Norwegian University
of Life Sciences

**Masters Thesis 2024 30 ECTS**
Faculty of Science and Technology

# A comparative study of soil temperature models, including machine learning models

Mats Hoem Olsen
Master of Science in Data Science

# Forword

Jeg vil takke alle som var rundt meg under skrivingen, for uten dere så ville jeg ikke ha klart å fullføre.

# Contents

# List of Algorithms



# List of Figures



# List of Tables



# Todo list

# Acronyms

**K | L | M | N | S**

**K**

**Kilden**

Norwegian Institute of Bioeconomy Research Kilden. 6

**L**

**LMT**

Norwegian Institute of Bioeconomy Research LandbruksMeteorologisk service. 6, 9

**M**

**MET**

The Norwegian Meteorological Institute. 6, 9

**MSTL**

Multiple Seasonal-Trend decomposition using LOESS. 9

**N**

**NMBU**

The Norwegian University of Life Science. 12

**S**

**STL**

Seasonal-Trend decomposition using LOESS. 9

**Abstract**
This study focuses on 3 models that have been used in the literature to predict soil temperatures. The depths chosen as targets are 10cm, and 20cm in 4 regions; Innlandet, Østfold, Vestfold, and Trøndelag. In each region there are 4 stations with self-draining or saturated soils.

**Oppsummering**
Denne studien ser på modeler som predikerer jordtemperaturer

Fill out when soon done

*Keywords*: Soil temperature, Machine learning, regression,hour, ...

# 1 Introduction

In agriculture soil temperature is one of the important parameters to put into consideration when thinking about pest prevention, conservation, and yield prediction. The reasoning for this is that knowing the soil temperature is knowing climate change [1], water management [2], yield [3], nitrogen processes [4] in the soil, calculation of plant-growth [5], when seeds start to sprout [5], potential flooding and erosions[6], and predicting when insect eggs hatch that were laid last winter. Being able to predict the soil temperature into the future will be a huge advantage for farmers, civilians, and scientists.

If it's important, why don't institutions measure it everywhere? There are several reasons for this, but a common reason is that it's expensive to install new equipment on old weather stations. Sometimes the weather station do have the sensors in the fields reading soil temperature at given levels, but due to technical misadventures and unforeseen phenomenons there might be gaps or misreadings that need to be replaced with approximations or NULL values[1]. There are algorithms, models, and statistical tools to interpolate these missing values but they have their drawbacks. For instance approximation by global mean, which is a common method used in timeseries[7]. This method is preserved global statistics, however does not represent local changes. Further more for a good estimation of soil temperature it is useful to include exogenous[2] features.

There has been done research into heat conductivity in soil that has lead to differential equations[8], however these equations[8, 9] are computationally expensive and difficult to simulate, or calculate[4]. To add to the complexity the heat dynamics change depending on soil temperature as they change the physical stucture of the soil

Find source!, if not remove

There are also massive developments in types of models, one of which are ASPER[10] that combines logical statments with deep learning models to aceve better or similar results to "non-logical" deep learning models, but on fewer samples. A study has been preformed with the model which shows that the model can reduse the number of samples/observations needed by a factor of 1/1000[11]. After an intervju with the study resercher (Machot), although the model needs a strict ruleset it is possible to incoporate baysioan statitics to make the model more geneneral for more applications by weaking the ruleset and imply that the rulse given might not be 100% accurate and can be relaxed.

Furter more there are investigation into introducing randomness to the model to improve prediction. As an example is a study into Brownian fractional motion to predict temperature fluxiations at the Campi Flegrei Caldera area.

[1]These values are different from 0 as they represent "no data" and can't be used to do calculations.
[2]Variable that can affect the model, but is not not directly described by the model.

In this study 4 methods will be compared and evaluated for the sake of further research into interpolation of missing data in northic countries based on as few features as possible. This study has chosen 2 types of models; Analytical, and Data-Driven models. There will also be base models to compare against, one for each model type.

# 2 Theory

This section discusses the theory behind the models used in the study, there are a few base models (section 2.1, and 2.4) based on the interpretation of the papers hand picked for this study

## 2.1 Linear regression

The regression model will be for the sake of convenience be expressed as the following expression

$$\left(\vec{F}(\mathbf{A})\right)\vec{\beta} = \vec{y} + \vec{\varepsilon}$$

Where $\vec{F}$ is a vector function with following domain $\vec{F}: \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times p}$ where $m, n, p \in \mathbb{N}$, $\mathbf{A}$ is the data in matrix form with dimensions $\mathbb{R}^{m\times n}$, $\vec{\beta}$ is the regression terms, $\vec{y}$ is the target (TJM10 or TJM20), and $\vec{\varepsilon}$ is the residual error.

The $\vec{F}$ is not important, just that your data is shaped by a function.

Remove

This basic model to express the linearity of the components to soil temperature. This will function as the base model for regression models.

## 2.2 Plauborg linear regression model with Fourier terms

An improvemnt over an time independent linear regression model would be a time dependent linear regression model that takes not only current time into account of the calculations but also previus messurments. It is current knowlage that soil temperatures depends on previous temperatures and metrological phenomenons. In the paper "Simple model for 10 cm soil temperature in different soils with short grass" the author chose to extend the features from only air temperature at current time to include also previus days of year and the air temperature from those days. This means the following F function that Plauborg used would be

$$\vec{F} := [air_t, air_{t-1}, air_{t-2}, air_{t-3}, \sin(\omega t), \cos(\omega t), \sin(2*\omega t), \cos(2*\omega t)]^T$$

Where $air_t$ is the air temperature at time $t$ expressed in day of the year (0-365), $\omega$ is the angular frequency in radians per hour or radians per day, depending on the time unit. The sine/cosine elements in the F function represent the variations through the day by fitting $\vec{\beta}$ to the yearly variation. To adapt the authors model to an hourly time unit would be to either

1. Extend the F function to include a larger $\omega$ coefficient to reflect hourly oscillations in conjunction with daily fluxiations

2. Refit the Fourier terms with a larger $\omega$ coefficient to make the oscillations more representative of daily temperature changes.

The larger coefficient could be expressed as $\pi/12$ while the smaller $\omega$ for daily values would be rescaled to $2\pi/365$.

The problem with this approsh would be Fouriers Sine-Cosine series approximation which would suggest that Plauborg's method could be subject to overfitting with addition of more terms. On the other hand it gives us a way to compute the coefficients $\alpha_i$ and $\gamma_i$ for sine and cosine terms respectively, though it would be more numerically stable with a pseudo-inverse computation or a max log likelihood approach.

## 2.3 Rankin's finite difference method of simplified heat flow in snow covered soil

Consider removing since not implemented correctly

A more direct method based on laws of physics develop by Karvonen involves forming a Finite Difference Method (FDM) around point of interest with simplifications to the equations described in *A model for predicting the effect of drainage on soil moisture, soil temperature and crop yield*. A team of researchers collaborating with the original author found an algorithm by making simplifications to the general differential equations forming a iterative 2-step procedure seen at the procedure 1.

