Irrigation of Sesame

Michal Mannes

Introduction

Sesame (Sesamum indicum L.) stands as a summer crop of major importance within Israeli agriculture, particularly in the context of crop rotation, due to its extensive root system. Despite its global significance as a source of oil and nutrients, sesame remains under-researched.

Israel's semi-arid climate poses significant challenges and opportunities for agricultural innovation, particularly in the optimization of water resources. Investigating how sesame plants respond to varying irrigation levels can provide critical insights that go beyond surface-level understanding, delving into both aboveground and underground responses.

The current project will focus on the data measured in a field experiment conducted at the summer of 2023 in which sesame plots were grown under different irrigation levels. All plots were irrigated at the same time – once a week. The amount of water was determined according to Penman method, with 6 coefficients. We will analyze an hourly data obtained from dendrometers (3 in each plot) and volumetric soil moisture meters (measured in 6 depths levels). Using this data, we will try to characterize the differences between a sufficiently watered plant and a plant that is in a water stress – aboveground, and underground.

Exploratory analysis

Fig.1 - Dendrometer data structure

```
dendrometer data:
----------------------
shape: (127588, 4)
----------------------
treatment  species
0.0        ms-3       3
           ms-4       3
0.5        ms-3       3
           ms-4       3
0.7        ms-3       3
           ms-4       3
1.0        ms-3       3
           ms-4       3
1.2        ms-3       3
           ms-4       3
1.4        ms-3       3
           ms-4       3
```

As mentioned above, the structure of the dendrometer data and volumetric soil moisture data is different, although they were all measred as part of the same experiment.
When we group the dendrometer data, we can see (Fig.1) the 6 level of irrigation ("treatment") and the 2 species of sesame. Each combination has 3 dendrometers. This results in 127,588 rows of data.

As can be seen in Fig. 1, there were 6 initial treatments- "0.0", "0.5", "0.7", "1.0", "1.2" and "1.4". Each of these tratment names represent the used coefficiant number in the Penman method during their reproductive stage. Meaning, "0.0" representing no irrigation at all throughout the reproductive stage of plants, and treatment "1.4" representing 140% of advised water amount by said method. After examining the curves of dendrometers in the different irrigation levels, a noticeable difference wasn't found between the group of treatments consisting of "0.0", "0.5", "0.7" and the rest as a group. Therefore, for the purpuses of this project the treatments that we will analyze to best represent the differences above ground will be "0.0" and "1.0". Moreover, from the initial species ("ms-3" and "ms-4") the final data was used and represented only through the "ms-3" species.

From an initial observation of both treatments, we see that there are gaps in the data that will need attending before we can calculate and compare the mean curve of the two treatments. (fig.3, fig.5)
Even so, from observing the raw data, it is already possible to identify very different growth trends between the two treatments (fig.2, fig.4).

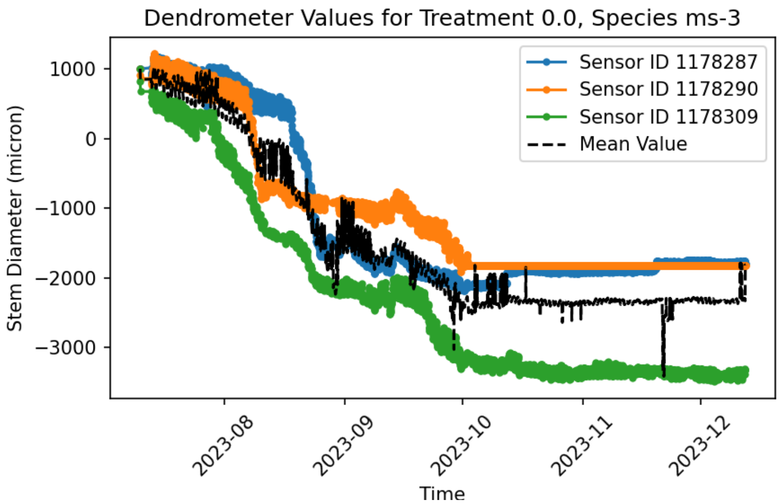

Fig.2 – Stem diameter curve of treatment "0.0"

