Royal Netherlands
Meteorological Institute
*Ministry of Infrastructure
and Water Management*

Data Quality Control Procedure for Cabauw Lysimeter:
Precipitation, Evapotranspiration, and Latent Energy

J.M.I. Strickland

De Bilt, 2026| Technical report; TR 26-02

Data Quality Control Procedure for Cabauw Lysimeter:
Precipitation, Evapotranspiration, and Latent Energy

Jessica M. I. Strickland

Contact: jessica.strickland@knmi.nl
Koninklijk Nederlands Meteorologisch Instituut (KNMI),
3731AE, De Bilt, The Netherlands

March, 2026

Abstract

The Royal Netherlands Meteorological Institute (KNMI) has operated an Eijkelkamp Smart Lysimeter at the Cabauw observational site since 2020. This instrument measures evapotranspiration and precipitation at the surface directly with a high temporal resolution based on changes in weight. This document describes the current post-processing and quality control which is applied to validated lysimeter datasets as well as the calculation of latent energy derived from these measurements.

Contents

1 Introduction

Evapotranspiration (ET) is the process of water leaving the surface due to both evaporation and plant-released transpiration. Better understanding and monitoring of these hydrological processes is crucial as the climate changes and extreme weather conditions such as droughts are increasingly common in areas such as the Netherlands. Currently, the Royal Netherlands Meteorological Institute (KNMI) employs a simple Makkink model [1] based on temperature and incoming solar radiation to estimate the potential evaporation and their satellite observations also provide some insight. However, in-situ evaporation observations throughout the Netherlands are limited and are often acquired using indirect methods. Moreover, many established observational methods infer ET at the surface based on moisture fluxes detected above the ground. In contrast, KNMI has recently deployed a lysimeter which measures the water that enters and leaves the surface directly.

Since 2020, a Eijkelkamp Smart Lysimeter (described in Section 2.1) has measured ET at the Cabauw observational site for atmospheric research (details in Section 2.2) as part of the surface energy balance observation array. The lysimeter is a column of soil from which ET and precipitation are calculated based on the changes in weight. These values can be compared to nearby precipitation sensors and the long-standing established indirect methods for measuring ET, such as eddy covariance (EC), bowen ratio (BR), and the residual method as described in Reference [2]. While the lysimeter measurements are directly at the surface and have a high temporal resolution, this apparatus is subject to inherent limitations, making the Cabauw site a great location for inter-comparison and evaluating its performance. In addition to the fact that the lysimeter is a sensitive instrument with a small spatial representation, the observations can be heavily impacted by varying maintenance and environmental conditions. Therefore, the data must be surveyed and requires quality control.

This report summarizes the current post-processing applied to the validated lysimeter datasets. Overall, quality control of these datasets is complex and can be improved with further research and development. Currently, a python code has been developed and is used to process the raw lysimeter output delivered by Eijkelkamp (refer to Section 2.3). This code applies some automatic flagging and filtering; however, manual validation by experts is still required at this stage to evaluate remaining suspicious data. The automatic filtering is based on the mechanical limitations of the instrument and is described in Section 3. The manual filtering that follows is explained in Section 4. Once the lysimeter output is validated, the Latent Energy (LE) is also calculated (described in Section 5), which can then be used to investigate the surface energy and hydrological balance. A summary of the process, as well as limitations of the quality control procedure, are discussed in Section 6.

2 Lysimeter Setup

2.1 Smart lysimeter instrument

A lysimeter is a sub-surface instrument which measures weight changes due to moisture loss and accumulation. The smart lysimeter operating at the Cabauw observational site, shown in Figure 1, is manufactured by Royal Eijkelkamp [3]. The primary features of this lysimeter are a soil monolith and water reservoir which exchange water depending on the external soil moisture. The total weight of the enclosed system decreases and increases during ET or precipitation, respectively (see Figure 5). A pressure plate between the monolith and reservoir transports moisture to and from the soil column to regulate the moisture of the monolith corresponding to an external tensiometer which measures the soil moisture near the lysimeter. This smart feature ensures that the monolith is representative of the surrounding soil. Nonetheless, the lysimeter is a small instrument with limited spatial representativeness.

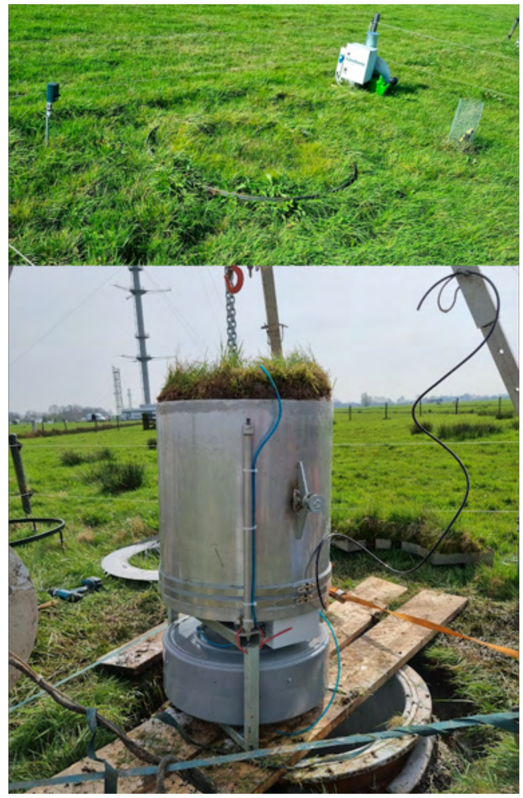

Figure 1: **Upper**: Top-view of undisturbed lysimeter plot with external tensiometer located on the far right. **Lower**: Unearthed lysimeter showcasing the soil column and lower reservoir.