**Algorithm 1:** Rankin algorithm

```
Data: D, f_d
Result: T_Z
α_t ← (∂T/∂t)/(∂²T/∂z²);
for t ∈ T do
    T_*^{t+1} ← T_Z^t + Δt × α_t/(2Z)² × (T_{air}^t − T_Z^t);
    T_Z^{t+1} ← T_*^{t+1} ∗ e^{−f_d × D};
end for
```

Where $\alpha_t = K_T/C_A$ is the Thermal diffusivity from Fourier's law in thermodynamics, $K_T$ is average soil thermal conductivity, $C_A$ is the apparent heat capacity, and $f_d$ is the damping parameter that has to be empirically derived however for this study it will be estimated from the data through the following estimation

$$f_d \approx \frac{-\ln\left(\frac{T_Z^{t+1}}{T_Z^t + \Delta t \frac{\alpha_t}{(2Z)^2}(T_{air}^t - T_Z^t)}\right)}{2D}$$

The approximation used in the algorithmn 1 assumes that $K_T$ is not dependend on depth . To make the approximation of $\alpha_t$ more accurate the inclusion of rain ($\theta$) to introduce variation can be approximated with

$$\alpha_t \approx \frac{b_1 + b_2\theta + b_3\sqrt{\theta}}{a_1 + a_2\theta}$$

proposed by Kodešová *et al.*[13][3]. To make the computation easier of this Padé-Puiseux[4] approximation hybrid we will realize that $\alpha_t$ is expressed by

$$\frac{b_1 + b_2\theta + b_3\sqrt{\theta}}{a_1 + a_2\theta} \approx \alpha_t \approx \frac{(T_z^{t+1} - T_{air}) * (2z)^2}{(T_{air} - T_z^t) * \Delta t}$$

Thereby only needing a linear regression of two F-functions; $F_1 = [1, \theta, \sqrt{\theta}]^T$ and $F_2 = [1, \theta]^T$ rather than a three step approximation. This algorithm (algorithm 1) will approximate the following integral

[3]This representation was not proposed by the author however the linear approximations was proposed to approximate $K_T$ and $C_A$ respectfully. Since $\theta \propto m_w$ we can substitute water content with rain in mm since the area is constant and during all messurement the soil type will be the same, however this would need to be resestimated if a station contains a different soil type as the constant has a wide range of values[13].

[4]Padé Approximation is a of the form $\frac{\sum_{i=0}^{\infty} c_i x^i}{\sum_{j=0}^{\infty} c_j x^j}$ and a Puiseux series is a $\sum_{j=N}^{\infty} c_j x^{j/N}$

$$T = \int_{t_0}^{t_{max}} \frac{K_T}{C_A} \frac{\partial^2 T}{\partial z^2} dt$$

via a Finite Difference Method, although other methods are possible with higher accuracy[5].Must verify for this case! This study will use the FDM used by the author for the purpose of making the results in this study comparable with the study presented in the paper “A simple model for predicting soil temperature in snow-covered and seasonally frozen soil.”

For inital values this study are utelizing 2 methods under different assumtions:

$$T_z^0 \approx \frac{k \exp(D)}{1 + \exp(D) \times (k-1)} \times T_{air}$$

Where k is $K_T * \Delta t/(C_A * (2Z)^2)$, and D is $-f_d * Snow_{Depth}$. This assumes constant air temperature above a constant layer of snow, though unrealistic since air temperature has a tendensy to change during the day due to solar radiation and other climate factors that can cool down or heat up the air. Another problem is the fact that the snow level ramins the same which is also untrue.

## 2.4 Long Short Term Memory model


Missing diagram of LSTM cell

When modeling soil temperature it is important to know the previus hours or days to predict the next timestep, for this a natural selection for a data driven model is a recurent network. This type of network makes prediction based on previus timesteps in the data, however the longer timespan the model takse into account the less important are the erlier timesteps in the data. To combat this there was developt an imporoved model called Long-Short Term Memory model[15] that deploys a memory cell that feeds information from erlier timesteps to the late ones. To make sure that redundant information or unimportant informantion dont get feed forward there are also forgetting gates that removes some of the newly learned patterns and integrates it into the memory cell.

Write more

The most common problem in Neural networks is the vanishing gradient problem where updating the first few layers of a large network becomes exponentially more difficult since the adjustments gets smaller and smaller for each layer towards the start rather than the reverse. Long Short Term-Memory changes this by caring information from the previous cells forward thereby allowing updating earlier cells with bigger impact than the standard approach[15]. LSTM is part of a family of Recurent Neural Network’s that passes information to other cells in the same layer.

Do proper credit of picture

[5] For example fourth degree Runge-Kutta method[14] which converges quicker than forward-Euler method or FDM.

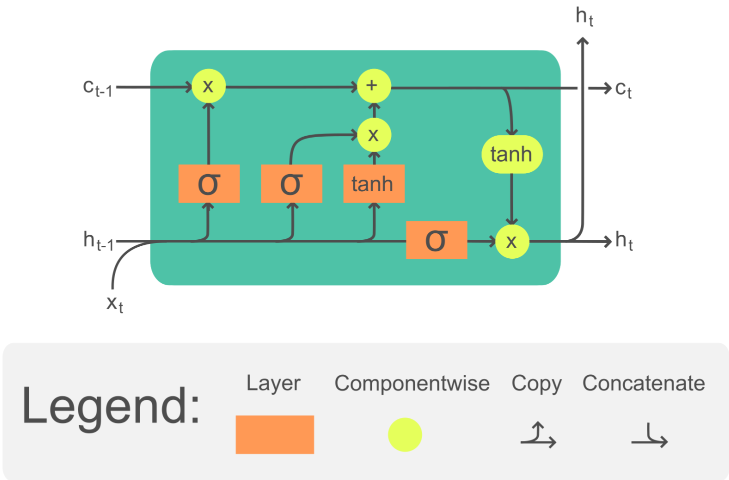


Figure 1: LSTM cell Artist: Chevalier [16]

## 2.5 Attention aware LSTM model

Consider removing if not model

# 3 Method

## 3.1 Source of data

For this comparative study the following data sources will be used
1. LMT 2. Xgeo 3. Kilden 4. MET

## 3.2 Dataset

The dataset is chosen from four regions in Norway; Innlandet, Vestfold, Trøndelag, and Østfold. From each region are four stations picked:

| Region | Name | ID | Drain type | Soile category | Texture | MET name | Latitude | Longdetude |
|---|---|---|---|---|---|---|---|---|
| Innlandet | Apelsvoll | 11 | Selvdrenert | CM | 17 | SN11500 | 60,70024 | 10,86952 |
| Innlandet | Fåvang | 17 | Selvdrenert | CM | 15 | SN13150 | 61,45822 | 10,1872 |
| Innlandet | Ilseng | 26 | Selvdrenert | PH | 17 | SN12180 | 60,80264 | 11,20298 |
| Innlandet | Kise | 27 | Vannmettet | GL | 99 | SN12550 | 60,77324 | 10,80569 |
| Trøndelag | Kvithamar | 57 | Vannmettet | ST | 16 | SN69150 | 63,48795 | 10,87994 |
| Trøndelag | Frosta | 15 | Selvdrenert | LP | 13 | SN69655 | 63,56502 | 10,69298 |
| Trøndelag | Mære | 34 | Selvdrenert | RG | 14 | SN71320 | 63,94244 | 11,42527 |
| Trøndelag | Rissa | 39 | Vannmettet | PL | 13 | SN71320 | 63,58569 | 9,97007 |
| Vestfold | Lier | 30 | Vannmettet | ST | 16 | SN19940 | 59,79084 | 10,25962 |
| Vestfold | Sande | 42 | Vannmettet | ST | 16 | SN26990 | 59,6162 | 10,22339 |
| Vestfold | Tjølling | 50 | Selvdrenert | AR | 13 | SN27780 | 59,04641 | 10,12513 |
| Vestfold | Ramnes | 38 | Vannmettet | ST | 16 | SN27315 | 59,38081 | 10,2397 |
| Østfold | Rakkestad | 37 | Vannmettet | ST | 18 | SN3290 | 59,38824 | 11,39042 |
| Østfold | Rygge | 41 | Selvdrenert | AR | 13 | SN17380 | 59,39805 | 10,75427 |
| Østfold | Tomb | 52 | Vannmettet | ST | 16 | SN17050 | 59,31893 | 10,81449 |
| Østfold | Øsaker | 118 | Vannmettet | ST | 18 | SN3370 | 59,31936 | 11,04221 |