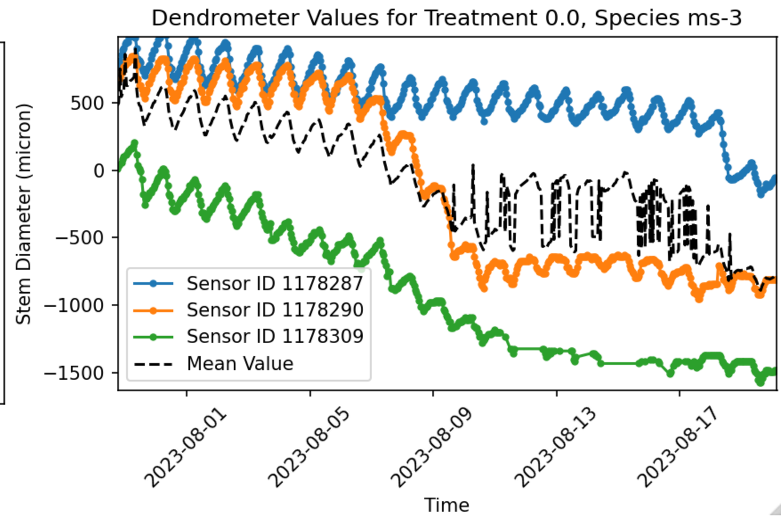

Fig.3 – Zoom in of Fig.2

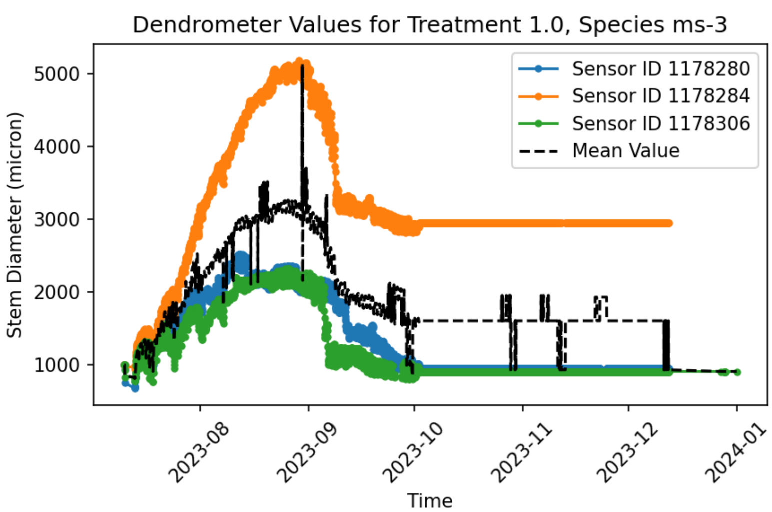

Fig.4 – Stem diameter curve of treatment "1.0"

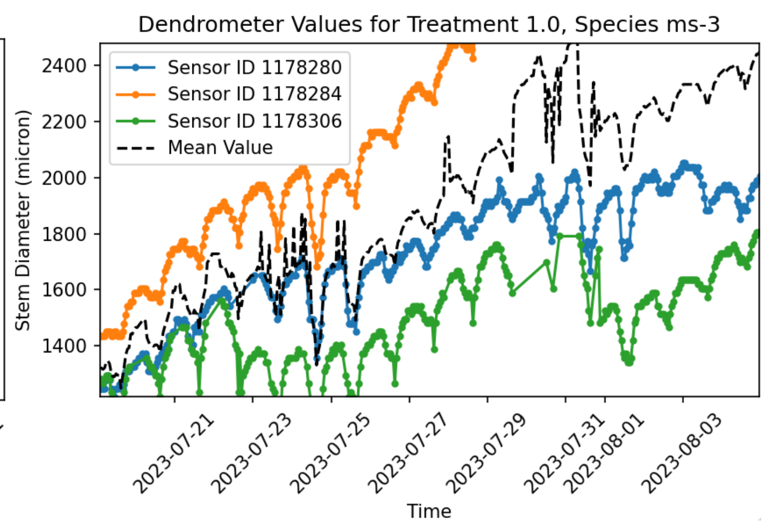

Fig.5 – Zoom in of Fig.4

Fig.6 – Volumetric soil moisture structure and shape

```
volumetric soil moisture data:
shape: (414536, 7)
```

date_time	project_id	depth_cm	measurement_type	inner_index	sensor_id	format	project_name
2023-06-01 13:00:00	878405	15	moisture	0	1174921	10.80	106
2023-06-01 13:00:00	892693	30	moisture	1	1174921	40.55	106
2023-06-01 13:00:00	878405	15	Temp	6	1174921	28.30	106
2023-06-01 13:00:00	892693	30	Temp	7	1174921	24.90	106
2023-06-01 13:00:00	892693	45	moisture	2	1174921	41.20	106

As for the volumetric soil moisture data, instead of examining differences between normal irrigation treatment and no irrigation at all, we will focus on the

common irrigation according to the Penman method (treatment “1.0”) and try to evaluate if said treatment can be enough for sesame. (cool spoiler allert – it is not!)
To acheive this, we use the data we have in various depths of soil- varying from 15 cm to 90 cm below ground surface. This can give us more insight and a better idea of the root system structure.

