Czech University of Life Sciences Prague

Faculty of Agrobiology, Food and Natural Resources

Department of Water Resources

Data filtration methods of field weighing lysimeters with high temporal resolution for soil water balance components determination

Diploma Thesis

Faisal Zeidalkilani

Thesis Supervisor: Ing. Kamila Báťková, MSc., PhD.

Declaration

I declare that the Diploma Thesis “Data filtration methods of field weighing lysimeters with high temporal resolution for soil water balance components determination “is my own work and all the sources I cited in it are listed in Bibliography.

Prague, Date of submission Signature _____

Acknowledgement

I would like to express my sincere gratitude first and foremost to my supervisor, Dr. Kamila Báťková, for her continuous guidance and support. Without her persistent help and constant motivation, the goal of this thesis would not have been realized.

I would also like to extend my gratitude to the department of water resources at CULS for their support and consultations.

Finally, I must express my very profound gratitude and love to my family and friends for providing me with unfailing support and continuous encouragement throughout my years of study.

Data filtration methods of field weighing lysimeters with high temporal resolution for soil water balance components determination

Summary

Weighing lysimeters are state of the art equipment that allow for direct measurement of the terrestrial water balance components. Their temporal and weighing resolution makes them prone to an array of errors, both systematic and random. Random errors can significantly affect lysimetric measurements, where external mechanical disturbances result in stochastic distortions that mask the magnitude of the true measurement signal. Such errors manifest as noise that requires proper filtration prior to utilizing the measured data.

This study aims to adopt a suitable filtering scheme and to optimize its parameters by tailoring them to the site's soil properties, meteorological conditions, and lysimeter specifications. An investigation period of 164 days has been chosen to ensure the process's representativeness. The lysimeter utilized for the purposes of this study is the SFL-300 (METER Group AG), located at the experimental field of CULS. The comprehensive filtering routine by Hannes et al. (2015) has been employed to filter the lysimeter measurements.

The comprehensive filtering routine consists of 5 consecutive filtration steps. The manual, threshold, and median filters address discrete measurement errors. While the subsequent smoothing and oscillation threshold filters tackle the remaining diffuse noise.

A prominent feature of the selected filtering scheme is the oscillation threshold filter. Its calculation yielded a value of 12.8 g, compared to the software's default value of 30 g. When the filtered summarized masses for both the default and calculated parameters are compared, the calculated parameters show a better fit to the raw data than the default ones.

Furthermore, in order to validate both parameters results, their cumulative precipitation values were compared to that of a reference rain gauge. The calculated parameters display slightly better correlation with the rain gauge, with a coefficient of determination R^2= 0.9993 compared to R^2=0.9990 for the default parameters. Additionally, the normalized root mean square error for the calculated parameters was 1.86%, compared to 1.98% for the default ones. While the standard deviation for the calculated parameters was 118.8 mm compared to 116.5 mm for the default parameters.

Key words: Weighing Lysimeters, Noise Filtration Methods, Oscillation Threshold Filter, Comprehensive Filtering Routine, Water Balance Components

Content

1 Introduction

Understanding the dynamics of the different components governing the soil water balance is crucial for quantifying the crop available water, and the inverse effect of plants and vegetation on evapotranspiration, interception and uptake (Johnsson et Jansson, 1991). On the basis of the water balance approach it is possible to make a quantitative evaluation of water resources and their change under the influence of man's activities (Sokolov et Chapman, 1974). Such understanding allows for the prediction of the consequences of different changes in the study medium. Furthermore, such awareness of the soil water balance components allows for measuring losses arising from drainage and evaporation in arid and semi-arid regions (Soldevilla-Martinez et al., 2014). Several modelling and measurement techniques have been developed over the years to investigate the soil water balance components, one of the most prominent instruments used are the lysimeters.

Lysimeters can yield data that is beneficial to further understand the subterranean hydrological conditions. Such data is significant for many disciplines like agriculture, hydrology, and climate sciences among several others (Hannes et al., 2015). Lysimeter measurements allow for a more explicit understanding of the continuous fluxes between the soil-plant-atmosphere (SPA) system and enable the modelling of water and groundwater recharge, solute dynamics and energy transfer in said system (Peters et al., 2014; Zupanc et al., 2012).

Lysimeters have proven to be fundamental instruments in measuring the various water balance components (Hannes et al., 2015). With the current technological advancements that enhanced these instruments' precision and resolution, their ability to mimic the actual soil conditions has become more representational. In addition, lysimeters have proven to be one of the most reliable options for measuring the soil water balance components, as their measurements are used as a reference, especially when measuring evaporation and transpiration, which are collectively investigated as evapotranspiration (Pütz et al., 2018). Lysimeter data in this regard can be used as a benchmark for the validation of data from other instrumentation (Hannes et al., 2015).

Generally speaking, a lysimeter is a terrain block of soil monolith that is held within a container through which several measurements are conducted. Lysimeters can be broadly divided into weighing and non-weighing lysimeters, where the first are used for agricultural purposes such as the determination of soil water content, estimation of evapotranspiration, precipitation and other hydraulic fluxes. The latter are used mostly for chemical analysis (Parisi et al., 2009; Pütz et al., 2018).

While some older weighing lysimeters still operate with lever arm counterbalance weighing systems (Nolz et al., 2013a), the use of high precision balance systems in more sophisticated, allowing modern lysimeters to conduct rigorous measurements of precipitated and condensed water on soil surface, including dew and hoarfrost (METER Group AG, 2013). Nonetheless, such precision comes at an expense. This increased accuracy allows for distortion in the data caused by mechanical factors such as tillage, maintenance, grass cutting, and most importantly wind.

Raw evapotranspiration and precipitation data from weighing lysimeters can rarely be used without further processing. Due to the increased precision of the lysimeters' weighing systems,

the smallest distortions such as low wind speeds or animal steps can affect the accuracy of the data. Therefore, prior to inferring relevant information from the data, adequate processing is vital. Manual processing of several thousand, or even millions of data points especially in high temporally resolved lysimeters is almost impossible and time consuming. Hence, such processing is done by means of data filters.

Lysimeter data filtration is relatively a recent topic of scientific research. And one that is still under hot debate. Weighing lysimeters with appropriate data filtering have the potential to yield the most precise and yet unbiased information regarding the hydrological fluxes (Peters et al., 2016). Therefore, several filters have been tested and new filtering schemes were developed over the years.

The aim of this thesis is to find a suitable data filter and to determine its settings for the Smart Field Lysimeter SFL-300 (UMS GmbH, now METER Group AG) that is located at the experimental field belonging to the Czech University of Life Sciences in Prague, Suchdol. It is also expected that this thesis will result in finding the optimal parameters for the filter's settings, and test the magnitude of the effect of such parameters on the filtering scheme.

2 Hypothesis and Objectives

Hypothesis: The following hypothesis is tested through this study: “Parameter settings of selected filtration procedure do not significantly affect the magnitude of resulting water balance components".

Objectives: Weighing lysimeters are basic instruments that measure soil water balance components. Such instruments are typically containers of undisturbed soil. Hence, the soil, vegetation, and climatic conditions of a given experimental area are respected. There are different sources of measurement errors and noise that can affect data interpretation and application (evapotranspiration or precipitation determination).

Therefore, the thesis is aimed at finding an applicable data filter and its settings for the Smart Field Lysimeter SFL-300 (UMS GmbH, now METER Group AG), reflecting recent developments and enabling to achieve the following requirements:

- Determination and filtration of unreliable data records,
- Bridging the data gaps appearing in the raw data file,
- Reduction of measurement noise while maintaining the fast mass changes attributable to even short precipitation events,

3 Literature review

Weighing lysimeters are state of the art instruments that measure fluxes in the mass of the soil column which are respective to events such as precipitation and evapotranspiration (Allen et al., 2011). Lysimeters provide real situation measurements of the hydraulic field conditions, allowing the interpretation of the various water balance components. Lysimeters with high precision and temporal resolution assist in the separation of precipitation events from evapotranspiration, and the determination of their fluxes across the soil-plant-atmosphere continuum (Hannes et al., 2015).

Weighing lysimeters resting on weighing cells are the product of recent developments in lysimetry, such instruments with a surface area of 1 m^2, which is the area a substantial number of the relevant studies in terms of lysimeters data processing have used, have a resolution of up to 0.01 mm (Peters et al., 2014), (or 10 g, taking into account the density of water ρ_w = ~ 1000 kg/m^3, 1 kg ≈ 1l/m^2 = 1 mm (Meißner et al., 2010). Such a resolution enables lysimeters to detect the most sensitive mass changes such as those caused by fog, dew and rime (Meissner et al., 2007). Nonetheless, such an enhancement at precision and resolution comes at a cost. Mechanical vibrations due to external factors cause distortions in the measurement data, which in turn results in measurement noise (Hannes et al., 2015; Peters et al., 2014).

Several data filtration methods have been suggested through the available literature, ranging between the moving average (MA), the more sophisticated Savitsky – Golay (SG) filter at several points, and the relatively recent Adaptive Width Adaptive Threshold (AWAT) filter by Peters et al. (2014). With each having its advantages and limitations, the determination of the most appropriate filtering scheme depends on the purpose of the evaluation, the temporal resolution, size of the data set, and the period of evaluation among other influencing factors.

3.1 Weighing Lysimeters

The weighability of a lysimeter is only one of the characteristics that can be used to classify these instruments. According to weighability, they are classified to either weighing or non-weighing lysimeters. Other classification characteristics include, size, depth, type of construction (disturbed/ undisturbed), excess water drainage mechanism and soil classification (Kohnke et Dreirelbis, 1940). In terms of weighing lysimeters, the earlier models included lever-arm counterbalance systems, which are still in use until today (Peters et al., 2014).

It's assumed that the first extensive research including lysimeters was conducted in France in the late 17th century by Phillipe de la Hire, who used the instrument to study precipitation's influence on springs' formation (Pütz et al., 2018). Lysimetry have been used to study water uptake by vegetation for over 300 years, while the use of precision lysimeters has seen development in the past 50 years (Howell et al., 1991).

The most advanced lysimeters are containers accommodating undisturbed soil within, they rest on porous plates that drain water to allow a realistic representation of the hydraulic pressure conditions of the soil (Pütz et al., 2018). And while the lysimeter is physically separated from the surrounding soil, it is covered in a way making it indistinguishable from the surrounding terrain (Khan et al., 1993). The precision of weighing lysimeters comes from the sensitivity of the scales that are placed below the lysimeter container and the seepage tank to

measure the changes in both masses that correspond to the hydraulic fluxes. With technological developments, several types of scales such as mechanical, floating, hydraulic, and electronic have been used in weighing lysimeters (Howell et al., 1991).

The advancements in lysimetry have made the instrument more sophisticated, and hence expensive. And while it is considered to give the most accurate representations to field conditions in terms of evapotranspiration and precipitation (Peters et al., 2016), and depending on the purpose of the observation, lysimeters might not always be the most feasible option. In addition, lysimeters require constant operation and maintenance and continuous data processing efforts, which are factors in considering its installation. In 2005, around 2400 lysimeter units installed in Europe were surveyed, out of which 400 were weighing lysimeters (Hannes et al., 2015). In 2006, the number of lysimeters surveyed in Europe was around 2450, 1260 of which are located in Germany. Of the total number, 63% are field lysimeters and 21% are grassland lysimeters (Harsch et al., 2009). Unfortunately, within the scope of the literature reviewed, an updated number has not been found. The importance of such a number of units arises in assisting in regional and even continental studies regarding hydroclimatic conditions such as the 2003 drought in Europe, and also the study of large scale evaporation drivers and characteristics (Hannes et al., 2015).

3.1.1 Operation and Working Principle

While the working principle behind the weighing lysimeter is straight forward, it indirectly measures the hydraulic fluxes, out of which the soil water balance components can be determined (Hannes et al., 2015). Such determination is done through continuous measurement of the mass changes of the lysimeter soil column and the drainage tank. A positive mass change denotes a precipitation event, while a negative change or a decrease in measured mass represents an evapotranspiration event (Nolz et al., 2013b; Peters et al., 2017).

Weighing lysimeters function on the basis that water from precipitation events such as rain, snow, and even dew will register an increase in the lysimeter mass, water leaving the lower boundary will percolate through the soil column and is collected in the seepage tank (Meissner et al., 2007). Modern lysimeters, with the use of sensors, try to mimic the adjoining soil conditions as accurately as possible. Hence, an accuracy enhancing mechanism is done by measuring the soil water content in the lysimeter and soil, and then pumping water back to the lysimeter from the seepage tank by the use of suction cups if the conditions in the lysimeter were drier than those of the surrounding soil (Doležal et al., 2018; Hannes et al., 2015). Noting that such a capability in terms of bi-directional water pumping is not available in all weighing lysimeters (Schrader et al., 2013).

Such sensors assist lysimeters to correctly reflect the surrounding field conditions. Lysimeters are usually equipped with various sensors at several depths as seen in figure 1. These sensors measure the characteristics of the adjacent soil such as soil moisture content, temperature, matric potential, chemical parameters, and electrical conductivity(Gebler et al., 2015; Hannes et al., 2015; Pütz et al., 2018).

Major advancements occurred over the past half century in terms of the registration and handling of weighing lysimeter data (Howell et al., 1991). The enhancements of data loggers and control units allowed for the magnification of the temporal resolution of weighing

lysimeters, where nowadays several studies conducted measured changes in lysimeter data by one minute intervals (Doležal et al., 2018; Hannes et al., 2015; Peters et al., 2014; Zupanc et al., 2012).

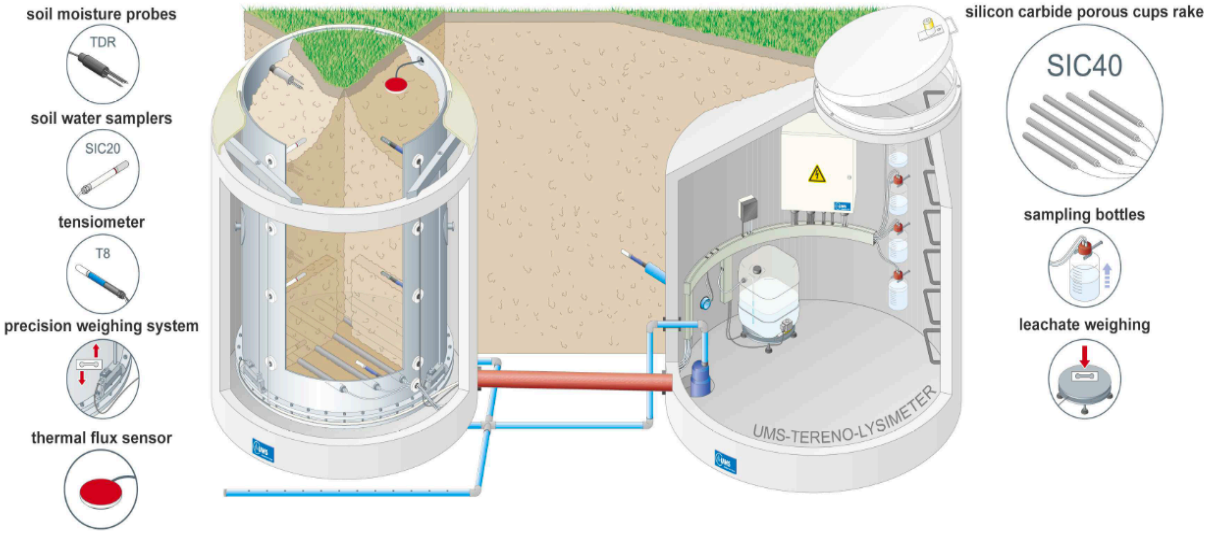

Figure 1: Cross section example of a smart field weighing lysimeter and the seepage tank with examples of sensors used such as soil moisture content, matric potential sensors (tensiometers), temperature, and soil thermal flux sensors (Gebler et al., 2015).

3.2 Water Balance Components Calculation

The determination of the net water balance, that is the sum of precipitation (P) and evapotranspiration (ET) is quite robust. However, the process of distinguishing between the fluxes of such P and ET events is mostly problematic due to the stochastic fluctuations resulting from external factors in the form of mechanical vibrations, affecting the weight measurements in specific time intervals (Schrader et al., 2013).

The water balance equation can be expanded to include its various components such as dew, biomass changes, and irrigation water. Nonetheless, the following denotes the basic water balance equation (Zupanc et al., 2012);

$$P - ET - O - \Delta S = 0 \quad (1)$$

Where (P) is precipitation, (ET) portrays both evaporation and transpiration processes, which are collectively measured as Evapotranspiration (Nolz, 2016), (O) is outflow, and (ΔS) represents the change of water storage in the soil.

The high precision and temporal resolution of weighing lysimeters allow for the determination of the different components of the water balance. In addition, the sensors fitted to the instrument also help differentiate between the components (Pütz et al., 2018);

$$\Delta W = P + I + D - (A + SW + B + ET) + CR \quad (2)$$

Here (ΔW) stands for the change in the lysimeter mass, (P) is precipitation, (I) is irrigation water, (D) is dew; these processes have a positive mass change effect to the lysimeter. On the

other hand, (A) is runoff, (SW) is seepage water, (B) represents biomass changes, ET is evapotranspiration, and finally CR is capillary rise (Pütz et al., 2018). Since lysimeters contain the complete soil– plant system, interception is part of the precipitation term if lysimeters are vegetated (Hoffmann et al., 2016).

In most studies, such as the one undertaken here, such comprehensive and specific separation of the water balance components is not of much concern. Rather, it is usually of interest to only differentiate precipitation events (P) from evapotranspiration (ET), or simply mass increases from decreases. Such events must be analyzed over short periods, hence the required high temporal resolution of weighing lysimeters, where the positive mass changes of the lysimeter and the cumulative mass of the seepage tank are regarded as precipitation and negative ones as evapotranspiration (Herbrich et Gerke, 2016). And since the aforementioned events are investigated over relatively short periods, it is considered that they do not happen simultaneously (Peters et al., 2014).

Since most literature is interested in the determination of the water balance components to derive the evapotranspiration for several purposes, the water balance can be shortened as the following (Nolz et al., 2013b; Nolz, 2016);

$$ET = P + I - SW \pm \Delta W \tag{3}$$

Where (SW) is the seepage water, and the others are denoted as in the first and second equations. The equation can be rearranged as follows (Nolz et al., 2013a; Von Unold et Fank, 2008);

$$P + I - ET - SW \pm \Delta W = 0 \tag{4}$$

3.3 Measurement Noise and Errors

Despite having their own sources of errors, lysimeters largely avoid errors made by traditional measurement systems, such as the wind error of Hellmann rain gauges or the island error of class-A pans (Schrader et al., 2013). Measurement errors can have various sources, whether systematic or random. However, out of the various sources, what is of concern for the purposes of this thesis is the data noise. Such noise leads to fundamental impacts on the interpretation of the data, while also reducing the general accuracy (Zhu et Wu, 2004). Noise can be thought of as what remains after removing the true signal from a data point. Where in the case of a lysimeter data point, the signal is the actual mass change due to an evapotranspiration or precipitation event, and the noise is what remains due to distortions and errors such as mechanical vibrations, data logging errors, and other.

3.3.1 Measurement Errors

Lysimeters are always directly exposed to environmental conditions and therefore prone to multiple error sources (Hannes et al., 2015). Such data errors are typically specific to the datasets measured, they are dependent on the magnitude, duration, and frequency of such erroneous events.

Errors can result from the different components of the lysimeter (systematic errors), and the processes of handling the data (random errors). It's weighing mechanism, like other measurement systems, is subject to both random and systematic errors (Nolz, 2016). While systematic errors do exist, they can be avoided or decreased by adequate calibration and inserting correction coefficients with the use of, for example, meteorological data. In addition, systematic errors are substantially eliminated due to the method of installation of lysimeters at ground surface, while random errors are reduced due to the relatively large size of lysimeters compared to other devices and instruments (Peters et al., 2016). What is of concern for the purposes of this thesis is the portion relating to noise (noise error) resulting from mechanical and external environmental disturbances such as wind, and the (mixing error) when a part evapotranspiration and precipitation fluxes are mixed due to a reduction in the temporal resolution by for example increasing the averaging window (Hannes et al., 2015).

3.3.2 Sources of Noise

With the enhanced precision of weighing lysimeters and the ability to register and examine data at high temporal resolutions comes greater sensitivity to external disturbances, these disturbances are depicted as correlated data and noise (Herbrich et Gerke, 2016).

Noise in general is an inherent random error or set of errors that are indistinguishable and overlaying actual, real measured data. Noise originates from different sources, such as at the signal processing of the data acquisition system, or later on during data management and filtering processes (Marek et al., 2014).

Such disturbances are in the form of mechanical vibrations caused by animal or human induced weight when stepping or standing on the lysimeter monolith, the maintenance of the lysimeter including the regular emptying of the seepage bottle, errors in data transfer, biomass changes, and most importantly noise caused by wind of various speeds, durations and magnitudes (Gebler et al., 2015; Hannes et al., 2015; Hoffmann et al., 2016; Schrader et al., 2013). Wind for instance has a significant influence on the noise in the measurements; as the wind speed and gusts vary in time, so does the magnitude of the noise in the raw data (Peters et al., 2014). Nolz et al. (2013a) reported that wind has a substantial impact that it could lower the accuracy of the system by up to 3 times.

Therefore, such data noise is crucial to be filtered before inferring any information about the hydraulic conditions in the site. Noise filtration is a strenuous process, but it is necessary to separate such noise from true evapotranspiration and precipitation events and may lead to either overestimating or underestimating their fluxes (Hoffmann et al., 2016; Schrader et al., 2013).

3.4 Chronological Review of Noise Filtration Methods Development

While lysimeters serve as an alternative, and to some extent a reference, to rain gauges for the determination of precipitation and as a benchmark for actual evapotranspiration (ET_a) quantification in methods such as eddy covariance or the Bowen ratio energy balance (BREB) (Gebler et al., 2015; Hoffmann et al., 2016), they also provide the most accurate data when examining a relatively small surface area (Vaughan et Ayars, 2009). Furthermore, recent

literature views lysimeters as reference instruments for evapotranspiration measurement, by which results from other methods are authenticated (Hannes et al., 2015).

Nonetheless, the raw data from lysimeters cannot give accurate representations of the hydraulic fluxes, as noise needs to be separated from true mass changes before inferring any information about the existent field conditions (Hoffmann et al., 2016; Peters et al., 2017; Schrader et al., 2013). This issue arises from the enhanced temporal resolution and weighing sensitivity, making the noise not only visible in a substantial magnitude, but also temporally variable (Peters et al., 2017).

It is apparent that finding an accepted or even an adequate filtering scheme is still proving to be a challenging task (Gebler et al., 2015; Schrader et al., 2013), and the investigation of comprehensive filtering routines form the premise on which several recent studies have been published (Hannes et al., 2015; Nolz, 2016).

3.4.1 Data Filtration Through Basic Smoothing Functions

There have been several publications dealing with noise filtration from lysimeter measurements. Vaughan et al. (2007) applied a filtering scheme on a set of two-year lysimeter data with 1- hour resolution in California, USA. They eliminated questionable data and measurements registered on rainy days to evaluate and improve evapotranspiration measurement through lysimeters. This was done by applying a seven-point Savitsky – Golay (SG) quadratic time derivative filter, and comparing it to hourly meteorological data of evapotranspiration rate by using Root Mean Square Error (RMSE). Vaughan et al. (2007) developed a computer software to validate the automated SG filter output with the lysimeter data to manually ensure that no noise was included in the filtering process that would affect the validity of the procedure. While the filter helped in smoothing the data for the purposes of ET determination, there still were some discrepancies between the measured and meteorological data in the form of a non-linear relationship.

Vaughan et Ayars. (2009) implemented noise smoothing methods to lysimeters' noise, particularly in terms of wind loadings, in the same vicinity in California. Their method included the application of a simple arithmetic averaging using means and medians of the data points and addressing the noise with a 7-, 11-, and 15- point SG filter to the data at 1-minute temporal resolution in order to smoothen the noise that resulted in a reduction of evapotranspiration data quality. Their study confirmed the relation of wind loadings to noise in lysimeter data, in addition to deducing that the use of arithmetic mean rather than median is more preferable.

On the other hand, it was argued that the 7-point SG filter, although fit for the reduction of high frequency noise, it is not as effective in eliminating outliers represented by low frequencies but high magnitudes (Huang et al., 2012).

Recent studies have been focused on the improvement of data processing of modern lysimeters (Nolz, 2016). The first study to evaluate filtering schemes for high temporal resolution lysimeters with synthetic data was presented by Schrader et al. (2013). In their publication, the authors tried a different approach through evaluating different smoothing and filtering processes using both real measurement data and synthetic ones. The synthetic data were derived from numerical models using a simulation software in order to produce errorless data which are then intentionally disturbed in order to test the performance of different filtering

algorithms. The authors provided 2 steps to smoothing the data, starting with the correction of systematic errors and outliers, then dealing with random errors by applying smoothing algorithms such as the moving average (MA) or the SG filter. It was also argued that despite the application of various filtering procedures, the noise cannot be fully extinguished. Moreover, Schrader et al. (2013), based on Fank's (2013) temporally adaptive threshold, included the application a specified threshold value where for any measurement to be registered as precipitation or evapotranspiration event, the threshold value had to be exceeded. The publication also demonstrated the effects of the selection of an averaging window on the smoothing of the data noise, and showed that large averaging windows, typically over an hour, can affect the integrity of the filtering scheme.

With more research being conducted on the filtering methods of data noise, it became evident that filtering procedures such as the simple moving average alone, although could produce acceptable smoothing in some cases, are not sufficient especially with severely noisy data (Nolz et al., 2013b). Through their second paper, Nolz et al. (2013b), studied the effects of two smoothing techniques on noisy lysimeter data. The first, a basic piecewise sigmoid function, was easy to fit and gave satisfactory results over diurnal periodicities without rainfall and their evapotranspiration variations. The second, a polynomial spline function, was more practical as it can be applied for days with precipitation events and can be extended to process data from longer periods. Furthermore, when evaluated, the spline smoothing returned lower RMSE values compared to the sigmoid function.

Marek et al. (2014), published a comprehensive work on post processing techniques, they discussed sources of error, data processing and Quality Assurance / Quality Control. The paper evaluated the importance of the length of the averaging window and its effect on data smoothing, while also arguing that relatively long term mean values can produce errors of their own and render the data obscure. Moreover, the research compared three simple moving average (MA) functions to three SG functions using 5-min averaged data to evaluate evapotranspiration. The study showed the effects of increasing the points of the function on the overestimation or underestimation of the events.

3.4.2 The Development of Advanced Filtering Routines (multi-step)

As the importance of the averaging window width and the threshold value became more evident (Schrader et al., 2013), more comprehensive filtering routines that included several steps were being developed. Peters et al. (2014), established the Adaptive Window Adaptive Threshold (AWAT) filter. The principle of the approach is to apply a certain smoothing routine with a defined averaging window, denoted as (w), and certain measurement threshold (δ). The reason behind adapting the window width and threshold value is due to the fact that in case of a fixed small averaging window, noise might be translated into real measurements. While if the window is too large, rapid changes might be filtered out. This principle is identical for the application of a threshold value as well. Moreover, when the authors compared the new developed AWAT filtering routine on a data set of 4.5 months to the (SG) filter and the moving average (MA), they discovered that only the AWAT filter provided acceptable results for all data in various events (wind, rain, etc.).

The AWAT method provided the basis for subsequent research regarding high temporally resolved data filtration of weighing lysimeters (Gebler et al., 2015; Hannes et al., 2015; Hoffmann et al., 2016). Gebler et al. (2015) applied the AWAT filter to a set of data from the measurements of 6 weighing lysimeters. The process included 3 consecutive steps in order to calculate hourly precipitation and evapotranspiration;

1- Eliminating outliers by automatic threshold filter
2- Applying the AWAT filter to 1-minute temporally resolved datasets
3- Calculation of hourly precipitation and evapotranspiration

Hannes et al. (2015) proposed a comprehensive multi step filtration routine to winnow data errors ranging from large disturbances in the raw data to noise caused by factors such as wind. The study also discussed the effects of averaging windows and thresholds on the filtering steps. Furthermore, an important outcome of this study was the development of the Synchro filter, which is used to deal with data sets from several lysimeters, assisting in studying the data and filter behavior with regard to multiple lysimeters, and assisting in reducing filtering errors. The study made use of 2 months measurements from 18 lysimeters located in Bad Lauchstadt, Germany. Moreover, the authors developed a computer software that applies the suggested filtering routine and is available for download as a supplement to their publication (Hannes et al., 2015; Vogel, 2015).

It seems that the innovative idea behind the AWAT filter and its adaptive function is based on the work of Fank (2013), who not only portrayed the effects of different averaging windows, but was the first to suggest applying temporally adaptive threshold values (Hannes et al., 2015).

3.5 Noise Filtration Methods

3.5.1 Simple Moving Average

A moving average is a type of mathematical convolution, creating a new time series from subsequent values of another time series. Thereby, the new one is based on the averaged values of the previous (Butler, 2008). If smoothing a data set consisting of n values is considered, the simple moving average has the following basic equation (Herbrich et Gerke, 2016; Nolz et al., 2013b);

$$X_t = \frac{1}{n} \sum_{i=0}^{n-1} x_{t-i} \tag{5}$$

Where X_t is the smoothed value, n is the number of data points, x_t is a data point in the set.

Therefore, applying a point MA function (e.g. 3-point and 5-point) would take the form of the following equations consecutively (Marek et al., 2014);

$$X_i = \frac{1}{3}\,(x_{i-1} + x_i + x_{i+1}) \tag{6}$$

$$X_i = \frac{1}{5}\,(x_{i-2} + x_{i-1} + x_i + x_{i+1} + x_{i+2}) \tag{7}$$

The reason behind using an odd number for the MA points is to ensure that the average is centered around the middle value in the data set being observed (Butler, 2008). After choosing a fixed number of points, taking measurements' sum and dividing it by their count as aforementioned, the point at one end of the set is dropped and a point at the other end is added, then the process is repeated (Savitzky et Golay, 1964).

Nonetheless, the MA faces some disadvantages, mainly due to the fact that the last observation needs to be dropped and hence lagging it. Therefore, the MA can be disproportionally affected by old observations dropping out of the average, and some effects of such data points do not manifest (Ostertagová et Ostertag, 2016).

3.5.2 Savitzky – Golay Filter

The Savitzky – Golay method, named after its designers, is a convolution function for fitting data sets through the use of the least squares polynomial approximation technique (Peters et al., 2014; Savitzky et Golay, 1964). The method hence provides an estimate of the function value, or its derivatives, at the central point of a measurement (Vaughan et Ayars, 2009). This means that in principle of application, the SG filter behaves in an analogous manner to the simple moving average, in terms of treating a central measurement with respect to adjoining ones.

Hence, to smoothen or differentiate a data set, each value will have to be replaced with one obtained from fitting an n order polynomial to the 2m+1 neighboring points. The method also depends on using convolution weights that were tabulated by Savitzky and Golay (Gorry, 1991).

Therefore, a 5-point quadratic polynomial S-G convolution function Y_i, for y_i raw data set would look as follows (Marek et al., 2014):

$$Y_i = \frac{1}{35}\,(-3y_{i-2} + 12y_{i-1} + 17y_i + 12y_{i+1} - 3y_{i+2}) \tag{8}$$

3.5.3 Adaptive Window Adaptive Threshold Filter

The AWAT filter developed by Peters et al. (2014) is relatively recent. The principle behind this filter is to firstly apply a smoothing routine with a particular averaging window (w) and to account for measurement accuracy by applying a certain threshold (δ). This ensures that only significant mass changes are considered if they exceed a certain threshold value (δ). The adaptivity function of this method is its most prominent feature, its importance comes from the variability of the noise that needs to be smoothed, such as strong signal events over short periods, or light but constant noise over longer periods. Therefore, using fixed window and

threshold values rather than adaptive ones would lead to data loss due to discarding some rapid but intense signals, or introducing noise to the data set as real measurements.

As was suggested by Fank (2013) and what was reaffirmed by Schrader et al. (2013), in filtering routines like the AWAT and subsequent multi-step filtering routines, the smoothing process is done by first applying a smoothing function such as the SG filter, or the MA with a specific window (w). Which is subsequently followed by the introduction of the threshold (δ) value, where any changes smaller than the predefined (δ) value is discarded (Peters et al., 2016).