Table 1: Station information from stations used in this study. The texture class is defined in this article: https://nibio.no/tema/jord/jordkartlegging/jordsmonnkart/dominerende-tekstur-i-overflatesjikt

All stations are sampled from the date[6] 03-01 to 10-31 from 2016 to 2020. The features rain (RR), mean soil temperature at 10cm (TJM10), mean soil temperature at 20cm (TJM20), and air temperature at 2m (TM) are sampled from the LMT database. The snow parameter is sampled from MET via Xgeo for imputed values in areas where there are no messured values. The soil type, and soil texture is sampled from Kilden from Norwegian Institute of Bioeconomy Research.

### 3.2.1 Selection process

The selection process for finding these station can be compiled into these steps

1. Recommendation from Norwegian Institute of Bioeconomy Research
2. Compute the missing values in the data
3. Missing values analyse

[6]Format month-day

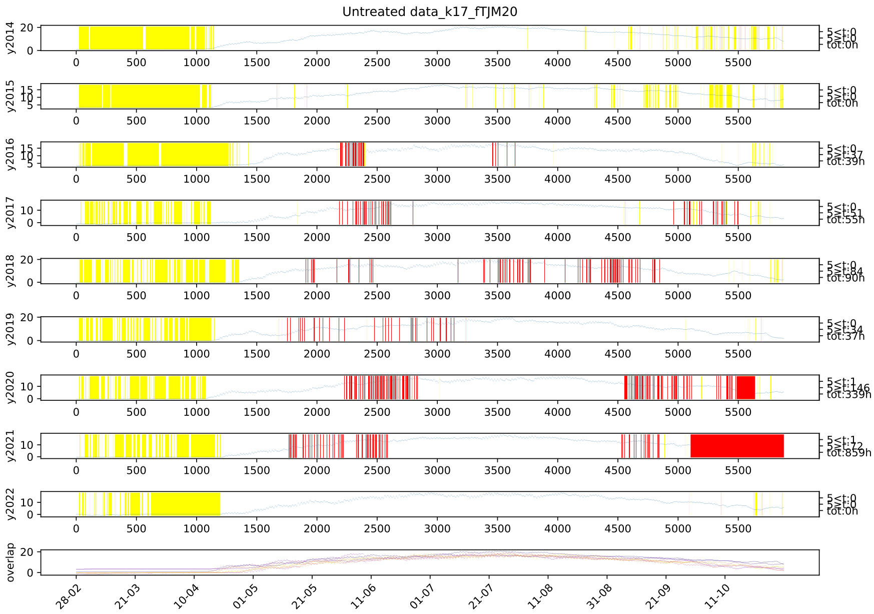


Figure 2: Visual representation of missing values at station 17 from 2014 to 2022 at the parameter "TJM20". The left numbers indicated how many hours that are missing and how many of them are shorter than or longer than 5 hours. The yellow markings indicate possible outliers based on the given year, all markings was checked if they were actual outliers. The red colouring indicate missing values in the data (represented in the data with code "NULL").

4. Searching LMT database for alternative station candidates if current data is insufficient
5. If some station was replaced the repeat step 2

The plots of stations follow a simple representation where the y-axis represent the year and the x-axis represent the index of the data as all tables are taken from the same period. A circle represent a singluar na values, while a band represent a series of 2 or more missing values. The colours represents the features used in this comperative study. This representation of the missing values will indicate sesonal, and systematic removal of data and give an overall indication of how much data is missing. To get further insight into the data a report is generated in parallel to the plots describing precise date and time of all values and which other parameter values is also missing values in the same period. See appendix **??** for the full detail of the report generation and appendix A for na-plots of the station chosen for this study.

### 3.2.2 Collection of data

The method used was a powershell[7] script that called the respective institutions servers using the "curl" program[8] to send an http request for the timeseries starting from 2014 to 2020 in the interval 1 of May to 31 of October. Code for data collection can be viewed in appendix **??**. The data is stores as an either a csv file or a json file for easy retrieval and manual control of values.

[7]Version 7.3.11
[8]curl 8.4.0 (Windows) libcurl/8.4.0 Schannel WinIDN

| FROST | Description |
|---|---|
| Stations with rain | Requested rain data in millimeters. |
| Station ID | Sendt a request to LMT for station information using their remote API. |
| LMT | Description |
| Meteorological data | Requested soil temperature from 10cm depth, and 20cm depth and air temperature (2m), from 2014-03-01 to 2022-10-31. |

Table 2: Description of what was requested from each server (FROST part of MET, LMT).

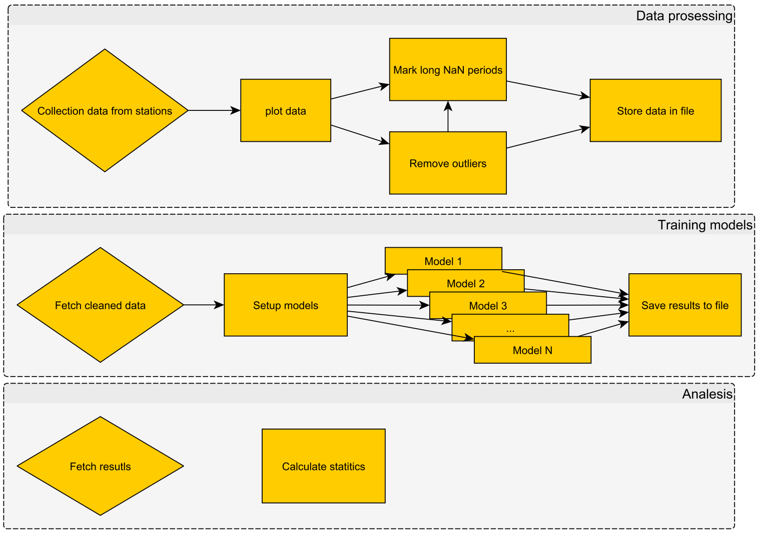


Figure 3: A surface level diagram of the

### 3.2.3 Labeling of stations between Nibio and MET

Since Nibio and MET have different names for the same stations one must compile a list that converts Nibio ID to MET ID. This was performed with these requests shown in table 2 where ID is the Nibio Id for the given station, Frost.ID is the MET id, ID. latitude is the latitude gathered from Nibio, ID.longitude is the longitude gathered from Nibio. These variables can be swaped out for the relevant station.