The small size of the lysimeter is convenient but is also imposes limitations. When installed, a soil column is carefully extracted from the location of interest to minimize disturbance. Specifically, the diameter and depth of the monolith are 50 cm with a sample volume of 98 L. Grass beds (visible in Figure 3a) are installed around the column to cover parts of the apparatus with ground surface, however, their depth is only 15 cm which can limit water uptake for the grass. During extremely wet conditions, if the monolith weight exceeds 18 kg or the reservoir exceeds 15 kg, the lysimeter will automatically pump water outside (and is no longer a closed system). Ultimately, completely drying out or becoming saturated during extreme or extended events is possible and compromises functionality and data quality.

While ET and precipitation (and therefore lysimeter weight) are the core focus, additional lysimeter observations include soil moisture and temperature sensors which are located near both the top and bottom of the monolith. The instrument's output data is described in Section 2.3.

2.2 Cabauw observational site and maintenance

A smart lysimeter has been deployed at KNMI's Cabauw observational site since July, 2020 to present. This site is a mid-latitude grassland located in Lopik, the Netherlands (51.972 N, 4.927 E), for its limited disturbance from surrounding infrastructure and geographic representativeness. For decades, an impressive range of in-situ observations have been measured at this site, including various methods for determining latent energy, making this a valuable location to investigate the data quality generated by a lysimeter.

The lysimeter is located in a subplot of the energy balance (EB) field (see Figure 2a), roughly 300 m North of the 213 m tall mast that characterizes the observational site since 1972 [2, 4, 5]. Along this mast, various meteorological parameters such as temperature, relative humidity, visibility, wind speed and direction, are measured at multiple heights and 10-minute averaged values are made available to the public via the KNMI Data Platform. Precipitation intensity is measured by a nearby KNMI precipitation gauge and a present weather sensor. These continuous observations provide useful insight for more specialized instruments and scientific campaigns. For instance, an array of instruments are deployed in the EB-field to measure each component of the surface energy balance (SEB) such as radiation, heat, and moisture fluxes. Many of these observations span decades and detailed information can be found in Reference [6].

ET has been measured at this site using a variety of established methods. The variety of meteorological observations are also used to model the potential ET which can be used as a reference for the various in-situ methods [1, 7, 8]. Since roughly 1986, ET has been measured using the eddy covariance (EC) method based on turbulence measurements from a 3D sonic anemometer in combination with an open path sensor (shown in Figure 2b) to determine the vertical moisture fluxes. This method was found to underestimate dewfall [9]. Also, since 2007, ET has been measured using the bowen ratio (BR) method, based on the relative moisture fluxes from a psychrometer tower (shown in Figure 2c) that measures the air and dew-point temperature at 1, 2, and 4 m [10]. Uniquely, ET can also be determined using the residual method, where all other components of the SEB are measured and (with some assumptions) the remainder can be attributed to latent energy. In contrast to these indirect methods, the lysimeter measures ET directly at the surface and has a smaller spatial representation. However, the sensitivity of these lysimeter measurements requires careful handling and interpretation. Moreover, the lysimeter data is only available intermittently due to various issues and data has been impacted by varying site conditions.

The site's grass length influences both the surface's ability to evaporate water and particularly the amount of transpiration. The vegetation cover is close to 100% year-round and the EB-field grass is maintained by a combination of sheep grazing, and mowing if the grass length exceeds 15 cm. However, the subplots are (electric) fenced to prevent disturbance from the sheep and are only mowed and maintained at 8-12 cm. Mowing the lysimeter must be performed with a smaller tool and with care to minimize disturbance and prevent damage. In Figure 1, a metal ring is visible which is placed to indicate the location of the lysimeter and a small cage protects the external tensiometer. Nonetheless, inconsistency with maintenance and varying grass condition due to environmental factors (examples shown in Figure 3) remain as issues for data quality control.

Figure 2: **(a)** Energy balance field with a variety of related instruments described in Reference [6]. **(b)** Turbulence measured at 3 m height using a 3D sonic anemometer and open path sensor installed together on a rotator which aligns with the wind direction. **(c)** Psychrometer measuring both the air and dew-point temperature at 1, 2 and 4 m. **(d)** KNMI precipitation gauge located at the automatic weather station near the main mast.