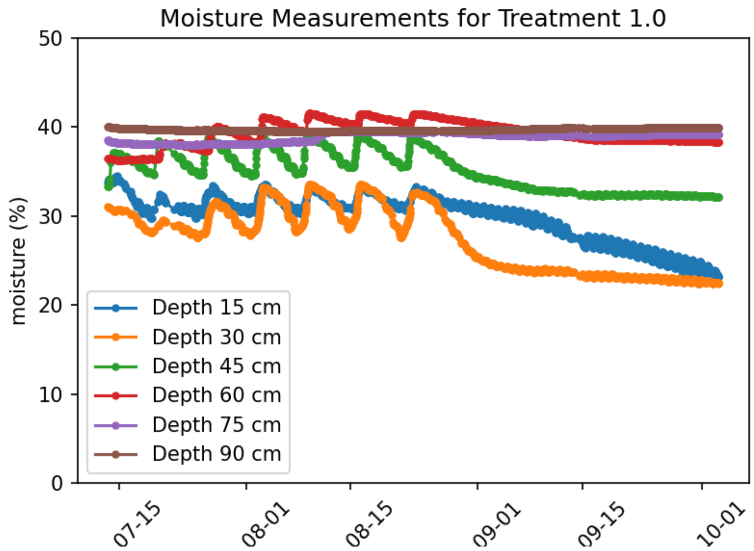

Fig.7 – Soil moisture level in 6 depths

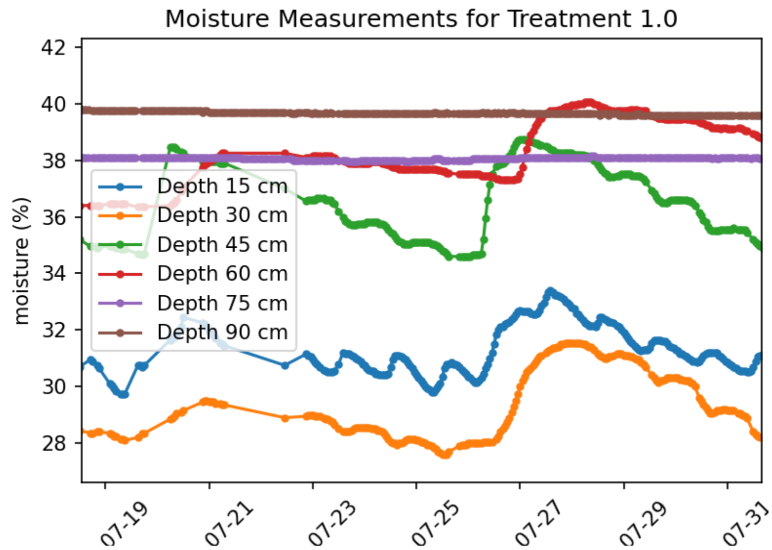

Fig.8 – Zoom in of Fig.7

Once more, the initial examination of the data shows some gaps (fig. 7, fig.8). Nevrertheless, interestingly we can see flactuations of two kinds- both visible to the naked eye: Weekly flactuations caused by the irrigation, and daily flactuations caused by water absorbtion of roots. We definitely have some interesting stuff to research here.

Detailed analysis

We will start with the basic question of – what are the differences between plants experiencinge sufficient and insufficient irrigation througout their reproductive stage?
we can expect that the plants will look very different, but can we quantify this difference?
Can we see the difference not only in the plant’s structure, but also through ploted numbers?

Since we want to quantify differences that are visible to the naked eye, we will use the dendrometers data. We want to create a single (rather then three) curve per treatment, so it will be possible to compare the treatments more afficianatly.

For that purpose, first we need to fill the gaps in the data. The method that yielded the best results is “Random Forest” (fig. 9). A technical detail - Before we filled the gaps in the data, we first needed to add rows in places where we identified missing indices (of time). Naturally, when the device failed to transmit a reading, it did not transmit a NaN, it simply did not add a record for that time.

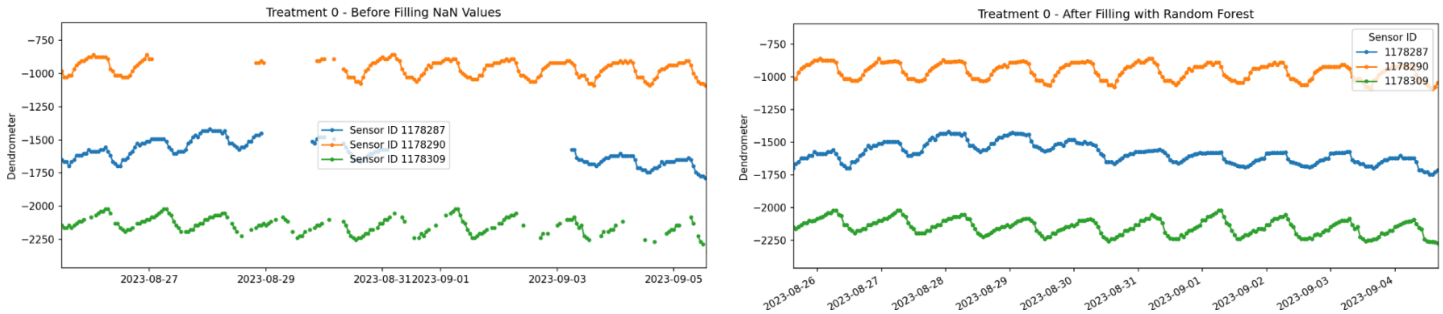

Fig.9 - Example of succesfull gap filling using Random Forest method, showing the sensetivity of data.

Once we have a continouos curve for the two treatment - 3 sensors each - we need to identify outliares. We did so using roling mean +- 2 std. We did not identify many outliers, and combined with the desire to maintain the data's sensitivity, we chose to smooth the data locally using moving average with a window of 6 hours —only where outliares were found.

Finaly, we can now create a mean curve for each treatment, which will allow us to start addressing the questions we posed.