Move to caption

### 3.2.4 Storage of data

The storage of the data is done through two data structures; Hashmap and DataFrame from the package pandas. The transformation of data is done with a costume datatype called "DataFileHandler" which is converted to a module for convenience. The keys for the hashmap is chosen by the naming of the data files and the pattern given to the class. To escalete modeling the data will also be exported to a binary file for faster retrieval.

MORE DETAIL

Complete diagram

**Technical overview of custom data structure** The data structure used to store the data from the different stations is called "DataFileHandler" and stores the data in a tree-structure where indexes are dictated by the filename. It has several built-in functions to assist with data partitioning, and merging of data. This makes it easier to move and store all 846 720 observations from 16 station from all 4 regions[9]. Further more, the data structure has iteration functions so it is compatible with python's built in loops, and print functions.

## 3.3 Data cleaning and treatment

To use the data in this study it must be cleaned and treated for training. The following methods were picked common practice in litterateur with new methods based on the decomposition of the data in the from of Seasonal-Trend decomposition using LOESS (STL)[17][10].

### 3.3.1 Outlier detection and removal

Revise so it reflects what you actialy do

Though the data fetched from LMT is treated and controlled the external data from MET might not be, and this research project incorporated raw, untreated data from LMT to fill inn missing values.

The method to quickly find obvious outliers was to look at the following condition

$$|\frac{|\Delta T| - (|\Delta T|)}{var(|\Delta T|)}| > 4$$

Make more clear

This condition looks at the absolute difference between consecutive measurements and calculates the z-score for each observation. It is expected that the change in temperature can't be too rapid. Further methods used to highlight potential outliers is

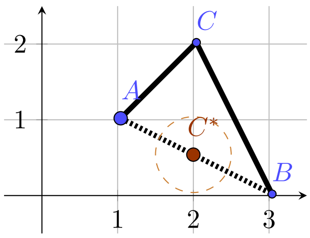


Figure 4: An simple outlier detection method utilizing a simple line to estimate where the expected point ($C^*$) is supposed to be. If observed point C falls outside the tolerance level (red dotted circle) then it is marked as an outlier.

### 3.3.2 Missing value imputation

The data has missing values, in particular during early Fall when there were sub-zero temperatures meaning any rain measurements done during this period would have unpredictable fluctuations since at negative temperatures water can freeze, get clogged up with residual bio-material from the surrounding area

Rewrite this part to reflect what is going on

[9]there are 4 stations per region.

[10]In this study we expand this for multiple seasons using Multiple Seasonal-Trend decomposition using LOESS (MSTL)[18], but the theory of this imputation method remains the same.

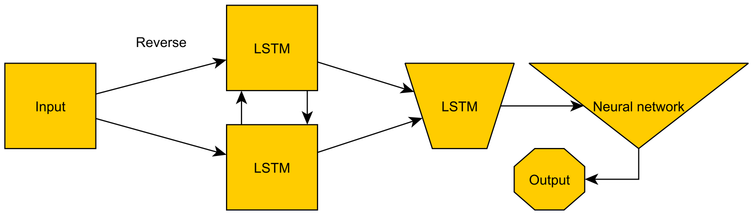


Figure 5

## 3.4 Setup of models

In general, write more on all subsections

The models are set up in according to the relevant paper the model is fetched from, alternatively reuse the code made by the author. When importing the data to the model there will be modifying to the original code to facilitate for the model as far as it goes. Any modifications will be in the appendix under section **??**. For the convenience of the reader all code is using the sklearn estimator class to make all the models discuses in this study more user friendly and compatible with sklearns other functions. The details of the models will be discussed in section 2, this section discusses the setup and implementation of the models.[11]

### 3.4.1 Basic Linear model

The linear model (sec 2.1) utilises in the study is created from the python model sklearn (or scikit-learn according to pythons package manager)

### 3.4.2 Plauborg

The Plauborg regression will be formulated as a linear regression problem so that the LinearRegression function in the Sci-kit module can be used. For the parameters used in the paper[12] the F function defined in section 2.2 will be formulated with loops to give rise 3 more parameters for fine-tuning the model.

### 3.4.3 BiLSTM

Complete caption, and improve figure

Soil temperatures are dependent on earlier timestemps meaning that to make a good prediction one needs to include temperature from $t, t-1, \ldots, t-k$ to make a decent prediction. As a base model it is usefull to evalueate the data both forwards and backwards to find features that is only notisable in a sertan direction. The BiLSTM defined is crafted with Tensorflow's Keras module for ease of use.

### 3.4.4 Attention-aware ILSTM

It is a know fact, since 1849[12], that to know previous weather patterns will greatly improve prediction accuracy. To improve the accurasy even more a model can focus one specific patterns that has a big impact on the prediction. The ILSTM attempts to do this by including a new thecnick in data science called attention[19] that takes a collection of data and gives each element a weight associated with importance. When that paper was published it was focused

[11] Caution to the reader; The code used was run on the Linux subsystem (Debian) on windows due to the fact that the current version of tensorflow can't run on Windows.

[12] First weather prediction made by Joseph Henry in 1849

on translation between English and German, however the paper published by Li *et al.* uses this novel teknick to do both time and feature importance and from that make a prediction.

## 3.5 Metrics

The metrics used in this study are
• Mean Squere Error • Mean Absolute Error • Explained Variance • bias • Log Condition number • digit sensitivity

Soil temperatur as a differnet behavor than air temperature since energy (temperature) though the soil gets dampen and delayed. Since the data used in this study has outliers that was not cought during datatreatment, which has been addresed, the author of this study desided to include two more metrics that are not usually included in the evaluation; The log condition number, and digit sensitivity. Both metrics are based on the calculation of the condition number defined as

$$\kappa = \lim_{\varepsilon \to 0^+} \sup_{|\partial x| \leq \varepsilon} \frac{|f(x + \partial x) - f(x)|}{|f(x)|} * \frac{|x|}{|\partial x|} \tag{1}$$

This is not feasible to calculate since infinite calculations with infinitesimal numbers is not possible as per April 29, 2024for simulation approach[13]. Therefore this paper uses algorithm 2 to approximate $\kappa$ for all the models.

**Algorithm 2:** Method for calculating $\kappa$. $\mathcal{U}$ is a uniform random distrebution in a range.

**Data:** Data
**Result:** $\log(\kappa)$

Let $\kappa_f$ be the function 1;
$\kappa \leftarrow 0$;
**for** $i \in 1 \ldots |Data|$ **do**
    $\partial x \leftarrow \mathcal{U}_{[-\sqrt{\varepsilon/|Data|}, \sqrt{\varepsilon/|Data|}]}$;
    $k \leftarrow$ calculate with $\kappa_f$ from $x$ and $x + \partial x$;
    **if** $k > \kappa$ **then**
        $\kappa \leftarrow k$;
    **end if**
**end for**
**return** $\kappa$

The digit sensitivity is included to give an intitive understanding of $\kappa$ and is computed simply as $\log_e(\kappa) + 1$. This number tells us the significant digit generated from the model. If the number is less than 0 then its the ith digit after the decimal point.