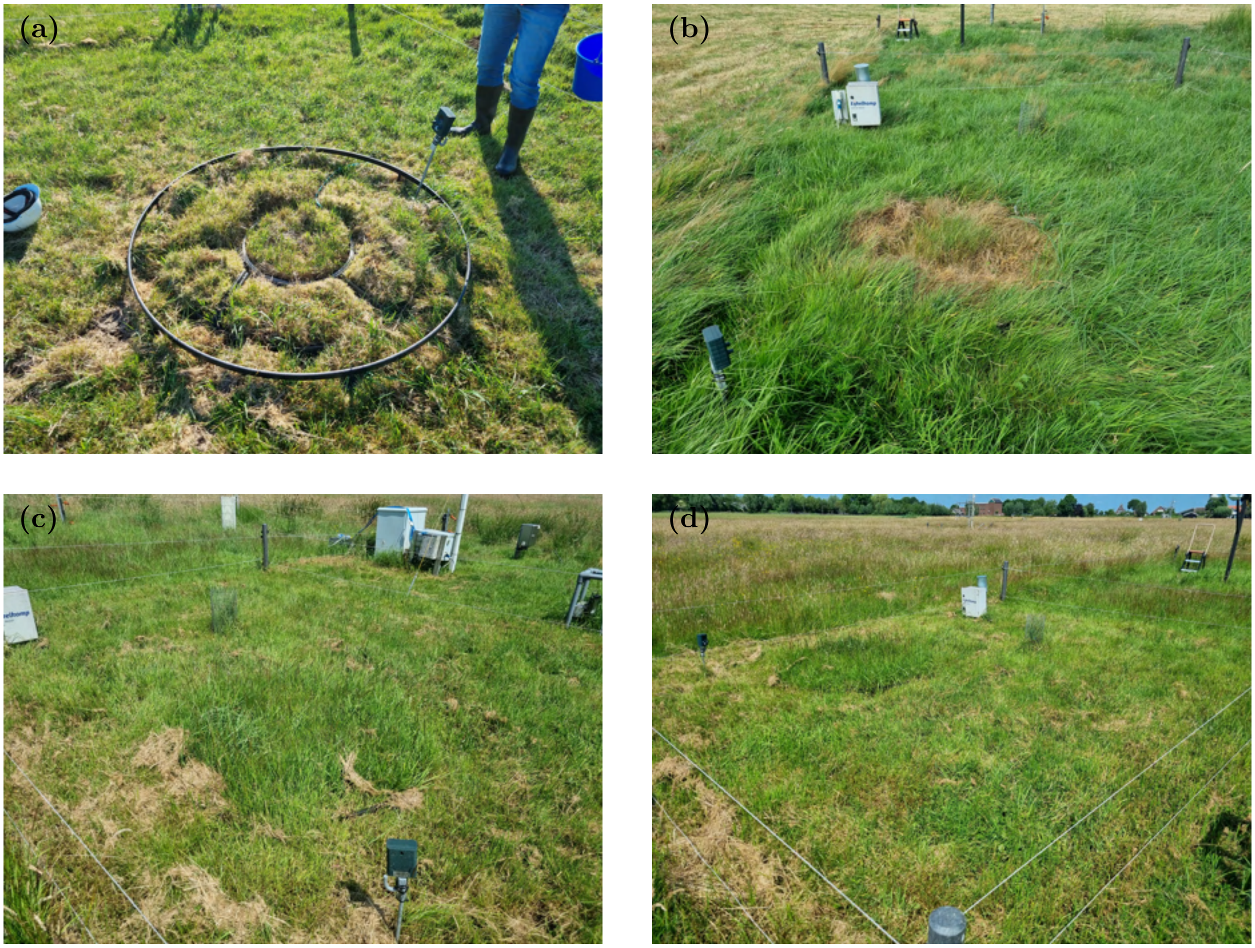

Figure 3: Grass condition examples. **(a)** Lysimeter reinstalled after maintenance exhibiting the outlines of the column and surrounding grass beds. **(b)** Grass in the surrounding grass beds has dried out and is discolored compared to lysimeter column and surrounding subplot. **(c)** Cut brown grass laying on top of lysimeter subplot. **(d)** Grass in lysimeter subplot is a different height than the grass in the surrounding field.

Besides grass length, the condition of the grass and soil impact the ET and representativeness of the soil column. The soil is characterized by organic matter, clay until about 75 cm, which then transitions to peat. In the winter, the soil can often become over-saturated due to the wet climate and high water water table. Meanwhile, the hot dry summers cause the clay soil to harden and crack (shown in Figure 4a), which causes preferential water pathways as well as grass discoloration. Vegetation in the grass beds surrounding the lysimeter is even more susceptible (see Figure 3b) due to the shallow 15 cm depth which prevents the grass roots from penetrating as deep to access water. Therefore, occasionally, the grass beds must be watered to reflect the surrounding environment.

Furthermore, animal interference has been observed, as shown in Figure 4b-d. However, the extent to which the data is impacted by their direct interaction and tunnels is not easily determined. A humane acoustic rodent deterrent has been placed by the lysimeter, but the effectiveness has proven to be questionable at best. Beyond that, we have observed interference such as electric fence failure allowing the sheep into the plot, rodents have made nearby nests, and human activity has occurred both documented and undocumented. These impacts and other sources of interference contribute to the complexity of validating the lysimeter data sets.