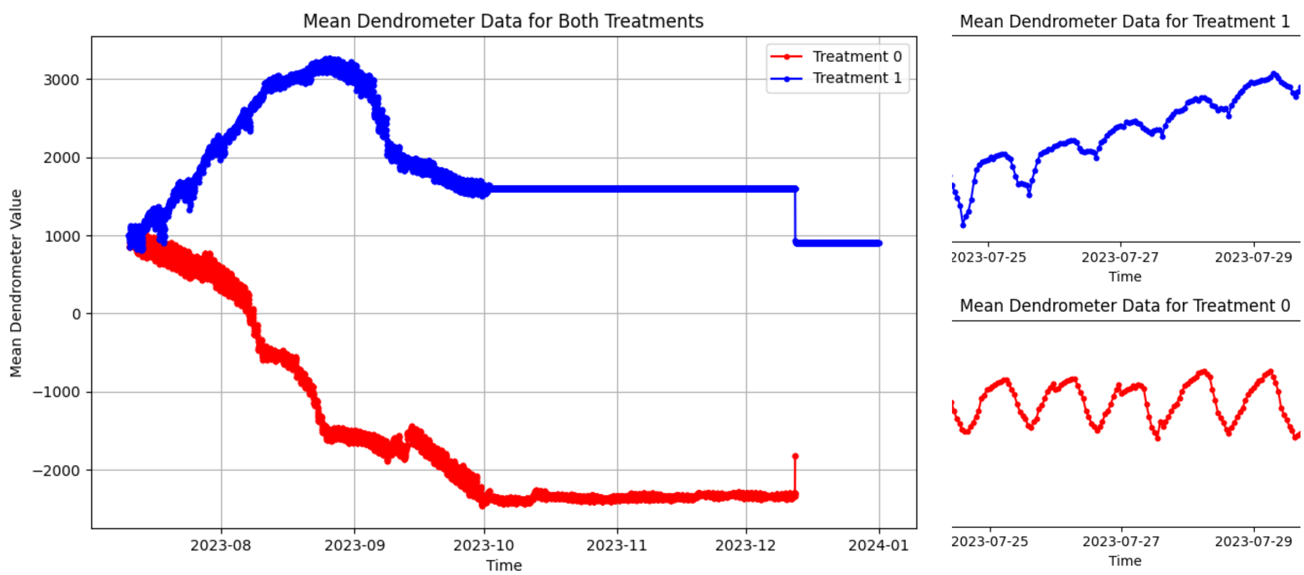

Fig.10 – Mean curve per treatment

With a vissual inspection we can determine the the averaging of the 3 sensors for each treatment went well, and the curve represent the treatment with beautiful sensetivity (fig. 10).

To address the question we posed, we will use the Maximum Daily Shrinkage ("MDS") index, that is commonly used as an indicator of water stress. MDS is calculated by subtracting the daily minimum from the daily maximum stem diameter. Higher values of MDS indicates water stress.
For better visualization we applied Savitzky-Golay filter with a window of 7 days.

From observing fig. 11, we can detect for each treatment a different trend of the MDS, actually - a different behaviour that can be quantified as different MDS patterns.
Up until slightly before september, the plants of treatment "0.0" is at greater stress then treatment "1.0" (remember – the higher the MDS index, the greater the stress). Since treatment "0.0" plants didn't get any

water, it make sence that they are in greater stress.
The trend is then reversing abruptly!
This is due to the cessation of watering as preparation for the plant's drying out prior to harvesting. We can clearly see that treatment “1.0” plants are shocked, and their stress level is rising sharply because of the sudden deterioration in water conditions, until after about two weeks when the first trend is returning.

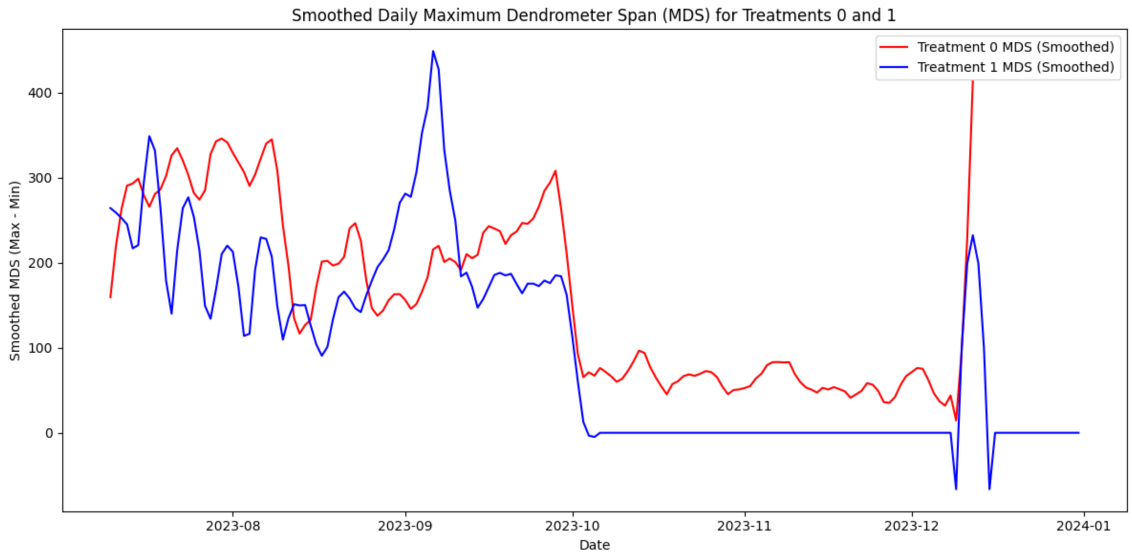

Fig.11 – MDS for treatment “0.0” and “1.0”, smoothed by savgol filter

Apparently, Treatment “1.0” seems ideal, at least compared to Treatment “0.0”. But here is the place to ask -

what's happening below the surface?
How deep does the root system get?
How does the flactuation look like below ground, and what is revield to us through that channel that we couldn’t see above ground? Is 100% according to Penman realy enough?

Initial look at the plot (fig.12) immidiatly tells us that the different depths looks different – in the moisture level, as in the amplitude of flactuations. We can also see that in 75 and 90 cm there arent any flactuation! Does that mean that there arent any roots there?
By comparing with treatment “1.4” curves (140% of Penman recommendation)(fig.13), we can guess that this is exactly the situation. apperantly, the 100% according to Penman method is only enough water to sustain root system 60 cm deep. Not very impresive, sadly.

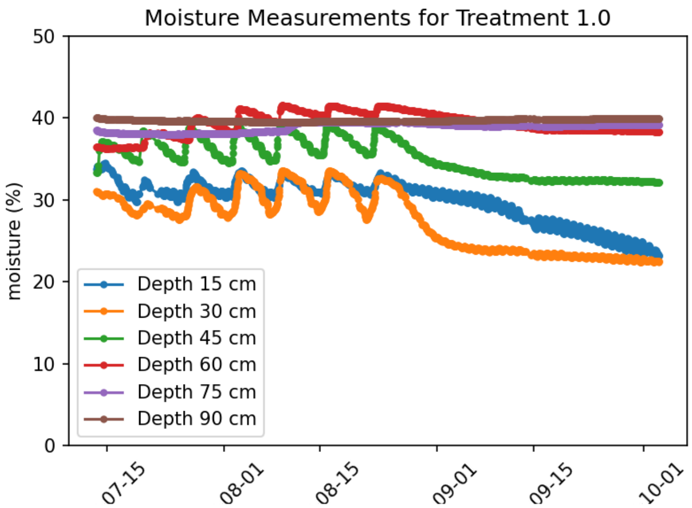

Fig.12 - Soil moisture of treatment “1.0”