The AWAT filter and its operating principle will be described elaborately over this section. Such an explanation is necessary as the AWAT filter shows competitive advantages to be chosen as the filtering scheme for the purposes of this thesis compared to the comprehensive filtering routine which is to be discussed over the next sections. Hence, it is of value to describe its operating principles and define its strengths and weaknesses. Furthermore, the research of Peters et al. (2014) contains significant comparison in terms of smoothing functions, which are important to understand when choosing the appropriate filtering routine.

This filtering routine is sensitively dependent on the choice of both the averaging window width (w) and the threshold value (δ). Such values need to be valid for a variety of events and data noise. A rapid but heavy precipitation event would require a narrow averaging window, while a smooth evapotranspiration event with moderate wind speeds (low noise) would require a larger window. Moreover, periods with constant evapotranspiration or precipitation events with relatively low wind speeds and no further disturbances would generate data that does not require filtering or processing. Applying filters to such data may produce unwanted losses to accuracy.

The AWAT filter defines firstly a maximum window width (w_{max}). Setting such a temporal value is done manually. Within such a defined window, the signal strength and noise for each data point are identified by fitting a moving polynomial function. Considering having a data point, i, a polynomial function of the order (k) is fitted to its neighboring points within a defined maximum window, (w_{max}). The order of the polynomial should be high enough to guarantee that it can describe the data in the time window reasonably well. However, it should be low enough to avoid the noise being described by the polynomial as well (Peters et al., 2014).

Having a time interval between $t_{i-w_{max}/2}$ and $t_{i+w_{max}/2}$, the polynomial for data point i, $Y_i(t)$, is given with;

$$Y_i(t) = \sum_{j=0}^{j=k} a_j \, t^j \tag{9}$$

This fitting function is not ideal to rely on with smoothing the data, it is rather used to infer information regarding the strength of the signal and noise. This information is of particular importance in order to define the window width (w) with which a smoothing filter is applied, as (w) is a function of the signal strength. Nevertheless, using a spline function with k knots is better than the polynomial function, but it would require non-linear regression which is consuming for large data sets (Peters et al., 2014).

Following, the smoothing routine is applied with a defined (w), this filter could be an MA or an SG filter as both have the ability to filter the data to a decent degree (Peters et al., 2014). Nonetheless, their study has relied on the use of MA since it is easier to apply and does not consume much computational power while still yielding a similar satisfactory output. When

testing the AWAT filter against the MA and SG, it was observed that the SG filter does not smooth heavy precipitation data like the MA. In addition, it was noticed that SG filter tends to oscillate, resulting in overestimation of precipitation and evapotranspiration events. Furthermore, the study concluded that in the cases of strong wind events, the MA and the AWAT filter yield better results than the SG filter.

Since (w) is a function of the signal strength, (w) is dynamically dependent on it. Therefore, the window width allowable boundaries need to be defined, which are w_{min} and w_{max}. it is recommended for w_{min} to use the temporal resolution of the data (Peters et al., 2014).

The final step in the AWAT filter is to apply an adaptive threshold (δ). The reason for the use of (δ) in the culminating step is that while the smoothing function does remove a majority of the noise, there remains some noise which the application of a threshold value aims to eliminate. Hence, (δ) aims to prevent the overestimation of any event as a result of the interpretation of noise as real mass changes in the lysimeter. Moreover, to preserve the threshold's adaptive function, the maximum and minimum parameters need to be delineated, namely δ_{min} and δ_{max}. Where δ_{min} is suggested to be equivalent to the scale's lowest resolution in mm equivalent (Peters et al., 2014).

Therefore, the variability of (w) and (δ) is based on the signal strength. Strong signals require small (w), while the opposite is true. Equivalently, a high value of (δ) is needed to exclude high noise, and a low (δ) for low noise.

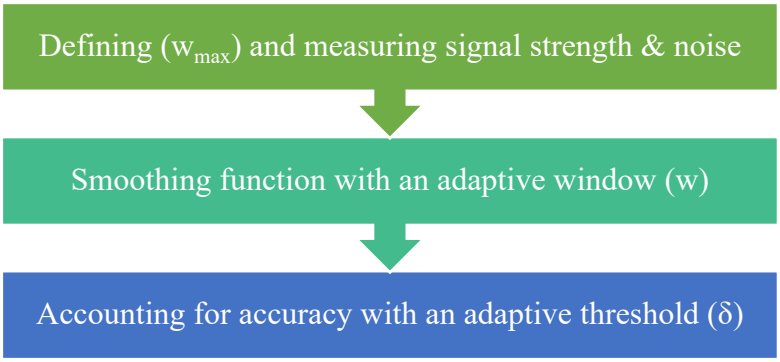

Figure 2: Schematic overview of the AWAT filter process(Peters et al., 2014).

3.5.3.1 An Experiment with the AWAT Filter

To validate the AWAT filter's performance against existing filtering schemes, Peters et al. (2014) tested the AWAT with its adaptive w and δ was tested against MA and SG filters with fixed (w) and (δ). 18 variations were constructed as a result of testing 3 (w) values for the three filters against 2 (δ) values. The (w) values tested were 11, 31, and 61 minutes, while the (δ) values were 0.081and 0.24. For the (δ) values, 0.081 resembles the minimum resolution of the lysimeter used in the experiment in mm equivalent, hence 0.081 can be denoted as (δ_{min}).

The test was concentrated on 3 particular days, each with a different distinctive event. A day with heavy precipitation, a day with smooth evapotranspiration, and a day with high wind speeds and low evapotranspiration. The filters were tested firstly by using an adaptive (w) but a fixed (δ_{min}). The 11 min averaging window for the MA and SG filters returned decent results

for the heavy precipitation and the smooth evapotranspiration events. Nonetheless, the strong wind data were interpreted as a series of small precipitation and evapotranspiration events (due to the fluctuation of the mass measurements). From meteorological measurements, it was determined that there was no precipitation registered on that day. Higher values for (w), e.g. 31 or 61, returned less noise but at the cost of the data accuracy for the heavy precipitation event.

Testing the AWAT filter with varying (w) and (δ) yielded the most acceptable results. It was able to handle the three different events' data and to separate it from noise, unlike the MA and SG filters that provided acceptable results for some events, but could not treat all of them while maintaining accuracy.

3.5.3.2 Enhancing the AWAT Filter

As the AWAT filter underwent further testing and application, some shortcomings of the filter were drawn into attention. Such deficiencies in the filter led to inaccurate estimation of the precipitation and evapotranspiration events especially at high temporal resolutions, particularly at the last step of the filtering routine, the application of the adaptive threshold (δ). The threshold approach makes sure that significant weight changes are separated from insignificant changes and leads to a step-like course of the calculated cumulative upper boundary flux (Peters et al., 2016). Hence, the time where such steps occur are referred to as "anchor points", while the remainder of the points are products of interpolation.

In the publication of Peters et al. (2016), an issue with respect to step interpolation was highlighted, which affected the integrity of the AWAT filter proposed by Peters et al. (2014). The step interpolation provides reasonable estimates for the precipitation and evapotranspiration events as upper boundary fluxes if the required resolution is a daily one. However, using step interpolation for any smaller resolution such as an hourly one would generate erroneous values that would misestimate the aforementioned fluxes. The use of step interpolation leads to discontinuous fluxes, occurring as single peaks at the steps and then zero flux in between (Peters et al., 2017).

Therefore, alternative interpolation functions were tested against the step interpolation scheme. Such alternatives investigation included linear and cubic Hermitian spline interpolations, aiming to smooth the steps without hindering the values of the cumulative fluxes.

After the MA is calculated, the threshold separates significant fluxes from insignificant ones, starting with the first MA value at $t = 0$. This value can then be denoted as the first anchor point, or ap_0. This value is kept until the corresponding time step with respect to ap_0 is larger than the designated threshold value (δ), marking the second anchor point, ap_1. Therefore, anchor points coincide with the MA, while all points in-between are interpolation products. This leads to a stepwise interpolation of the cumulative upper boundary flux, guaranteeing that the oscillations that are probable to occur even after the smoothing function is applied are not regarded as real fluxes. Figure 3 below describes the interpolation scheme, and the issue associated with the step interpolation function.

After testing the three interpolation schemes, it was evident that the spline interpolation function yields better results than the linear one, and by default than the step interpolation scheme. As the spline interpolation scheme yields smooth curves even at the anchor points, unlike the linear one.

Nonetheless, there were still issues associated with strong signal events, e.g. heavy rainfall, after a period of low signal, e.g. low evapotranspiration or precipitation. The interpolation scheme in such cases would smooth the data excessively leading to erroneous interpretation. Hence, Peters et al. (2016) suggested applying the interpolation function only to negative mass changes, i.e. evapotranspiration events, or if the mass increase between anchor points is under a specified value, denoted as (aδ), where a must have a value >1. This enables small mass changes resulting from events such as dew to be smoothed as well. Thus, the interpolation scheme will only be used if the mass change (ΔM) is bigger than (aδ), representing medium to strong signal precipitation events.

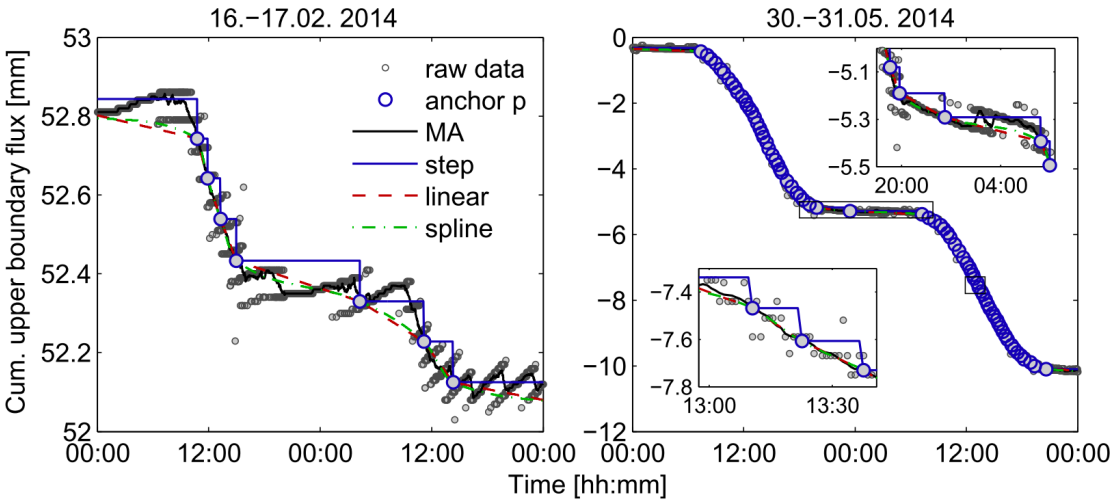

Figure 3: Raw data of two evapotranspiration events, filtered with the original AWAT filter (steps) and linear as well as spline interpolation schemes. Left: low evapotranspiration on 16–17 February 2014; right: high evapotranspiration rates on 30–31 May 2014 (Peters et al., 2016).

Both the linear and spline interpolation lead to smoother cumulative fluxes that are closer to the MA anchor points. Both techniques have no noticeable difference in terms of cumulative fluxes. Nonetheless, the spline interpolation leads to slightly better smoothing. Therefore, it is suggested to always use the cubic Hermitian spline function as its addition computational capacity requirement is insignificant (Peters et al., 2016).

Nonetheless, even after resolving the interpolation issue the AWAT filter still suffered from some downfalls. Peters et al. (2017) identified an issue in the filter with regard to a systematic error responsible for underestimation of changes in flow direction due to the threshold (δ) scheme. The issue is correlated to the size of the signal strength of the event, thereby it is pertinent to small fluxes, e.g. dew, where the systematic error can reach a degree equivalent to the magnitude of the flux itself.

Therefore, in order to produce an unbiased representation of the real signals, Peters et al. (2017) formulated a so-called “snap routine” calibrated and tested with synthetic data and real measurement data collected over a period of 10 months.

As the issue pertains to the flux direction change from evapotranspiration (ET) to precipitation (P) and vice versa, a systematic underestimation of P and ET occurs. The extent of such an underestimation relates to the threshold (δ) value. The higher the (δ) value, the higher

the underestimation is. Nonetheless, this manifestation is not dependent on the magnitude of the P or ET event. Therefore, the effect can be seen more evidently in small flux events.

The snap routine designed to resolve the aforementioned issue detects two anchor points (ap) in the MA that represent a change in flow direction, and then inserts two additional (ap) between them. The first (ap) can be assigned to the maximum or minimum values between the original ones. The second additional anchor point is set at a time (t) when the MA is back to the level of the first original (ap).

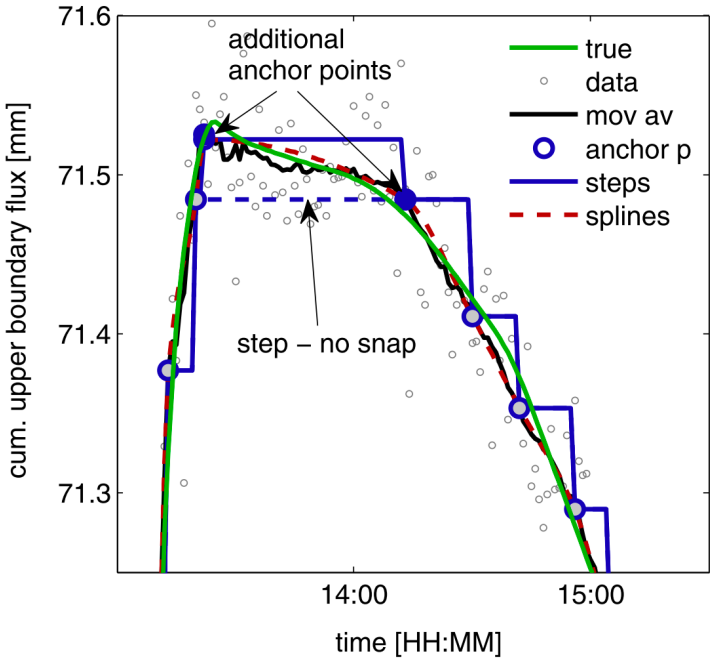

Figure 4: A demonstration of the snap routine function on synthetic data (Peters et al., 2017).

As the MA can still be influenced by random fluctuations, setting the first additional (ap) to the maximum can result in an overestimation of the flux by the MA oscillation's amplitude. Therefore, in the cases of change from P to ET, a certain quantile (Q_{MA}) of the adjacent MA data is used. Such adjacent data are the ones lying between the old anchor point and the second additional one. On the other hand, for cases of ET to P changes, the minimum is substituted for maximum and a quantile of (1 - Q_{MA}) is used.

Despite that the use of MA quantiles avoids the overestimation of P and ET events, there yet lies some unknown number of fluxes that are below scale resolution, making them undetectable. In order to account for these undetectable flux events, the detectable ones would have to be overestimated to compensate for them, jeopardizing the functionality of the snap routine. Therefore, for the best agreement between the true and identified P and ET flux events, such undetectable events need to be disregarded.

3.5.4 The Comprehensive Filtering Routine and the Synchro Filter

After the removal of outliers, and in order to conduct an efficacious smoothing of the raw data, recent studies have suggested a two phase filtering procedure, a smoothing function (e.g.

MA or SG filter), and a threshold scheme (Fank, 2013; Hannes et al., 2015; Peters et al., 2014; Schrader et al., 2013), a process that is implemented in the AWAT filter. Hannes et al. (2015) developed a comprehensive filtering routine that integrates the aforementioned principles into a smoothing filter with other steps to deal with a variety of possible errors related to high precision lysimeters. The study analyzed the effects of different values for the temporal window and threshold functions. In addition, an outcome of the study was the development of the Synchro filter, a filter applicable to a set of several lysimeter, allowing a more detailed and comparative analysis of the filtering schemes, and leading to further reduction of filtering errors. The Synchro filter offers an advantage to research facilities with more than one lysimeter to investigate the sources of errors at each data point, where it is supposed that each lysimeter should display coherent reaction and values to the upper boundary fluxes, but evidently, each will record different values due to external factors resulting in errors that are randomly distributed.

The processing method proposed by Hannes et al. (2015) introduces additional steps to the already implemented and tested smoothing functions (e.g. MA), and to the threshold value (δ). The suggested additional steps include a manual filter, a median filter, and an oscillation threshold filter. Resulting in a 5 total filtering steps, when including the smoothing filter and the threshold filter. It is imperative that the filtering steps are conducted in the proposed order to yield the expected smoothing results. Therefore, the steps are divided into discrete events filtration, which is conducted through the first 3 steps, namely the manual filter, threshold filter and the median filter. Followed by diffuse noise filtration, which is conducted in the last 2 steps; the smoothing filter (MA) and the oscillation threshold filter. If this is not strictly followed, the discrete events will be blurred by smoothing, making it difficult to treat later.

The prominent feature of this filtering routine is its ability to filter out noise based on their magnitude compared to the signal on consequent phases. Furthermore, the introduction of the oscillation threshold filter is perhaps its most significant advantage compared to other filtering schemes. As trying to retain a short temporal window, all fluctuations in the raw data cannot be adequately eliminated without tampering with the short-term effects, affecting even cumulative values of the fluxes. Hence, the oscillation threshold filter differs from the typical threshold filter in that its algorithm sets an initial data point and determines the next point where the cumulative mass of the lysimeter exceeds a predefined value, interpolating the intermediate data points.

When compared to the AWAT filter, the comprehensive filtering scheme seems less complex. It's ability to utilize several lysimeters' measurements (up to 6) is a distinctive function that helps understand the causes and effects of different sources of errors in. Moreover, this filtering routine is free to use and is readily available as a supplement to Hannes et al. (2015) research publication.

The following figure demonstrates the effect of each of the filtering steps on the raw data and their corresponding smoothing effect (Hannes et al., 2015);

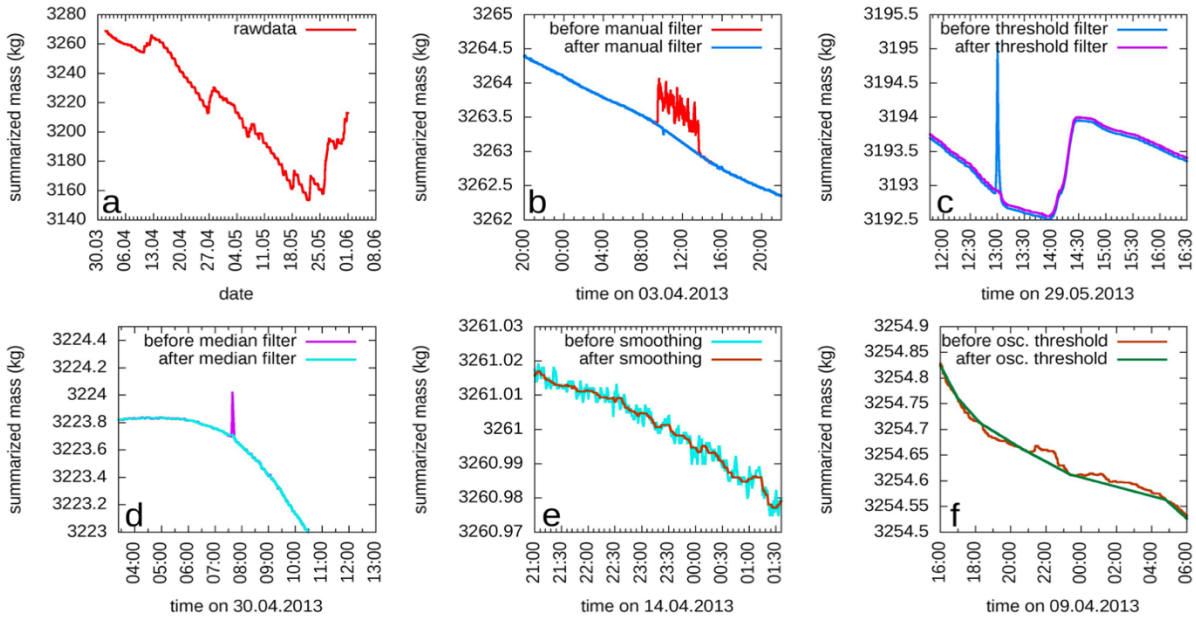

Figure 5: Effect of each of the filtering routine's steps on the raw data. a: raw data. b: manual filter. c: threshold filter. d: median filter. e: MA. f: oscillation filter (Hannes et al., 2015).

4 Materials and Methods

With respect to the determination of the magnitude of the effects of certain filtering parameters on the resulting water balance components, real measurements from a high precision weighing lysimeter and high temporally resolved measurements in an experimental field were used for the evaluation. A certain filtering scheme that has been deemed to be the fittest and most suitable to use based on the literature review has been employed to filter the data set. The scheme in use is based on Hannes et al. (2015) that also provided a computer software for handling the data, setting up initial parameters, and calculating the resulting water balance components.

4.1 Experimental site description

The description of the site at which the lysimeter is installed is based on two publications by Doležal et al. (2015, 2018). The weighing lysimeter that is used to measure field hydrological fluxes is located at the experimental field of the Czech University of Life Sciences in Prague (CULS), Suchdol, Prague (50°8' N, 14°23' E, 286 m asl).

Figure 6: The location of the Lysimeter of the CULS expiremental field in Prague (www.geoportalpraha.cz, accessed March, 2020).

The region is characterized by a warm and moderately dry climate, with an average annual precipitation of 495 mm and an annual temperature mean of 9.1 C°. The soil at the experimental and demonstration field is a loamy carbonate Chernozem on loess (22–33% sand, 40–54% silt and 22–28% clay) of loamy texture on an Aeolic loessial substrate with a boundary between the A and C horizons at about 35 cm, having the layer laying between 15 and 25 cm being more compacted than the rest of the profile. The topsoil contains about 2.5% (dry-matter basis) total organic C and 7.8% calcium carbonate ($CaCO_3$).

Furthermore, the soil's permanent saturated zone does not exist until a few meters below the surface. The soil is prone to moderate swelling and shrinkage, with cracks having a width of 1 to 3 mm appearing during dry spills 15 to 20 cm apart.

With respect to the soil structure, granular formations can be witnessed in the A horizon and subpolyhedric ones in the loessial C horizon. At the higher level of organization, the structure is prismatic. Biopores are common. Moreover, in terms of the physical properties of the site's soil, the saturated hydraulic conductivity tested with 100-cm^3 cores varies roughly between 6 x 10^{-4} and 4 x 10^{-1} $cm.min^{-1}$. While the dry bulk density varying between 1.20 and 1.55 g/cm^3, being 1.43 g/cm^3 on average. The total porosity based within the topsoil varies between 40% volumetric (plow sole) and 54% volumetric (topsoil), with a mean value of 45.7% volumetric (0–100 cm). The field capacity varies between 30 and 35% (v/v). The average particle density is 2.63 g/cm3 and does not vary significantly. The soil water content at field capacity varies between 0.30 and 0.35 m^3/m^3. The total organic carbon content is about 2.5% in the top layer.

The site does not seem to contain any surface water or surface run-off. The soil had been plowed for several centuries, but grass (commercial park lawn mixture) was sown in the spring of 2009 and has been maintained since without any irrigation or fertilization. The lawn often suffers from water stress and has already slightly degraded because of drought and N deficiency, as witnessed by the local appearance of legumes and mosses. The lawn is cut about four times a year. The average height of the canopy in the lysimeter and around it is about 6 cm. The size of the grass field around the lysimeter is about 20 by 20 m. It is surrounded by a mosaic of agricultural research parcels and facilities, containing croplands, orchards, trees, hedges, roads, and occasional buildings. Only a small part of the surrounding area is irrigated. Because of the high thermal conductivity of the metallic case of the lysimeter cylinder, the soil temperatures inside the lysimeter are somewhat higher than those in the natural soil. However, the soil water contents and matrix potential values are statistically in good accordance with those in the surrounding soil (Báťková et al., 2013).

4.2 SFL- 300 Weighing Lysimeter

This experiment was designed to deal with real measurements recorded by the Smart Field-Lysimeter-300 (SFL-300). The SFL-300 is manufactured by the METER Group AG, previously known as UMS GmbH. The SFL-300 at the CULS experimental field was installed on the 25th of April, 2013 (Doležal et al., 2018). The Lysimeter system consists of the lysimeter itself and a field box which contains the seepage water bottle, the drain water balance, and the pump and distribution box. The field box works as a supply station, it contains the bi-directional pump which is used to dispose of excess water from the lysimeter to the drain bottle, and to pump water into the lysimeter column when the conditions of the lysimeter are drier than the surrounding soil to maintain an accurate reflection of the field conditions. The drain water bottle in the SFL-300 is situated on a PL-10 balance, which has a recommended resolution of 1 g with 0.1 g as its smallest resolution, and a weighing accuracy of ±1.7 g (METER Group AG, 2013).

As for the main instrument, the lysimeter itself, it consists of the cylinder and all equipment connected to it. According to the manufacturer, the field box must be installed a distance of 1 m away from the lysimeter (METER Group AG, 2013), unlike the logger box which can be installed at any further distance. The SFL comes at many variations, 300, 600, and 900. The SFL-300, the smallest in size, contains a stainless-steel cylinder with a height of 30 cm and an inner diameter of 30 cm. It contains six sensor ports, two each at 5, 15 and 25 cm

(METER Group AG, 2013). Ideally, each size of the lysimeter requires a different weighing platform. The SFL-300 balance is the PL-50, the balance's function is to measure the changes in the lysimeter column which correspond to either water flow from or to it, representing evapotranspiration or precipitation respectively. The PL-50 has a nominal measurement range of 0 to 50 kg, with a maximum 75 kg. It's weighing accuracy is ±7 g, with a recommended resolution of 1g and the smallest of 0.3 g.

The use of such smaller lysimeters is relatively recent as their conception is also novel. Therefore, their application for studying different water balance components is yet contemporary. It is more common to find literature dealing with lysimeters >1m^2 in terms of surface area.

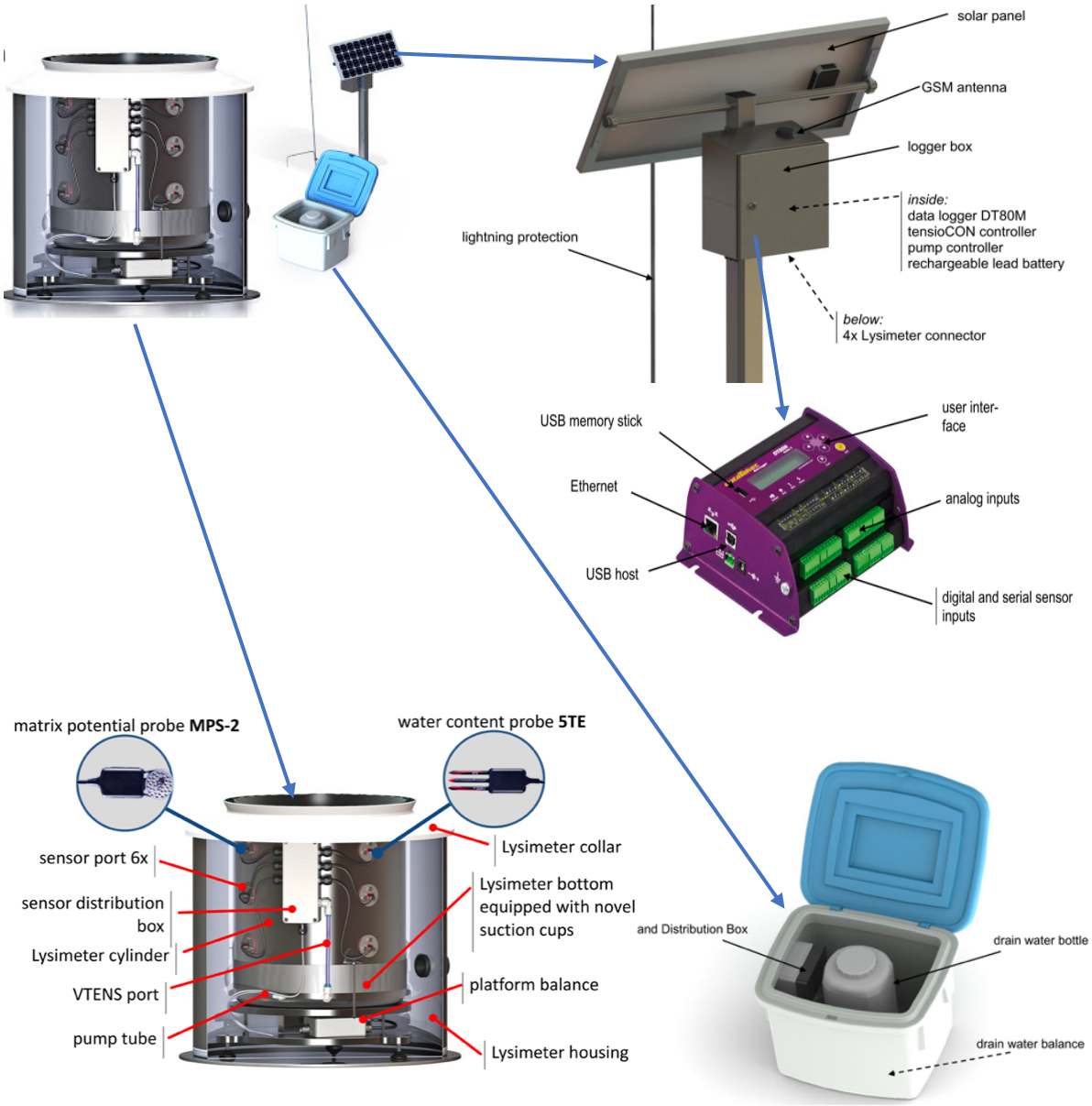

Figure 7: The components of the SFL-300 (METER Group AG, 2013).

The SFL-300 was installed at the experimental field by cutting a soil monolith of the same size. The cylinder with the soil was then placed in a sunken barrel on an electronic balance. The suction at the bottom of the monolith is automatically maintained corresponding to the suction of water in the native soil nearby (at about 1-meter distance) at the same depth (30 cm), measured by a T8 tensiometer from the METER Groups AG. The shallow underground collar of the lysimeter makes it difficult for grass to grow within an annular strip about 15 cm wide around the lysimeter. The strip was mostly bare and probably contributed to the increase in evapotranspiration from the lysimeter (Doležal et al., 2018).

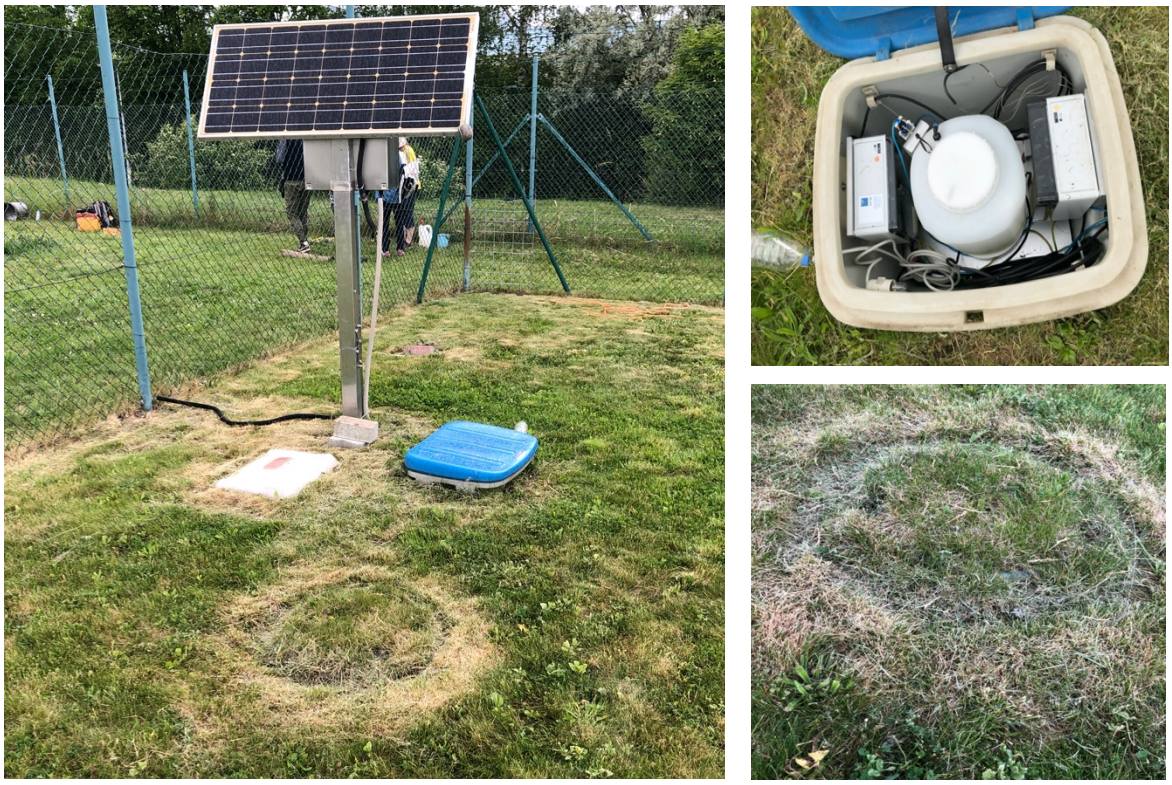

Figure 8: Photos of the SFL - 300 at the CULS experimental field in Suchdol, Prague. Left: Complete lysimeter station. Right top: Seepage Water tank in plastic compartment. Right down: Lysimeter covered in grass. Photos taken in June, 2020.