For the rest of the metrics, they are defined as follows

- RMSE $= \sqrt{\frac{\sum(y_{\text{pred}} - y_{\text{truth}})^2}{n}}$
- MAE $= \frac{\sum |y_{\text{pred}} - y_{\text{truth}}|}{n}$
- bias $= \frac{\sum(y_{\text{pred}} - y_{\text{truth}})}{n}$
- Explained variance $= 1 - \frac{\sum(y_{\text{pred}} - y_{\text{truth}})^2}{\sum(y_{\text{pred}} - \bar{y})^2}$

[13] This calculation is possible for some models, for instance linear regression models when converted to the form $A\vec{\beta} = \vec{y}$.

Where $\vec{y}$ is the mean of the target, $y_{\text{pred}}$ is the predicted data, and $y_{\text{truth}}$ is the observed soil temperature.

### 3.6 Use of Artificial Intelligence in this paper

In this paper there has been used Artificial Intelligence (AI), specifically Bing Chat / Copilot hosted by Microsoft Cooperation with special agreement with The Norwegian University of Life Science (NMBU), for the following purposes:

1. Formalising sentences and rephrasing sentences.
2. Spellchecking
3. Code generation of basic consepts and structures (tree traversal, template for generic classes)

It is important to emphasize that our engagement with AI have been actively curated and verified with known information. All code underwent rigorous manual inspection within a dedicated testing environment. Furthermore, no confidential or sensitive information was shared with the AI; our interactions focused solely on broad topics and general inquiries. To validate the accuracy of AI-generated responses, we cross-referenced them with established research papers and textbooks.

# 4 Results

Summerize results

Make sections

Include tables

## 4.1 Linear regression vs Plauborg

The result of the modelling (table 5 to 8) show that modelig soil temperature without the inclution of time is an inneffichent, and innaccurat method of predicting soil temperatures. The global measure for the linear regression has an average error of $2.3C^{\circ} \pm 4.23C^{\circ}$ while the global messure of the Plauborg daily model has an average error of $0.6C^{\circ} \pm 1.96C^{\circ}$. Further more Plauborg has an hight $R^2$ value indicating that it follows the changes in the soil better than just scaling the air temperature.

## 4.2 Modification of Plauborg

IN this study the original model, that was trained for daily values was converted to predict hourly data to see if the same formulation could be used to make predictions. When comparing the results shown in table 4 and table 3 to their daily counterpart it shows similar values showing that the model proposed in Plauborg can be extended to hourliy timeseries.

## 4.3 Deep learning models

If not gotten to work by 6.May, then delete. Try to do the stat (without model) to have results.

# 5 Discussion

Write a discussion

## 5.1 The Fall descrepensy

metion annomelies and resons

A phenomenon that arrose during performence evaluations was that the linear models struggles with the Fall season. The difference graphs shows a clear over or under estimation that are larger than $10\sigma$. When inverstigating the coeffisents to the model this decrepsensy can be contributed to the intercept that during low temperature ($< 5c°$) giving either an over estimation or an under estimation. Further more when removing the calculation of the intersept the same phenomenon is still precent.

Show table of parameters

## 5.2 Basis for results

The inclution of previous temperatures gives an improved estimation, even on hourly basis. The coefficents for both daily and hourly are observed to be <1 making it a mean temperature and the fourier terms would estimate the function[20]

$$e^{z/D}\sin(\omega t - z/D + \phi)$$

Since the term $exp(z/D)$ is constant we would be estimating $C\sin(\omega t - Q) = \sin(\omega t)\cos(Q) - \sin(Q)\cos(\omega t)$, where Q is $z/D - \phi$ and is constant. This will be extrapolated to a simple sum of sines and cosines as the model does. Together the Plauborg model would estimate

$$E_{\text{year}}(T)+e^{z/D}\sin(\omega t - z/D+\phi) \approx E_{\text{period}}(T)+\sum e^{z/D}\sin(Q_i)\cos(i\omega t_i)+\sum e^{z/D}\cos(Q_i)\sin(i\omega t_i)$$

A possible reason for the residuals could be the soiltypes at the stations making so a universal, high accurasy model would not be feasible unless including other metrological metrics (air pressure, humidity, soil type, soil texture, etc...) or including other non-linear features ($\sqrt{\text{temperature}}$,2D temperature change in depth and time, etc...).

## 5.3 Future work

The models chosen in this study is not a representative sample of current knowledge of soil temperature modelling, and this study did not aim for optimizing the models beyond what the original authors have already done with the exception for base models used for comparison puposus. A more comprehensive is needed of more complex models that utelises cutting edge technologies, techniques, and theory. One of which is logic based models, for instance ASPER[10] that tries to incoerate logical descriptions of the problem and limits the model for better or equal results based on fewer samples[11]. Another approch is to incorporate randomness into the deterministic models to explain the variation in the data, for instance fractional Brownian motion[21].

move som of this to introduction

# 6 Conclution

Everything is okay

# 7 Bibliography

## References

[1] Qingliang Li, Yuheng Zhu, Wei Shangguan, Xuezhi Wang, Lu Li, and Fanhua Yu, "An attention-aware LSTM model for soil moisture and soil temperature prediction," *Geoderma*, volume 409, page 115 651, Mar. 2022, ISSN: 00167061. DOI: 10.1016/j.geoderma.2021.115651. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S001670612100731X (visited on 10/05/2023).

[2] Meysam Alizamir, Ozgur Kisi, Ali Najah Ahmed, Cihan Mert, Chow Ming Fai, Sungwon Kim, Nam Won Kim, and Ahmed El-Shafie, "Advanced machine learning model for better prediction accuracy of soil temperature at different depths," *PLOS ONE*, volume 15, number 4, Lei Lin, Ed., page 25, Apr. 14, 2020, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0231055. [Online]. Available: https://dx.plos.org/10.1371/journal.pone.0231055 (visited on 09/29/2023).

[3] Ha Seon Sim, Dong Sub Kim, Min Gyu Ahn, Su Ran Ahn, and Sung Kyeom Kim, "Prediction of strawberry growth and fruit yield based on environmental and growth data in a greenhouse for soil cultivation with applied autonomous facilities," *Korean Journal of Horticultural Science and Technology*, volume 38, number 6, pages 840–849, Dec. 31, 2020, ISSN: 1226-8763, 2465-8588. DOI: 10.7235/HORT.20200076. [Online]. Available: https://www.hst-j.org/articles/doi/10.7235/HORT.20200076 (visited on 10/05/2023).

[4] Katri Rankinen, Tuomo Karvonen, and D. Butterfield, "A simple model for predicting soil temperature in snow-covered and seasonally frozen soil: Model description and testing," *Hydrology and Earth System Sciences*, volume 8, number 4, pages 706–716, Aug. 31, 2004, ISSN: 1607-7938. DOI: 10.5194/hess-8-706-2004. [Online]. Available: https://hess.copernicus.org/articles/8/706/2004/ (visited on 03/17/2023).

[5] Cong Li, Yaonan Zhang, and Xupeng Ren, "Modeling hourly soil temperature using deep BiLSTM neural network," *Algorithms*, volume 13, number 7, page 173, Jul. 17, 2020, ISSN: 1999-4893. DOI: 10.3390/a13070173. [Online]. Available: https://www.mdpi.com/1999-4893/13/7/173 (visited on 03/17/2023).