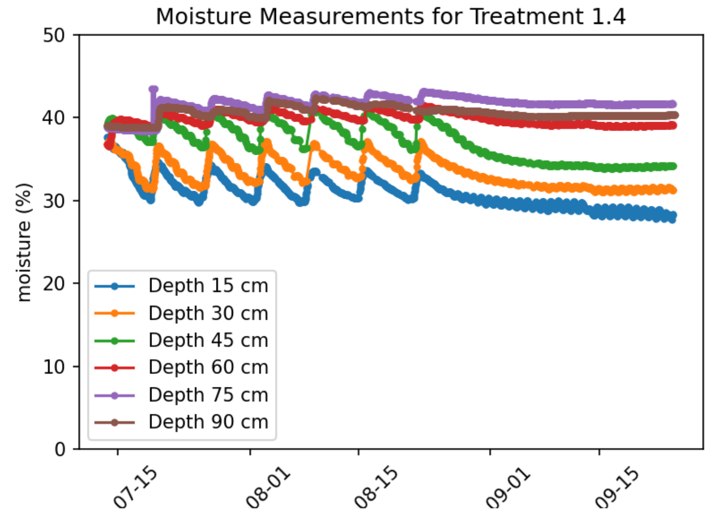

Fig.13 – Soil moisture of treatment “1.4”

Stil, we want to further investigate whats happening underground, and for that, again we need to fill the gaps in the data (using Random Forest).

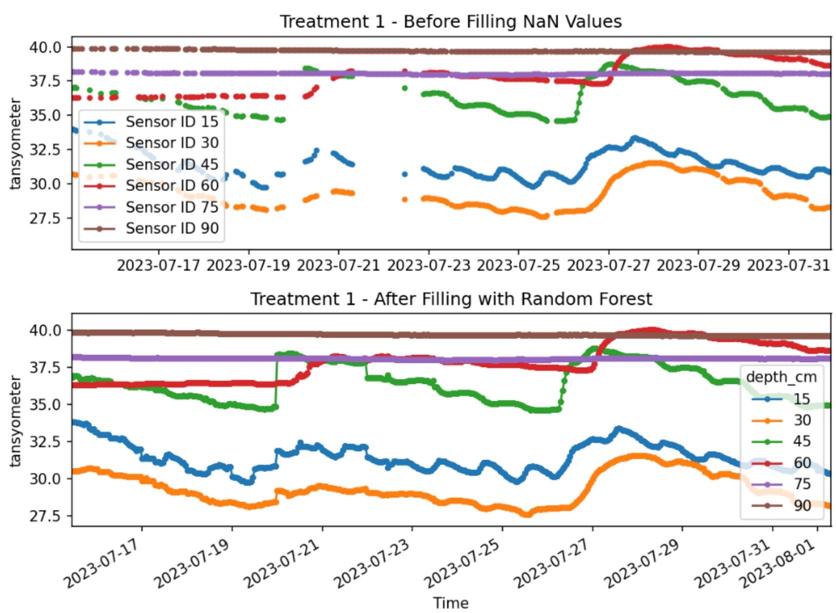

Fig.14 – Treatment “1.0” at all deapths, befor and after gap filling (zoom in)

The outcome (Fig. 14) is a bit rough, and because this time the sensitivity is less inportant that preserving the flactuation – we will smooth the hole curve using low pass filter (FFT)

We define the the cycle as daily, and smooth the frequensies below it – so we can clearly see the daily and weekly flactuations (Fig. 15).

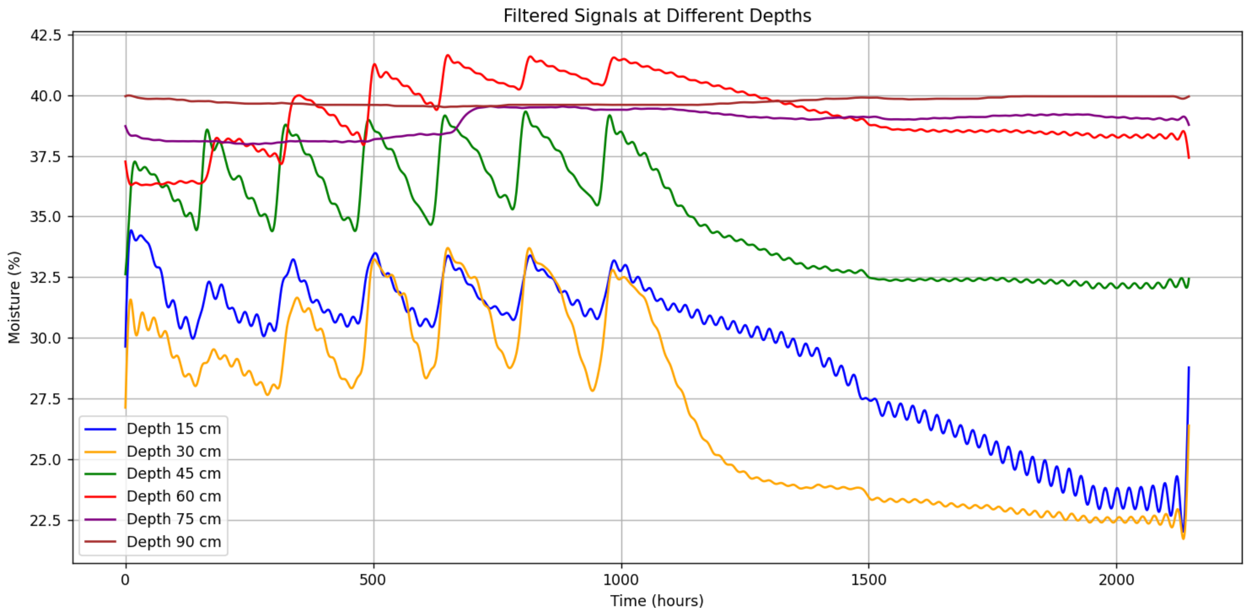

Fig.15 – Treatment "1.0" signal after low-pass filter.