Furthermore, the lysimeter itself houses 6 sensors, 3 matrix potential sensors MPS2, and 3 volumetric water content 5TE sensors. The MPS2 determines the water content of a porous matrix, which correlates with the matrix potential. The 5TE sensor measures the volumetric water content (% v/v), the electrical conductivity and the temperature of the soil. The 5TE is based on Frequency Domain Response (FDR) technology. The sensors installed record readings at 10 min intervals, and were not each individually calibrated (METER Group AG, 2013).

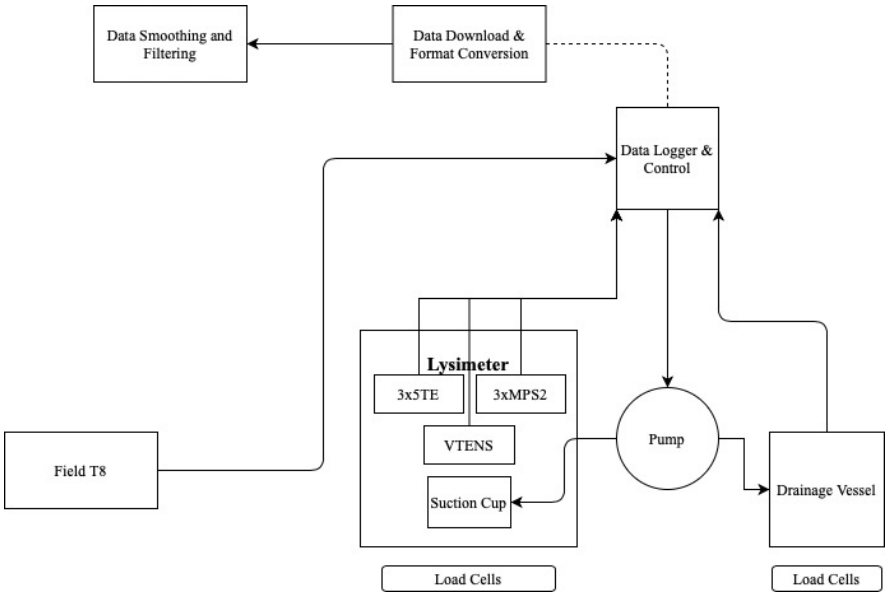

Figure 9: Network of the working principle of the SFL-300 weighing lysimeter.

Moreover, the lysimeter cylinder contains a virtual Tensiometer (VTENS) which is mainly used to adjust the matrix potential in the bottom of the Lysimeter, the lower boundary. The value of a reference Tensiometer is controlled at a ceramic surface by a vacuum controller. VTENS therefore is the actual measurement value directly within the ceramic. In principle VTENS is a 4th matrix potential value directly above the Lysimeter bottom. The VTENS is not a sensor located at the lower part of the lysimeter. Instead, a reference tube is led out, which leads to the suction cups in the Lysimeter bottom (METER Group AG, 2013).

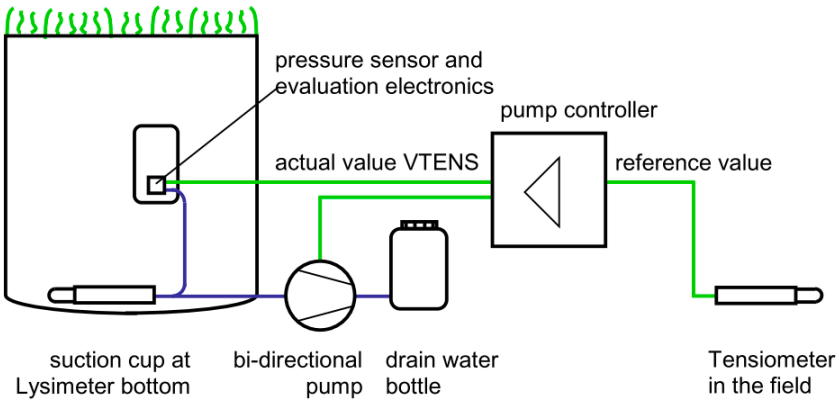

Figure 10: The bi-directional pumping mechanism (METER Group AG, 2013).

4.3 Data Logging

The SFL-300 has a high temporal resolution of 1 min, meaning that the PL-50 and the PL-10 register measurements at 1 min intervals. The data logger in the SFL-300 is the DT80M which functions as both a data logger from which logged data can be read automatically or manually, and also a controller of the lysimeter.

The data logger can store data either in .CSV or .DBD formats. In the case of the SFL-300 in the CULS experimental field, the data were stored in .DBD format. The reason for choosing the .DBD over the .CSV is that the .CSV format is country specific as no international standard exists, so transferring them can render erroneous da. In addition, the .CSV format is memory space consuming, and require laborious work of rearranging the data prior to making use of it as well as consuming significant data when transferring the readings virtually. Therefore, especially when transferring the data over a File Transfer Protocol (FTP), it is more convenient to use the .DBD format. The data hence is compressed and stored independent of the platform, country and format. On the other hand, the use of the .DBD format requires a conversion software to convert .DBD to .CSV or to, more conveniently, .XLS.

4.4 Weather Station

Meteorological information is critical in the process of filtering the lysimeter measurements as many parameters are dependent upon them. The CULS experimental field is equipped with a weather station, where several measurements such as wind speed, temperature, precipitation, and evaporation are conducted. For the purposes of this thesis, precipitation measurements are the most important of the meteorological data as some of the filtering parameters, and subsequent comparisons are dependent on them. Information regarding wind speeds, temperatures, and others are also useful for studying the impact of such parameters on the measurement errors.

Rainfall amount, with a temporal resolution of 10 min, was measured by the TBR-MR3H (MR3H-METEOSERVIS v.o.s., Czech Republic). This rain gauge has a 1 meter elevation above ground and has a catching area of 500 cm^2 (Mekonnen et al., 2015). The tipping bucket rain gauge has a resolution of 0.1 mm per flip, and has an error of 2% of the measured amount for rain intensity at 30 mm/hr.

Other measurements were taken at a horizontal distance of several meters from the lysimeter. The data was recorded at 10-min intervals. The air temperature and humidity were measured by a combined probe HMP 45A/D (Vaisala), the solar radiation by the pyranometer LP02 (Hukseflux), the wind speed by an ultrasonic sensor Windsonic (Gill Instruments Ltd.) (Doležal et al., 2018).

4.5 Experiment Design

4.5.1 Data Sets

A representative time period needs to be selected where the registered measurements for each of the lysimeter (LYW) mass changes and the seepage water vessel (SWW) will be treated by the suggested filtering routine. In order to obtain data sets that can be opened with software like Excel, the conversion of the logged data format to .CSV or .XLS, if not already available in such formats is required. The data sets in this time period will be smoothed with the use of the computer software provided by Hannes et al. (2015). The filtered results will allow for the calculation of the fluxes and the evaluation of the effects of the selected parameters in the software on the cumulative fluxes.

The raw data, which represents instantaneous mass records of both the lysimeter (LYW) and the drainage water vessel (SWW) were recorded at 1-min intervals. To ensure a good representative characteristic of the data set chosen, and to cover a range of meteorological conditions and seasons. A data set consisting of 1-minute measurements for both LYW and SWW over a period of 165 days was chosen, extending between the 12th of April 06:50, until the 23rd of September 09:59.

This investigation period contained 236,351 datapoints for each the LYW and SWW. The reason behind not extending this period of measurements beyond, or before, the selected range is that the lysimeter at the CULS experimental field had problems with registering measurements before the 12th of April. While after the 23rd of September there was a reoccurring issue of missing measurements on the 10th minute, with other missing measurements (e.g. every 10th minute), which lasted until the subsequent day. Hence, and since the period already covers 5 months and 12 days, occurring within the range of the growing season. It was speculated that it would provide an adequate representativeness of the fluxes and hydrological processes.

The lysimeter measurements do not only provide the interval masses of the LYW and SWW, the logged file contains also time of pumping in and out in seconds for each minute log. Moreover, the file provides electrical components information such the wattage, amperage and the temperature of the logger. The file also provides field measurements and information from the sensors, such as the 3 matric potential sensors (MPS2), VTENS virtual tensiometer in and the field tensiometer in (KPa). In addition to temperature, volumetric soil water content and conductivity.

4.5.2 Data Filtration Methodology

Prior to the calculation of the water balance and the derivation of the fluxes, adequate processing of the raw data is necessary (Hannes et al., 2015). For the purposes of this thesis, the comprehensive processing routine published by Hannes et al. (2015) was utilized. Throughout the literature review, it became evident that more complex filtering routines are necessary to yield smoother data while not affecting true signals especially with high temporally resolved and high precision lysimeter measurements. Moreover, and despite the fact that there has not been much research dealing with the perfection of the filtering process of raw lysimeter data, two methods seem to yield very satisfactory results when compared to the simple filtering schemes such as the MA and SG. Those two are the AWAT filter by Peters et al. (2014) and the Comprehensive filtering routine and the Synchro filter by Hannes et al. (2015).

Nonetheless, the work in this thesis is hinged on the comprehensive filtering routine by Hannes et al. (2015). The justification for the use of this routine rather than the AWAT is based on the accessibility of the Waterbalance software offered by Hannes et al. (2015) as a free open source supplement to their publication (Hannes et al., 2015; Vogel, 2015), in addition to being relatively straightforward and not laborious or highly complicated. Moreover, the AWAT filter seems to be yet undergoing a process of perfecting some of its shortcomings as can be seen in the subsequent papers published by Peters et al. (2016, 2017) complimenting the original work of (2014). Additionally, and despite that there has not been a recent comparison of the two filters especially after the enhancements the AWAT filter has undergone, the results of the filtering routine designed by Hannes et al. (2015) are still of quality and are satisfactory. Figure

11shows slightly better results when comparing the standard deviations of the AWAT filter and the basic filtering techniques with the oscillation threshold filter, the scheme which this thesis shall utilize. While the Synchro filter, which employs the data of several lysimeters, exhibit the least deviations.

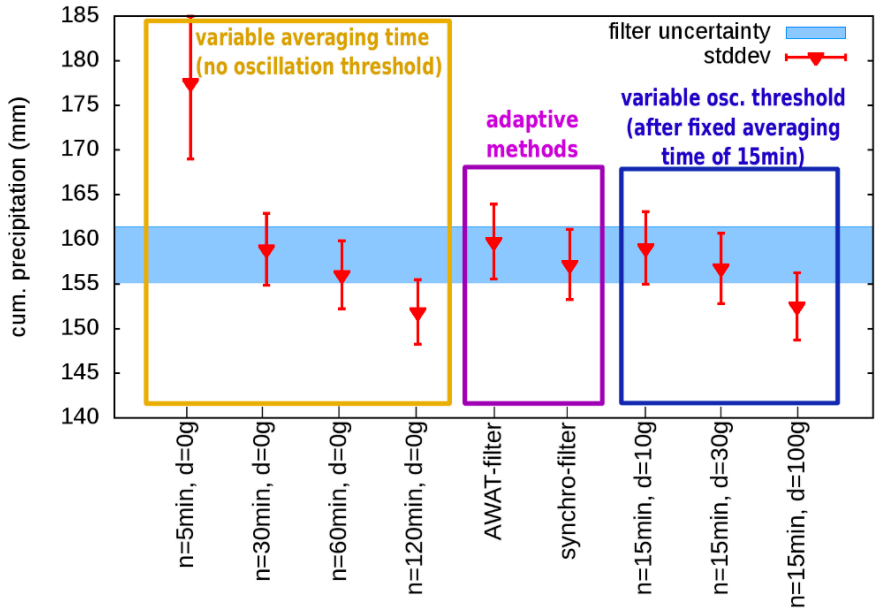

Figure 11: Sample of the standard deviation of cumulative precipitation using the basic filtering technique (yellow), the adaptive methods AWAT & Synchro filters (purple), and the basic techniques of fixed window with an additional varying oscillation threshold filters (blue) (Hannes et al., 2015).

The comprehensive procedure of the filtering routine of Hannes et al. (2015) is based on 5 consequent steps;

1. **Manual Filter:**

The first step is the application of a manual filter, typically following an interpolation scheme to fill missing data points. The manual filter is the only step applied to the mass data sets of the seepage tank and lysimeter before summarizing them, whereas all the other steps are applied to the seepage tank mass data, which represents the lower boundary flux, as well as to the summarized lysimeter and seepage tank mass data. Initiating the filtering routine with the manual filter is crucial in order to remove defective data sets that are either too strong or hard to recognize, hence compromising the integrity of the consequent filtering steps. Such data can include periods of known regular maintenance and emptying of the drainage vessel. This property of such defective data gives potential to automate this step by connecting it to the logs of the known maintenance works.

2. **Threshold Filter:**

The second step is the threshold filter, aimed to remove unrealistic data points. Such data is characterized by being strong over short durations (Nolz et al., 2013a). Most common sources of such data are the external disturbances affecting the lysimeter including maintenance activities, emptying of the seepage tank, mechanical pressure due to stepping on the lysimeter, malfunctions in data transfer, and other possible sources of errors defined in previous sections. Small noise values resulting from moderate or low wind cannot be removed in this step as not to remove low signals unintentionally, causing another source of error.

The idea behind the threshold filter is attributed to Fank (2013) and has been studied by Schrader et al. (2013). It has also been implemented in the AWAT filter of Peters et al. (2014) in addition to the filtering routine of Hannes et al. (2015). The purpose of defining a threshold value (δ) is to ensure the accurate separation of significant mass changes from insignificant ones, which are translated to fluxes. A difference in measurements between t and t_{+n} needs to exceed the (δ) value (kg) in order to be registered (Schrader et al., 2013). Hence allowing for the removal of unrealistic data points and substituting them with means of linear interpolation (Hannes et al., 2015).

3. Median Filter:

This step designates the first measure towards the filtration of noise with residual values smaller than what the threshold filter in the previous step can filter, eliminating short spikes that are below the threshold value. This step functions as a complementary step to the threshold filter in order to remove unrealistic data points and single outliers. Hannes et al. (2015) applied the median filter with a window of 15 mins.

The median filter is particularly helpful in the reduction and removal of random noise, it is classified as a non-linear filter, used for various types of data processes sing. Median filters are handy in the elimination of data that displays odd patterns compared to the rest of the data set. Such peculiar data have stronger signal than the rest of the data set, also known as long tails (Ohki et al., 1995). The median filter is designed to eliminate local errors of only some data points in the set and is also used for subsequent calculations in the filter. Finally, the median filter functions in a similar manner to the MA, by taking the median over a sliding window of a fixed size, where in the case of the comprehensive filtering routine, such window width is 15 mins (Arias-Castro et Donoho, 2009).

4. Smoothing Filter (MA):

This process denotes the first step towards the treatment of diffuse noise. Several smoothing methods have been discussed in the literature, such as the SG filter and the MA (Huang et al., 2012; Marek et al., 2014; Schrader et al., 2013;Vaughan et al., 2007) . Nonetheless, each of these filtering schemes suffer from their own drawbacks, particularly when it comes to mixing errors and the blurring of short time patterns of P and ET. Therefore, it was suggested to limit the averaging window width to eliminate such flaws (Hannes et al., 2015). One of the reasons of choosing applying the MA filter over the SG filter, is that The SG filter does not smooth the heavy precipitation data as much as the MA filter does. Moreover, the SG filter tends to oscillate, which will lead to an overestimation of both precipitation and evapotranspiration (Peters et al., 2014).

The smoothing filter applied by Hannes et al. (2015) for this step is the MA, which shall be also the filtering scheme implemented for the purposes of this thesis as it is integrated in the software used. The MA smoothens time series by averaging a fixed number of consecutive terms, whether weighted or non-weighted, the MA is considered to be one of the simplest methods of extrapolation and is widely used in various disciplines (Ostertagová et Ostertag, 2016). Hannes et al. (2015) implemented a simple moving average (MA) with an averaging window of 15 mins to avoid blurring impacts of larger periods. The simple moving average is unweighted and is the most common of the moving average methods, it calculates the arithmetic mean of the data points in the time window from $t_{i-(n-1)/2}$ to $t_{i+(n-1)/2}$ for each data point at time t_i (Hannes et al., 2015).

Therefore, for n=15 min averaging period, and for data point t_i, $t_{i-(n-1)/2} = t_{i-7}$, and $t_{i+(n-1)/2} = t_{i+7}$, the smoothed T_i value would yield;

$$T_i = \frac{1}{15}\,(t_{i-7} + t_{i-6} + t_{i-5} + t_{i-4} + t_{i-3} + t_{i-2} + t_{i-1} + t_i + t_{i+1} + t_{i+2} + t_{i+3} + t_{i+4} + t_{i+5} + t_{i+5} + t_{i+7}) \quad (10)$$

5. Oscillation Threshold Filter:

Finally, in order to remove the remaining oscillations that lead to misestimating the precipitation and evapotranspiration cumulative values and fluxes, and cause errors that manifest in the form of mixing errors and noise, an oscillation filter is introduced. This description of the oscillation threshold filter is based on the publication of Hannes et al. (2015).

In measurements with low signals, remaining noise can substantially affect the estimation of P and ET. While such residual fluxes are still small, especially after the use of the MA, they can carry effects that are observable in the cumulative P and ET values. The application of the standard threshold mechanism would assist in achieving the purpose of this step, however slow occurring changes such as ET or dew would not be registered. Therefore, the oscillation threshold filter functions with an algorithm on the basis that if the sum of such aforementioned slow changes exceeds a predefined threshold, then such values will be registered as real measurements.

Setting an initial data point, this filter determines the next point in time where the cumulative readings reach the threshold value. Once this is achieved, the intermediary points are interpolated as follows;

$$M_k = M_i + \frac{M_l - M_i}{t_l - t_i}\,.\,(t_k - t_i), \quad where\ i < k < l - 1 \quad (11)$$

Here (M) is the sum of the mass of the lysimeter and the drainage water vessel at time (t), (k) is the starting point, and (l) is when the filter registers the first instance of surpassing the threshold value.

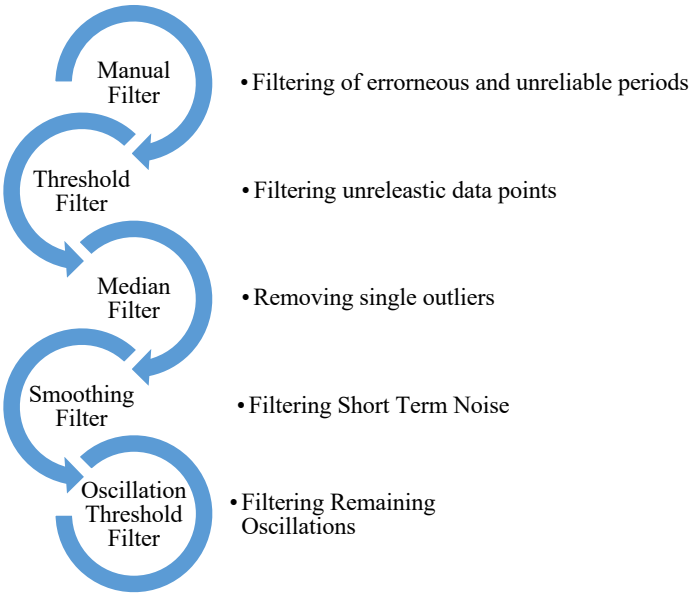

Figure 12: Flow Chart of the comprehensive filtering routine (Hannes et al., 2015).

4.5.3 Calculation of Fluxes

Once the comprehensive filtering routine is concluded, the fluxes can be calculated. The calculation of the lower boundary fluxes (seepage flux) is straightforward, and relies on the mass of the drainage water vessel (SWW). Therefore, the fluxes in the SWW (kg) can be determined by the following calculation;

$$\mathrm{F(SWW)} = \frac{SWW_{t_{i+1}} - SWW_{t_i}}{t_{i+1} - t_i} \tag{12}$$

Where F(SWW) is the flux of the drainage water vessel mass (kg), (SWW_{ti+1}) is the mass of the vessel at the consequent time point (kg), and (SWW_{ti}) is the mass measurement at time point (t_i) (kg) (Hannes et al., 2015).

A positive lower boundary flux would mean that water is being drained from the lysimeter column into the drainage vessel, while a negative flux would mean that water is being repumped to the lysimeter column to represent the true field conditions measured with the T8 sensor and the VTENS.

Moreover, if the amount of seepage water is added to the lysimeter mass, temporal changes of the lysimeter mass can be used to solve the water balance equation for atmospheric fluxes. The masses of both the LYW + SWW are used, collectively denoted as ΔM;

$$\Delta\mathrm{M} = \Delta\mathrm{M}_{LYW} + \Delta\mathrm{M}_{LSWW} \tag{13}$$

Where (ΔM_{LYW}) and (ΔM_{SWW}) (kg) are consecutively the masses of the Lysimeter and the drainage vessel between two consecutive time points. Breaking it down further; the equation would then represent;

$$\Delta\mathrm{M} = (\mathrm{M}_{LYW\,t_{i+1}} - \mathrm{M}_{LYW\,ti}) + (\mathrm{M}_{SWW\,t_{i+1}} - \mathrm{M}_{SWW\,ti}) \tag{14}$$

Therefore, any positive change in mass (ΔM) theoretically represents a precipitation event, while a decrease represents an evapotranspiration event. Furthermore, since the time window is relatively small (here being at 1-min resolution), it can be assumed that P and ET do not occur simultaneously, meaning that when one occurs the other has a value of 0, (Hannes et al., 2015). Nonetheless, during precipitation events, the air aloft need not be completely saturated, meaning that ET could actually occur during a precipitation event, but during such a P event and small temporal window , ET can be assumed to be negligible, P>>ET (Peters et al., 2014).

Hence, applying a mass balance equation for the determination of ET and P from the cumulative mass change (ΔM) and breaking down the mass changes between the two-time steps would yield;

$$\Delta\mathrm{M} = \frac{\Delta\mathrm{M}_{LYW} + \Delta\mathrm{M}_{LSWW}}{t_{i+1-}t_i} \tag{15}$$

$$\Delta\mathrm{P} = \begin{cases} \Delta\mathrm{M}, & \Delta\mathrm{M} > 0 \\ 0, & \Delta\mathrm{M} \leq 0 \end{cases} \tag{16}$$

$$\Delta \mathrm{ET} = \begin{cases} 0, & \Delta \mathrm{M} \geq 0 \\ \Delta \mathrm{M}, & \Delta \mathrm{M} < 0 \end{cases} \quad (17)$$

Where (ΔM) is the cumulative mass change of both LYW and SWW between two consecutive time-steps (kg).(ΔP) (kg) is the sum of precipitation that descended on the lysimeter column and covering vegetation during the same time interval, and (ΔET) (kg) is the occurring evapotranspiration and interception loss during the same interval (Doležal et al., 2018; Hannes et al., 2015; Peters et al., 2014; Schrader et al., 2013). Nonetheless, such interpretation of the upper boundary flux in equations 15-17 is only valid after appropriate filtering. As such mass changes possibly contain noise or mixing error, strong filtering will lead to the underestimation of such events, while weak or no filtering of the raw data will lead to overestimating the strength of such signals (Peters et al., 2014; Schrader et al., 2013).

Therefore, and since ET and P theoretically do not occur at the same time, the cumulative upper boundary flux (F) is simply stated as;

$$F = P - ET \quad (18)$$

With one of the fluxes always being 0. It is also worth noting that most of the literature encountered used lysimeters with a surface area of 1m^2 (Hannes et al., 2015;Peters et al., 2014; Schrader et al., 2013). Hence, the conversion of the kg equivalent of the mass change to (mm) or (l) does not require any further processing, meaning that every 1 kg mass change equals to 1 mm of water equivalent or 1 l change if having a lysimeter with a surface area of 1 m^2. This conversion is necessary due to the fact that P and ET events are represented in L/T rather than M/T.

$$F(t_i) = \frac{\Delta M_{t_i}}{\rho_w A_s} \quad (19)$$

Where F(t_i) is the cumulative flux at any time value (t_i) in mm or l, (ΔM_{t_i}) is the corresponding mass change at such time in kg or g, (ρ_w) is the density of water (g.cm^{-3}), and (A_s) is the surface area of the lysimeter (m^2) (Peters et al., 2017).

Therefore, for the SFL-300 with a diameter of 0.3 m, hence a radius (r) of 0.15 m, each ΔM (kg) measurement at any time t_i would have to be divided by 70.686 to yield the F(t_i) value in m, then multiplying by x1000 to get the value in mm;

$$F(t_i) = \frac{\Delta M_{t_i}\ (kg)}{\pm 1000\ (\mathrm{kg.\,m^{-3}}) \times\ \pi \times 0.15^2 (m^2)} = \frac{\Delta M_{t_i}}{70.686}\ (m) = \frac{\Delta M_{t_i}}{0.0707}\ (mm) \quad (20)$$

Shortly, since the surface area (A_s) of the SFL-300 is 0.707 m^2, any 1 kg increase is equal to an additional 14.15 mm of precipitation, and vice versa of ET in cases of 1 kg mass decrease.

It is also apparent that when examining the cumulative fluxes for a relatively long period of time (i.e. semi-annual or annual), the biomass changes in terms of growth or mowing the grass such as in the experimental field in CULS have also influence on the fluxes registered.

This is treated by keeping the grass at a constant height all year long to minimize the effect of such a process, where the grass is reported to be cut 4 times per year (Doležal et al., 2018).

4.5.4 Computer Software Description

The filtering computer software designed by Hannes et al. (2015) is free and available as a supplement to the publication. The software is available for Windows and Linux operational systems, this might constitute a limitation for users of different operation systems, such as MacOS. For the purposes of this thesis, the software was operated on a Windows system.

The software is straightforward despite having no user interface. The file containing the software has three main components; the input folder, the output folder and the parameters .Dat file (param.Dat).

The input folder contains two .Dat files, first is the mass.Dat file. The mass.Dat, herein shortened as the mass file contains the raw measurements of the lysimeter and seepage water vessel. The advantage of this software is that it can support the filtering process for up to 6 lysimeters. It is by default more convenient to open the mass.Dat file in excel and work with the data within the Excel software. The mass.Dat file for 1 lysimeter, such as in the case of this thesis, shall include 4 columns. The first column contains the Julian date, denoted as "Time(days)", taken from a specified starting date, in this instance the first counting date was chosen to be the 1st of January, 2017. Hence the first Julian date of measurements which is the 12th of April 2017 is 101. Moreover, the second column, "Date" should include the Gregorian date with the hour, minute and second on which the measurement was taken for both LYW and SWW. In order for the filter software to run, the format of the Gregorian date needs to be exactly as what the authors used, which for the same date aforementioned needs to be as 2019/09/18_15:03:00. The issue with this format is that it is not a template time format in excel and the "_" once a new template is created is displayed as a space; 2019/09/18 15:03:00. Therefore, the Date column's cells were costume formatted with the following type: yyyy/mm/dd"_"hh:mm:ss. The use of the quotation marks around the underscore prevents the software from viewing it as a space contrary to its default setting. Following, the two remaining columns contain the SWW mass (kg) and the LYW mass (kg) consecutively.

The second .Dat file in the input folder is related to manual filter step. This file is only necessary if there are unreliable or erroneous measurement periods that need to be removed. The file contains 4 columns, the first indicates the number of the lysimeter, the second must be assigned a value between (0-2). 0 removes measurements of SWW, 1 removes LYW measurements, and 2 removes both. The following two columns are regarding the start and end period of operation of the manual filter. In the data set dealt with in this thesis, the missing measurements between 06:00 and 15:12 on the 7th of July were inserted to the manual filter, as such values might be unrepresentative if interpolated.

Moreover, the Output folder contains 31 files, containing the filtered output for the bottom flux, top flux, seepage vessel, lysimeter and the cumulative mass. For each of them, the output folder contains a result file by each of the filtering steps (i.e. median_bottom flux).

Nonetheless, the most important files in the output folder are the filtered ones;

1. Filtered_01_Sw_mass.dat - contains the two time columns followed by the n columns with the filtered masses for the seepage water in kg.

2. Filtered_02_masssum.dat - contains the two time columns followed by the n columns with the filtered masses. For the summarized mass of Lysimeter and Seepage water vessel (corresponding to the upper boundary) in kg.
3. Filtered_03_topflux.dat - contains the two time columns followed by the n columns with the actual fluxes at the upper boundary of the filtered dataset in mm/h.
4. Filtered_04_bottomflux.dat- contains the two time columns followed by the n columns with the actual fluxes at the lower boundary of the filtered dataset in mm/h.
5. waterbalance.dat - contains one time column for the date followed by n columns for each lysimeter with cumulative values for Evapotranspiration (ET), Precipitation (P), storage, and seepage given in [mm].
6. Waterbalance_end_values.dat - contains the water balance values from waterbalance.dat at the end of the dataset for each lysimeter for a quick overview.
7. Dailyprecipitation.dat - contains the cumulative precipitation at each day of the dataset for each lysimeter in [mm].

4.5.5 Selection of the Initial Filtering Parameters

The crucial part in the experiment is setting up the filter parameters. These parameters determine the smoothing process of the raw data, including the threshold values for the lysimeter and seepage tank fluxes, the median filter window, the MA window and most importantly the oscillation threshold value. The basic processes remove noise and provide smoothed data to some extent, yet setting different values for the parameters control the quality of the output fluxes.

The parameter file of the software (param.Dat) provides several parameters which the user may tamper with and set based on the experiment's properties. The first parameters are related to the data sets, defining the number of time columns and number of lysimeters, up to 6. With 6 lysimeters' data sets, the Synchro filter can be used. Nonetheless, for the purposes of this thesis using only one lysimeter, the value of the Synchro filter is set to =0, with 0 indicating that it is off.

The following parameters set the values of the filtering process. These are the parameters that can be adapted based on the characteristics of the lysimeter and the design of the experiment;

1- Maximal Seepagewater-flux (g/min) (for automatic threshold filter) (default=150)
2- Minimal flux in Seepagewater (g/min) (for automatic threshold filter) (default=-150)
3- Maximal flux at upper boundary (g/min) (for automatic threshold filter) (default=1000)
4- Minimal flux at upper boundary (g/min) (for automatic threshold filter) (default= -100)
5- Time window for median filter (min) (default=15)
6- Time window for moving average (min) (default=20)
7- Threshold for oscillation threshold filter (kg) (default=0.030)

The above parameters are the default settings of the Waterbalance software. They seem as they were suggested to be able to fit and process different data sets. As the aforementioned

parameters seem to be general, during the experiment of Hannes et al. (2015) the following parameters were used and suggested;

Table 1: Parameters for the different filters in the basic processing approach (Hannes et al., 2015).

Standard parameters	
threshold for lysimeter mass changes	±60 mm h^{-1}
threshold for seepage mass changes	±9 mm h^{-1}
median filter window	15 min
smoothing filter window	15 min
oscillation threshold	50 g = 0.05 kg = 0.05 mm

Since the first step is done manually in the software. For the second filtering step, the threshold filter, the threshold values needs to be assigned based on the maximal pumping rate at the lower boundary of the lysimeter, maximal seepage rate, the maximal P and maximal ET rates, while including a safety factor (Hannes et al., 2015). The threshold value for the drainage or seepage mass increase represents a reasonable expected drainage rate, while for the reverse process of pumping back to the lysimeter, the threshold value can be set to equal the maximal pumping rate. The same is applied with respect to the upper boundary fluxes, an increase should lie below the maximal expected P value, while a decrease should lie below an expected maximal ET rate (Schrader et al., 2013).