Figure 4: Environmental issue examples. **(a)** Exposed cracked clay during prolonged hot dry periods. **(b)** Upper soil profile where surrounding grass beds are removed showcasing the extent of the grass roots and rodent-made tunnels. **(c)** Rodent nest next to lysimeter monolith exposed by removing grass bed. **(d)** Damage to lysimeter silicone ring which may be due to rodents as droppings were present.

2.3 Lysimeter data

The lysimeter reports multiple variables each minute, including the monolith and reservoir weight in grams and the pressure of the suction plate between them which moderates the moisture of the soil column. Evapotranspiration (ET) and precipitation are determined from changes in weight as described below. The pressure of the external tensiometer is also reported which can be compared to the lysimeter pressure to ensure environmental representativeness (refer to Section 2.1). Additionally, the monolith temperature and soil moisture are measured at the top and bottom of the monolith. However, the soil moisture is reported in dielectric permittivity of the soil (ϵ) and must be converted to a volumetric measurement. Specifically, soil moisture is converted from ϵ to m^3/m^3 using the empirical calibration equation by Topp et al. [11]. Unfortunately, the nearby soil moisture profile observations have been out of commission during the lysimeter's operation and cannot be used to corroborate the data. The lysimeter also reports informative parameters such as external pumping when water is expelled, as well as a variety of alerts which can be used for flagging and filtering erroneous data (refer to Section 3). Ultimately, we focus on ET and precipitation as we describe the quality control procedure.

The lysimeter automatically calculates the hourly ET and precipitation in mm based on the difference in the total lysimeter weight over time ($\Delta W_{lys}(t) = W_{lys}(t = n) - W_{lys}(t = n - 1)$). As water enters and leaves the closed system, an increase in lysimeter weight is classified as precipitation and a decrease in weight is classified as ET:

$$\begin{aligned} \Delta W_{lys}(t) > 0 &= \text{Precipitation} \\ \Delta W_{lys}(t) < 0 &= \text{Evapotranspiration} \end{aligned} \tag{1}$$

We determine the ET and precipitation at a higher frequency. In post-processing, we calculate the one-minute values of ET and precipitation directly using ΔW_{lys}. First, we calculate the water weight in mm from reported values in grams (Equation 2) based on the density of water ρ_w in g/mm^3 and the lysimeter surface area (where radius $R = 250$ mm). An example of the total lysimeter weight in mm and changes over time are shown in Figure 5. These values are ultimately resampled to 10-minute averages (as described in Section 4).

$$W\ [\text{mm}] = W\ [\text{g}]/\rho_w \pi R^2 \tag{2}$$

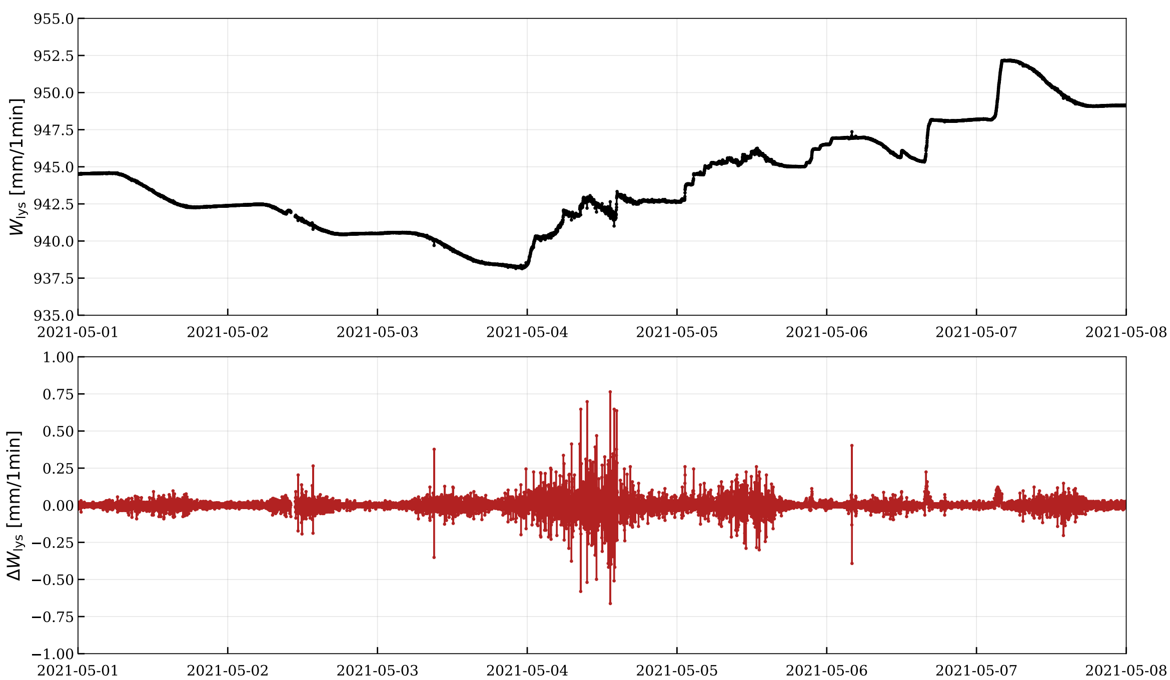

Figure 5: Lysimeter weight equivalent in water (W_{lys}) and the difference in weight (ΔW_{lys}) throughout one week in May, 2021.