We can see now that there are different behaviours in the different deapths – the amplitude of flactuations is highest in the middle (30 and 45 cm), we will adress that later on. We can also see trends (fig.16) which is pretty interesting due to the fact that they are different in each deapth.

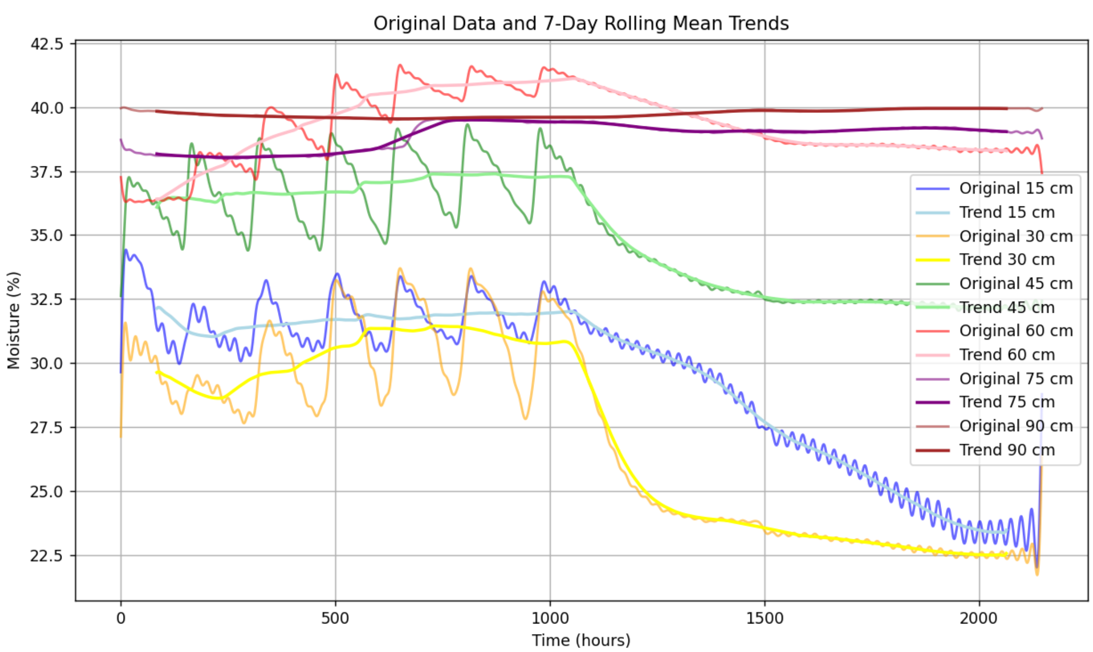

Fig.16 – Trend visualisation using 7D window for rolling mean

For example - it seems that in 60 cm depth (where the deepest roots are) there is accumulation of water durring the season – which is beneficial to the ground, although we cant see that accumulation deeper then where the roots are. This is what tipically happens after growing wheat (as a crab rotation).

So, we discovered what is the trend in each depth.

Now, we can address the flactuations, and so we need to detrend the data (fig.17), as well as leaving only the period of irrigation (first 1100 hour), so we can analyse with FFT.

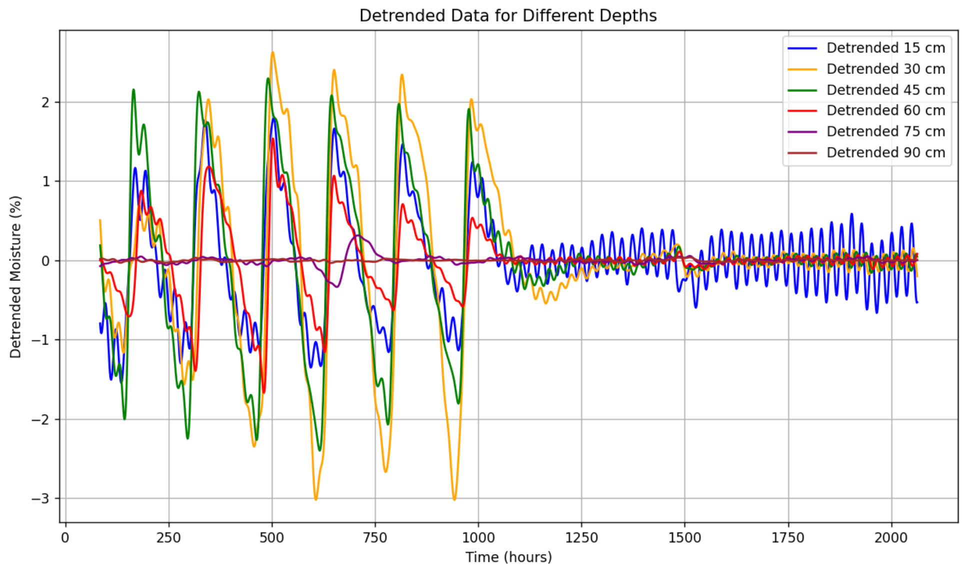

Fig.17 – Detrended data for treatment “1.0”, daily cycle

We can see (fig. 18) clearly that in 90 cm deap there is no flactuation at all – the water of the irrigation don’t get that deep, as well as the roots.
at 15 and 45 cm we can clearly see the daily flactuation (caused by roots), as well as piriodicity of 7 hours, 10.5 hours, and 14 hours. We can guess that this is in correlation to daylight time and temperature changes along day.
We can note here that the daily piriodicity is stronger in 15cm then in 45cm, probably due to evaporation of water from the top soil.