For the median filter and smoothing filter, the chief parameter is the setting of an appropriate time window. A time window (w) of a reasonable value ensures the high temporal resolution of the processed data sets. Available literature show that smoothing functions can still yield acceptable results even when the averaging window is set up to 60 minutes, although heavy precipitation filtering might be less representative to some degree (Peters et al., 2014). Nonetheless, for the purposes of this thesis, and based on the recommendations of the authors of the software, the window will be kept at 15 mins for both the median filter and the MA.

4.5.6 Settings of the Oscillation Threshold Filter

The choice of the oscillation threshold value requires a thorough determination in order to avoid eliminating real signals, or interpreting unrealistic measurement errors as real readings. It has to be chosen to be large enough to remove residuals and small enough to maintain slow processes and to avoid the underestimation of precipitation events, a problem that accompanies applying this filter with relatively small averaging windows that is comparable to the effect mixing errors have on the smoothing process (Hannes et al., 2015).

For the purpose of reducing the aforementioned errors, the software provided by Hannes et al. (2015) utilizes the adaptive method for the determination of the oscillation threshold value suggested by Fank (2013) and elaborated by Schrader et al. (2013). Therefore, the minimum oscillation threshold value is set to a lowest possible value, the value could be zero indicating that any oscillation is a real event. Nonetheless, Peters et al. (2014) suggested setting the minimum oscillation threshold value to the lysimeter scale's maximal resolution value. Moreover, the determination of the maximum oscillation threshold value has the paramount effect on the filter's operation. In order to obtain this value, the measurements of the lysimeter

during nighttime are monitored, and the oscillation threshold maximum value is chosen at a height where nearly all the oscillations disappear. Hence, as the minimum oscillation threshold value leads to overestimating P and ET events. The maximum oscillation threshold typically will lead to underestimating them. Hence, the resultant combination shall generate results with proximity to the true values.

For an accurate determination of the threshold value of the oscillation filter, the dataset is divided into periods with precipitation and periods without precipitation; separating periods of positive fluxes from others. This is done in order to assess the possibility of the overestimation effect of the oscillatory noise is compensated by the underestimation effect of the mixing error. as in the cumulative fluxes these effects might cancel out each other, in the temporally resolved fluxes such deviations will be evident (Hannes et al., 2015). Periods of precipitation exceeding 1 mm/hr will only be utilized for the rain periods, as the noise errors in such intervals is less relevant than the mixing between ET and P (mixing error). Periods with precipitation under 1 mm/hr will be excluded as both errors are relevant. For the noise error investigation, the periods with no registered precipitation will be employed as the mixing error is negligible in such cases.

This complication of the underestimation and overestimation of the fluxes can be used to calculate the oscillation threshold value. The separation of periods of precipitation from the periods of no precipitation can help distinguish between the effects of noise and mixing errors, and finding the oscillation threshold value in between. For the periods with precipitation, each interval that has a corresponding precipitation measurement of > 1 mm/hr will be chosen with 15 minutes before and after measurements as well, even if they did not register precipitation. This method was highlighted by Hannes et al. (2015) to eliminate the possibility of blurring out the fluxes due to the smoothing window. In their case, 200-250 minutes were chosen since their study included investigating the effects of longer averaging time. Hence, for the purposes of this thesis, and the 15 minutes window selected, 15 minutes before and after each measurement shall suffice.

Since in rainy periods the noise has relatively low impact on the total residual value, the size of the mixing error can be estimated as the fluxes are typically increasing and do not oscillate between positive and negative values (Hannes et al., 2015). For the periods of no precipitation, the mixing error plays relatively no role, and the major type of residual is the noise.

Hence, the aforementioned separation will lead to assessment of the contribution of each of the investigated errors. Firstly, during the precipitation periods, the filtering routine will be run with the use of a minimum threshold value, which is typically the weighing accuracy of the SFL-300 scale, the PL-50. This value for the PL-50 is ± 7 g. The resulting cumulative precipitation will be used as the reference where other consecutive increased or decreased oscillation thresholds' cumulative precipitation will be compared to. For the periods without recorded precipitation, the threshold will be set to a value where when investigating the nighttime fluxes, oscillations disappear, denoted previously as the maximum oscillation threshold value. This value will be as well the reference for the cumulative precipitation, where other oscillation threshold steps' cumulative precipitation will be compared to.

Once this step is completed, the deviations of the cumulative fluxes from the reference values will be calculated, and a value where such differences, represented as mixing error for periods with precipitation and noise error for periods without precipitation, will be determined

when both errors are the lowest and relatively equal in magnitude; meaning that the overestimation of the said fluxes will be compensated by the underestimation resulting from mixing errors and noise respectively. The following graph represents the results Hannes et al., (2015) derived. For this thesis, a similar graph with a specific value for the oscillation threshold filter for the SFL-300 at the CULS experimental field is expected;

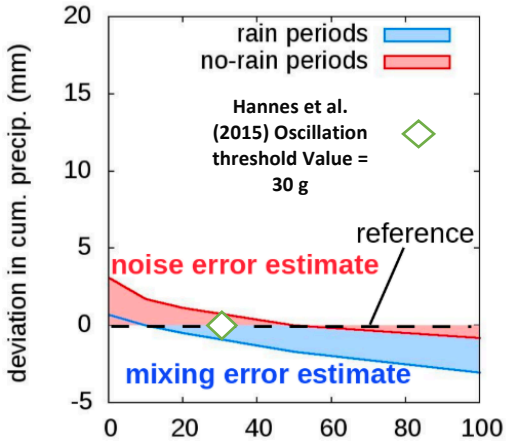

Figure 13: Representation of the effects of different oscillation threshold values on both types of noise (Hannes et al., 2015). *X- axis represents oscillation threshold steps in (g). On around 30 g, which is the value determined for the default filter settings, the magnitude of the noise error estimate is almost equal to the mixing error estimate, but opposite in magnitude. Speculating that they might compensate for each other's effects.*

4.6 Validation of the Filtering Process

In order to validate the filtering process, and to compare the resulting outputs of both the default and calculated filter parameters, a comparison must be conducted against reference values. Precipitation measurements by the 1 meter elevated TBR-MR3H tipping bucket rain gauge has been selected. The evaluation of the resulting cumulative precipitation is expected to provide a verification of the fitness of the default and calculated filtering parameters, a sense of which provides better results, and the magnitude of the difference if any. Precipitation has been chosen as the benchmark as evapotranspiration is complex and requires extensive modelling, in which's case lysimeters are typically used as the reference instrument for (ET_a).

It has to be also noted that despite choosing the rain gauge as the reference instrument, it yet still suffers from its own systematic and random errors which shall also be briefly investigated during the validation process. As wind under-catch is a significant source of systematic errors in these instruments (Mekonnen et al., 2015).

Nevertheless, the main statistical evaluation tools that will be employed to investigate relationships and properties between the datasets will include Root Square Mean Error (RMSE), Coefficient of Determination (R^2), and Standard Deviation (σ).

RMSE is one of the most used accuracy methods that are especially useful when dealing with sets using the same data but different methods. RMSE is often preferred as its result is typically on the same scale as the data. However, they show sensitivity to outliers (Hyndman et

Koehler, 2006). Typically, a reference set of data is compared to an observed one (O), the RMSE has the following equation (Nolz et al., 2013b);

$$RMSE(\hat{O}, O) = \sqrt{\frac{\sum_{i=1}^{n}(\hat{X} - X)^2}{n}} \tag{21}$$

Where ($\hat{X}$) is the reference data set and (X) is the observed, simulated, or calculated set. the RMSE has a unit similar to that of the measured set.

The normalized root mean square error NRMSE is a form of RMSE that has % as a unit, making it easier to understand. NRMSE is given by dividing the RMSE by the range of the measurements (Di Piazza et al., 2016);

$$NRMSE = \frac{\text{RMSD}}{X_{max} - X_{min}} \tag{22}$$

Where X_{max} and X_{min} are the largest and smallest values in the investigated series.

Standard deviation, SD, or (σ) is the dispersion of individual observations about the mean, characterizing the typical distance of an observation from distribution center or middle value. (σ) has the following formula (Barde et Barde, 2012);

$$\sigma = \sqrt{\frac{\sum(X_i - \mu)^2}{n - 1}} \tag{23}$$

Where X are individual values, μ is the sample mean, and n-1 is the Bessel correction, which the Excel STDEV.S function also calculates with.

Moreover, the coefficient of determination (R^2) describes the proportion of the total variance in the observed data that can be explained by the model. It ranges from 0.0 to 1.0, with higher values indicating better agreement. Assuming having two data sets, Y and X, (R^2) has the following formula (Legates et McCabe, 1999);

$$R^2 = \left\{ \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})(X_i - \bar{X})}{\sqrt{[\sum_{i=1}^{N}(Y_i - \bar{Y})^2][\sum_{i=1}^{N}(X_i - \bar{X})^2]}} \right\}^2 \tag{24}$$

Where the overbar denotes the mean of all the observations, for both data sets. And Y_i, X_i are each set's individual measurements.

5 Results

In order to examine whether different filtering parameters have any effect on the smoothing process, and the magnitude of the effect if existent, extensive raw data treatment and investigation of different filtering parameters was conducted. Tailored filtering parameters specific to the SFL-300 located at the experimental field were calculated, and several trials and runs of the filtering routine were executed to inspect the effects of the parameters. Leading to a deduction of an oscillation threshold value that is specific to the lysimeter used, soil conditions and properties, study period and climatology.

5.1 Raw Data Processing

5.1.1 Manual Filtration and Data Preprocessing

The lysimeter and its various components are prone to a variety of errors, including systematic ones such as logging errors. This results in some logging intervals missing their corresponding measurements. In the case of the examined period from the 12th of April 06:50 till the 23rd of September 09:59, several intervals were missing. The most prominent case was on the 7th of July, 2017 between 06:00 and 15:12, where the lysimeter did not log any measurement. More frequent cases were missing measurements on each 10th minute, which were the most noticeable on the 13th of June and the 23rd of September.

For the scattered missing intervals such as the ones not logged on the 10th minute, an interpolation function was applied to fill the missing data points. However, the 9 hours missing from the 7th of July data were removed using the manual filter of the software. Where this time period was inserted to the manual filter's .DAT file in the input folder of the software.

Furthermore, the maintenance logbook of the lysimeter was screened against the investigated period to identify grass cutting periods, maintenance, seepage tank emptying, and any other activities that would result in disturbances to the lysimeter measurements. Three main time periods displayed an incoherent trend that could be attributed to external disturbances. These events can be seen marked on figure 12, and are as follows;

1. 3rd of May: A sudden increase in a 1-minute interval in the mass of SWW from 5.17 kg to 6.16 kg,
2. 9th of May: A drop in the SWW mass from 5.94 Kg to 3.79 Kg between 15:41 and 15:43,
3. 29th of June: A significant increase in the LYW mass between 07:30 and 13:50,

After revising the logbook for any maintenance or grass cutting activities on those days, there was no record of any activities around the lysimeter on the 3rd of May, in addition to no precipitation events derived from the rain gauge. But an increase of 1 kg during a single minute interval is not possible, therefore this was considered an outlier and the value for that interval was interpolated. The 9th of may drop was registered in the logbook as an emptying of the seepage bottle. Hence, this drop in mass was added to all the subsequent data points to correct the data as seen in figures 14 and 15. A summary of the main events registered in the log book can be found in appendix 1.

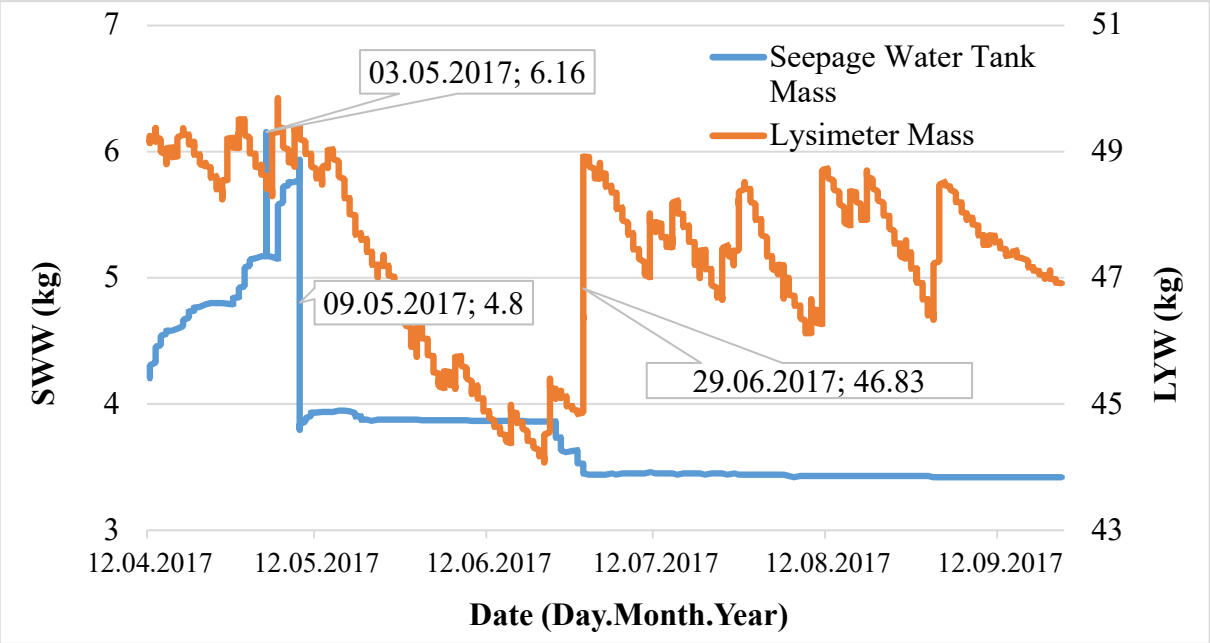

Figure 14: Lysimeter (LYW) unfiltered raw measurements and Seepage water tank (SWW) unfiltered raw measurements before any pre-processing or filtration for the entire investigation period (12th April - 23rd September).

The 29th of June significant increase was not registered in the maintenance logbook. However, after studying the meteorological data for the corresponding day, the increase was attributed to the longest rainy period in the investigated duration and was registered by the rain gauge in the experimental field. Therefore, the data for this date was kept and no pre-processing was needed.

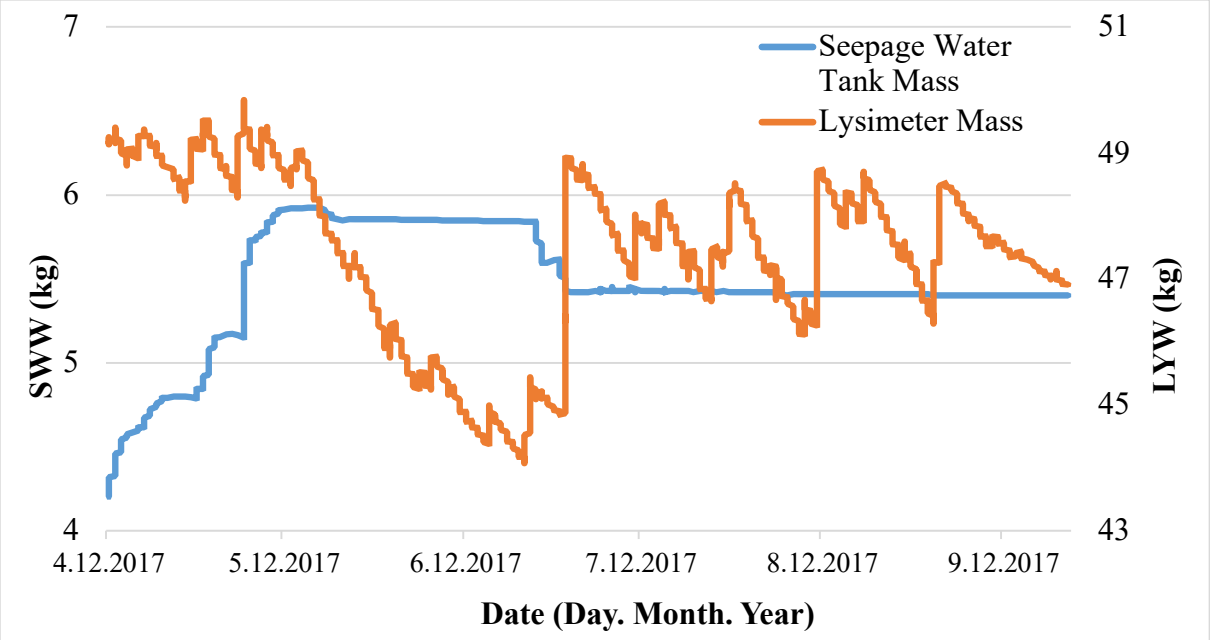

Figure 15: Lysimeter (LYW) and Seepage tank (SWW) raw data after removing outliers and applying the manual filter (Pre-processed).

This identification of heavy external influences on the measurements is crucial to the filtering process, as such outliers could have a strong influence on the filtering process and are also hard to recognize (Hannes et al., 2015). This effect and influence on the filtered data can be seen in figure 16, where the raw data was filtered without any pre-treatment and without the application of the manual filter. Hence, resulting in a mass difference of around 2 kg between the raw summarized (LYW+SWW) data, and the filtered summarized mass. This 2 kg difference well corresponds to the mass of water emptied from the seepage tank on the 9th of May.

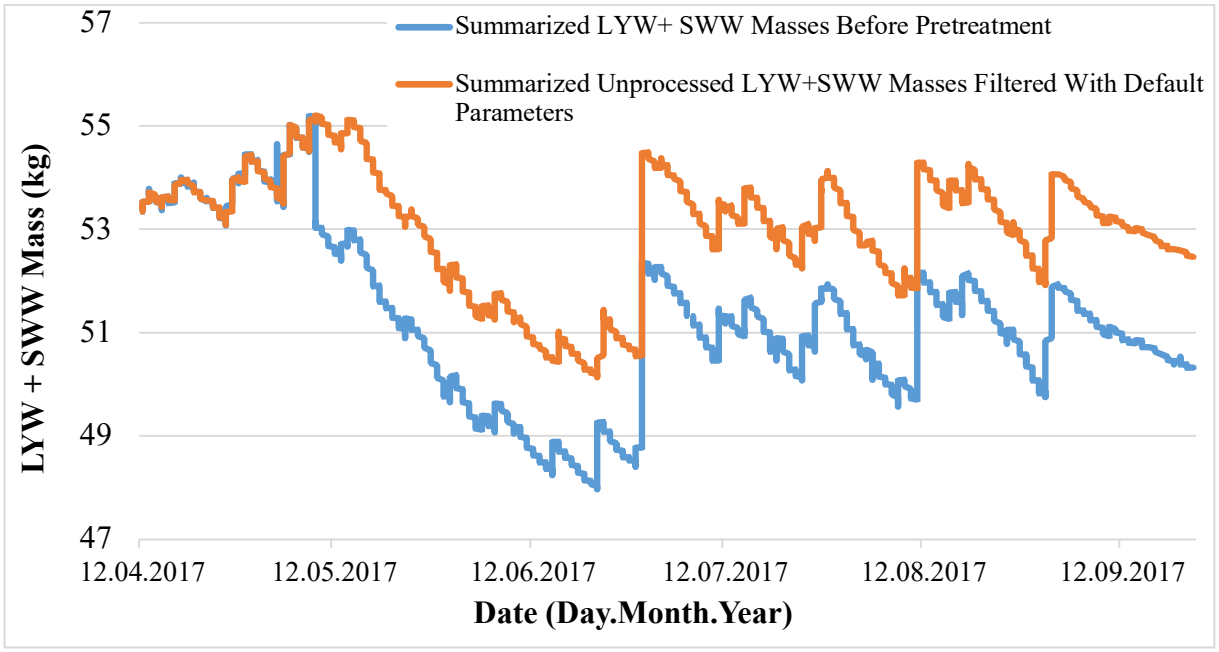

Figure 16:The difference in mass between Summarized unprocessed raw LYW + SWW data against their filtration output using the default filtration parameters of the software developed by Hannes et al.,(2015) as a result of not removing significant outliers and compensating for maintenance activities.

5.1.2 Meteorological Data

5.1.3 Temperature and Precipitation

After screening the LYW and SWW data for any outliers and irregularities requiring manual filtration, interpolating the missing data series, and compensating the masses of seepage bottle emptying activities in accordance with the field lysimeter maintenance log, the summarized masses of the lysimeter and seepage water tank were illustrated against the main governing meteorological data, namely rainfall and temperature, which have the most direct effect on the precipitation and evapotranspiration measurements of the lysimeter.

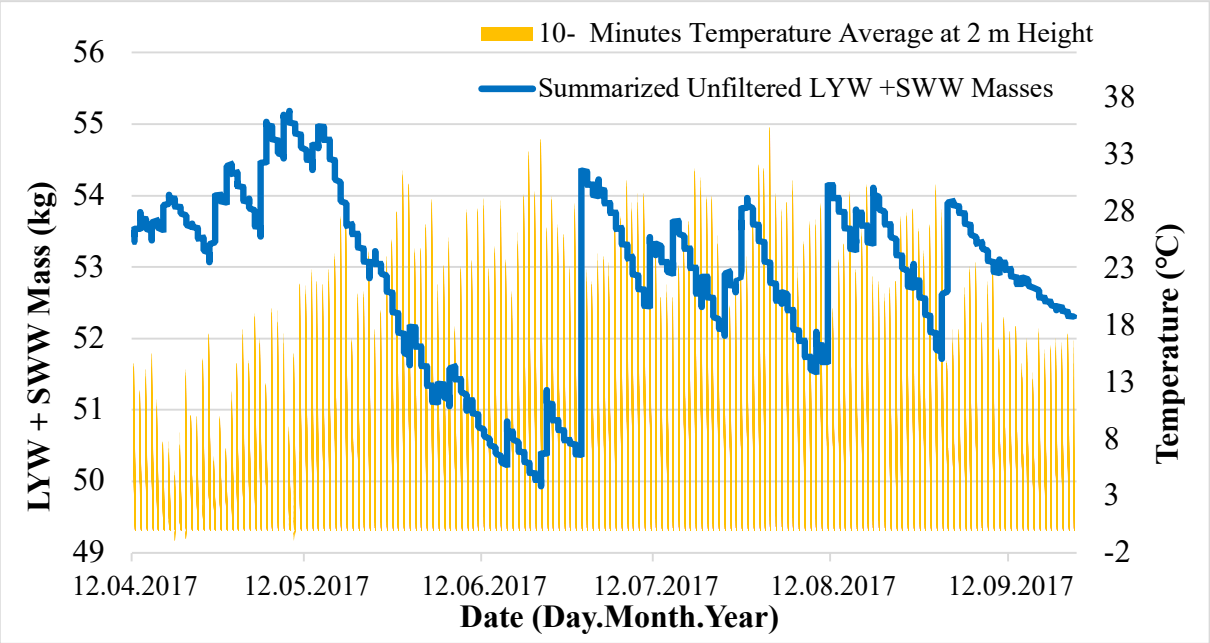

Figure 17: Entire examination period (12 April – 23 September) summarized unfiltered masses of LYW + SWW compared against 10- minute interval recorded temperature values.

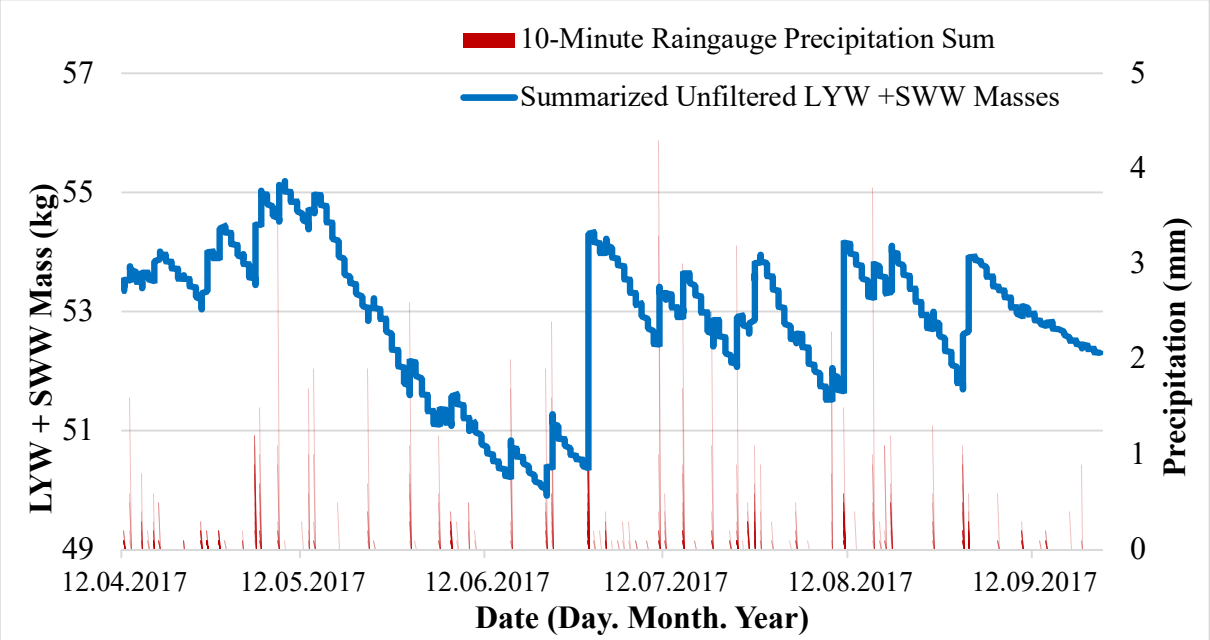

Figure 18: Entire examination period (12 April – 23 September) summarized unfiltered masses of LYW + SWW compared against 10-minute interval summed recorded precipitation using the rain gauge at the CULS experimental field.

Figures 17 and 18 above show the behavior of the LYW and SWW measurements as a summarized total with the alteration of the meteorological conditions. This was examined on a shorter monthly time scale to ensure that any jump, which is an increase of mass attributed to precipitation, is occurring due to a rainfall event or precipitation and not due to mechanical or external disturbances. The same is valid for mass drops, translating to evapotranspiration

events, where in figure 17 an evident inverse relationship can be seen between summarized mass declines and increasing temperatures.

Figures 19 and 20 below are an example of a closeup for the month of June. A gradual decrease in mass can be seen corresponding to the increasing and higher temperatures of the month, a good example is seen between the 15th and 22nd of June, where the summarized mass is gradually decreasing due to increasing temperatures. On the other hand, a spike in the raw summarized mass can be seen around the 16th of June despite the day having temperatures reaching around 25 ℃, this would normally be a strange phenomenon. Nonetheless, when this particular date's readings are screened against the precipitation data provided by the rain gauge as seen in figure 20, it can be clearly noticed that there was a series of precipitation events, with its highest reaching over 2 mm/ 10 minutes. The same is applicable to the 29th of June jump in recorded summarized mass, which is the biggest increase in the entire examination period. This jump also corresponds to the longest precipitation event recorded by the rain gauge, lasting between 07:30 and 13:50.

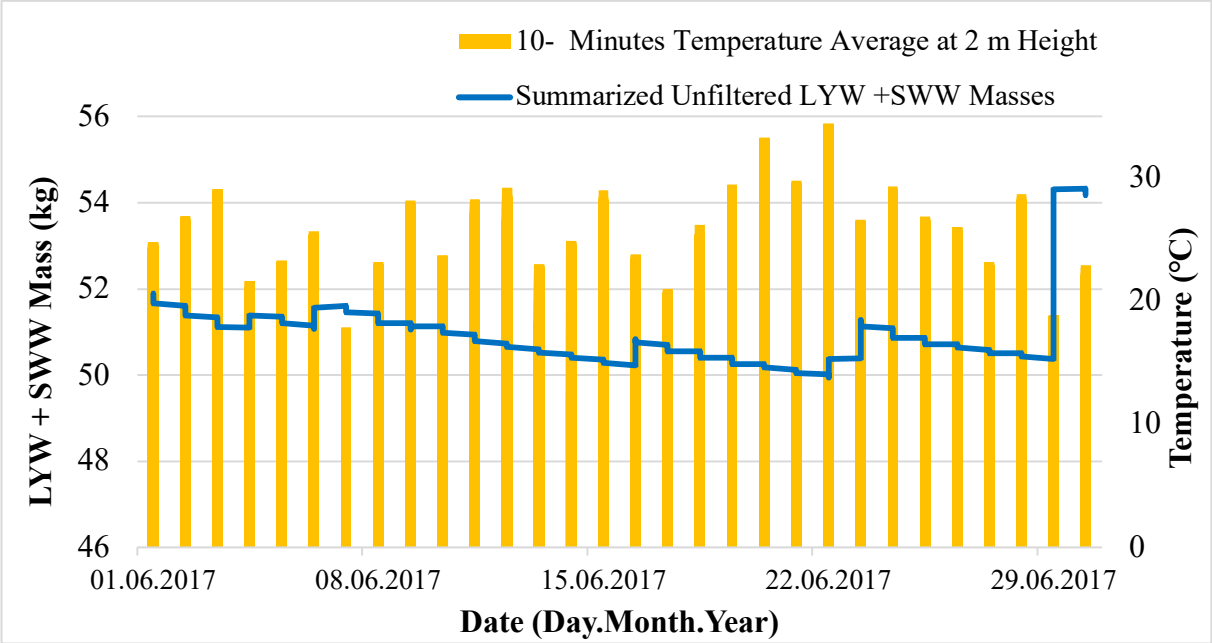

Figure 19: Summarized LYW+SWW unfiltered masses against 10-minute averaged temperature data for the month of June, 2017.

Having identified the outliers in the data, logging errors, and screened the SFL-300 field lysimeter maintenance log book against the data series to identify any defective or outlying data points, in addition to justifying significant drops or increases in masses through the use of the meteorological data, the raw data has been pre-processed and the first step of the comprehensive filtering routine developed by Hannes et al. (2015), the Manual Filter, have been applied.

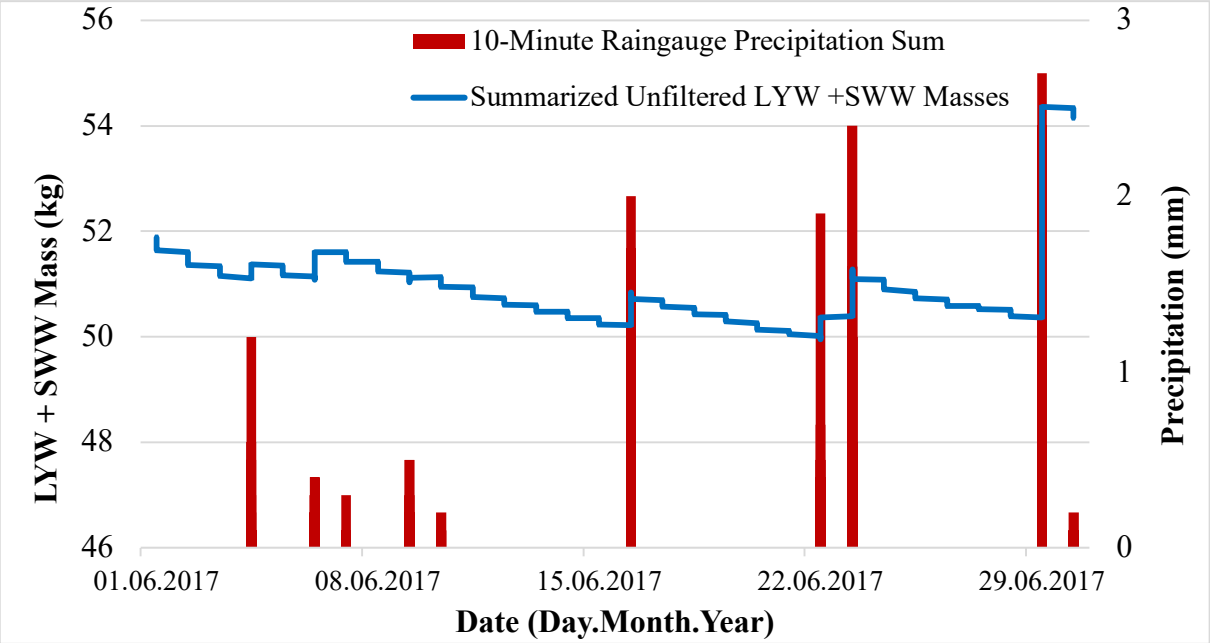

Figure 20: Summarized LYW+SWW unfiltered masses against 10 minutes summed precipitation data recorded by rain gauge for the month of June, 2017.