[6] Joris C. Stuurop, Sjoerd E.A.T.M. Van Der Zee, and Helen Kristine French, "The influence of soil texture and environmental conditions on frozen soil infiltration: A numerical investigation," *Cold Regions Science and Technology*, volume 194, page 103 456, Feb. 2022, ISSN: 0165232X. DOI: 10.1016/j.coldregions.2021.103456. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0165232X21002378 (visited on 01/31/2024).

[7] Mathieu Lepot, Jean-Baptiste Aubin, and François Clemens, "Interpolation in time series: An introductive overview of existing methods, their performance criteria and uncertainty assessment," *Water*, volume 9, number 10, page 796, Oct. 17, 2017, ISSN: 2073-4441. DOI: 10.3390/w9100796. [Online]. Available: http://www.mdpi.com/2073-4441/9/10/796 (visited on 02/17/2024).

[8] Tuomo Karvonen, *A model for predicting the effect of drainage on soil moisture, soil temperature and crop yield*. Otaniemi, Finland: Helsinki University of Technology, Laboratory of Hydrology and Water Resources Engineering, 1988, xvi, 215, Open Library ID: OL15197205M.

[9] Jean Baptiste Joseph Fourier and Alexander Freeman, *The analytical theory of heat*. New York: Cambridge University Press, 2009, OCLC: 880311398, ISBN: 978-1-108-00178-6.

[10] Trung Hoang Le, Huiping Cao, and Tran Cao Son, *ASPER: Answer set programming enhanced neural network models for joint entity-relation extraction*, version: 1, May 24, 2023. arXiv: `2305.15374[cs]`. [Online]. Available: `http://arxiv.org/abs/2305.15374` (visited on 03/12/2024).

[11] Fadi Al Machot, *Bridging logic and learning: A neural-symbolic approach for enhanced reasoning in neural models (ASPER)*, Dec. 18, 2023. arXiv: `2312.11651[cs]`. [Online]. Available: `http://arxiv.org/abs/2312.11651` (visited on 03/12/2024).

[12] Finn Plauborg, "Simple model for 10 cm soil temperature in different soils with short grass," *European Journal of Agronomy*, volume 17, number 3, pages 173–179, Oct. 2002, ISSN: 11610301. DOI: `10.1016/S1161-0301(02)00006-0`. [Online]. Available: `https://linkinghub.elsevier.com/retrieve/pii/S1161030102000060` (visited on 03/17/2023).

[13] Radka Kodešová, Miroslava Vlasáková, Miroslav Fér, Daniela Teplá, Ondřej Jakšík, Pavel Neuberger, and Radomír Adamovský, "Thermal properties of representative soils of the czech republic," *Soil and Water Research*, volume 8, number 4, pages 141–150, Dec. 31, 2013, ISSN: 18015395, 18059384. DOI: `10.17221/33/2013-SWR`. [Online]. Available: `http://swr.agriculturejournals.cz/doi/10.17221/33/2013-SWR.html` (visited on 02/29/2024).

[14] Carl Runge, "Ueber die numerische Aufl sung von Differentialgleichungen," *Mathematische Annalen*, volume 46, number 2, pages 167–178, Jun. 1895, ISSN: 0025-5831, 1432-1807. DOI: `10.1007/BF01446807`. [Online]. Available: `http://link.springer.com/10.1007/BF01446807` (visited on 02/29/2024).

[15] Sepp Hochreiter and Jürgen Schmidhuber, "Long short-term memory," *Neural Computation*, volume 9, number 8, pages 1735–1780, Nov. 1, 1997, ISSN: 0899-7667, 1530-888X. DOI: `10.1162/neco.1997.9.8.1735`. [Online]. Available: `https://direct.mit.edu/neco/article/9/8/1735-1780/6109` (visited on 10/18/2023).

[16] Guillaume Chevalier, *English: Schematic of the long-short term memory cell, a component of recurrent neural networks*, May 16, 2018. [Online]. Available: `https://commons.wikimedia.org/wiki/File:LSTM_Cell.svg` (visited on 12/10/2023).

[17] Robert B. Cleveland, William S. Cleveland, and Irma Terpenning, "STL: A seasonal-trend decomposition procedure based on loess," *Journal of Official Statistics*, volume 6, number 1, page 3, Mar. 1990, Num Pages: 3 Place: Stockholm, Sweden Publisher: Statistics Sweden (SCB), ISSN: 0282423X. [Online]. Available: `https://www.proquest.com/docview/1266805989/abstract/4DFACD0236364745PQ/1` (visited on 10/11/2023).

[18] Kasun Bandara, Rob J. Hyndman, and Christoph Bergmeir, *MSTL: A seasonal-trend decomposition algorithm for time series with multiple seasonal patterns*, Jul. 28, 2021. arXiv: `2107.13462[stat]`. [Online]. Available: `http://arxiv.org/abs/2107.13462` (visited on 03/19/2024).

[19] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin, *Attention is all you need*, 2017. arXiv: `1706.03762[cs]`. [Online]. Available: `http://arxiv.org/abs/1706.03762` (visited on 04/16/2024).

[20] T. R. H. Holmes, M. Owe, R. A. M. De Jeu, and H. Kooi, "Estimating the soil temperature profile from a single depth observation: A simple empirical heatflow solution," *Water Resources Research*, volume 44, number 2, 2007WR005994, Feb. 2008, ISSN: 0043-1397, 1944-7973. DOI: `10.1029/2007WR005994`. [Online]. Available: `https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2007WR005994` (visited on 04/29/2024).

[21] Antonio Di Crescenzo, Barbara Martinucci, and Verdiana Mustaro, "A model based on fractional brownian motion for temperature fluctuation in the campi flegrei caldera," *Fractal and Fractional*, volume 6, number 8, page 421, Jul. 30, 2022, ISSN: 2504-3110. DOI: 10.3390/fractalfract6080421. [Online]. Available: https://www.mdpi.com/2504-3110/6/8/421 (visited on 04/28/2024).

# Glossary

**D | H | M | S**

**D**

**DataFrame**

A table of values. The name is from the python library Pandas used in this study.. 8

**H**

**Hashmap**

A list of items where their unique placmnt in the list is detemend by their unique refrence key using a function that maps the key to a placement in the list.. 8

**M**

**Multiple Seasonal-Trend decomposition using LOESS**

Based in the traditional Seasonal-Trend decomposition using LOESS it decomposes a time-series into several seasons, trend, and residual[18].. 9, 18

**S**

**Seasonal-Trend decomposition using LOESS**

Takes a timeseries and decomposing it into a trend component, season component, and residual component using local regression for smoothing[17].. 9, 18