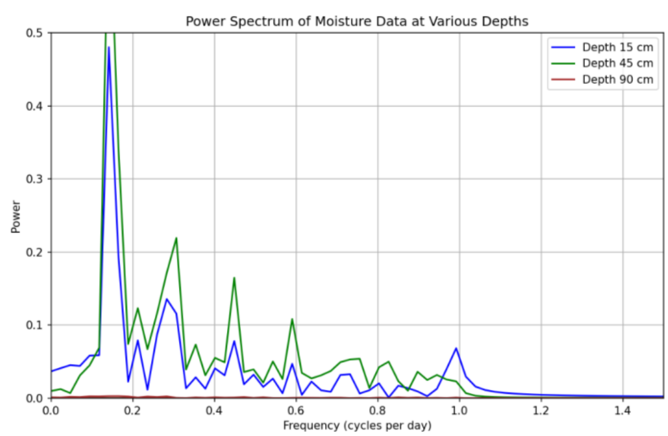

fig. 18 – FFT results, power spectrum of treatment “1.0” for 3 deapths.s

After we examined the flactuations, we can now return one step back, and validate the stationarity of the detrended unsmoothed data, focusing now on 45cm as a test case:

```
ADF Statistic: -4.784632
p-value: 0.000058
Critical Values:
	1%: -3.436
	5%: -2.864
	10%: -2.568
Reject the null hypothesis (H0), the data does not have a unit root and is stationary.
```

Fig.19 – Stationary check up, ADF test.

Once we validated the stationarity of the data we can plot the ACF and PACF of the data.
Again, we see (fig. 20) higher autocorrelation every 7 days - width of gridline.
With these parameters we can run a SARIMAX model (fig. 22), and generate a synthetic time series that will have the same parameters (fig.23).

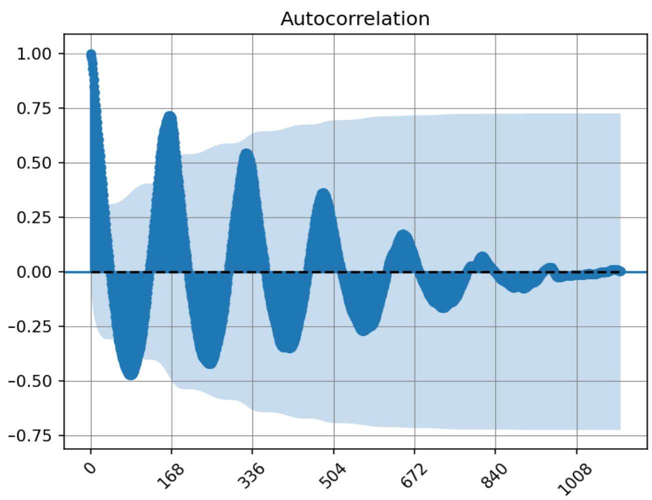

Fig.20 – ACF of 45cm, treatment "1.0"

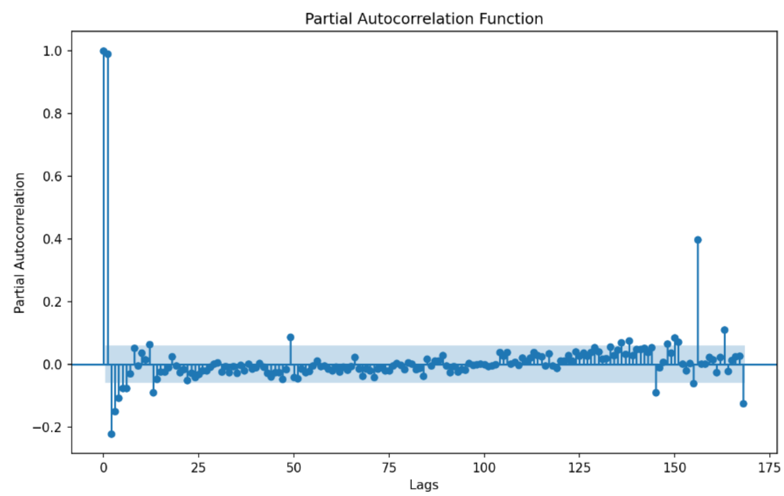

Fig.21 – PACF of 45cm, treatment "1.0"

```
                              SARIMAX Results
==============================================================================
Dep. Variable:                     45   No. Observations:                 1100
Model:                 ARIMA(1, 0, 1)   Log Likelihood                 445.563
Date:                Sun, 21 Apr 2024   AIC                           -883.126
Time:                        10:25:13   BIC                           -863.114
Sample:                    07-13-2023   HQIC                          -875.556
                         - 08-28-2023
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         36.5170      0.773     47.267      0.000      35.003      38.031
ar.L1          0.9910      0.006    173.953      0.000       0.980       1.002
ma.L1          0.2271      0.010     23.065      0.000       0.208       0.246
sigma2         0.0259      0.000    186.908      0.000       0.026       0.026
===================================================================================
Ljung-Box (L1) (Q):                   1.36   Jarque-Bera (JB):            1854617.12
Prob(Q):                              0.24   Prob(JB):                          0.00
Heteroskedasticity (H):               0.20   Skew:                             10.11
Prob(H) (two-sided):                  0.00   Kurtosis:                        203.14
===================================================================================


Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
```