5.1.4 Wind Speed

Examining the raw LYW and SWW data with reference to precipitation and temperature is crucial to determine the soundness of the raw data and to ensure removing or compensating for any outliers. Another principal factor affecting the data is the wind speed. As the effects of precipitation and temperature can be identified on a macro scale that is distinguishable when examining monthly data, or even the entire examination period, wind speed's effects must be examined on a micro scale, meaning on a scale averaging around an entire day to have a good visibility of the accompanying effects.

Figure 21 portrays the correlation between wind velocity and the noise in the raw data that usually would not be visible on a larger time scale. As wind velocity exceeds 2 m/s, noise in the raw data starts to be more visible and increasing in magnitude. This is clear between around 08:30 and 18:00 on the 12th of June. This determination is of a particular importance due to the effect of wind gusts and velocity on measurement accuracy, where wind speeds exceeding 5 m/s could decrease measurement accuracy by a third. Making the interpretation of the water balance components in shorter time intervals a difficult task (Nolz et al., 2013a; 2013b).

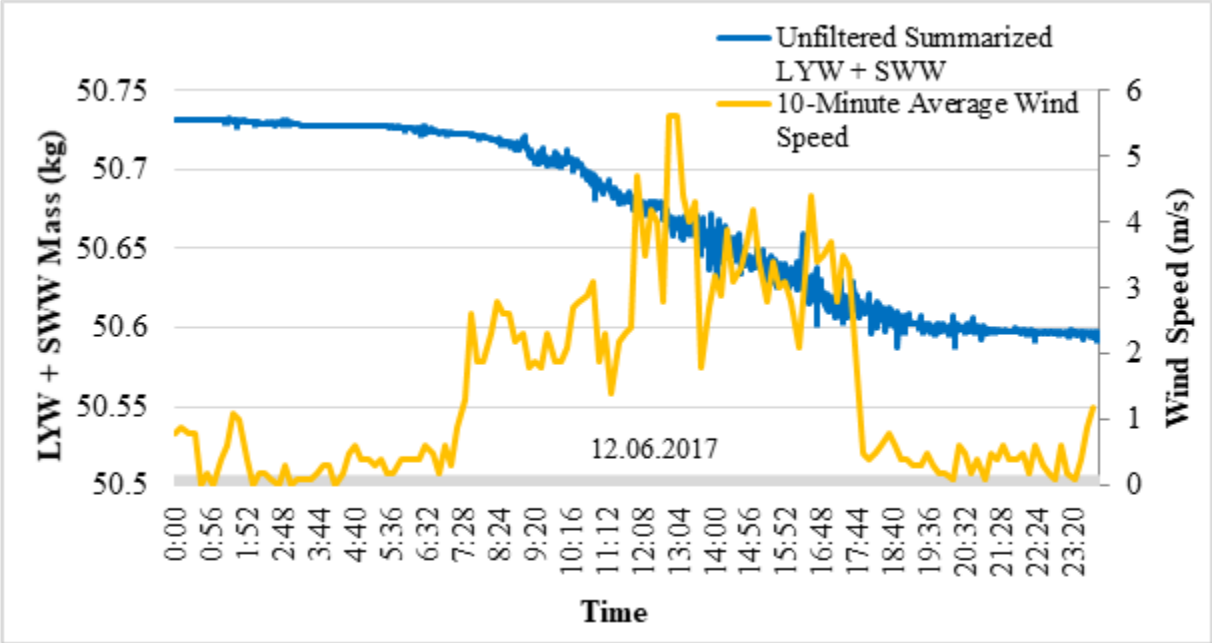

Figure 21: The effect of wind on the noise in the summarized raw data. Displayed in 24 hours time scale, June 12th, 2017.

5.2 The Calculation of Threshold Filter's Parameters

The software defines two spectrums for the threshold function, an upper one consisting of the LYW mass, and a lower one for the SWW mass. With each one having a maximum (δmax) and a minimum (δmin). Resulting in a total of 4 threshold values for the LYW and SWW masses. Meaning that for the upper boundary fluxes, which translate to the mass changes of the LYW masses over each recorded interval, only increases smaller than the LYW (δmax) values will be recorded, and any drops smaller than the LYW (δmin) values will also be recorded. The same is applicable for the SWW measurements.

Figure 22 is a demonstrative example to show the relevancy of such threshold parameters. It is based on April 12th data between 06:50 and 23:20 with a few synthetic data points added as spikes. The arrows on the synthetic spikes as seen in the figure represent merely an illustrative example of how the threshold filter will function, removing any data points having values higher than the permissible LYW or SWW (δmax) for mass increases, and LYW or SWW (δmin) for mass decreases. When such threshold value is exceeded, the filter removes these data points and substitutes for them by linear interpolation (Hannes et al., 2015).

Such limits for the upper and lower boundary fluxes are not arbitrarily chosen or inferred from the raw data. Each limit, maximum and minimum, for the upper and lower boundaries is based on physical characteristics and properties specific to the site, lysimeter's components specifications, and meteorological parameters.

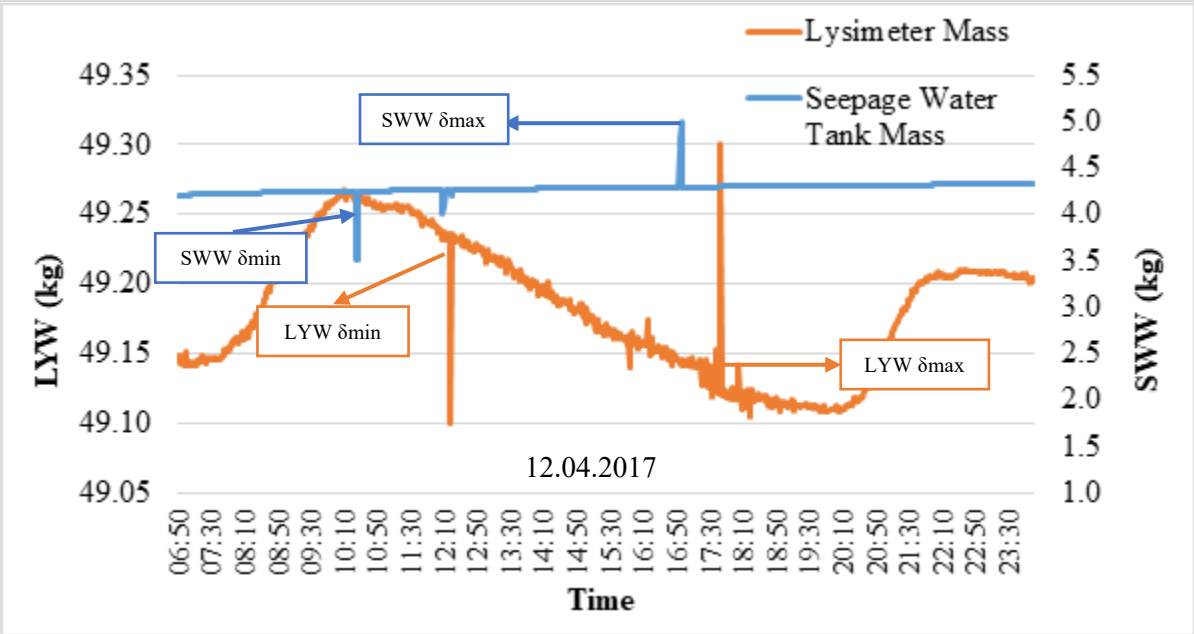

Figure 22: An illustrative example of the process of the threshold filter for LYW and SWW mass measurements. The data is based on April 12th measurements with a few synthetic data points added that can be seen as spikes. Arrows represent an illustrative limit where if the data point's value is exceeded the measurement will be discarded and a new value will be interpolated.

As any mass increase in the seepage water tank between any two measurement intervals, or any flux, must lie below a specific reasonable value. Any measurement exceeding this value is considered an erroneous measurement. For this maximal lower boundary limit value, the soil's maximal drainage rate, or maximum saturated hydraulic conductivity is the threshold value. As the SFL-300 is equipped with a bi-directional pump that returns water to the soil monolith of the lysimeter if it becomes drier than surrounding soil, any decrease in the seepage water tank's mass can be attributed to this, which is a mimicking process of the natural capillary rise (Schrader et al., 2013)

As for the upper boundary fluxes, a similar principle is applicable. The maximal allowable increase value for the LYW mass should lie below the maximum expected precipitation (P) rate. Such a value can be inferred from the rain gauge readings for the examination period. Where if any registered flux by the lysimeter exceeds this maximum (P) reading from the rain gauge, it would be discarded. Moreover, any decrease in the upper boundary cumulative flux would be due evapotranspiration (ET), hence the minimal flux at the upper boundary value should reflect a maximal (ET) rate. Such a value can be extracted from evapotranspiration models or statistics provided for the examination year, or period if available (Schrader et al., 2013).

Table 2: Threshold filter parameters as denoted by the Software developed by Hannes et al.,(2015), their recommended values by the authors, and their corresponding physical property.

Threshold Filter's Parameters			
Hannes et al.,(2015) Parameter Name	**Software Defualt Paramter Value**	**Unit**	**Governing Physical Property**
Maximal Seepagewater-Flux	150	g/min	Maximum Saturated Hydraulic Conductivity Value at the Lysimeter's Location
Minimal Flux in Seepagewater	-150	g/min	Bi-directional Pump's Maximal Pumping Rate
Maximal Flux at Upper Boundary	1000	g/min	Biggest Precipitation value at a Measurement Interval
Minimal Flux at Upper Boundary	-100	g/min	Biggest Evapotranspiration value at a Measurement Interval

5.2.1 Maximal Seepage Water Flux – Saturated Hydraulic Conductivity

As the SFL-300 lysimeter utilized in this research is located at the experimental field vicinity of the Czech university of Life Sciences in Prague (CULS). Several experiments and measurements have been carried out over the years with respect to measuring the saturated hydraulic conductivity. Available literature from CULS indicate that the loamy carbonate Chernozem on loess soil at the field has a saturated hydraulic conductivity (Ksat) value ranging between 1 x 10^{-7} and 7 x 10^{-5} m/s (Doležal et al., 2018;Doležal et al., 2015)

Since the maximal seepage water flux represents the biggest permissible drainage rate (Schrader et al., 2013), the largest (Ksat) value is taken, which is 7 x 10^{-5} m/s. To convert this value to g/min, the area of the lysimeter which equals 0.0707 m^2 needs to be accounted for, and the following equation is used;

$$7 \times 10^{-5} \frac{\mathrm{m}}{\mathrm{s}} \times 0.0707\, m^2 = 4.95 \times 10^{-6} \frac{m^3}{s} = 4.95 \times 10^{-3} \frac{kg}{s} = 296.94 \frac{g}{min} \quad (25)$$

To add a safety factor as suggested by Hannes et al. (2015), and to make subsequent calculations easier, this value is rounded up to 300 g/ min.

Furthermore, to study the individual effect and significance of the Ksat parameter on the filtering process. The filter was run using the lowest Ksat value measured in the field, which equals 1 x 10^{-7} m/s, while maintaining all other parameter values as calculated. The effect is portrayed in figure 23.

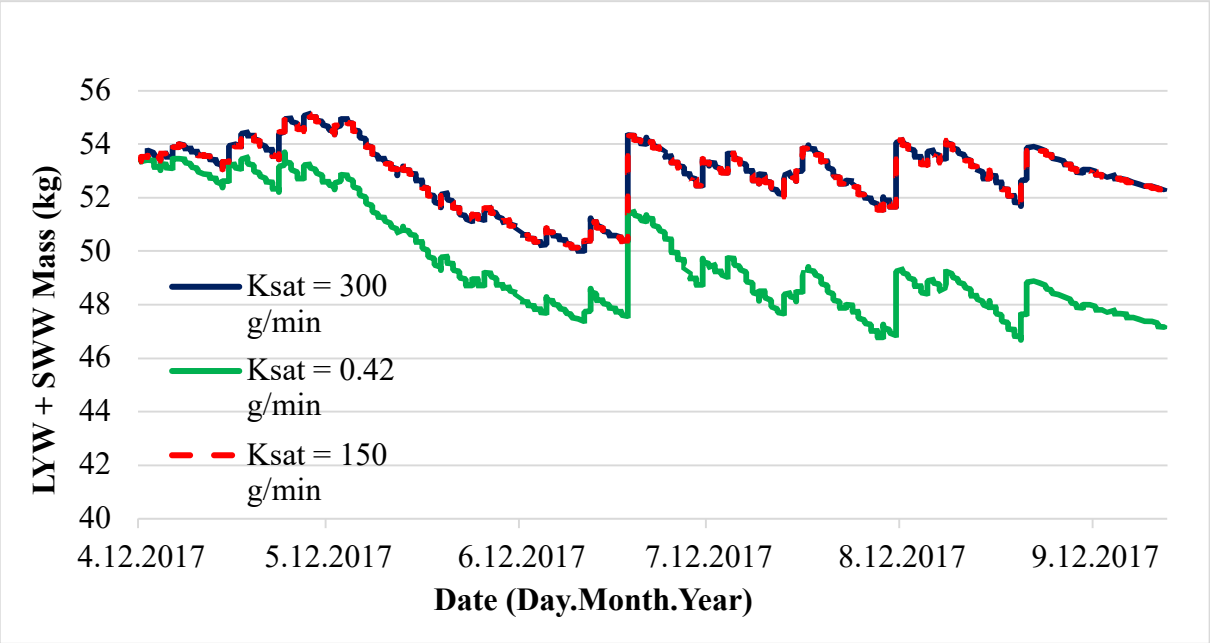

Figure 23: Effects of using different Ksat values on the filtered summarized mass while maintaining other parameters as calculated. Ksat= 300 g/min is as calculated using the highest Ksat value measured in the field. Ksat = 150 g/min is the default software parameter value. And Ksat= 0.42 g/min is based on the lowest Ksat value measured at the field.

The effect of using the lower value in Ksat does not only result in a substantial decrease in the summarized lysimeter mass, but also in translating unrealistic drops as real measurements. Which in turn is the cause for such continuing decrease in the summarized (LYW + SWW) masses.

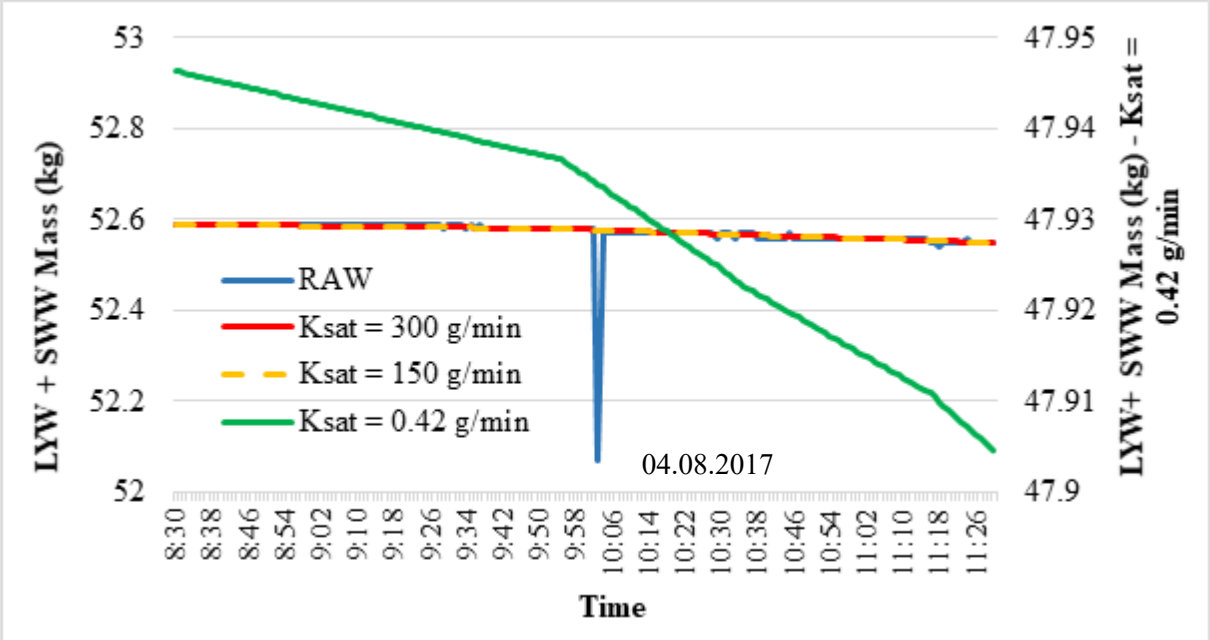

Figure 24: The effect of different Ksat values in treating unrealistic outliers, other filter parameters are maintained as calculated. 4th of August, 2017. 08:30 – 11:30.

5.2.2 Minimal Flux in Seepage Water – Maximal Pumping Rate

In older gravity lysimeters, weight decrease of the seepage water tank is practically impossible (Schrader et al., 2013). However, as recent advancements allow for mimicking the natural capillary rise phenomenon, a bi-directional water pump can pump water back into the lysimeter when needed. Hence the maximum possible increase corresponds to the bi-directional pump's pumping rate.

As the SFL-300's User Manual did not contain any information about the pump's specification, a request for this information was communicated to Meter Group AG, the manufacturers of the SFL-300, through their website. A response was received on the 3rd of June 2020 at 16:32 via email (personal communication) stating that the bi-directional pump has a maximum pumping rate of 0.0132 m3/hr. In order to be able to use this value in the software, it must be converted to g/min. Unlike the process in equation 25, no compensation for the area of the lysimeter is needed. Hence, the remaining steps can be followed as in equation 25 resulting in a value equal to 220 g/min. As recommended, the value used in the calculation shall be 250 g/min, as an addition of a safety factor.

5.2.3 Maximal Flux at Upper Boundary – Precipitation

While rain gauges still suffer from their own systematic and random errors such as wind errors (Hoffmann et al., 2016; Schrader et al., 2013), for the purpose of the determination of the maximum flux in a defined examination period, the rain gauge is a decent source. The rain gauge of the weather station that is situated in CULS's experimental field provides precipitation sums over 10-minute intervals. When the meteorological data is examined for the largest precipitation event for the period between the 12th of April and the 23rd of September, the 11th of July had the biggest precipitation sum for the time interval between 08:20 and 08:30 of a value that equals 4.3 mm/ 10 minutes, or 25.8 mm/ hr.

In order to convert this value's unit to the software's g/min unit, the following conversion is applied;

$$25.8\ \frac{\text{mm}}{\text{hr}} = 25.8\ \frac{\frac{L}{m^2}}{hr} \times 0.0707\ m^2 = 1.824\ \frac{Kg}{hr} = 30.4\ \frac{g}{min} \tag{26}$$

As aforementioned, to round up the number and to add a safety factor to the calculation, 35 g/min was used in the software parameters.

5.2.4 Minimal Flux at Upper Boundary – Evapotranspiration

As the source of possible loss of water from the upper flux typically occurs through evapotranspiration, choosing a maximum possible (ET) value specific to the investigated period defines the limit for any decrease in the upper boundary fluxes. Nonetheless, such a value is not easy to retrieve or calculate compared to the previous parameters. There exists models such as the Penman-Montieth procedure, eddy covariance, and even remote sensing that could provide evapotranspiration information (Allen et al., 2011; Gebler et al., 2015).

Hence, extracting a maximum evapotranspiration rate is laborious and does not lie within the scope of this thesis. Instead, the software was run using the default value for evapotranspiration as suggested by Hannes et al. (2015) , where the filtered data was showed good accordance with the summarized raw data.

Furthermore, and to investigate the effect of the maximum ET value on the filtration process, the comprehensive filtering routine using the calculated threshold parameters was run with different ET values to ensure that the filter is not sensitive to the choice of the ET value. The selected examination ET values were chosen based on average ET values for the Czech Republic measured on a long duration just to obtain an idea of a possible range for ET_{max}. In addition to adding other arbitrary values to inspect its impact on the filtered output.

Figure 25 is from a report on the drought in the Czech Republic and its effects. Using Penman's relationship, the report provides the average ET values based on data from 1961-2010 (Rožnovsky, 2012). Furthermore, it can be seen that Prague in particular is one of the driest areas of the Czech Republic (Smith et al., 1996). A drought prone trend in the western part of the country has been noticed through 1961 – 2000 (Manatsa et al., 2008). Moreover, figure 25 displays an annual average ET value slightly higher than 750 mm (Rožnovsky, 2012). Which when used to calculate the ET value in g/min using equation 25 above, results in a value of 0.1 g/min. To round up the value and consider a safety factor, a value of 1 g/min was used.

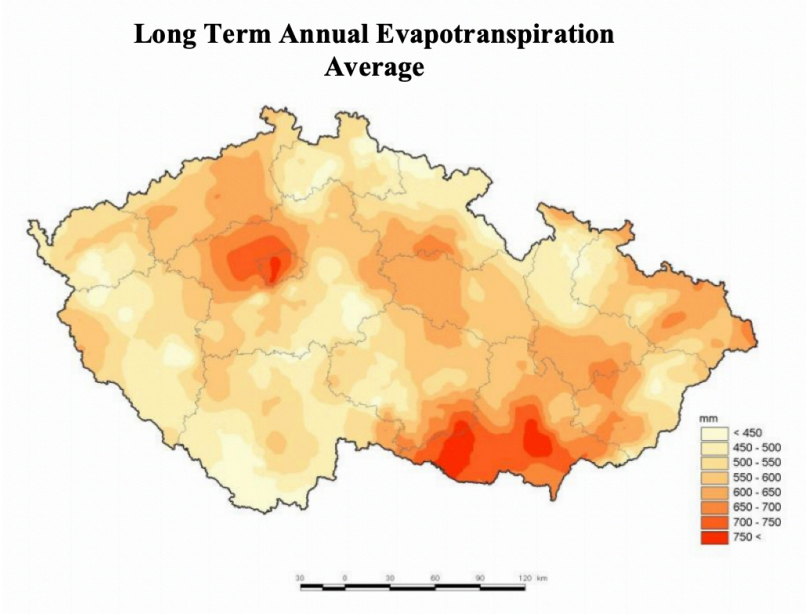

Figure 25: Evapotranspiration long-term annual averages (1961-2010) in the Czech Republic (Rožnovsky, 2012).

When examining the effect the ET value has on the entire studied period, it becomes hard to distinguish if there is any effect at all as all the data series overlay each other. Therefore, the time scale has been reduced to only 24 hours to inspect if there is any effect, and if so, it's magnitude. 3 random days for each month were investigated, figure 26 below is an example of the behavior of the filter when different ET values are used, having all other parameters as calculated.

June 12th data was also used for this illustration as it contains a substantial amount of noise, making it easier to distinguish between different ET values effects. It can be noticed that high ET values, e.g. 300 g/min lead to serious misestimation of the occurring fluxes. Moreover, when examining lower ET values such as the calculated 1 g/min, it can be seen that they show closer behavior to that of the raw data, until the raw data starts becoming noisy. During noisy period, they begin oscillating with the noise and around 16:00 where there is an evident spike, these ET values overestimate the fluxes resulting in filtration errors.

ET= 3 g/min is showing a good accordance with the raw data, however during noisy periods it also causes the filter to oscillate with the noise, even when the oscillation threshold filter is applied. This sight is noticed on other investigated days, contradicting the purpose of this filtration process and leading to misestimating the occurring fluxes when studied over shorter time scales.

Despite some ET values showing better accordance with the data on some periods, they also show distinctive disassociation on others. The only value that is providing a satisfactory trend in the filtered output in all investigated periods is the ET value of 100 g/min, which is the default software value suggested by Hannes et al. (2015). Therefore, the subsequent filtering steps and final filter results will rely in its minimal upper boundary flux on this value.

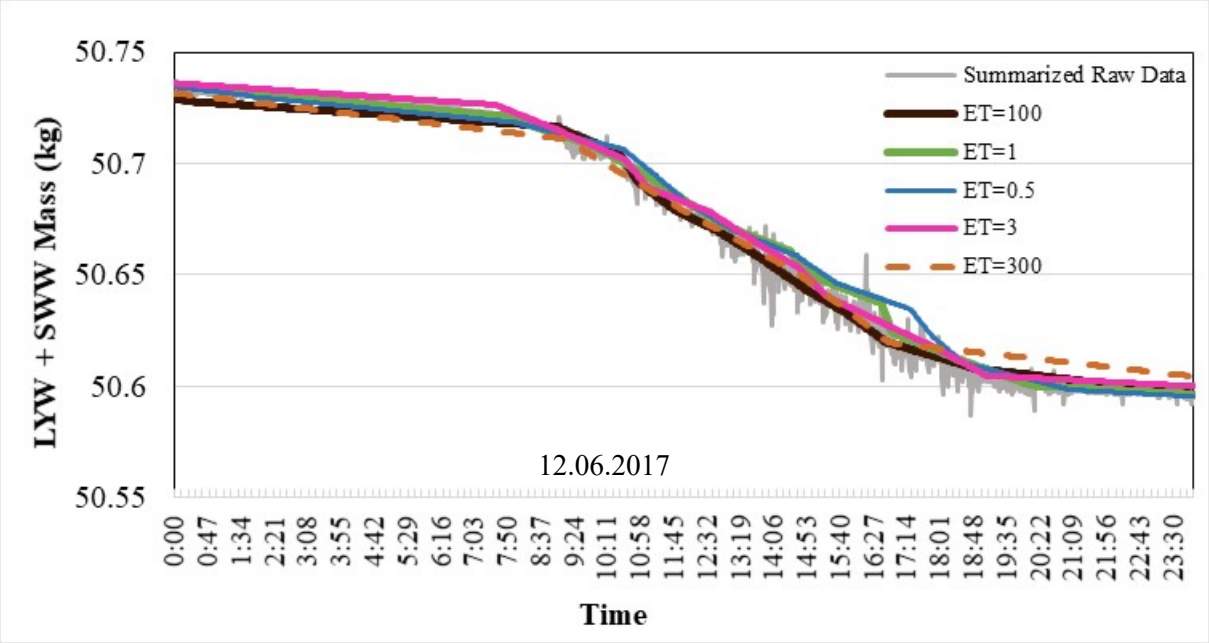

Figure 26: The filtered summarized LYW+SWW mass using the calculated parameters and different ET values plotted against the raw summarized data (12th of June, 2017. 00:00-23:59).

5.3 Choosing the Width of the Time Window

The choice of a suitable window width for the subsequent filtering steps, the median average and the smoothing filter, is more central to the filtration process than the previous steps (Hannes et al., 2015). Unsuitable time windows could carry a systematic misfit in the filtered output when sudden and rapid changes happen, e.g. a short intense precipitation event (Schrader

et al., 2013). Schrader et al. (2013) demonstrated this effect when the moving average time window changes from 10 minutes (11 data points) to 30 minutes (31 data points).

For the median filter's time window, Hannes et al. (2015) have used a value of 15 minutes and stated that it is sufficient. However, the software's default value is 20 minutes as it is tailored to be used with different data sets. Where for this step the difference should be negligible. Nonetheless, 15 minutes window was kept and used for the purposes of this thesis based on the authors' recommendation.

As for the smoothing filter, the authors recommended a window of 15 minutes at a maximum as the subsequent oscillation threshold filter allows the preservice of the high temporal resolution of the process. Which is crucial to maintain the high dynamics of precipitation events of less than 20 minutes in high evapotranspiration periods (Hannes et al., 2015).

There is a substantial sensitivity and an interrelation between the calculated fluxes by the smoothing algorithm and the smoothing window width (Schrader et al., 2013). This window width needs to be able to capture rapid changes while avoiding the misestimation of the occurring fluxes (Doležal et al., 2014). The effect of only using the MA to filter the data can be seen in figure 27. A smoothing window of 15 minutes filters the data relatively well. However, the filtered mass still oscillates with the noise to some extent. During sudden spikes or drops in the raw data, the MA seems to follow their trend. Moreover, higher MA time windows lead to serious misestimation and deviation of the resulting data.

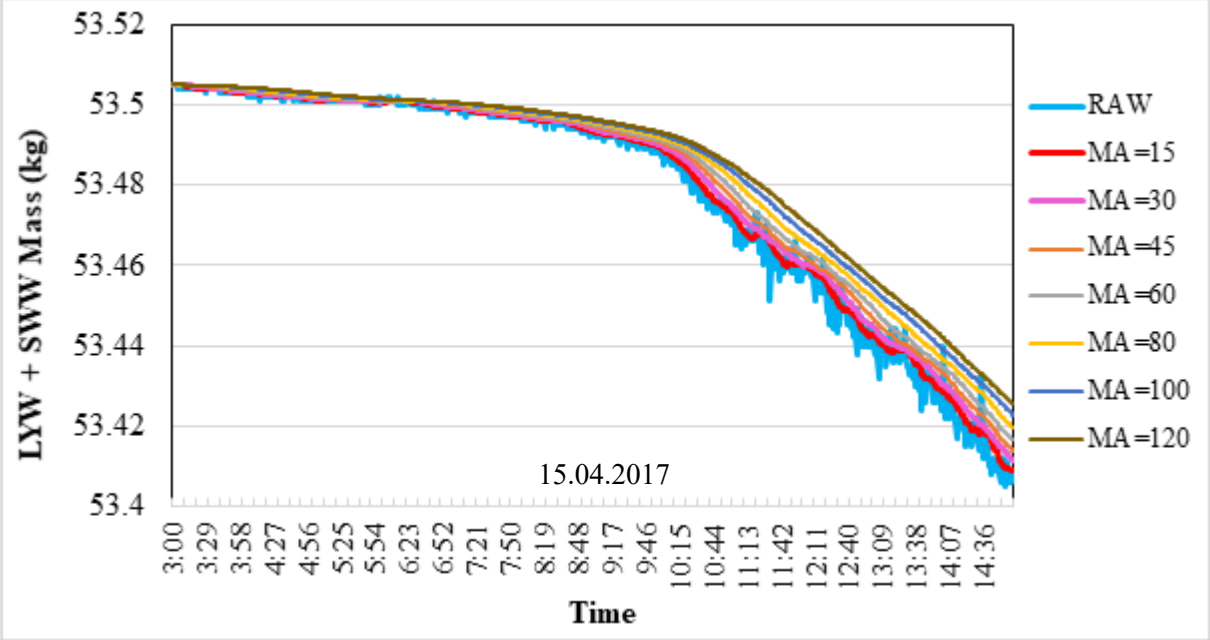

Figure 27: The effect of different smoothing time windows using the MA filter (units in minutes) on the resulting summarized LYW +SWW masses, without any other filtering steps. 15 April, 2017, 03:00 – 15:00.

Figures 28 and 29 confirm that once unrealistic disturbances and outliers are removed via the manual filter and the subsequent threshold filter, the MA provides an acceptable outcome conditional to the application of an oscillation threshold filter to remove noise on a high temporally resolved scale.

