansen et al 1984.

Arrenius 1896 og tyndal 1861 fourier 1827.

cloud regimes from satelite images. What to expect in Europe.

Figur forrig ipcc rapport.

- Sitere the earth machine. Si at machinlæring for stadig mer oppmerksomhet innenfor klimaforsning.
- Snakke generelt om machine læring og artikkelen the measure of intelligence.
- We'll come back to this in section 2.1.5.
- Is out of the scope of this thesis.
- To abbreviate the notation for the cross-entropy the loss function is used, so we can rewrite eq. (6) into
- The distribution of values by feature is shown in fig. 2.
- Create correlation matrix using

- We will keep in mind both the scores of 0.743 and 0.782, serving as fair baselines for a very simple, and slightly more sophisticated regression fit, respectively.
- Plotting R2 vs epoch, set ylim to 0,1. Not interresting to see where it learns.
- Summaries the best five architecture
- A drawback of introducing more layers is that it increases the complexity, and thereby the chance of

Advances in making deep learning possible:

1. Each layer learn the representation of the previous layer.
2. Deep networks can find relations in complex datasets this removed the need for feature representations and
3. classification + natural language processing –> image captioning. Getting the machine to write a sentence describing the image.
4. Bottlenecks - rediscoveries backpropagation algorithmn and gradient descent. Can take any loss surface and move down the gradient toward a minimum.
5. k nearest neighbour, easy to explain - one of the first classification algoritmns.
6. convolution - 2010 - works with every perceptual task - task involving learning something
7. solving pde's
8. emulating other – increasing speed while mimicking other models.

Good phases

- In its most simple form, the diffusion equation is given by
- By using a finer grid one can usually get better approximations
- The comparison will focus on computational time and accuracy.
- This was chosen after studying the development of the loss function as a function of number of iterations

- and is mandatory in more advanced architectures(e.g. residual nets) where a constant spatial dimension is demanded.
- a list of examples and so forth.
- (I will reference to source code/project part where relevant!)
- out of sample precision.
- Learn how to create plots with a zoomed in view.
- sufficiently large
- Viktig poeng. *However, both academic researchers and practitioners alike acknowledge the need to make tests on the actual data set that is subject of interest, as well as dedicating time and resources to tune hyper parameters*
- methods for tabular data vs images
- I have opted to use
- We have also modified our own code for a dense feed forward neural network produced for Project 2, see

Appendix

Section title

An additional appendix