Fig.22 – SARIMAX results of 45cm, treatment “1.0”

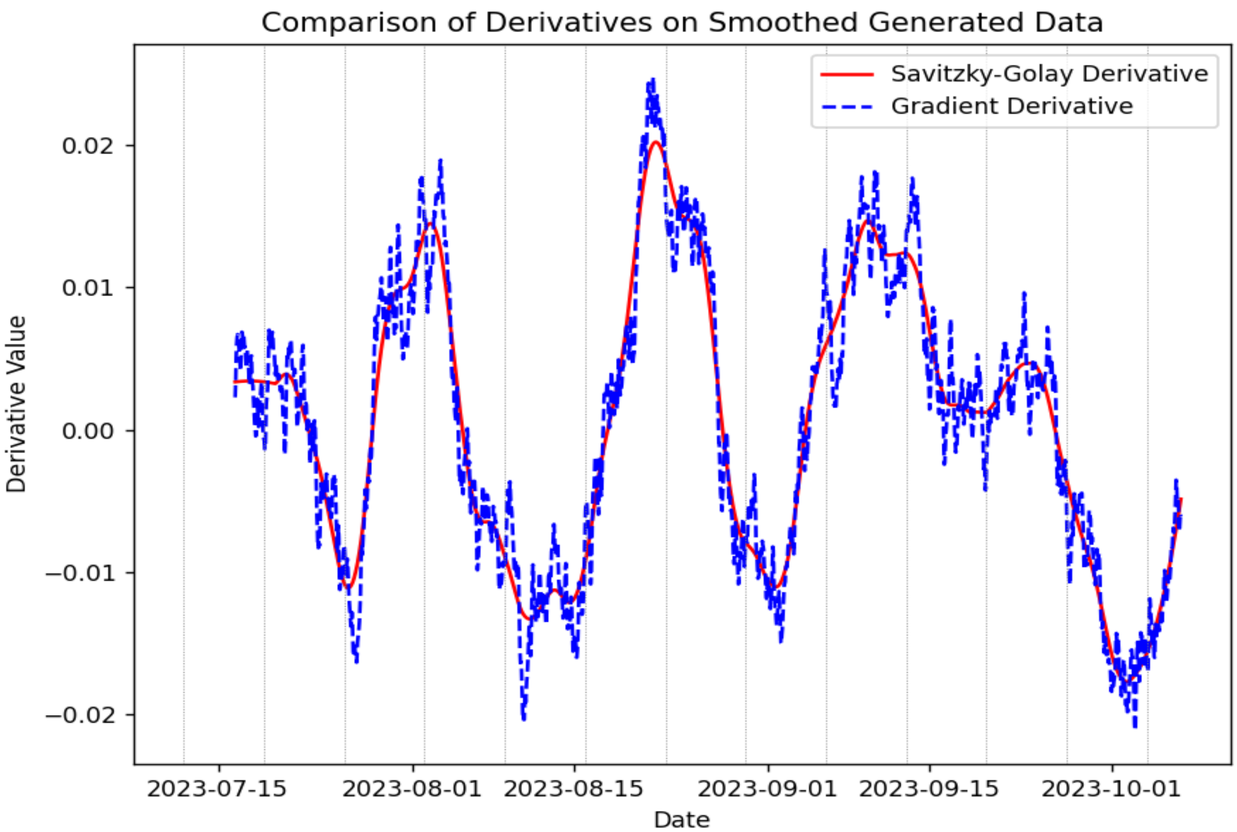

Fig.23 – Two types of derivative on the generated time series based on SARIMAX variables

The generated time series, while not so similar to our time series, still has a hint of weekly periodicity. Savitzky-Golay method did a better job smoothing things up, but the message is similar bothways.

Conclusion and pitfalls

As we probably all assumed, it is clearly visible that in terms of stress levels as measured by the MDS, irrigation treatment “1.0” is preferable to no irrigation treatment (treatment “0.0”). What is interesting to notice is that although the treatment “0.0” was not irrigated at all durring the reproductive period – we can still see the MDS changing in weekly flactuation. This points to the limitations of the experimental system, the treatments are not fully seperated, Therefore we should be very careful with the conclusions we draw from the results.

Keeping that in mind, we explore the recommended irrigation according to Penman method, treatment “1.0”, to see whether we find draw-backs that were not revealed to us by MDS. And we find to our surprise, that the root system of threatment “1.0” extends to only 60 cm below the surfes. I would expect that a vigor plant like sesame will have much dipper root system. But this is also a question of enviromental conditions, and definition of plant cultivation objectives.

It was also interesting to see the trend of humidity level durring the season – at 60 cm below serfuse there was accumilation of water, considered as a good thing for the soil.

When we dove into the flactuations of humidity levels we revealed more then just binary daily flactuations (such that used to calculate MDS), but a 7, 10, and 14 hours periodicity – that can be further investigated.

The attempt of generating a synthetic time series that will resemble the one that we analysed did not come up to good.. According to the ACF plot there are 27 significant lags, and 6 lags by the PACF plot. Generating synthetic time series by that order (6, 0, 27) yielded a nonstationary time series, probably due to the very large amount of roots. And so we generated a series based on the kind of resampled values - resample by day – to turn 27 to 1 - and resample by week – to turn 6 to 1 - so the order is (1, 0, 1). This is a major pitfall that I couldn't overcome in an ellegant way.

The main challenges I encountered revolved around focusing the research questions. I was overwhelmed by the huge volume of data and the endless available options, which made it difficult to decide what to concentrate on. What helped me was an initial review and characterization of the data.
No dought that there are a lot of more questions and petterns to be investigated and found – a lot of science to do.