The 10-minute averaged precipitation and ET values provide more detail than the automatically reported hourly values and an example for comparison is shown in Figure 6. As expected, the calculated values show more variation than the hourly reported values. Comparing the precipitation intensity, we observe that the hourly reported values may have better agreement with the nearby KNMI precipitation gauge during relatively larger precipitation events. However, this precipitation gauge is less sensitive to sudden and small changes in precipitation [12]. Furthermore, the hourly reported values sometimes exhibit non-zero values when the precipitation gauge and calculated lysimeter values do not report these sudden smaller spikes. Therefore, the calculated lysimeter value is likely more realistic than the hourly reported values. Comparing the hourly reported and calculated values of ET to the Makkink model for potential evaporation, we observe better agreement overall from the calculated values. Therefore, our findings show that both the precipitation intensity and evaporation should be calculated from the one-minute samples while the hourly reported values can provide a coarse indication. These findings also exhibit the potential of the lysimeter as a reference device for less responsive observational equipment.

While these raw lysimeter data exhibit relatively good agreement with references, post-processing and quality control is still required to produce validated datasets. Flagging and filtering procedures are described in Sections 3 and 4. Then, the latent energy is calculated from the validated lysimeter data and nearby precipitation gauge measurements (refer to Section 5). Finally, these values are stored alongside the raw lysimeter data and published.

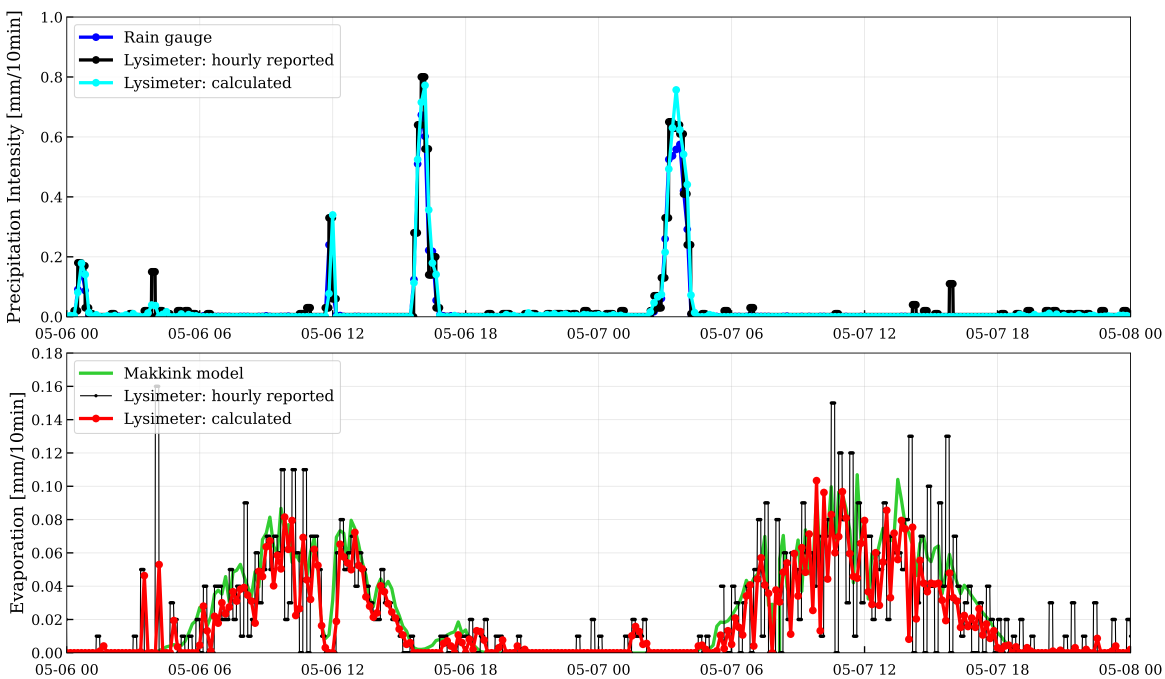

Figure 6: Precipitation intensity and evapotranspiration calculated using unfiltered ΔW_{lys} compared to the automatically reported hourly values. For reference, the precipitation intensity reported by the nearby precipitation gauge is also shown as well as potential evaporation estimates from the Makkink model [1].

3 Mechanical Filters

The lysimeter mechanisms have known limitations and corresponding erroneous data can be identified, flagged, and then removed for the validated datasets. The mechanical filters below are automatically applied to the raw one-minute data, removing samples characterized by any the following:

- The lysimeter reported errors/warnings.
- The difference in monolith weight exceeds 400 g between one minute samples.
- The difference in reservoir weight exceeds 200 g between one minute samples.
- The pressure difference between the lysimeter pressure plate and the external tensiometer exceeds |0.2| kPa.
- A weight limit is reached and the pump which expels water outside of the system is engaged.

A flag array is created (initiated with 0.0) for each of the above criteria and when met, the sample is flagged (1.0) for automatic removal by replacing the values with NaN. Each filter can be individually turned on and off and the thresholds can be adjusted with ease. Currently, these mechanical filters are applied to the one-minute data and then further processing and manual filters are applied as described in the following Section 4.