# Acronyms

**K | L | M | N | S**

**K**

**Kilden**

Norwegian Institute of Bioeconomy Research Kilden. 6

**L**

**LMT**

Norwegian Institute of Bioeconomy Research LandbruksMeteorologisk service. 6, 9

**M**

**MET**

The Norwegian Meteorological Institute. 6, 9

**MSTL**

Multiple Seasonal-Trend decomposition using LOESS. 9

**N**

**NMBU**
: The Norwegian University of Life Science. 12

**S**

**STL**
: Seasonal-Trend decomposition using LOESS. 9

# A Plots

# B Tables

| scope | spesific scope | RMSE °C | MAE°C | bias °C | log($\kappa$(model)) | digit sensitivity | $R^2$ |
|---|---|---|---|---|---|---|---|
| global | — | 2.677 | 2.061 | 0.527 | -0.322 | -1 | 0.756 |
| region | Østfold | 2.564 | 2 | 0.176 | -0.33 | -1 | 0.8 |
| region | Vestfold | 2.565 | 1.958 | 0.785 | -0.33 | -1 | 0.81 |
| region | Trøndelag | 2.937 | 2.279 | 0.753 | -0.32 | -1 | 0.484 |
| region | Innlandet | 2.62 | 2.001 | 0.374 | -0.324 | -1 | 0.798 |
| local | 52 | 2.504 | 1.976 | -1.2 | -0.328 | -1 | 0.803 |
| local | 41 | 2.135 | 1.665 | 0.13 | -0.321 | -1 | 0.872 |
| local | 37 | 2.513 | 1.938 | 0.067 | -0.326 | -1 | 0.83 |
| local | 118 | 3.029 | 2.422 | 1.722 | -0.323 | -1 | 0.656 |
| local | 50 | 2.176 | 1.7 | 0.815 | -0.318 | -1 | 0.836 |
| local | 42 | 2.739 | 2.099 | 0.746 | -0.323 | -1 | 0.807 |
| local | 38 | 2.983 | 2.323 | 1.149 | -0.323 | -1 | 0.736 |
| local | 30 | 2.276 | 1.708 | 0.428 | -0.327 | -1 | 0.859 |
| local | 57 | 3.079 | 2.419 | 0.744 | -0.33 | -1 | 0.617 |
| local | 39 | 2.79 | 2.186 | 0.633 | -0.316 | -1 | 0.615 |
| local | 34 | 3.158 | 2.424 | 0.811 | -0.325 | -1 | -0.342 |
| local | 15 | 2.706 | 2.094 | 0.827 | -0.323 | -1 | 0.484 |
| local | 27 | 2.455 | 1.885 | 0.335 | -0.33 | -1 | 0.839 |
| local | 26 | 2.757 | 2.105 | 0.892 | -0.324 | -1 | 0.801 |
| local | 17 | 3.063 | 2.29 | 0.102 | -0.329 | -1 | 0.754 |
| local | 11 | 2.346 | 1.847 | -0.023 | -0.325 | -1 | 0.755 |

Table 3: Results from hourly version of the Plauborg model for 20cm depth.

| scope | spesific scope | RMSE °C | MAE°C | bias °C | log($\kappa$(model)) | digit sensitivity | $R^2$ |
|---|---|---|---|---|---|---|---|
| global | — | 2.531 | 1.927 | 0.596 | -0.439 | -1 | 0.794 |
| region | Østfold | 2.448 | 1.894 | 0.512 | -0.447 | -1 | 0.816 |
| region | Vestfold | 2.412 | 1.81 | 0.733 | -0.445 | -1 | 0.846 |
| region | Trøndelag | 2.821 | 2.176 | 0.784 | -0.442 | -1 | 0.554 |
| region | Innlandet | 2.391 | 1.809 | 0.306 | -0.449 | -1 | 0.845 |
| local | 52 | 2.514 | 1.964 | -0.349 | -0.444 | -1 | 0.636 |
| local | 41 | 1.938 | 1.519 | 0.151 | -0.444 | -1 | 0.903 |
| local | 37 | 2.344 | 1.804 | 0.237 | -0.441 | -1 | 0.857 |
| local | 118 | 2.928 | 2.322 | 1.639 | -0.44 | -1 | 0.706 |
| local | 50 | 1.908 | 1.472 | 0.558 | -0.444 | -1 | 0.884 |
| local | 42 | 2.501 | 1.885 | 0.703 | -0.445 | -1 | 0.852 |
| local | 38 | 3.055 | 2.368 | 1.363 | -0.433 | -1 | 0.754 |
| local | 30 | 2.072 | 1.555 | 0.354 | -0.448 | -1 | 0.892 |
| local | 57 | 2.906 | 2.263 | 0.677 | -0.441 | -1 | 0.677 |
| local | 39 | 2.77 | 2.151 | 0.701 | -0.442 | -1 | 0.633 |
| local | 34 | 3.009 | 2.303 | 0.857 | -0.439 | -1 | -0.036 |
| local | 15 | 2.589 | 1.991 | 0.903 | -0.447 | -1 | 0.562 |
| local | 27 | 2.277 | 1.724 | 0.163 | -0.445 | -1 | 0.872 |
| local | 26 | 2.532 | 1.918 | 0.821 | -0.444 | -1 | 0.843 |
| local | 17 | 2.705 | 2.008 | 0.027 | -0.445 | -1 | 0.82 |
| local | 11 | 2.146 | 1.666 | 0.038 | -0.439 | -1 | 0.823 |

Table 4: Results from hourly version of the Plauborg model for 10cm depth.

| scope | spesific scope | RMSE °C | MAE°C | bias °C | log($\kappa$(model)) | digit sensitivity | $R^2$ |
|---|---|---|---|---|---|---|---|
| global | — | 1.91 | 1.536 | 0.643 | -1.909 | -2 | 0.876 |
| region | Østfold | 1.94 | 1.541 | -0.073 | -1.918 | -2 | 0.885 |
| region | Vestfold | 1.71 | 1.341 | 0.236 | -1.918 | -2 | 0.915 |
| region | Trøndelag | 1.843 | 1.56 | 1.461 | -1.91 | -2 | 0.797 |
| region | Innlandet | 2.16 | 1.735 | 1.019 | -1.915 | -2 | 0.863 |
| local | 52 | 2.33 | 1.873 | -1.402 | -1.914 | -2 | 0.83 |
| local | 41 | 1.748 | 1.409 | -0.371 | -1.918 | -2 | 0.914 |
| local | 37 | 1.877 | 1.496 | 0.353 | -1.904 | -2 | 0.905 |
| local | 118 | 1.742 | 1.384 | 1.139 | -1.914 | -2 | 0.886 |
| local | 50 | 1.251 | 0.985 | 0.096 | -1.912 | -2 | 0.946 |
| local | 42 | 1.966 | 1.54 | 0.346 | -1.91 | -2 | 0.901 |
| local | 38 | 1.721 | 1.367 | 0.515 | -1.915 | -2 | 0.912 |
| local | 30 | 1.817 | 1.471 | -0.014 | -1.904 | -2 | 0.91 |
| local | 57 | 1.841 | 1.538 | 1.427 | -1.91 | -2 | 0.863 |
| local | 39 | 1.729 | 1.468 | 1.402 | -1.913 | -2 | 0.852 |
| local | 34 | 2.102 | 1.834 | 1.815 | -1.91 | -2 | 0.406 |
| local | 15 | 1.681 | 1.411 | 1.215 | -1.912 | -2 | 0.801 |
| local | 27 | 1.924 | 1.534 | 0.753 | -1.91 | -2 | 0.901 |
| local | 26 | 2.528 | 2.101 | 1.578 | -1.91 | -2 | 0.833 |
| local | 17 | 2.445 | 1.907 | 1.39 | -1.905 | -2 | 0.844 |
| local | 11 | 1.735 | 1.443 | 0.463 | -1.919 | -2 | 0.866 |

Table 5: Results from daily version of the Plauborg model for 20cm depth.