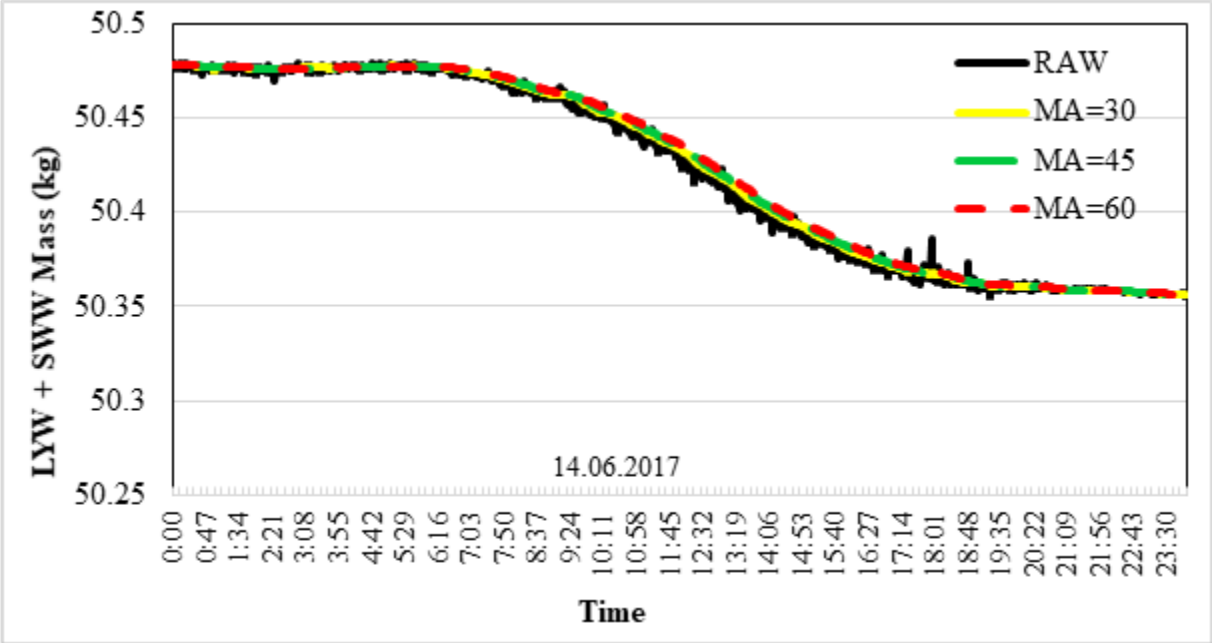

Figure 28: Effects of filtering the data with only the MA. The effects of overestimating the data when increasing the window width. 14th of June, 2017, 00:00 – 23:59

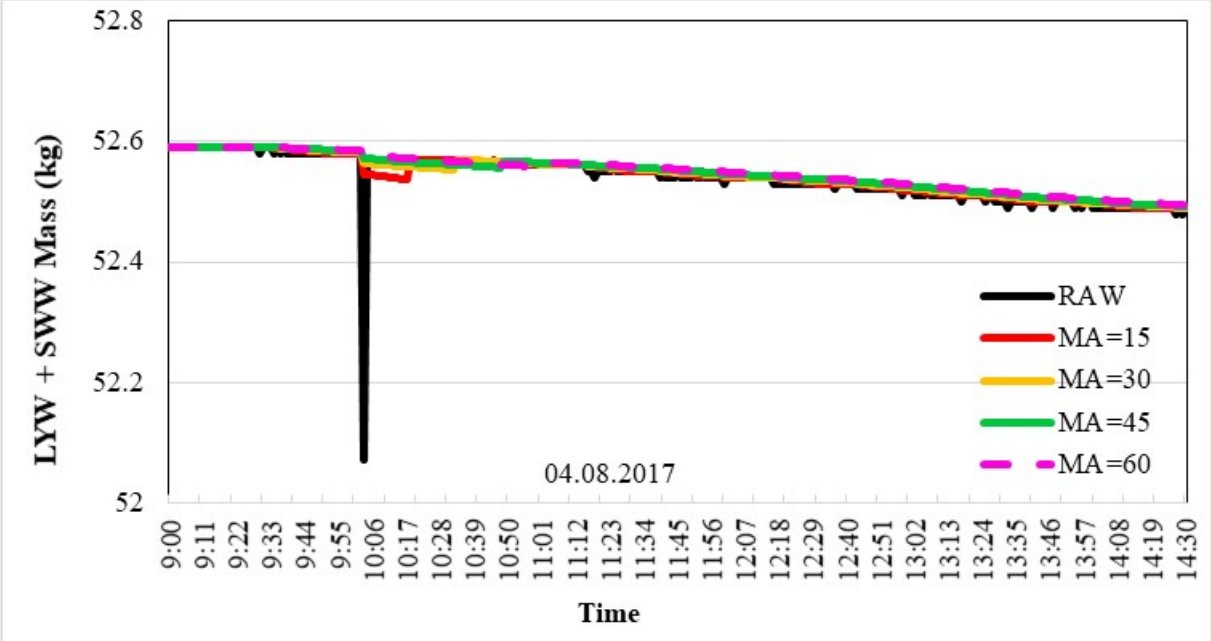

Figure 29: Effects of filtering the data with only the MA. Shorter MA windows (minutes), maintain outliers in the filtered output. 4th of August, 2017, 09:00 – 14:20

5.4 Parameter Selection for the Oscillation Threshold Filter

The oscillation threshold filter needs to be taken as large as required to eliminate the noise, and as small as possible to ensure that real fluxes are not filtered with the noise. Therefore, the choice of an oscillation threshold value needs to be precise in order to retain slow

processes and avoid the underestimation of precipitation (Hannes et al., 2015). To reach this value, its range based on the lysimeter used and data set examined needs to be defined.

5.4.1 Oscillation Threshold Minimum and Maximum

The minimum possible value for the oscillation threshold filter has been recommended by Hannes et al. (2015) to be assigned as 0 g. implying that each and every remaining oscillation after the previous filter steps is registered as an actual flux. Nevertheless, the determination of the maximum value is not quite as direct. The maximum oscillation threshold value is obtained through observing nighttime data over a span of hours and choosing the oscillation threshold step at a height where nearly all of the nighttime oscillations disappear.

Having obtained a value for the other filter parameters, the filtering scheme is run multiple times with varying values for the oscillation threshold filter to observe the nighttime fluxes. Since the minimum value has been assigned as 0 g, this value will be the first step. The subsequent steps will increase by a magnitude of 5 g with an exception of using an additional step of 7 g as this value reflects the scale accuracy. Hence, resulting in a total of 8 steps until 30 g, as it is not expected that the SFL-300 with its considerably smaller area would allow for real fluxes of more than this value.

As the determination of the maximum oscillation threshold value reflects the maximum points at which such oscillations vanish, it is hence dealing with the magnitude of the noise error rather than the mixing error which occurs during precipitation events. Therefore, the nights that must be chosen need to be ones without recorded precipitation events. In order to obtain the dates with no recorded precipitation at night, the precipitation data from the CULS experimental field's rain gauge is examined against the raw data. The time scale of the investigation starts at 21:30 of day n_x, until 06:30 of day n_{x+1} due to different months having different sunrise and sunset times.

Evaluating a single or a few nights is not sufficient as a low oscillation threshold value, e.g. 5 g, could remove all the oscillations at during the course of one night, but still display oscillations in another with more noise in its measurements. Thus, for each month 2 random days were chosen to be examined, with an additional one chosen at a night where the wind records show substantial wind velocities. Resulting in a total of 3 examined nights per month, and 18 in total. Constituting around 11% of the days of the investigation period.

Out of the 18 examined days 4 displayed vanishing oscillations at an oscillation threshold value of 5 g, 8 at 10 g, and 6 at 15 g including the 20 – 21 April night displayed in figure 30.

Table 3: The range of the Maximum and Minimum Oscillation Threshold value.

Range of the Oscillation Threshold Value		
	Hannes et al., (2015)	**CULS's SFL-300 2017 Data**
	(g)	**(g)**
Minimum Oscillation Threshold	0	0
Maximum Oscillation Threshold	50	15

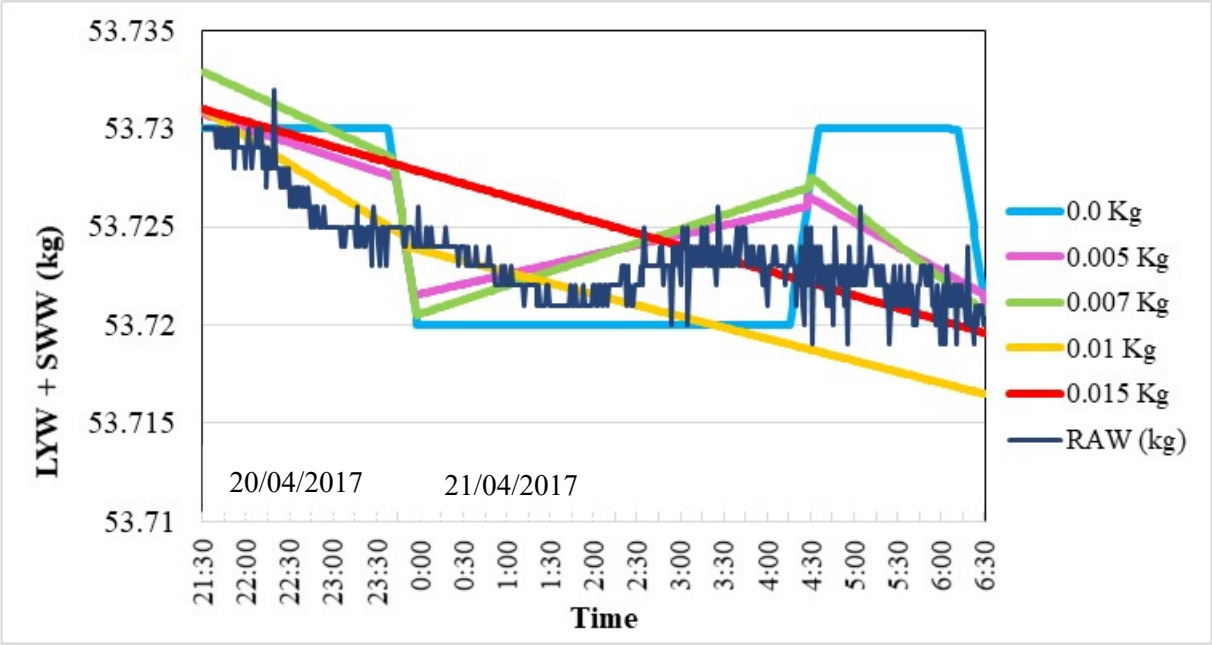

Figure 30: Investigating the maximum oscillation threshold value based on filtering the data with increasing oscillation threshold steps until the disappearance of nighttime oscillations. 20-21/04/2017 21:30 – 06:30.

5.4.2 Partitioning of the Data Set Based on Precipitation Features

It is speculated that the mixing error's effect in underestimating the cumulative precipitation during precipitation events could be compensated by the noise error's effect in overestimating it. Nonetheless, such errors would still manifest in the data series in the time resolved fluxes (Hannes et al., 2015).

In order to investigate this effect, and to find an appropriate oscillation threshold value that lies somewhere in between these errors and between the maximum and minimum values of the oscillation threshold values, the data set was divided into periods with precipitation and periods without. The segregation was made based on choosing times when precipitation exceeds 1 mm/h, as during such periods the fluxes are typically positive and do not oscillate from positive to negative values. Therefore, they could provide an acceptable indication of the size of the mixing error. Any precipitation event with values lower than 1 mm/hr have been excluded as such times still contain relevant mixing and noise errors. On the other hand, periods with no registered precipitation display a dominant effect of the noise error, where the contribution to precipitation is very minimal such as fog, and dew formations (Hannes et al., 2015).

Each data point with rainfall fluxes exceeding 1mm/hr were chosen while also taking 15 minutes data points before and after their occurrence. This is to ensure that MA's averaging time does not blur out the occurring fluxes and lead to spreading their effects outside the assigned time window. This process was applied by Hannes et al. (2015) by taking between 200-250 minutes before and after the registration of such positive fluxes, as their investigation included the effects of larger averaging times, up to 180 minutes. But since such an investigation is not within the purpose of this thesis, 15 minutes before and after each event shall suffice.

During the investigated period, 58 days contain precipitation fluxes exceeding 1 mm/hr, with a total of 6,879 data points including the 15 minutes intervals taken before and after.

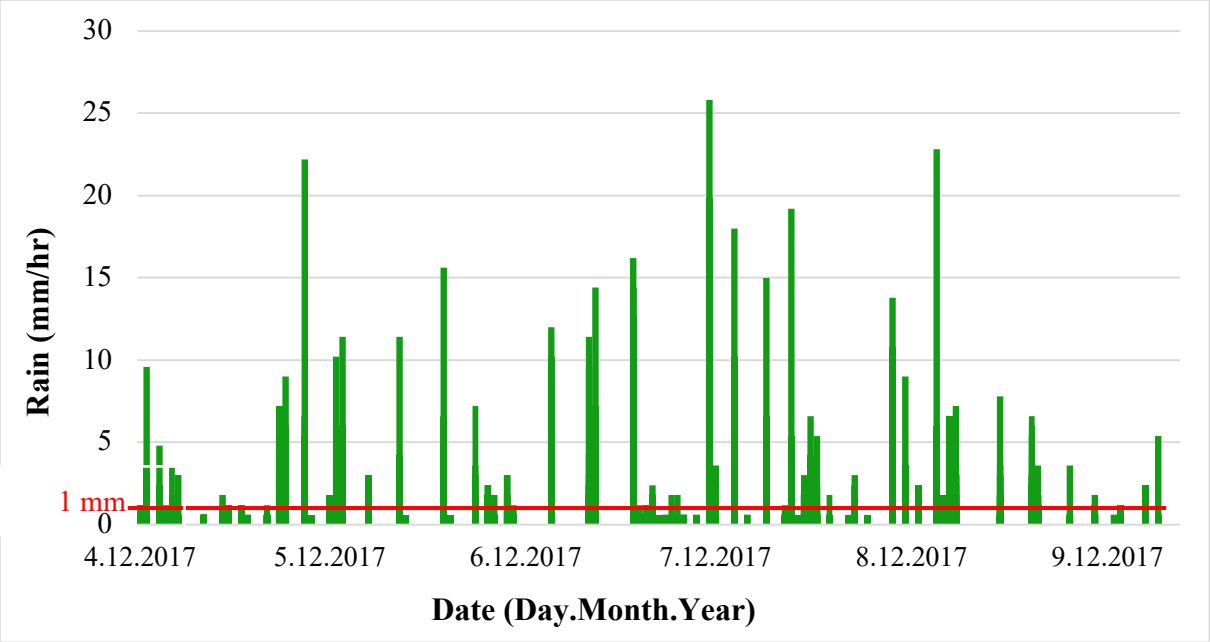

Figure 31:Precipitation fluxes based on the CULS Experimental field rain gauge used to define periods of precipitation exceeding 1 mm/hr. Periods exceeding 1 mm/hr (crossing the red line) are used for the investigation of the mixing error, while periods with precipitation less than 1 mm/hr are excluded since both mixing and noise errors are valid.

As for the no rain periods, they were strictly chosen with no registered precipitation. Resulting in a total of 235,541 measurements.

5.4.3 Deviation of Partitioned Data from Reference Values

To explore the magnitude of the effects of the noise and mixing errors on the resulting cumulative precipitation, the deviation of each data set's cumulative precipitation from a reference oscillation threshold value is calculated. For the rain periods data set, Hannes et al. (2015) recommend using the lysimeter's weighing accuracy as the reference oscillation threshold value since noise does not have a major impact in this case. Therefore, as the SFL-300 weighing accuracy is ±7 g, this is chosen as the reference value to which other oscillation threshold steps' cumulative precipitation are compared.

Furthermore, for the no-rain periods, where noise error is dominant, the maximum oscillation threshold value derived from the investigation of nighttime oscillations is used, which for this case equals 15 g.

Table 4: Cumulative precipitation values of reference oscillation thresholds and the deviation of each oscillation threshold step from the reference value, for both rainy and no rain periods. Using the Hannes et al. (2015) filtering software with the calculated parameters and a time window of 15 minutes.

Cumulative Precipitation Deviation from Reference Values				
	Mixing Error		Noise Error	
	Reference Oscillation Threshold (kg)	Cumulative Precipitation Value (mm)	Reference Oscillation Threshold (kg)	Cumulative Precipitation Value (mm)
REFERENCE VALUES	0.007	295.859	0.015	376.214
Osc. Threshold Value (kg)	Cumulative (P) (mm)	Deviation from Reference Value (mm)	Cumulative (P) (mm)	Deviation from Reference Value (mm)
0	300.681	4.822	420.320	44.106
0.005	297.684	1.825	396.142	19.928
0.01	294.226	-1.633	381.120	4.906
0.015	293.041	-2.817	376.214	0.000
0.02	290.375	-5.484	372.750	-3.464
0.025	288.279	-7.580	364.238	-11.976
0.03	286.818	-9.040	363.456	-12.758

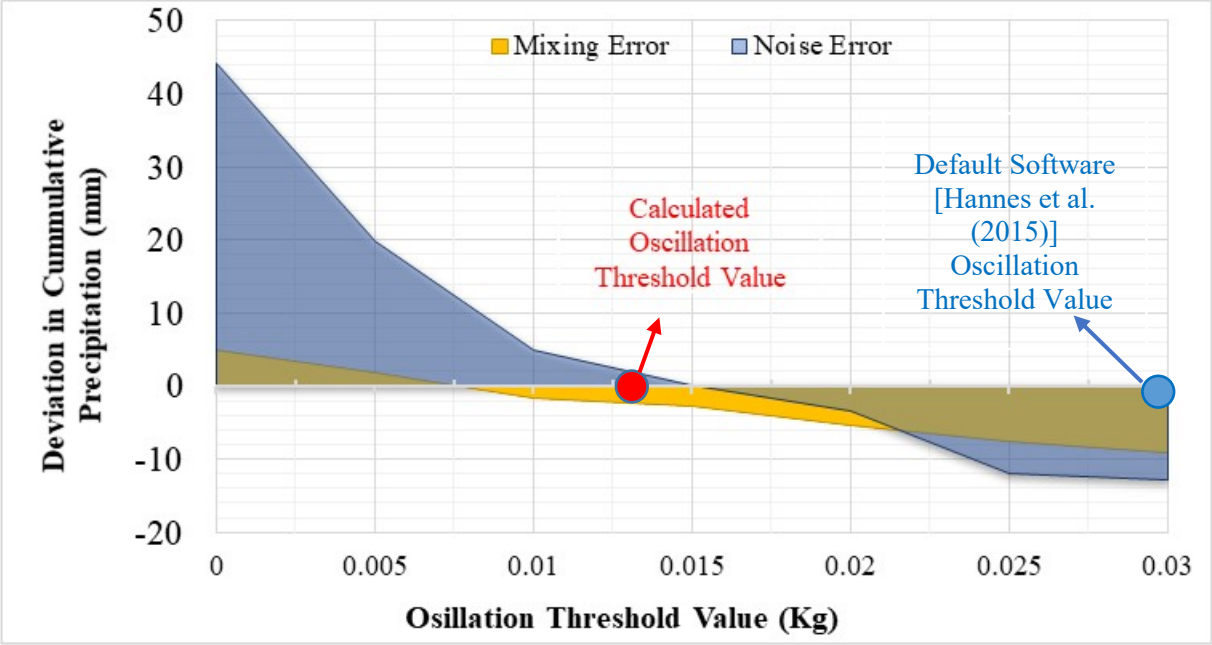

Figure 32: Effects of different values of the oscillation threshold (kg) on the deviation of the cumulative precipitation from the reference values. The red line represents the oscillation threshold value where the mixing error and noise error have equal but reciprocal effect.

Increasing oscillation threshold steps of 5 grams were applied. Through a graphical representation of the results of table 4 as seen in figure 32, for the mixing error (rain periods), higher threshold values result in higher mixing error resulting in an underestimation of the cumulative flux. While for the noise error (no-rain periods), lower threshold values show substantial increase in the overestimation of the cumulative precipitation. Yet, the noise

reduction is still relatively acceptable due to the effect of the previous filtering steps including the smoothing filter.

An oscillation threshold value of 0.0128 kg yields a deviation from the reference cumulative precipitation that is the lowest possible value for both errors to be equal in magnitude. Such value gives around + 2.5 mm deviation for the mixing error and -2.5 mm for the noise error. Hence, as speculated by Hannes et al. (2015), the combined effect of both errors when equal in magnitude compensate for each other and could give an actual representation of the true cumulative value.

In the same figure, the effect of using the default oscillation threshold value of 0.03 kg is also denoted. It can be noticed that both errors, mixing and noise, result in an underestimation of the cumulative precipitation value for the investigated period. The noise error seems to underestimate cumulative P by around 12 mm, while the mixing error by around 9 mm.

Table 5: Summary of the default and calculated parameters for the comprehensive filtering routine software developed by Hannes et al. (2015).

Comprehensive Filtering Routine Paramters			
Parameter Name	**Software Defualt Paramter Value (Hannes et al. 2015)**	**CULS's SFL-300 2017 Calculated Paramters**	**Unit**
Maximal Seepagewater-Flux	150	300	g/min
Minimal Flux in Seepagewater	-150	-250	g/min
Maximal Flux at Upper Boundary	1000	35	g/min
Minimal Flux at Upper Boundary	-100	-100	g/min
Minimum Oscillation Threshold	0	50	g
Maximum Oscillation Threshold	0	15	g
Smoothing Filter Time Window	15	15	min
Median Filter Time Window	15	15	min
Oscillation Threshold Value	30	12.8	g

5.5 Comparison of the Default and Calculated Filter Parameters

The default and calculated parameters show good accordance with the raw data, their resulting filtered summarized masses overlay the raw data perfectly. Therefore, in order to investigate their effects on the resulting summarized masses and fluxes when filtered with the default software parameters and the calculated ones, a close-up look at shorter time scales should be done.

The following graphs display the effects of both filtering process through different conditions, ranging from windy periods, periods with short precipitation followed by evapotranspiration, periods with outliers, and others.

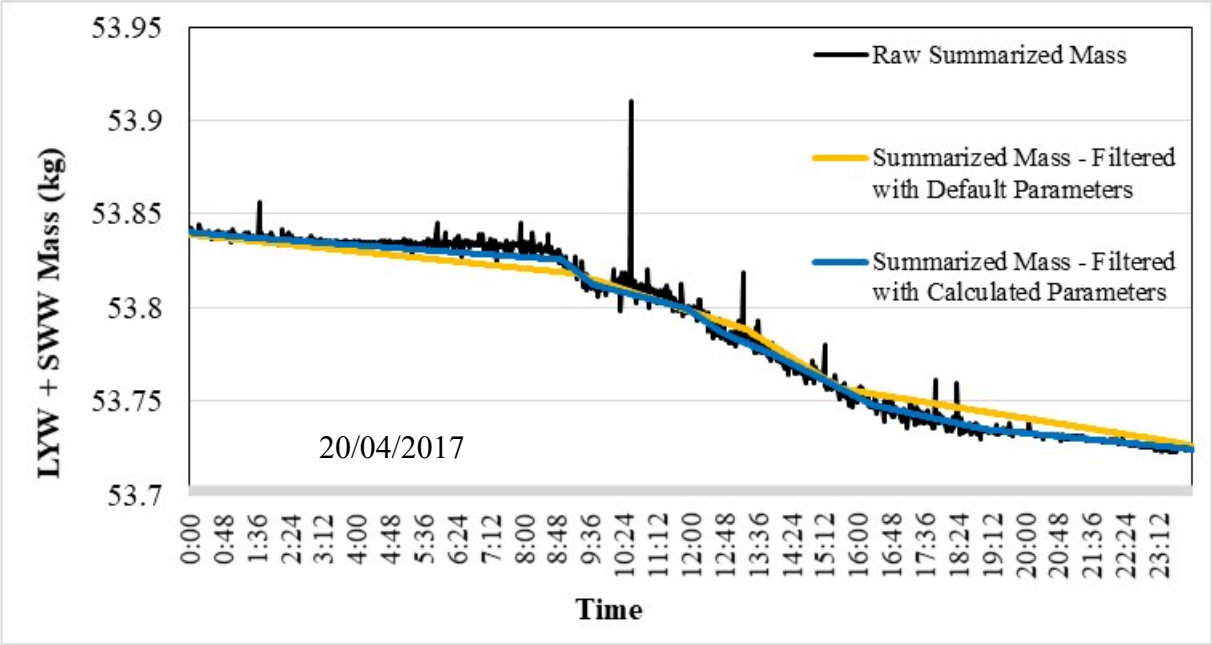

Figure 33: Effects of the default and calculated parameters on the filtration of the raw data on the 20th of April between 00:00 and 23:59. This day contains measurement errors displayed as positive spikes as well as substantial noise.

Figure 33 displays substantial noise with high erroneous fluxes displayed as large spikes. The threshold filter for both parameters seem to filter these spikes relatively well. The same is applicable for the oscillation threshold filter for both values. Nonetheless, the default parameters seem to misestimate the fluxes as seen at the beginning and end of the time scale, while the calculated oscillation threshold seems to show closer accordance with the raw data.

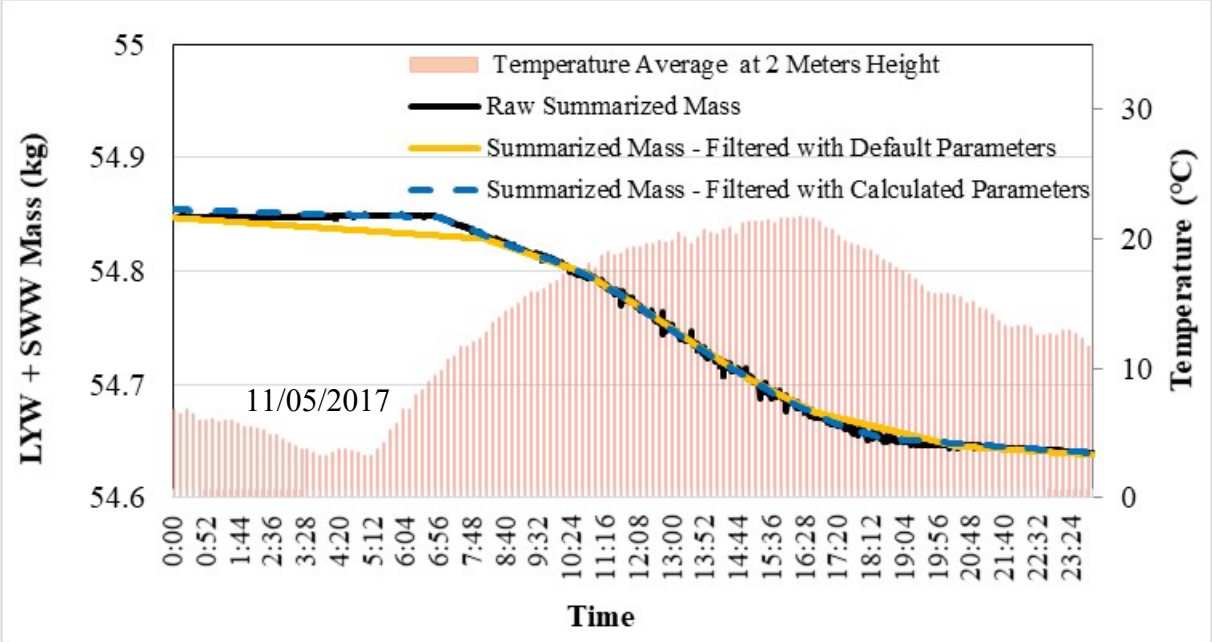

Figure 34: Effects of the default and calculated parameters on the raw data on the 11th of May 00:00 – 23:59. There are no major fluxes at night time, but during day time there can be noticed increasing temperatures that result in a decrease in the summarized mass representing ET.

Figure 35 displays the effect of precipitation events on the increase of the summarized LYW and SWW masses. A short low intensity rain increases the mass of the lysimeter lightly during nighttime, followed by a short period of evapotranspiration, hence weight loss during the early morning. Followed by around 7 hours of rain that increase the mass of the lysimeter. The effect of the noise error in this case is minor as the mixing error's impact prevails. This is noticed at the end of the time period where the default settings seem to underestimate the precipitation event.

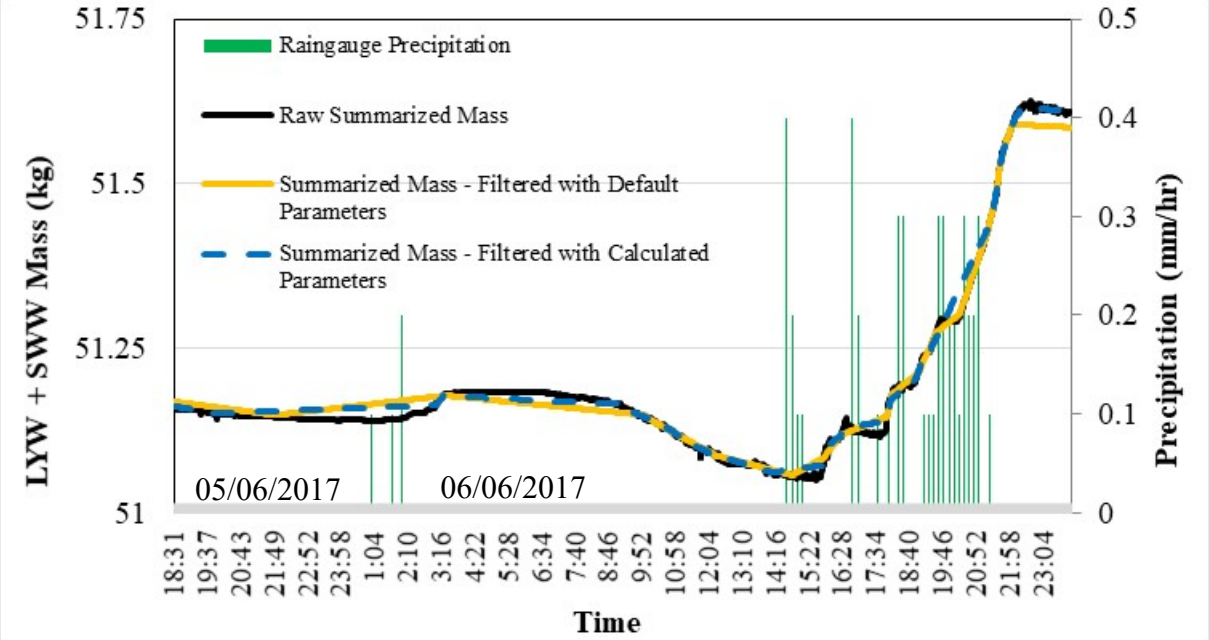

Figure 35: Effects of the default and calculated parameters on the raw data between the 5th of June 18:30 – 6th of June 23:59. Precipitation events during this period correlate with the mass increases.

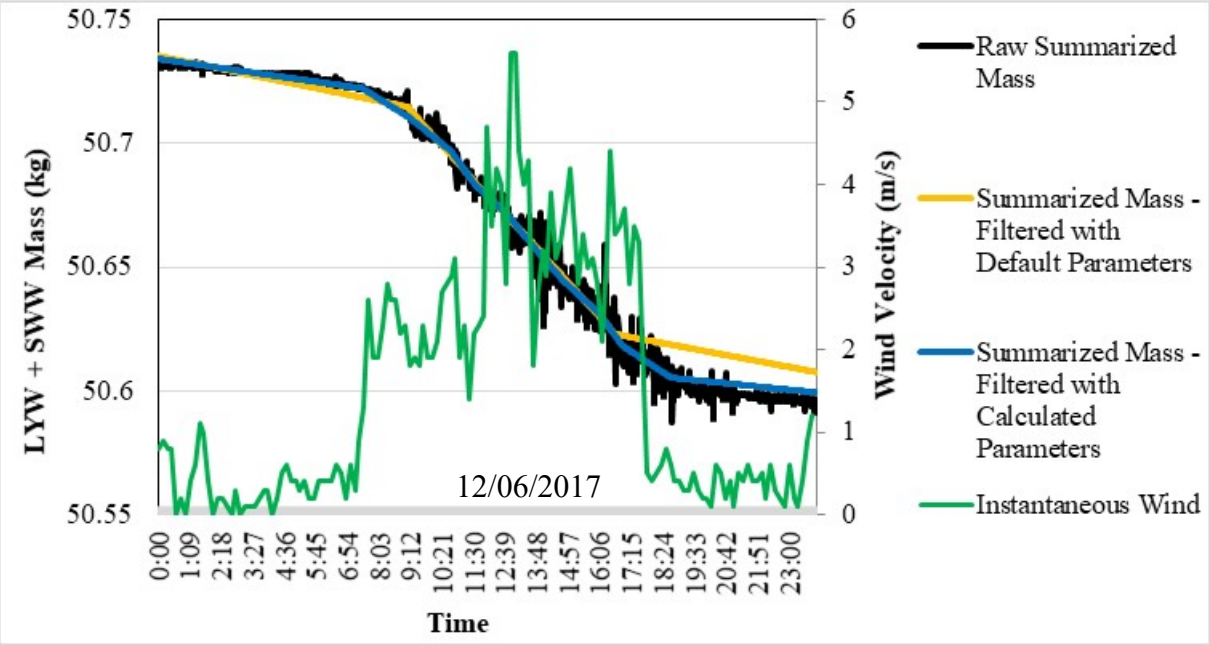

Figure 36: Effects of the default and calculated parameters on the raw data on the 12th of June, 00:00 – 23:59. Shows the effects of both filtering parameters on the noise during a day with high wind velocities.