For example, in Figure 7, the difference between the lysimeter's pressure plate (which is responsible for adjusting the column's soil moisture) and the external tensiometer's reported pressure is shown alongside the flag array. As shown, when the limit is exceeded and the column's soil moisture is not representative of the surrounding environment, the data point is flagged for removal. Differences in soil moisture often occur when the lysimeter cannot adapt as quickly as the surrounding environment, such as during large rain events. The other criteria are handled similarly and an example of the flagged and filtered lysimeter weight is shown in Figure 8 alongside the flag arrays. In this case, we can see that a heavy precipitation event caused the lysimeter monolith weight to increase too quickly and eventually its upper weight limit is reached causing the water to be pumped outside of the lysimeter, temporarily making it an unclosed system.

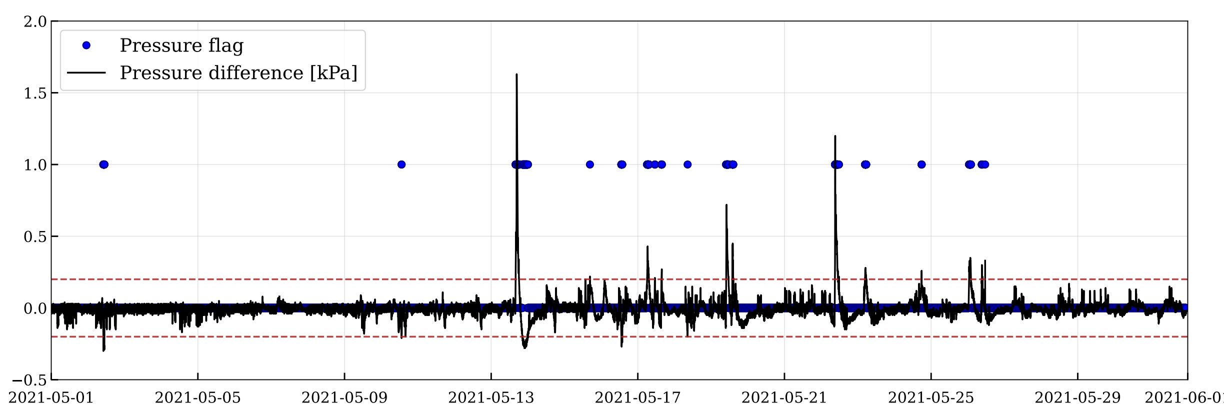

Figure 7: Pressure difference between the lysimeter pressure plate and the external tensiometer throughout May, 2021. Data characterized by a pressure difference >|0.2| kPa (red dashed lines) is flagged (1.0).

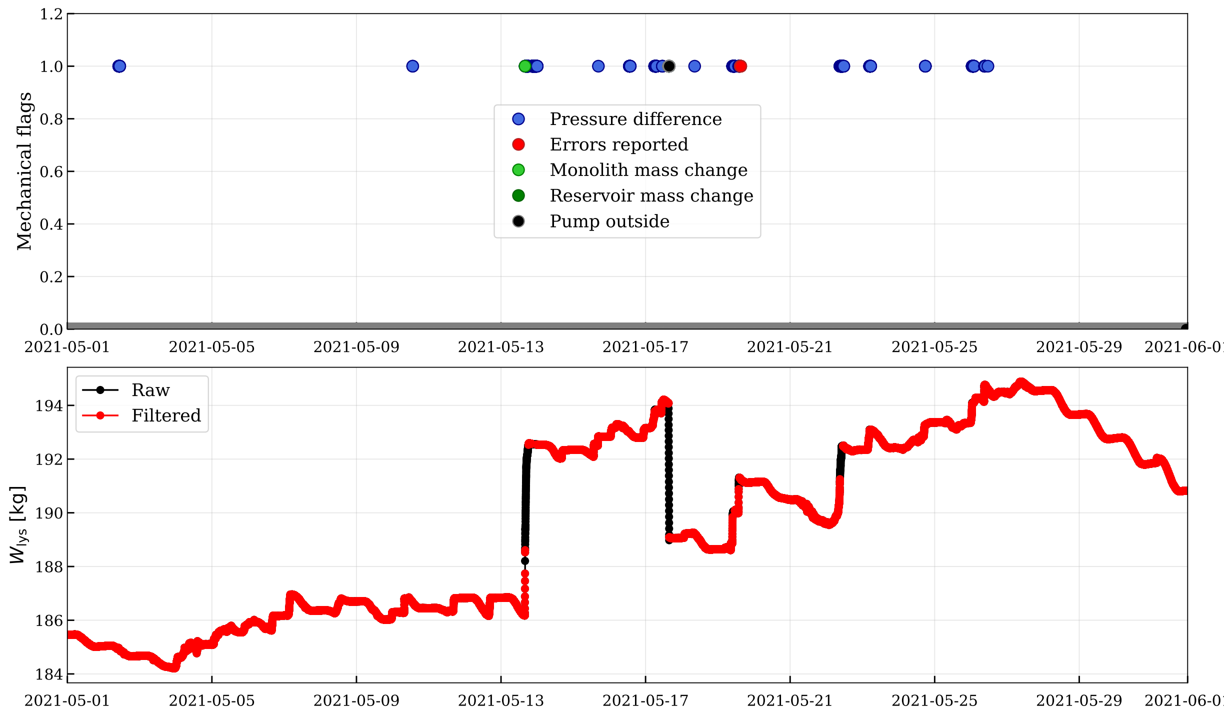

Figure 8: Example of flagged one-minute data points (1.0) in May, 2021, due to mechanical limitations shown with the raw unfiltered and resulting filtered lysimeter weight.