| scope | spesific scope | RMSE °C | MAE°C | bias °C | log(κ(model)) | digit sensitivity | $R^2$ |
|---|---|---|---|---|---|---|---|
| global | — | 2.075 | 1.622 | 0.607 | -1.269 | -2 | 0.861 |
| region | Østfold | 2.168 | 1.704 | 0.24 | -1.256 | -2 | 0.856 |
| region | Vestfold | 2.022 | 1.564 | 0.219 | -1.265 | -2 | 0.892 |
| region | Trøndelag | 1.957 | 1.528 | 1.235 | -1.26 | -2 | 0.785 |
| region | Innlandet | 2.167 | 1.712 | 0.711 | -1.261 | -2 | 0.873 |
| local | 52 | 2.418 | 1.837 | -0.636 | -1.266 | -2 | 0.664 |
| local | 41 | 1.975 | 1.587 | -0.293 | -1.26 | -2 | 0.9 |
| local | 37 | 2.206 | 1.755 | 0.373 | -1.264 | -2 | 0.873 |
| local | 118 | 2.165 | 1.697 | 1.137 | -1.267 | -2 | 0.839 |
| local | 50 | 1.395 | 1.105 | -0.046 | -1.262 | -2 | 0.938 |
| local | 42 | 2.239 | 1.75 | 0.333 | -1.262 | -2 | 0.881 |
| local | 38 | 2.42 | 1.908 | 0.667 | -1.26 | -2 | 0.845 |
| local | 30 | 1.914 | 1.519 | -0.046 | -1.267 | -2 | 0.908 |
| local | 57 | 1.978 | 1.547 | 1.108 | -1.263 | -2 | 0.85 |
| local | 39 | 1.896 | 1.455 | 1.193 | -1.26 | -2 | 0.828 |
| local | 34 | 2.142 | 1.687 | 1.535 | -1.263 | -2 | 0.475 |
| local | 15 | 1.806 | 1.428 | 1.114 | -1.269 | -2 | 0.787 |
| local | 27 | 2.063 | 1.627 | 0.396 | -1.265 | -2 | 0.895 |
| local | 26 | 2.43 | 1.937 | 1.251 | -1.259 | -2 | 0.855 |
| local | 17 | 2.273 | 1.789 | 0.9 | -1.259 | -2 | 0.873 |
| local | 11 | 1.879 | 1.504 | 0.339 | -1.254 | -2 | 0.864 |

Table 6: Results from daily version of the Plauborg model for 10cm depth.

| scope | spesific scope | RMSE °C | MAE°C | bias °C | $\log(\kappa(\text{model}))$ | digit sensitivity | $R^2$ |
|---|---|---|---|---|---|---|---|
| global | — | 4.504 | 3.473 | 2.486 | -0.796 | -1 | 0.309 |
| region | Østfold | 4.348 | 3.363 | 1.901 | -0.796 | -1 | 0.424 |
| region | Vestfold | 4.564 | 3.47 | 2.297 | -0.796 | -1 | 0.397 |
| region | Trøndelag | 4.438 | 3.508 | 3.175 | -0.796 | -1 | -0.178 |
| region | Innlandet | 4.687 | 3.567 | 2.597 | -0.796 | -1 | 0.353 |
| local | 52 | 3.556 | 2.841 | 0.559 | -0.796 | -1 | 0.604 |
| local | 41 | 4.248 | 3.286 | 1.677 | -0.796 | -1 | 0.491 |
| local | 37 | 4.754 | 3.675 | 2.174 | -0.796 | -1 | 0.391 |
| local | 118 | 4.726 | 3.654 | 3.208 | -0.796 | -1 | 0.162 |
| local | 50 | 4.048 | 3.025 | 2.207 | -0.796 | -1 | 0.434 |
| local | 42 | 4.863 | 3.741 | 2.364 | -0.796 | -1 | 0.393 |
| local | 38 | 4.832 | 3.682 | 2.601 | -0.796 | -1 | 0.308 |
| local | 30 | 4.465 | 3.433 | 2.015 | -0.796 | -1 | 0.456 |
| local | 57 | 4.655 | 3.636 | 3.153 | -0.796 | -1 | 0.125 |
| local | 39 | 4.31 | 3.39 | 3.083 | -0.796 | -1 | 0.081 |
| local | 34 | 4.581 | 3.673 | 3.469 | -0.796 | -1 | -1.822 |
| local | 15 | 4.198 | 3.342 | 3.006 | -0.796 | -1 | -0.241 |
| local | 27 | 4.672 | 3.547 | 2.535 | -0.796 | -1 | 0.415 |
| local | 26 | 5.17 | 4.009 | 3.282 | -0.796 | -1 | 0.302 |
| local | 17 | 5.042 | 3.835 | 2.913 | -0.796 | -1 | 0.335 |
| local | 11 | 3.821 | 2.924 | 1.692 | -0.796 | -1 | 0.35 |

Table 7: Results from the linear regression model for 20cm depth.

| scope | spesific scope | RMSE °C | MAE°C | bias °C | log($\kappa$(model)) | digit sensitivity | $R^2$ |
|---|---|---|---|---|---|---|---|
| global | — | 4.23 | 3.267 | 2.302 | -0.638 | -1 | 0.424 |
| region | Østfold | 4.236 | 3.28 | 2.015 | -0.638 | -1 | 0.45 |
| region | Vestfold | 4.277 | 3.26 | 2.019 | -0.638 | -1 | 0.517 |
| region | Trøndelag | 4.133 | 3.274 | 2.893 | -0.638 | -1 | 0.043 |
| region | Innlandet | 4.282 | 3.255 | 2.242 | -0.638 | -1 | 0.504 |
| local | 52 | 3.679 | 2.889 | 1.226 | -0.638 | -1 | 0.221 |
| local | 41 | 3.976 | 3.07 | 1.494 | -0.638 | -1 | 0.593 |
| local | 37 | 4.501 | 3.503 | 2.07 | -0.638 | -1 | 0.473 |
| local | 118 | 4.5 | 3.486 | 2.93 | -0.638 | -1 | 0.306 |
| local | 50 | 3.611 | 2.702 | 1.766 | -0.638 | -1 | 0.584 |
| local | 42 | 4.571 | 3.525 | 2.109 | -0.638 | -1 | 0.506 |
| local | 38 | 4.815 | 3.741 | 2.502 | -0.638 | -1 | 0.388 |
| local | 30 | 4.053 | 3.106 | 1.733 | -0.638 | -1 | 0.588 |
| local | 57 | 4.293 | 3.356 | 2.775 | -0.638 | -1 | 0.295 |
| local | 39 | 4.057 | 3.193 | 2.835 | -0.638 | -1 | 0.213 |
| local | 34 | 4.259 | 3.414 | 3.175 | -0.638 | -1 | -1.076 |
| local | 15 | 3.918 | 3.141 | 2.799 | -0.638 | -1 | -0.003 |
| local | 27 | 4.272 | 3.236 | 2.078 | -0.638 | -1 | 0.551 |
| local | 26 | 4.714 | 3.651 | 2.902 | -0.638 | -1 | 0.456 |
| local | 17 | 4.518 | 3.435 | 2.478 | -0.638 | -1 | 0.499 |
| local | 11 | 3.567 | 2.713 | 1.529 | -0.638 | -1 | 0.51 |

Table 8: Results from the linear regression model for 10cm depth.

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

Postboks 5003
NO-1432 Ås
Norway