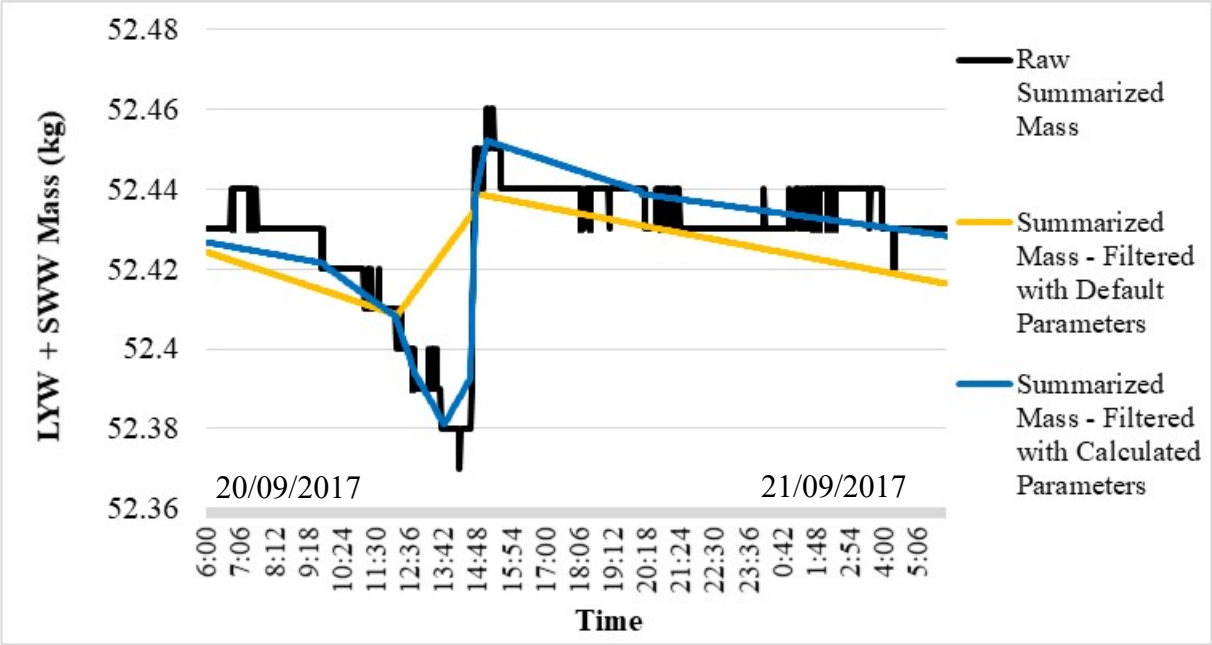

Figure 37: Effects of the default and calculated parameters on the raw data between the 20th of September, 06:00 and the 21st of September 06:00. A shorter time scale selected to display the effects of both filtering parameters on a short precipitation event.

The calculated filter parameters show closer accordance with the trend in the raw data. As the larger values in the default parameters disregard smaller fluxes that occur in this data set which might result from the smaller size of the SFL -300. Nonetheless, both the calculated and default parameters show good output when inspecting resulting fluxes.

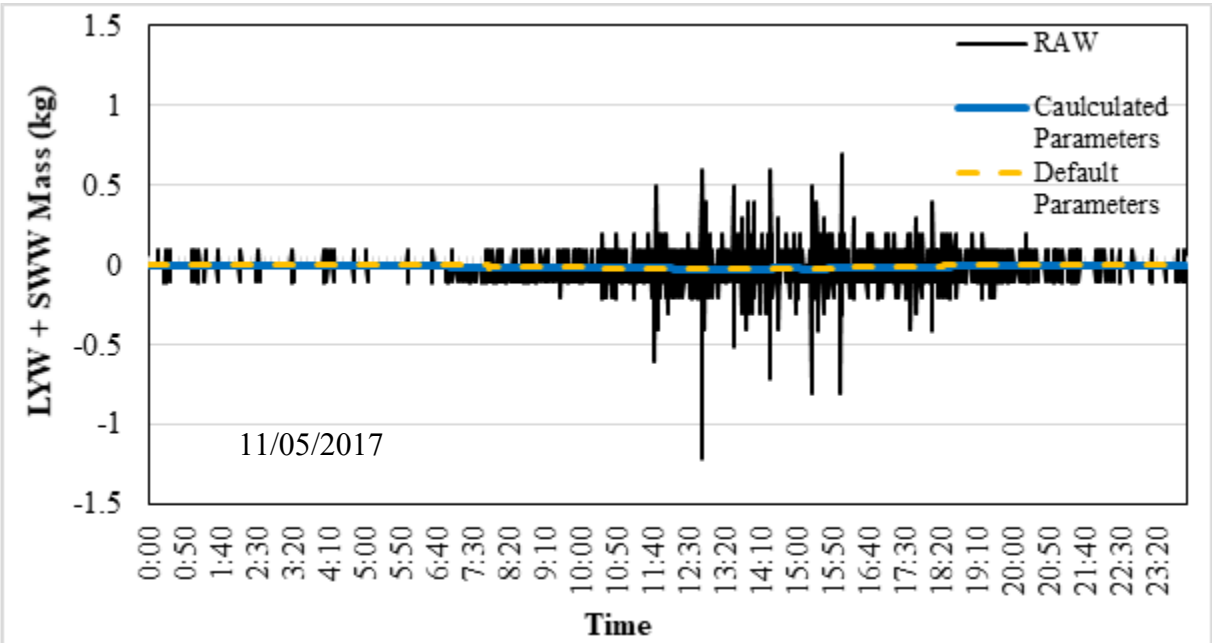

Figure 38: Effect of the calculated and default parameters on filtering noise in fluxes. 11th of May 2017, 00:00 – 23:59.

5.6 Water Balance Components Evaluation

As modern lysimeters, such as the SFL-300, log measurements with a high temporal resolution, the use of this filtering procedure and the application of the oscillation threshold filter preserve this resolution of the data set. The summary of the cumulative precipitation and the other soil water balance components once filtered with the calculated and default parameters are in table 6.

Table 6: Resulting water balance components for the time period between 2017/04/12_06:50:03 to 2017/09/23_09:59:03 Using the Comprehensive filtering routine developed by Hannes et al.(2015) with the modified parameters calculated as in table 5.

Waterbalance End Values			
Component	Abbreviation	Default Paramters (mm)	Calculated Parameters (mm)
Precipitation	(P)	361.90	378.72
Evapotranspiration	(ET)	377.08	393.90
Storage	(St)	-31.97	-31.97
Seepagewater	(SW)	16.79	16.79

It is apparent from the table above that both filtering parameters have no significant difference in terms of storage or seepage water. However, it seems that the default parameters underestimate the upper boundary fluxes, P and ET. The resulting difference between both parameters is -16.82 mm for both P and ET.

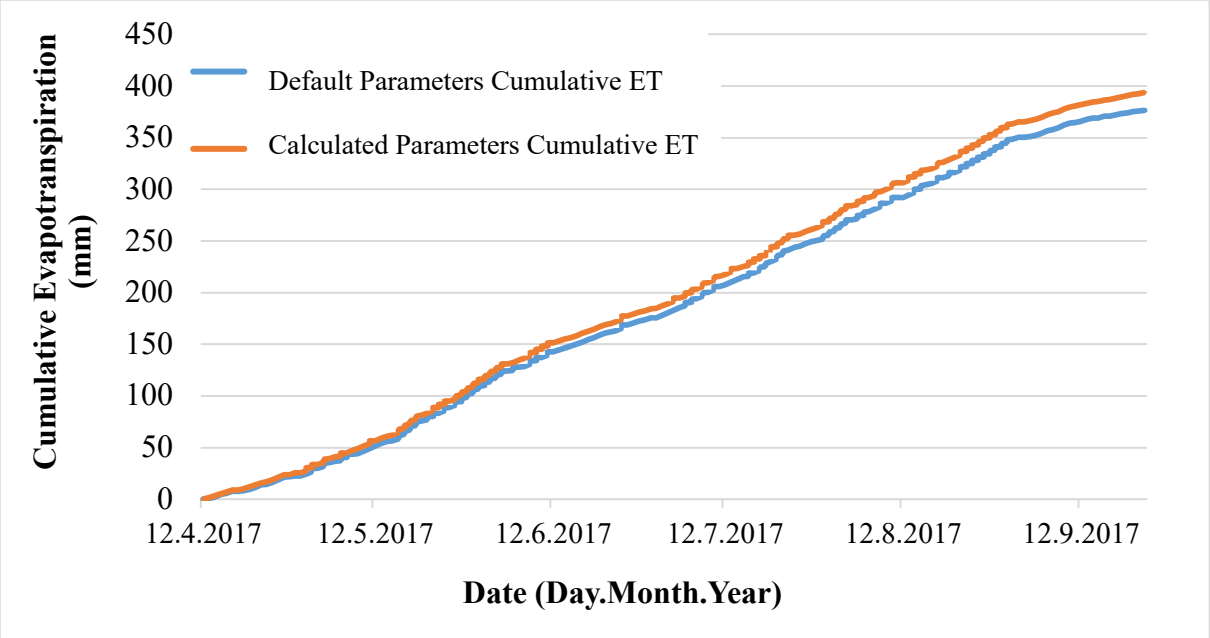

Figure 39: Cumulative evapotranspiration for both default and calculated filter parameters for the entire investigation period.

As lysimeters are typically used as reference instruments for the measurement of actual evapotranspiration, (ET_a), and in the absence of other ET measurements and models for this location and period it is difficult to assess the resulting water balance based on this component. Therefore, the validation of the calculated and default filtering parameters will be conducted against cumulative precipitation, using the TBR-MR3H 1-meter elevated rain gauge's measured precipitation.

5.6.1 Relationships between Rain Gauge and Lysimeter Precipitation

Rain gauges suffer from their own systematic and random errors. Tipping bucket rain gauges like the TBR-MR3H, seem to underestimate the cumulative precipitation due to random and systematic errors, with the systematic ones being of larger magnitude, especially wind induced under-catch. At the same experimental field, a study was conducted for the rain gauge measurements between 2011 and 2012, where an under-catch averaging around 46% for the TBR-MR3H rain gauge was detected (Mekonnen et al., 2015). Typically, ground level gauges give similar results to lysimeters and largely avoid such under-catch (Ruth et al., 2018).

During the investigated period in 2017, the rain gauge registered a cumulative precipitation of 258.4 mm, 31.7% less than the cumulative P of the lysimeter filtered with the calculated parameters. And 28.5% less than the lysimeter's cumulative P filtered with the default parameters.

Table 7: Different cumulative precipitation values for the calculated and default filtering parameters, and the TBR-MR3H rain gauge at the CULS experimental field.

Cumulative Precipitation (mm)	
Lysimeter Filtered with Calculated Paramters	378.7
Lysimeter Filtered with Default Paramters	361.9
Rain Gauge	258.4

There is a serious difference between the lysimeter's cumulative precipitation and the rain gauge's. Mekonnen et al. (2015), optimized a correction equation based on rainfall intensity (I_r) and wind speed (u). This is done by multiplying the registered rainfall sum by a correction coefficient (k), where typically ($k \geq 1$) (Mekonnen et al., 2015). The correction coefficient in this equation is not a constant, it varies depending on rainfall intensity and wind speed. Therefore, in order to derive a (k) value for a particular event, the catch ratio (C_R) needs to be calculated;

$$C_R = EXP\left[(a - b \times \ln(I_r) \times u - B\right] \quad (27)$$

Mekonnen et al. (2015) optimized values for the variables a, b, and B. yielding values of a=0.011, b=0.062 and B=0.776. Although it was argued that the optimized parameters are specific to the weather elements, rain gauge used, and the area's conditions, since the same rain gauge was used, the numbers initially were kept without any modifications and a correction coefficient was derived for each precipitation interval using the following equation;

$$k = \frac{1}{1 - C_R} \tag{28}$$

The resulting rain gauge cumulative precipitation after applying the corrections in formulas 27 and 28 lead to an overestimation of the cumulative flux, returning a value of 507.4 mm, around 134% of the calculated parameters lysimeter's cumulative P. On the other hand, the trend in the corrected cumulative rain gauge P shows accordance with the lysimeter's. Hence, it is speculated that by changing the variables, a more realistic result might be derived.

An attempt to correct the rain gauge measurements was done based on assuming the lysimeter's (calculated parameters) cumulative P as the true reference. The ratio between the lysimeter's cumulative P and the rain gauge's is 1.466. When each of the rain gauge's measurements is multiplied with this factor, the cumulative rain gauge measurements and the lysimeter's have a coefficient of determination of 0.9993 and an RMSE of 4.28 mm. Nevertheless, and despite this method providing a great fit with the lysimeter's cumulative P, it may introduce bias into the process and render the validation process obsolete.

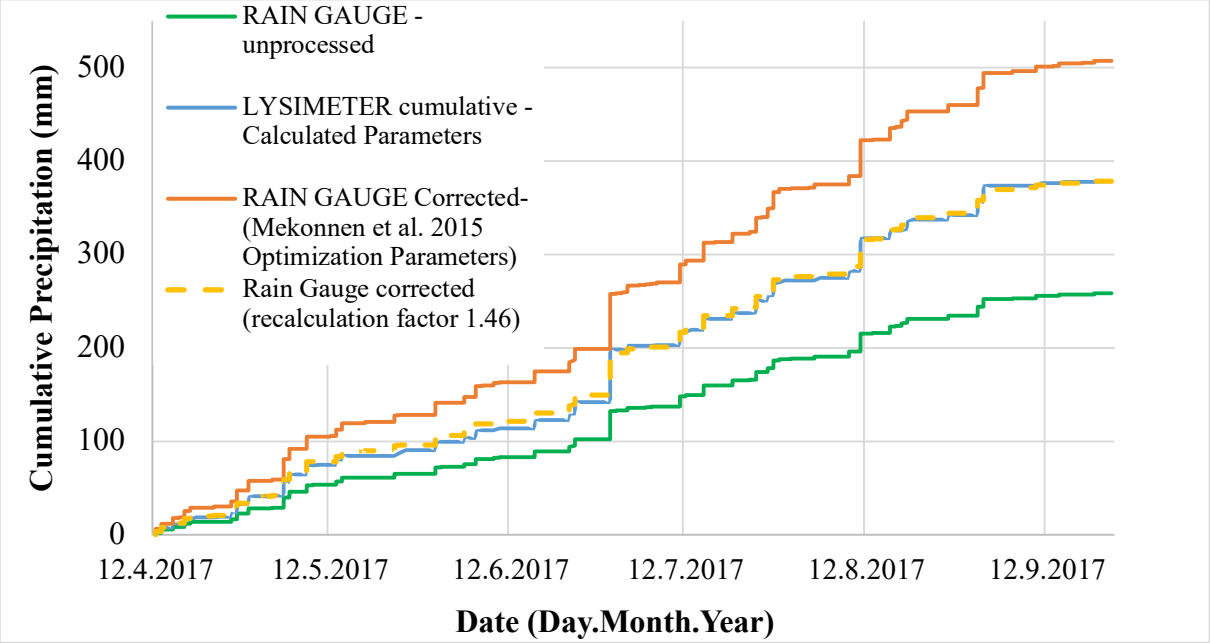

Figure 40: Rain gauge's raw and corrected measurements against the lysimeter's cumulative precipitation for the entire investigation period.

Since the correction of the rain gauge measurements is dependent on variables such as wind speed and rain intensity, a correction using the methodology developed by Mekonnen et al. (2015) was utilized. Using Excel's Solver function, (a) and (b) parameters were maintained

with as their default, while optimizing an offsetting value for (B). The resulting calculated (B) and the resulting cumulative precipitation are as in table 8.

Table 8: 2017 Equation 27's optimized variables using the default and corrected variables and the resulting rain gauge's corrected cumulative precipitation compared to the calculated parameters lysimeter cumulative precipitation.

2017 Rain Gauge Measurements Optimization			
	Mekonnen et al. (2015) Optimized Parameters [Default]	Corrected and Optimized Parameters for 2017	SFL - 300 Calculated Filter Parameters
a	0.011	0.011	
b	0.062	0.062	
B	0.776	1.27	
Cumulative Precipitation (mm)	**507.4**	**367.07**	**378.7**

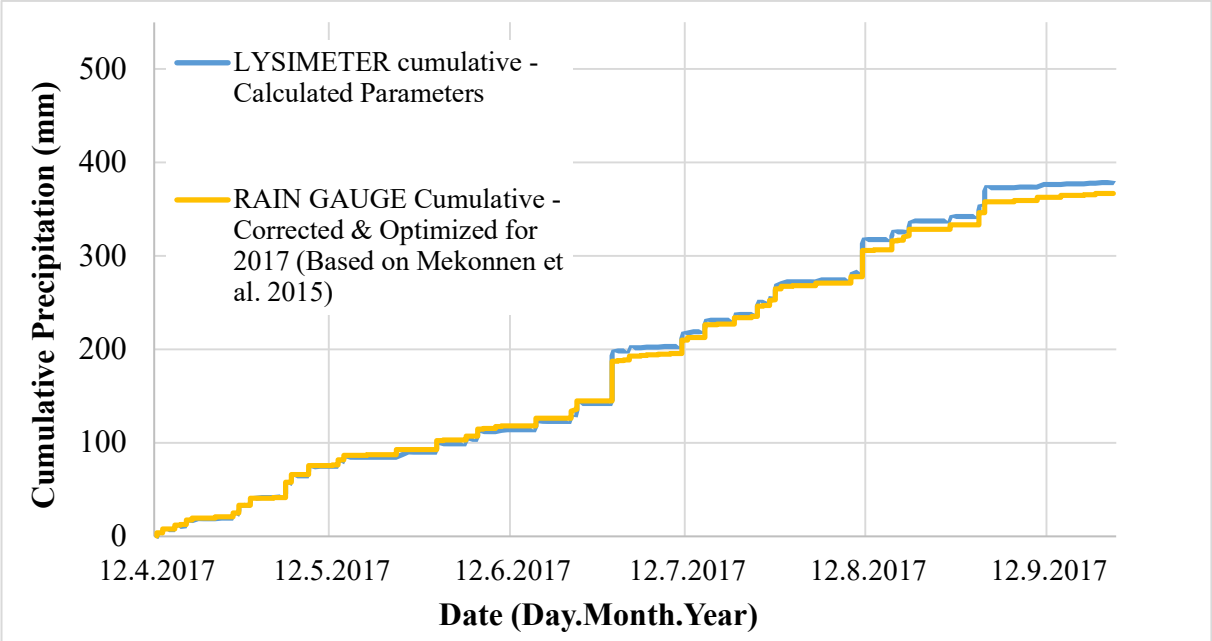

Figure 41: Rain gauge's corrected cumulative precipitation based on the 2017 data optimization using Mekonnen et al. (2015) method, and the lysimeter calculated parameters' cumulative precipitation for the entire investigation period.

As the rain gauge's measurements have been corrected, the default and calculated lysimeter parameters can be validated based on their cumulative precipitation values against those of the rain gauge.

5.6.2 Statistical Analysis

In order to validate which of the filtering parameters provides better cumulative precipitation output, and the magnitude of the difference if any, several tests have been carried

out to assess the relationships and differences between the different cumulative precipitation values. For some specific tests, the range of examined data was extended to include all of the following in order to spot any distinctive relationships;

1. Lysimeter - calculated parameters cumulative precipitation
2. Lysimeter - default parameters cumulative precipitation
3. Rain gauge - raw (untreated) cumulative precipitation
4. Rain gauge – corrected with Mekonnen et al. (2015) default optimization variables
5. Rain gauge – corrected with Mekonnen et al. (2015) calculated optimization variables of 2017
6. Rain gauge – multiplied by factor of 1.446

Table 9 provides a statistical overview of the main relationships between the main data sets investigated, the calculated and default lysimeter cumulative P compared to that of the corrected rain gauge.

Table 9: Summary of statistical relations between the rain gauge cumulative precipitation using the corrected and optimized Mekonnen et al. (2015) equation for 2017, compared to the cumulative precipitation of default and calculated lysimeter filtering parameters.

PARAMTER	Corrected Rain Gauge Cumulative (P) Vs. Calculated Parameters Lysimeter Cumulative (P)	Corrected Rain Gauge Cumulative (P) Vs. Calculated Parameters Lysimeter Cumulative (P)	UNITS
	VALUE	VALUE	
Count	23635	23635	
Total Difference	-80761.8	146266.7	mm
Total Difference²	1170864.5	1243543.1	mm²
SD (σ)	118.8	116.5	mm
RMSE	7.0	7.3	mm
NRMSE	1.86	1.98	%
R^2	0.9993	0.999	

As the rain gauge's measurements are recorded at 10-minute intervals, the lysimeter's 1-minute measurements were summed to 10-minutes as well. Hence the (Count) describes the number of measurements in each one's data set.

It can be seen from table 9 that the calculated parameters yield better correlation with the corrected rain gauge's cumulative precipitation, as well as a slightly lower RMSE and NRMSE. Nonetheless, the calculated parameters' cumulative P has a slightly higher standard deviation (σ) than the default ones in relation to the rain gauge.

In order to obtain a more general idea regarding each data set's own statistical characteristics, their specific attributes are investigated as portrayed in table 10. The table also shows the standard error of each mean, which is a measure of its sampling variability. The standard error is formed by dividing the pooled standard deviation by the square root of the number of observations at each level.

Table 10: Intrinsic statistical characteristics and parameters within the main data sets; the corrected rain gauge using Mekonnen et al. (2015) using 2017 derived optimization variables, the lysimeter calculated parameters, and the lysimeter default parameters

Parameters	Rain gauge – corrected with Mekonnen et al. (2015) using 2017 derived optimization variables	Lysimeter - Calculated Parameters' Cumulative (P) mm	Lysimeter - Default Parameters' Cumulative (P) mm
Mean	189.6	193.1	183.5
Standard Error	0.8	0.8	0.8
Median	193.2	202.1	193.1
Mode	328.3	84.7	78.9
Standard Deviation	116.1	121.4	116.9
Sample Variance	13471.3	14743.5	13657.7
Range	366.9	378.7	361.9
Minimum	0.2	0	0
Maximum	367.1	378.7	361.9
Sum	4482300.6	4563062.4	4336033.9
Count	23635	23635	23635

Moreover, when evaluating the cumulative precipitation data for the different data sets, they do not seem to show a normal distribution. Additional error is thus incorporated into the evaluation. The data transformation type for obtaining normal distribution needs to be further investigated, since standard transformation techniques such as logarithmic, exponential or 1/n transformation did not bring sufficient improvement.

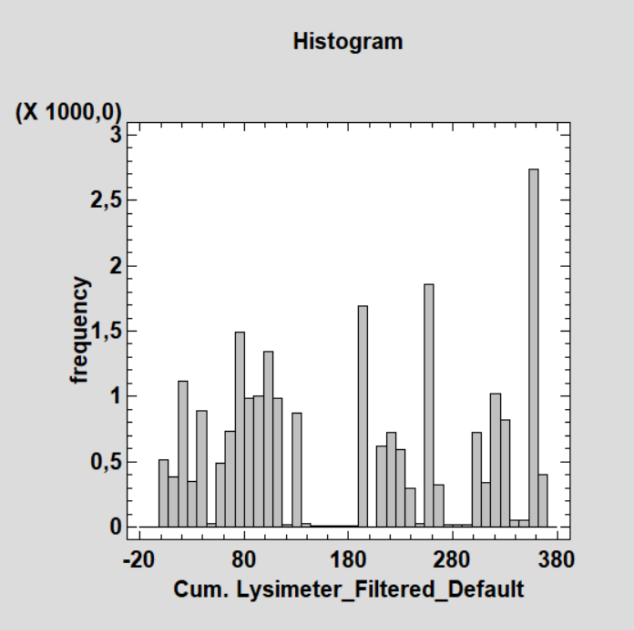

Figure 42: Frequency histogram for the default filter parameter's cumulative precipitation.

This evaluation was conducted using Statgraphics Centurion 18 software (version 18.1.13), by Statgraphics Technologies Inc. (Virginia, USA). Other data sets' frequency histograms can be found in Appendix 3.

Standard skewness and kurtosis tests were also conducted to further investigate the distribution of the data sets; summary of their results is as in table 11.

Table 11: Standard Skewness and Standard Kurtosis evaluation of the cumulative precipitation data sets.

	Range	*Stnd. skewness*	*Stnd. kurtosis*
Rain gauge - raw (untreated) cumulative precipitation	258.3	6.8	-42.7
Rain gauge – corrected with 2017 derived optimization variables for Mekonnen et al. (2015) equation	366.9	6.9	-42.7
Rain gauge – corrected with Mekonnen et al. (2015) optimization default variables	507.1	7.2	-42.9
Rain gauge – multiplied by factor of 1.446	378.6	6.8	-42.7
Lysimeter - calculated parameters cumulative precipitation	378.7	8.7	-43.2
Lysimeter - default parameters cumulative precipitation	361.9	9.2	-43.3
Total	507.4	-958.6	-60.7

The standardized skewness and/or kurtosis test indicates the significance of anomalies in the data, indicating whether they come from normal distributions or not. If the value is outside the range of -2 to +2 , then it can be assumed that they are not normally distributed, indicating some significant nonnormality. Further investigation of data transformation and using the Kruskal-Wallis test to compare the medians instead of the means can be found in Appendix 3. Where the results also still show significant statistical differences.

Moreover, an Analysis of variance (ANOVA) single factor test was conducted for the 6 pairs, the ANOVA table below segregates the variance of the sets into a measure between groups, and within. The F-ratio, which in this case equals 2691.81, is a ratio of the between-group estimate to the within-group estimate. Since the P-value of the F-test is less than 0.05 significance level, there is a statistically significant difference between the means of the 6 sets.

Table 12: Analysis of Variance (ANOVA).

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Between groups	1.98791E8	5	3.97582E7	2691.81	0.0000
Within groups	2.09445E9	141804	14770.0		
Total (Corr.)	2.29324E9	141809			

Table 13 applies a multiple comparison procedure to determine the significance of the sets' means from each other. The output shows the estimated difference between each pair of means. An asterisk has been placed next to 15 pairs, indicating that these pairs show statistically significant differences at the 95.0% confidence level. The method the software uses to discriminate among the means is Fisher's least significant difference (LSD) procedure.

Table 13: Significance of means difference analysis based on method: 95.0 percent LSD. The red pairs are highlighted due to their low difference significance value. The corrected rain gauge with the factor of 1.446 and the rain gauge 2017 corrected optimization variables for Mekonnen et al. (2015) show the lowest differences, almost reaching the ± limits.

Contrast	Sig.	Difference	+/- Limits
Rain gauge - raw (untreated) cumulative precipitation X Rain gauge – corrected with 2017 derived optimization variables for Mekonnen et al. (2015) equation	*	-56.1789	2.19117
Rain gauge - raw (untreated) cumulative precipitation X Rain gauge – corrected with Mekonnen et al. (2015) optimization default variables	*	-128.509	2.19117
Rain gauge - raw (untreated) cumulative precipitation X Rain gauge – multiplied by factor of 1.446	*	-62.1491	2.19117
Rain gauge - raw (untreated) cumulative precipitation X Lysimeter - calculated parameters cumulative precipitation	*	-59.5959	2.19117
Rain gauge - raw (untreated) cumulative precipitation X Lysimeter - default parameters cumulative precipitation	*	-49.9903	2.19117
Rain gauge – corrected with 2017 derived optimization variables for Mekonnen et al. (2015) equation X Rain gauge – corrected with Mekonnen et al. (2015) optimization default variables	*	-72.3297	2.19117
Rain gauge – corrected with 2017 derived optimization variables for Mekonnen et al. (2015) equation X Rain gauge – multiplied by factor of 1.446	*	-5.9702	2.19117
Rain gauge – corrected with 2017 derived optimization variables for Mekonnen et al. (2015) equation X Lysimeter - calculated parameters cumulative precipitation	*	-3.41704	2.19117
Rain gauge – corrected with 2017 derived optimization variables for Mekonnen et al. (2015) equation X Lysimeter - default parameters cumulative precipitation	*	6.18856	2.19117
Rain gauge – corrected with Mekonnen et al. (2015) optimization default variables X Rain gauge – multiplied by factor of 1.446	*	66.3595	2.19117
Rain gauge – corrected with Mekonnen et al. (2015) optimization default variables X Lysimeter - calculated parameters cumulative precipitation	*	68.9127	2.19117
Rain gauge – corrected with Mekonnen et al. (2015) optimization default variables X Lysimeter - default parameters cumulative precipitation	*	78.5183	2.19117
Rain gauge – multiplied by factor of 1.446 X Lysimeter - calculated parameters cumulative precipitation	*	2.55316	2.19117
Rain gauge – multiplied by factor of 1.446 X Lysimeter - default parameters cumulative precipitation	*	12.1588	2.19117
Lysimeter - calculated parameters cumulative precipitation X Lysimeter - default parameters cumulative precipitation	*	9.6056	2.19117
Sig. = significance level (* denotes a statistically significant difference			

6 Discussion

6.1 Critical Evaluation of the Comprehensive Filtering Routine

Both the default and calculated filtering parameters display good outputs when looking at the filtered fluxes and summarized masses. Table 6 in the previous section shows the water balance components' cumulative outputs for both parameters, where the default parameters have 16.82 mm lower cumulative precipitation and evapotranspiration. With negligible storage and seepage water differences. The differences in the upper boundary fluxes correspond well with the noise and mixing errors' misestimation values in figure 32. The default oscillation threshold value of 0.03 kg gives an underestimation for cumulative precipitation of around 12 mm for the noise error compared to around 9 mm for the mixing error. An opposite overestimation of cumulative precipitation is noticed at an oscillation threshold value of 0 kg, where the noise error overestimates the cumulative precipitation by around 45 mm compared to around 5 mm for the mixing error. Hence, this displays the impact of applying an oscillation threshold filter on the filtered resulting cumulative fluxes.

It was evident in the previous section that the pre-processing of the raw data has a significant impact on the output. This process could be considered one of the most important steps in filtering the data. Screening the raw data against outliers and logging errors is essential to the filtering process. The filter seems to well remove outliers surpassing the defined thresholds, but the manual identification and removal of such values is still important. However, this process could be time consuming depending on the size of the data set. It is important to keep a comprehensive logbook of all of the maintenance, grass cutting, seepage water tank emptying activities, and any other occurrences that could impact the measurements. It was suggested that the automation of the manual filtering by connecting the log to the filter code might be feasible (Hannes et al., 2015; Nolz et al., 2013b). Nonetheless, this might be complicated and would require it to be provided by the authors of the software. An alternative step could be the maintenance of an electronic logbook with a defined range of required details, such as the measurement of the mass of grass cut from the top of the lysimeter, quantity of water emptied from the seepage tank, exact periods of snow cover, amounts of irrigation water and so on. One suggestion that would help understand the causes of any outliers in the data is to install a CCTV camera directed at the lysimeter, for when any measurement that is not reflected in the logbook seems to not follow the data trend, such footage could be revised. Furthermore, as saving video footage is storage extensive, an alternative is to program the camera to take pictures every minute.

With regard to the threshold filter's parameters, having a weather station located at the vicinity of the experimental field is of great value. As the parameters are dependent on this information for the examined period and locality. However, such meteorological instruments suffer from their own systematic errors that need correction or to be accounted for. For the purposes of choosing reference maximum values for the threshold parameters, this information is fine as yet a safety factor is typically added to them. Nonetheless, this might result in another type of challenge, where such parameters need to be calculated for the duration of every data set filtered. Which in turn might yield a slightly different oscillation threshold value. This

assumption needs to be further studied, and different data sets' parameters need to be calculated in order to quantify the magnitude of difference in the resulting oscillation threshold filter values.

A certain challenge that arose during the course of this study was determining the appropriate value for the ET parameter. Where after extensive investigation, a discrete value was not reached and rather the default software parameter yielded the best effect compared to other values. Despite the fact that the impact of the ET threshold value on the filter is not of a large magnitude as the Ksat's impact, it is still an important parameter as it defines the acceptable limits of mass decreases at the upper boundary. Hence, deducing such a value for each investigation period is laborious and requires its own modelling, unless there exist some results from other studies. For this particular issue, it could be speculated that the use of the more complicated AWAT filter could extricate one from having to research such a value. Nevertheless, the output results still do show great improvement compared to the noisy and erroneous raw data, which for the deduction of the water balance components, or for studying the short time effects of the occurring processes and fluxes, cannot be considered representative.