4 Manual Filters

After data has been automatically filtered based on some of the mechanical limitations, manual filtering is still necessary to address other interference and dataset noise. Moreover, unphysical data that is identified via visual inspection is documented and removed from the dataset. Manual data rejection is applied conservatively and cautiously with a bias towards data retention particularly for historical data. Precipitation intensity reported by a nearby sensor as well as potential ET model values are input for the quality control code to better understand and evaluate the conditions of suspicious events. The manual filtering and remaining steps are described below.

- First, the (difference in) lysimeter weight values in grams are converted to equivalent millimeters of water, plotted, and visually inspected by experts for any remaining irregularities that indicate undetected issues. Such physically unrealistic outliers and periods of interference are removed manually by listing which time intervals to omit.

- Smoothing occurs after erroneous data has been removed so that these samples do not contaminate neighbouring data. Specifically, smoothing is applied using a triangle moving average with a ten minute window for simplicity.

- Next, the data is resampled to ten-minute averages, where a minimum number of seven one-minute samples is required. Then, the difference in weight is calculated.

- At this stage, an optional precipitation filter is available (flag array provided in data file but not applied directly) which reduces noise when studying ET. When activated, this flag omits data samples where precipitation is reported by the nearby KNMI precipitation gauge.

This filter has an adjustable window due to the differing sensitivity and proximity of the two instruments. Data is currently flagged ten minutes before and after the gauge measures precipitation intensity > 0.0 mm/10min.

- Finally, the difference in lysimeter weight (mm/10min) is used to determine the precipitation intensity and evaporation based on Equation 1 (example with raw data shown in Figure 6). The latent energy flux is also calculated but is described in Section 5.

Manual filtering based on visual inspection is not trivial nor absolute, and therefore, the manual flag array is also provided with the dataset. Identified irregularities beyond mechanical limitations occur and are treated for each case. Between one-minute samples, sudden changes where $|\Delta W_{lys}| < 0.5$ mm and are immediately followed by a similar magnitude change in the opposite direction, are viewed as minor disturbances which will be balanced in the 10-minute resampling. Meanwhile, outliers which are higher in magnitude and not characterised by a similar reaction require further inspection. For example, in Figure 9, we showcase outlier data that would not be appropriate to remove because it is likely condensation. Specifically, we observe an increase in weight before sunrise (5:25 CEST = 3:25 UTC), after a very clear night. In this case, condensation has accumulated on the lysimeter grass, a form of latent energy which will not be captured by the surrounding sensors such as the precipitation gauge. This example exhibits the complexity of developing and applying automatic outlier detection.

There are known disturbances which impact the overall data quality such as malfunctioning, extreme drought, saturation, and more. As a result, the dataset often exhibits noise, from known and also unobserved interference. Ultimately, even after removing obvious interference, the impact of irregular maintenance cannot be easily removed or addressed. For example, Figure 10 shows (unvalidated) daily LE estimates in July, 2022, with a break in continuity and sudden shift. On July 13th, the lysimeter grass was mowed, and we observe that prior to this date, the LE reported by the lysimeter (LELYS) displayed better agreement with the models for potential ET (MAK24H, LEPMJS) [13] because the long grass facilitated ET before it was cut. Meanwhile, after this date, the lysimeter agrees better with the eddy covariance method (LEEC). In this case, the impacts are clear, but nonetheless, improving the quality of these measurements is complex as the data represents different and inconsistent situations. Furthermore, maintenance information throughout the entire deployment is unfortunately limited and in other cases the impacts are far less obvious. Potential impacts of phenomena such as high winds with tall grass, freezing, and run-off, and other yet unidentified processes, remain uncertain. Therefore, we approach the manual data quality cautiously and all impacts to data quality cannot be mitigated.

After the data is inspected and erroneous data is removed, the data can be resampled and smoothed to reduce noise and capture the trends. An example of the smoothing and resampling is shown in Figure 11. Here, a moving triangle average is used for its simplicity while giving more weight to the middle of the data window compared to a simple moving average. Other, more complex and potentially suitable smoothing algorithms are available, such as the adaptive window adaptive threshold (AWAT) filter [14], which should be explored in future work. As expected, the data variations are reduced when the triangle averaging is applied and diminish further when the data is resampled to 10-minute averaged values. Moreover, the sudden increases or decreases in weight become less pronounced. Note, that some values are not present after resampling due to reduced 1-minute data availability (minimum seven per ten minutes), such as around 22:30 UTC in this plot.