As laid out by the authors of the software, the effect of the smoothing filter on the raw data is more important than the preceding threshold filter. This is due to the fact that the threshold filter is only able to remove discrete outliers exceeding a certain magnitude. Hence, any measurement error below this value manifests as either noise or mixing error causing a deviation of the recorded measurement from its true value. Different literature defines, to some extent, varying values for the smoothing window, that is of course governed by their used data sets, smoothing function, and the lysimeter used. Nolz et al. (2013a) for instance recommended an averaging window for the sole utilization of the MA not exceeding 30 minutes, which corresponds well with the investigated effects of different MA values in figures 28 and 29. is It is almost unanimously agreed that any additional increase in the MA window results in reduced noise, at the cost of the temporal resolution (Hannes et al., 2015; Marek et al., 2014; Nolz et al., 2013a).

The core of this study was investigating the parameters of the comprehensive filtering scheme developed by Hannes et al. (2015) and their impact on the resulting summarized masses and fluxes. The results section displays obvious improvements in the calculated parameters' filtered output when compared to the default parameters. This might be attributed largely to the tailored oscillation threshold value. The calculation of the oscillation threshold filter consists of a few steps. To be able to find a suitable oscillation threshold value, the minimum and maximum possible values need to be defined. A certain drawback of this filtering scheme is the process of finding the maximum value for the oscillation threshold as proposed by the authors. This is not associated with the software or the filtering algorithm, but the deduction methodology. The process was discussed thoroughly in section 5.4.1, and it could be speculated that the proposed method could expose the filtering process to possible random errors and bias. Hannes et al. (2015) ambiguously state that to find the maximum oscillation threshold value, nighttime oscillations are monitored until the oscillation threshold value where such oscillations nearly disappear is reached. This method is subject to the operator's definition and observation of the disappearing oscillations. And since such value is subsequently used as the reference value for the noise error, it could severely impact the final value of the oscillation threshold filter. Furthermore, in their original paper, Hannes et al. (2015) monitored the fluxes for

obtaining the maximum value at which nighttime oscillations disappeared. in this study, both the fluxes and summarized mass outputs were inspected for the disappearance of these nighttime oscillations. Investigating the fluxes yielded an unrealistic value of 0.045 kg, which is very high considering the smaller area of the SFL-300. Hence, when such value was applied as a reference for non-rainy periods in order to deduce the value of the oscillation threshold value, the result was an oscillation threshold value of around 32.5 g, which is higher than the default software's value that was calculated using a 1 m^2 lysimeter. When such value was applied to filter the data, the output was significantly less fitting compared to the calculated and even default parameters.

To avoid such a dilemma, an additional step in the filtration algorithm could be applied that is solely used to deduce the maximum oscillation threshold value. Such process could operate by defining a maximum allowable absolute flux value, absolute in terms that it applies to both positive and negative fluxes. Where the filter is run either automatically or manually until such value is reached, determining the step at which the nighttime oscillations almost vanish. Making the filtration process less prone to human errors.

Subsequently, the calculation of the oscillation threshold filter's value is relatively less complex. However, the process of graphically or manually deducing the oscillation threshold value from charts like graph 32 is subject to its own random errors. As speculated, the oscillation threshold filter's value lies in between the misestimations of the noise and mixing errors, where such a value result in both errors cancelling out each other's effects. This method is subject to random errors and it would be more accurate to deduce it mathematically or through the software.

Nevertheless, once the operator is familiar with the filter steps and the methodology, filtering any subsequent data set is not very time consuming. And it would be more convenient if the lysimeter data is filtered at constant time intervals, for example semi-annually or quarterly. This of course would require that meteorological data are always available, and templates of the spreadsheets and graphs used to calculate the maximum threshold value, precipitation fluxes, and oscillation threshold value to be saved. This would save the operator significant time and would unify the presentation of the data for record keeping.

The significance of the systematic errors of the meteorological instruments is faced when attempting to statistically compare the results of the calculated filtering parameters to the reference rain gauge's cumulative precipitation. Typically, a serious difference such as the one in this study would subject the results and the filtering scheme to several questions. Where the cumulative rain gauge precipitation prior to any corrections was 258.4 mm, compared to the lysimeter's 378.7 mm. Nevertheless, Mekonnen et al. (2015) investigated the effects of the wind caused under-catch, rainfall intensity, and the elevation of the rain gauge in underestimating rainfall measured by elevated tipping bucket rain gauges. Without the efforts of the authors in investigating such occurrences and deriving a simple mathematical equation for the correction of the elevated tipping bucket rain gauge, the results in this thesis would be difficult to validate. Initially, it was speculated that the optimization parameters for equation 27 laid out in Mekonnen et al. (2015) study would give acceptable results as their investigation was carried out in the same locality, using the same rain gauge. However, utilizing the correction with the original optimization parameters yielded an overestimation of cumulative precipitation.

Different time periods and several data sets need to be experimented on using the original optimization equation to study the resulting impact.

Further investigations are needed that are dedicated to comparing the advanced filtering schemes, such as the AWAT or the comprehensive filtering routine. The comprehensive filtering scheme developed by Hannes et al. (2015) show relatively significant improvement compared to the MA. The AWAT has undergone several improvements where Ruth et al. (2018) reported significant error reduction using the latest AWAT filter, and have validated it to a ground mounted rain gauge to avoid the under-catch issue. This improvement in the noise reduction using the AWAT filter has been reported in other literature as well (Hoffmann et al., 2016; Kohfahl et al., 2019; Saaltink et al., 2020).

With respect to the comprehensive filtering routine, a study using a 4 m^2 lysimeter implemented the filter to reduce noise and calculate the magnitude of effect of the non-rainfall precipitation. The resulting lysimeter measurements were reported to be suitable for subsequent measurements (Zhang et al., 2019). Hebrich et al. (2016) employed the synchro filter for dealing mainly with precipitation, but did not report the effect of the filter.

There is a number of studies that utilize the comprehensive filtering routine. Nonetheless, not many discuss the results of the application of the filter. Through the results section, and its statistical evaluation, it has been investigated to what extent the calculated and default filtering parameters impact the resulting soil water balance components. Precipitation validated with the corrected rain gauge show differences between the resulting cumulative precipitation. The calculated parameters show a better output when statistically evaluated, as it has less statistical differences from the rain gauge compared to the default. As a result, for this particular filtering scheme, location and time period, the null hypothesis stating that the parameter settings do not significantly affect the magnitude of the resulting water balance components can be rejected. An alternative hypothesis stating that the used parameters have a significant effect on the final soil water balance components is accepted.

6.2 Peculiar Observed Relationships

6.2.1 Effect of the Saturated Hydraulic Conductivity Threshold Value

Most of factors affecting the lysimeter measurements discussed in the literature review were observed and confirmed by this study as permitted. This includes the effects of temperatures, precipitation, and in particular, wind velocities. One interesting outcome was found while calculating the separate parameter values. In section 5.2.4, the effects of different maximum ET values were discussed, and as an outcome the varying investigated values did not have a visible effect when examining the entire period (12 April – 23 September). Hence, their effects had to be examined at shorter intervals to see if there were any significant effects. The effect of the saturated hydraulic conductivity (Ksat) parameter on the maximal seepage water flux is not as discrete. Since Ksat values are usually denoted as a range, the Ksat values for the CULS experimental field's Chernozem soil range between 7×10^{-5} - 1×10^{-7} (Dolezal et al., 2015, 2018). The maximum Ksat value was used for the maximal seepage water flux threshold since it defines the highest realistic flux that could occur at the lower boundary. Experimenting with the filter by only changing the Ksat value yielded significant differences in the summarized

mass. This particular effect was not mentioned in any of the revised literature and has not yet been discussed as to why the impact of the Ksat value is as dominant.

Potentially, the impact of high Ksat values indicates the fact that the fluxes or changes occurring in the lower boundary are rapid, meaning that the rate of the water movement becomes higher and water moves faster within the soil profile. Hence, bigger amounts of water would move in the soil profile per unit time, and vice versa if the Ksat value is lower. Therefore, assigning a high value for Ksat in the threshold filter parameters increases the permissible threshold of fluxes in the lower boundary and any sudden but big change would be interpreted as a real flux, where usually it would be removed. This particular phenomenon can be seen in figure 24.

Figure 23 displays the effects of different Ksat values on the summarized lysimeter mass. The calculated value of 300 g/min and the default 150 g/min seem to provide relatively the same result when looking at the entire period. This is relatively simple to explain as both values are apparently still higher that the real fluxes registered by the lysimeter, but still are smaller than the fluxes of outliers. This is also observed when looking at shorter intervals, as both values seem to perfectly overlay each other. Further analyzing the impact of using the default value for Ksat by looking at shorter time intervals and the water balance components, it is found that not only there is no distinguishable impact on the filtered summarized mass, but also in the cumulative water balance components. Indicating that there would have been no difference in using the default Ksat value instead of the calculated one in this case and soil. It can also be speculated that perhaps the default filtering parameters would yield better results for other periods since it sets a lower threshold that detects outliers of lower values. But of course, such a hypothesis needs to be tested over several periods before taking the default value for Ksat as a filter parameter for this particular location.

When using 0.42 g/min based on the lowest value for Ksat, the resulting effect does not need to be studied on shorter time scales to manifest. Figure 23 shows that at the end of the investigation period the biggest difference occurs, Reaching around 4.5 kg. This effect is also visible when looking at the water balance end values compared to the ones in table 6. Changing the Ksat value to the lowest one does not impact the cumulative precipitation, maintaining a value of 378.74 mm, which is close to the calculated parameters cumulative precipitation in table 6 (cumulative (P) = 378.72 mm). Storage is also not affected, with a value of -31.97 mm as in table 6. On the other hand, evapotranspiration has a significantly higher value of 461.5 mm, compared to -377.08 mm in table 6, and seepage water is -56.3 mm compared to 16.79 mm in table 6. Despite evapotranspiration being an upper flux parameter, a low Ksat value seems to have a significant effect on its cumulative output. This could be attributed to the lysimeter registering precipitation fluxes, but as some which exceed this low Ksat value reach the seepage water tank, they are removed as they exceed the threshold value. Hence, they are interpreted as being lost through evapotranspiration events. The same is applicable in terms of the low seepage water value, which could also be due to the fact that lower boundary fluxes exceeding this lower Ksat value are discredited.

6.2.2 Averaging Window Width

Through comprehensive raw data pre-processing, it could also be possible that filtering the data with a simple moving average in the absence of the time or technical means could yield relatively acceptable results. As aforementioned, it would be ideal when filtering the data with the basic approaches to combine a smoothing function with a threshold filter. Nevertheless, the use of a smoothing window with a reasonable width still could yield satisfactory results when looking at the summarized (LYW +SWW) masses.

Using a moving average with a window width of 15 minutes maintains the temporal resolution of the data. However, outliers and noise still do manifest, jeopardizing the integrity of the measurements, while very high windows result in underestimation of precipitation induced mass changes due to the mixing error.

To verify a suitable time window while using solely the moving average to filter the data, a 15 minutes window shows the closest correlation to the raw data as seen in figures 27, 28 and 29, yet the filtered summarized masses are still fluctuating due to noise, and the resulting filtered masses seem to reflect, to a lesser degree that is, drops or spikes in the raw data as seen in figure 29. A 30 minutes window shows better results especially in terms of filtering out outliers, yet some noise still manifest. On the other hand, windows of 45 – 60 minutes seem to achieve the best balance between reducing noise and avoiding large mixing errors.

6.2.3 Default Parameters with Calculated Oscillation Threshold

Another interesting observation was discovered when attempting to run the filter with its default parameters while changing only the oscillation threshold filter's value to be as calculated. Running the filter this way with the calculated oscillation threshold value seems to give better results than the default filtering parameters. In many observed days, the default parameters and calculated oscillation threshold give very similar results when observing the daily summarized masses. On the other hand, on days of precipitation events or severe noise, this method shows slight deviation from the calculated filter, but still better accordance with the raw data than using the default parameters and oscillation threshold value.

To further examine this effect, the water balance end values are compared to the ones in table 6. For seepage water and storage, the three filtering parameters display equal cumulative values. As for cumulative precipitation and evapotranspiration, the default parameters combined with the calculated oscillation threshold display better accordance with the calculated parameters.

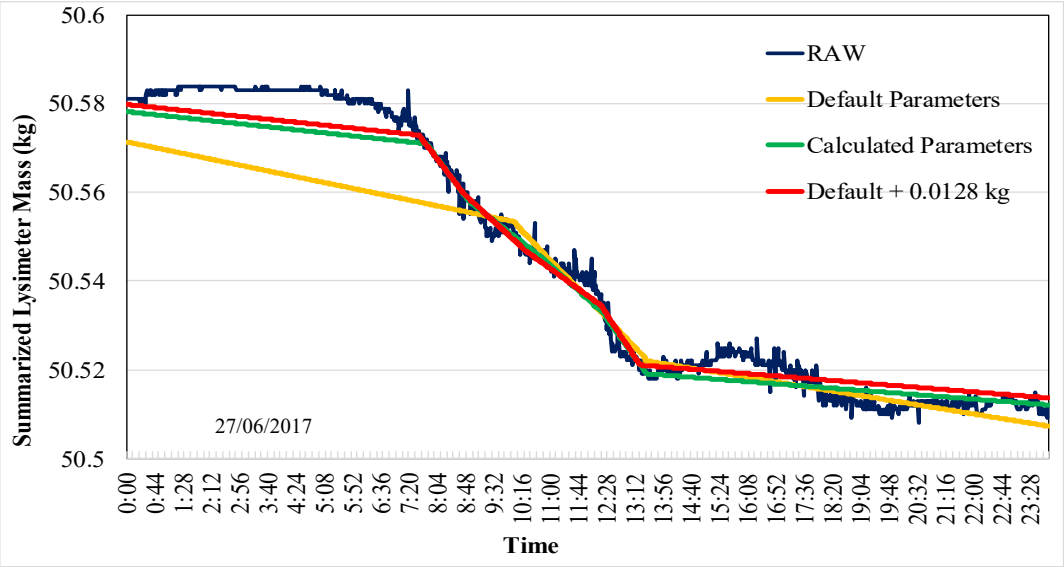

Figure 43: Effect of running the filter with the default parameters while changing the oscillation threshold filter as calculated compared to the calculated and default parameters. 27th of June, 00:00 – 23:59.

7 Conclusions

For the determination of the soil water balance components, lysimeters serve as a reference for the validation of other instruments and models' measurements, especially in terms of actual evapotranspiration. The high temporal resolution of modern lysimeters and their scales' increasing resolution subject their measurements to a range of error sources. Therefore, prior to utilizing the lysimeter measurements in any subsequent water balance investigations, treatment of the raw data must be undergone.

The comprehensive filtering routine was used to treat the lysimeter and seepage water tank measurement for the period between 12-April and 23-September, 2017. Each of the filter's parameters was calculated based on the site and investigation period's specific conditions, while also filtering the data with the filter's default parameters. The magnitudes of different parameters' impact on the filtration process were also compared, namely the effects of the saturated hydraulic conductivity and evapotranspiration. The effect of the saturated hydraulic conductivity was distinct, unlike the evapotranspiration effect that needed to be examined on a daily range to manifest. The filter's sensitivity to the saturated hydraulic conductivity value is a relationship that has not been investigated in the reviewed literature, and is a peculiar one that deserves further analysis.

Clear differences between the calculated and default filter parameters on the resulting fluxes and summarized masses are seen when investigating the filtered measurements in a high temporal resolution not exceeding a 24 hour period. Where the calculated parameters seem to provide better fitting to the raw data especially on periods of high noise caused by wind, periods of strong precipitation, and periods of evapotranspiration followed by precipitation and vice versa. While on periods of low wind velocities and normal evapotranspiration, the difference between both parameters is almost negligible.

The calculated oscillation threshold filter provided the balance between translating measurement noise into actual fluxes, and underestimating precipitation events due to the mixing error. The improvements it presented are in part due to it being tailored to the smaller SFL-300 lysimeter of 15 cm radius used in this study compared to the 1 m^2 lysimeter used to design the filtering software. Therefore, attempting to filter the data with the default filter's parameters while only changing the oscillation threshold's value to be as calculated seemed to yield better results than the default filter's parameters. Surely, different periods need to be filtered using the same methodology in order to firstly determine whether the oscillation threshold's value significantly changes with different periods and meteorological conditions, e.g. maximum precipitation flux, and to investigate whether using the default parameters with the calculated oscillation threshold can be accepted as an alternative filtering technique in the absence of some of the parameters' values.

Statistical evaluation of the cumulative precipitation of the lysimeter against that of a reference tipping bucket rain gauge was conducted. The rain gauge, like other measurement equipment, suffers from its own systematic and random errors. Hence, a correction formula for its measurements had to be applied prior to utilizing its measurements. When compared, the cumulative precipitation of the default filter parameters showed lesser accordance with the rain gauge's, while the calculated parameters showed a better one. Therefore, it can be accepted

that the calculated parameters result in a better representation of the water balance components compared to the default filter parameters.

Finally, some recommendations for future prospective research to be conducted to further elaborate and expand the results of this study are as follows;

- Investigating the difference between the oscillation threshold filter's value of this study and others during different investigation periods,
- Comparing the resulting water balance components of the selected procedure in this study with the recently modified AWAT filter, or other sophisticated techniques,
- Further investigation of the effect of saturated hydraulic conductivity on the filtration schemes,
- Investigating the relationship of wind speed and rain intensity on the optimization process for the correction tipping bucket rain gauge measurements,

8 References

METER GROUP AG. 2013. SmartField-Lysimeter User Manual. METER GROUP AG, Munich. Available from http://publications.metergroup.com/Manuals/UMS/SmartField-Lysimeter%20User%20Manual%20V12-2013.pdf (accessed December 2019).

Allen, R. G., Pereira, L. S., Howell, T. A., Jensen, M. E. 2011. Evapotranspiration information reporting: I. Factors governing measurement accuracy. Agricultural Water Management. 98 (6). 899–920. doi: 10.1016/j.agwat.2010.12.015.

Arias-Castro, E., Donoho, D. 2009. Does median filtering truly preserve edges better than linear filtering? Annals of Statistics. 37 (3). 1172–1206. doi: 10.1214/08-AOS604.

Barde, P., Barde, M. 2012. What to use to express the variability of data: Standard deviation or standard error of mean? Perspectives in Clinical Research. 3 (3). 113. doi: 10.4103/2229-3485.100662.

Báťková, K., Matula, S., Miháliková, M. 2013. Multimedial Study Guide of Field Hydropedological Measurements. 2nd revised edition [on-line]. English version. Czech University of Life Sciences Prague. Prague, Czech Republic. No pagination. Available at: http://hydropedologie.agrobiologie.cz. ISBN: 978-80-213-2434-3.(acessed March 2020).

Butler, S. 2008. Moving averages. Real Analysis Exchange. 33 (1). 29–30. doi: 10.5948/upo9780883859537.031.

Doležal, F., Hernandez-Gomis, R., Matula, S., Gulamov, M., Miháliková, M., Khodjaev, S. 2018. Actual evapotranspiration of unirrigated grass in a smart field lysimeter. Vadose Zone Journal. 17 (1). doi: 10.2136/vzj2017.09.0173.

Doležal, F., Matula, S., Moreira Barradas, J. M. 2015. Rapid percolation of water through soil macropores affects reading and calibration of large encapsulated TDR sensors. Soil and Water Research. 10 (3). 155–163. doi: 10.17221/177/2014-SWR.

Doležal, F., Mihalikova, M., Matula, S., Moreira Barradas, J.M., Mekonnen, G. B. 2014. The effect of the signal filtration method on the measurement of vertical water flux density using a small smart lysimeter SMF UMS. . 1–8.

Fank, J. 2013. Wasserbilanzauswertung aus Präzisionslysimeterdaten. Proceedings 15. Gumpensteiner Lysimetertagung. (Fank 2011). 85–92.

Gebler, S., Hendricks Franssen, H. J., Pütz, T., Post, H., Schmidt, M., Vereecken, H. 2015. Actual evapotranspiration and precipitation measured by lysimeters: A comparison with eddy covariance and tipping bucket. Hydrology and Earth System Sciences. 19 (5). 2145–2161. doi: 10.5194/hess-19-2145-2015.

Gorry, P. A. 1991. General Least-Squares Smoothing and Differentiation of Nonuniformly Spaced Data by the Convolution Method. Analytical Chemistry. 63 (5). 534–536. doi: 10.1021/ac00005a031.

Hannes, M., Wollschläger, U., Schrader, F., Durner, W., Gebler, S., Pütz, T., Fank, J., Von Unold, G., Vogel, H. J. 2015. A comprehensive filtering scheme for high-resolution estimation of the water balance components from high-precision lysimeters. Hydrology and Earth System Sciences. 19 (8). 3405–3418. doi: 10.5194/hess-19-3405-2015.

Harsch, N., Brandenburg, M., Klemm, O. 2009. Large-scale lysimeter site St. Arnold, Germany: Analysis of 40 years of precipitation, leachate and evapotranspiration. Hydrology and Earth System Sciences. 13 (3). 305–317. doi: 10.5194/hess-13-305-2009.

Herbrich, M., Gerke, H. H. 2016. Autocorrelation analysis of high resolution weighing lysimeter time series as a basis for determination of precipitation. Journal of Plant Nutrition and Soil Science. 179 (6). 784–798. doi: 10.1002/jpln.201600169.

Hoffmann, M., Schwartengräber, R., Wessolek, G., Peters, A. 2016. Comparison of simple rain gauge measurements with precision lysimeter data. Atmospheric Research. 174–175

. 120–123. doi: 10.1016/j.atmosres.2016.01.016.
Howell, T. A., Schneider, A. D., Jensen, M. E. 1991. History of lysimeter design and use for evapotranspiration measurements. Lysimeters for Evapotranspiration and Environmental Measurements. (January). 1–9.
Huang, W., Zhang, C., Xue, X., Chen, L. 2012. A data acquisition system based on outlier detection method for weighing lysimeters. IFIP Advances in Information and Communication Technology. 368 AICT (PART 1). 471–478. doi: 10.1007/978-3-642-27281-3_53.
Hyndman, R. J., Koehler, A. B. 2006. Another look at measures of forecast accuracy. International Journal of Forecasting. 22 (4). 679–688. doi: 10.1016/j.ijforecast.2006.03.001.
Johnsson, H., Jansson, P. E. 1991. Water balance and soil moisture dynamics of field plots with barley and grass ley. Journal of Hydrology. 129 (1–4). 149–173. doi: 10.1016/0022-1694(91)90049-N.
Khan, B. R., Mainuddin, M., Molla, M. N. 1993. Design, construction and testing of a lysimeter for a study of evapotranspiration of different crops. Agricultural Water Management. 23 (3). 183–197. doi: 10.1016/0378-3774(93)90027-8.
Kohfahl, C., Molano-Leno, L., Martínez, G., Vanderlinden, K., Guardiola-Albert, C., Moreno, L. 2019. Determining groundwater recharge and vapor flow in dune sediments using a weighable precision meteo lysimeter. Science of the Total Environment. 656 . 550–557. doi: 10.1016/j.scitotenv.2018.11.415.
Kohnke, H., Dreirelbis, F. R. 1940. a Survey and Discussion of Lysimeters a Bibliography on Their. . 44 (372).
Legates, D. R., McCabe, G. J. 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resources Research. 35 (1). 233–241. doi: 10.1029/1998WR900018.
Manatsa, D., Chingombe, W., Matarira, C. H. 2008. The impact of the positive Indian Ocean dipole on Zimbabwe droughts Tropical climate is understood to be dominated by. International Journal of Climatology. 2029 (March 2008). 2011–2029. doi: 10.1002/joc.
Marek, G. W., Evett, S. R., Gowda, P. H., Howell, T. H., Copeland, K. S., Baumhardt, R. L. 2014. Post-processing techniques for reducing errors in weighing lysimeter evapotranspiration (ET) datasets. Transactions of the ASABE. 57 (2). 499–515. doi: 10.13031/trans.57.10433.
Meißner, R., Prasad, M. N. V., Du Laing, G., Rinklebe, J. 2010. Lysimeter application for measuring the water and solute fluxes with high precision. Current Science. 99 (5). 601–607.
Meissner, R., Seeger, J., Rupp, H., Seyfarth, M., Borg, H. 2007. Measurement of dew, fog, and rime with a high-precision gravitation lysimeter. Journal of Plant Nutrition and Soil Science. 170 (3). 335–344. doi: 10.1002/jpln.200625002.
Mekonnen, G. B., Matula, S., Doležal, F., Fišák, J. 2015. Adjustment to rainfall measurement undercatch with a tipping-bucket rain gauge using ground-level manual gauges. Meteorology and Atmospheric Physics. 127 (3). 241–256. doi: 10.1007/s00703-014-0355-z.
Nolz, R., Kammerer, G., Cepuder, P. 2013b. Improving interpretation of lysimeter weighing data. Bodenkultur. 64 (1–2). 27–35.
Nolz, R., 2016. A review on the quantification of soil water balance components as a basis for agricultural water management with a focus on weighing lysimeters and soil water sensors. Bodenkultur. 67 (3). 133–144. doi: 10.1515/boku-2016-0012.
Nolz, R., Kammerer, G., Cepuder, P. 2013a. Interpretation of lysimeter weighing data affected by wind. Journal of Plant Nutrition and Soil Science. 176 (2). 200–208. doi:

10.1002/jpln.201200342.
Ohki, M., Zervakis, M. E., Venetsanopoulos, A. N. 1995. 3-D Digital filters. Control and Dynamic Systems. 69 (C). 49–88. doi: 10.1016/S0090-5267(05)80038-6.
Ostertagová, E., Ostertag, O. 2016. Methodology and application of Savitzky-Golay moving average polynomial smoother. Global Journal of Pure and Applied Mathematics. 12 (4). 3201–3210.
Parisi, S., Mariani, L., Cola, G., Maggiore, T. 2009. Mini-lysimeters evapotranspiration measurements on suburban environment. Italian Journal of Agrometeorology. (3). 13–16.
Peters, A., Nehls, T., Schonsky, H., Wessolek, G. 2014. Separating precipitation and evapotranspiration from noise - A new filter routine for high-resolution lysimeter data. Hydrology and Earth System Sciences. 18 (3). 1189–1198. doi: 10.5194/hess-18-1189-2014.
Peters, A., Groh, J., Schrader, F., Durner, W., Vereecken, H., Pütz, T. 2017. Towards an unbiased filter routine to determine precipitation and evapotranspiration from high precision lysimeter measurements. Journal of Hydrology. 549 . 731–740. doi: 10.1016/j.jhydrol.2017.04.015.
Peters, A., Nehls, T., Wessolek, G. 2016. Technical note: Improving the AWAT filter with interpolation schemes for advanced processing of high resolution data. Hydrology and Earth System Sciences. 20 (6). 2309–2315. doi: 10.5194/hess-20-2309-2016.
Pütz, T., Fank, J., Flury, M. 2018. Lysimeters in Vadose Zone Research. Vadose Zone Journal. 17 (1). 180035. doi: 10.2136/vzj2018.02.0035.
Rožnovsky, J. 2012. Sucho na území ČR a jeho dopady. Český hydrometeorologický ústav, pobočka Brno. Available from http://portal.chmi.cz/files/portal/docs/katastrofy/26zasedani/Roznovsky_sucho_230412.pdf. (Accessed April 2020).
Ruth, C. E., Michel, D., Hirschi, M., Seneviratne, S. I. 2018. Comparative Study of a Long-Established Large Weighing Lysimeter and a State-of-the-Art Mini-lysimeter. Vadose Zone Journal. 17 (1). 170026. doi: 10.2136/vzj2017.01.0026.
Saaltink, M. W., Kohfahl, C., Molano-Leno, L. 2020. Analysis of water vapor adsorption in soils by means of a lysimeter and numerical modeling. Vadose Zone Journal. 19 (1). 1–18. doi: 10.1002/vzj2.20012.
Savitzky, A., Golay, M. J. E. 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Analytical Chemistry. Analytical Chemistry. 36 (8). 1627–1639. doi: 10.1021/ac60214a047.
Schrader, F., Durner, W., Fank, J., Gebler, S., Pütz, T., Hannes, M., Wollschläger, U. 2013. Estimating Precipitation and Actual Evapotranspiration from Precision Lysimeter Measurements. Procedia Environmental Sciences. 19 . 543–552. doi: 10.1016/j.proenv.2013.06.061.
Smith, JB., Huq, S., Mata, LJ., Nemesova, I., Toure, S. 1996. Vulnerability and Adaptation to Climate Change, Interim Results from the U . S . Country Studies Program, Kluwer Academic Publishers. DOI 10.1007/978-94-017-3653-4.
Sokolov, A. A., Chapman, T. G. 1974. Methods for Water Balance Computations, an internation guide for research and practiceUNESCO Press. ISBN: 9231012274.
Soldevilla-Martinez, M., Quemada, M., López-Urrea, R., Muñoz-Carpena, R., Lizaso, J. I. 2014. Soil water balance: Comparing two simulation models of different levels of complexity with lysimeter observations. Agricultural Water Management. 139 . 53–63. doi: 10.1016/j.agwat.2014.03.011.
Vaughan, P. J., Trout, T. J., Ayars, J. E. 2007. A processing method for weighing lysimeter data and comparison to micrometeorological ETo predictions. Agricultural Water Management. 88 (1–3). 141–146. doi: 10.1016/j.agwat.2006.10.008.

Vaughan, Peter J., Ayars, J. E. 2009. Noise reduction methods for weighing lysimeters. Journal of Irrigation and Drainage Engineering. 135 (2). 235–240. doi: 10.1061/(ASCE)0733-9437(2009)135:2(235).

Vogel, H.-J. 2015. A comprehensive filtering scheme for high-resolution estimation of the water balance components from high-precision lysimeters. Hydrology and Earth System Sciences. 19 (8). 3405–3418. doi: 10.5194/hess-19-3405-2015-supplement.

Von Unold, G., Fank, J. 2008. Modular design of field lysimeters for specific application needs. Water, Air, and Soil Pollution: Focus. 8 (2). 233–242. doi: 10.1007/s11267-007-9172-4.

Zhang, Q., Wang, S., Yue, P., Wang, R. 2019. A measurement, quantitative identification and estimation method(QINRW) of non-rainfall water component by lysimeter. MethodsX. 6 . 2873–2881. doi: 10.1016/j.mex.2019.11.012.

Zhu, X., Wu, X. 2004. Class Noise vs. Attribute Noise: A Quantitative Study. Artificial Intelligence Review. 22 (3). 177–210. doi: 10.1007/s10462-004-0751-8.

Zupanc, V., Nolz, R., Cepuder, P., Bračič-Železnik, B., Pintar, M. 2012. Determination of water balance components with high precision weighing lysimeter in Kleče. Acta Agriculturae Slovenica. 99 (2). 165–173. doi: 10.2478/v10014-012-0016-1.

9 List of Figures

10 List of Tables

11 List of Abbreviations

A: *Runoff*
A_s: *Surface Area*
ANOVA: *Analysis of Variance*
A_P: *Anchor Point*
AWAT: *Adaptive Window Adaptive Threshold*
B: *Biomass Change*
BREB: *Bowen Ratio Energy Balance*
CR: *Capillary Rise*
C_R: *Catch Ratio*
CULS: *Czech University of Life Sciences in Prague*
D: *Dew*
ET: *Evapotranspiration*
ET_a: *Actual Evapotranspiration*
FDR: *Frequency Domain Reflectometry*
I*: Irrigation Water*
k: *Correction Coefficient*
KSAT: *Saturated Hydraulic Conductivity*
LSD: *Least Significant Difference*
LYW: *Lysimeter Column Mass*
MA: *Moving Average*
NRMSE: *Normalized Root Mean Square Error*
O: *Outflow*
P: *Precipitation*
Q: *Quantile*
R^2: *R- Squared, Coefficient of Determination*
RMSE: *Root Mean Square Error*
ΔS: *Storage*
SG: *Savitsky- Golay Filter*
SPA: *Soil-Plant-Atmosphere Continuum*
SW: *Seepage Water*
SWW: *Seepage Water Tank Mass*
u: *Wind*
ΔW: *Lysimeter Mass Change (kg)*
w: *Averaging Window*