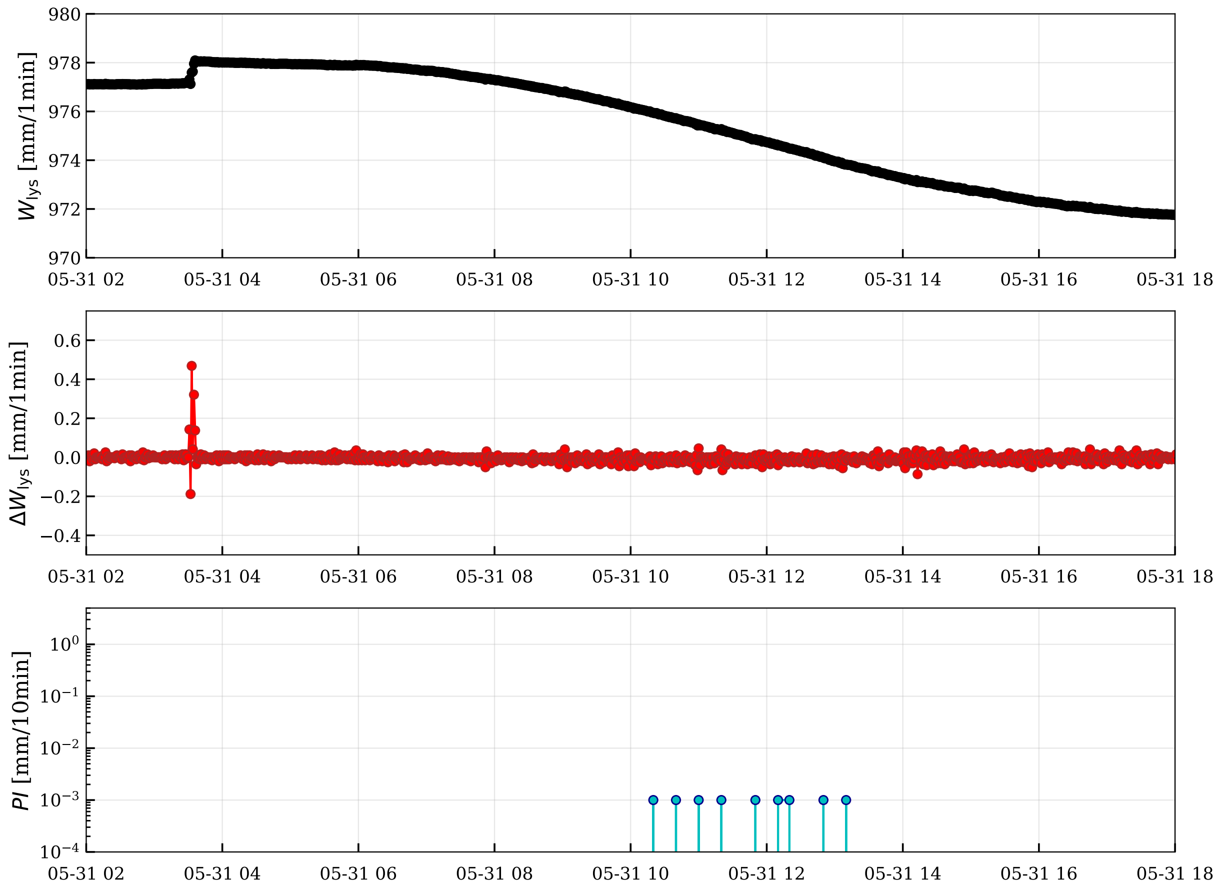

Figure 9: Example of condensation accumulating on the lysimeter before sunset on May 31st, 2021. Specifically, time series [MM-DD HH(UTC)] of lysimeter weight, the weight difference, and the precipitation intensity (PI) reported by a nearby KNMI precipitation gauge.

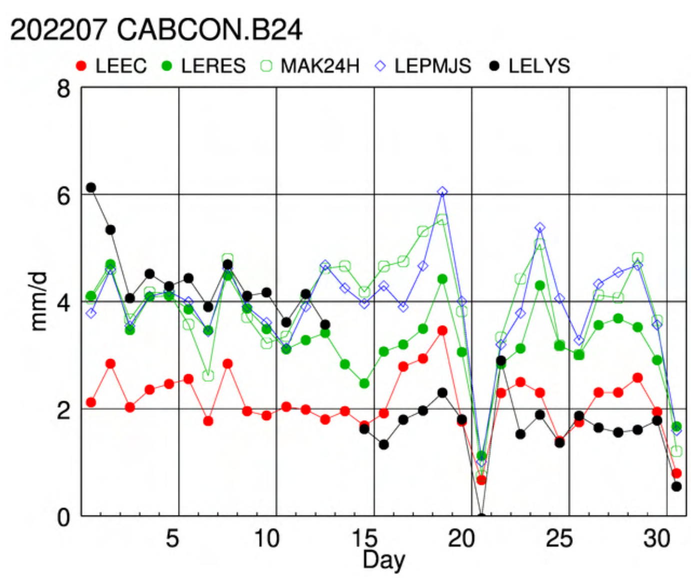

Figure 10: Daily Latent Energy (LE) in July, 2022, shown for the eddy covariance method (LEEC), residual method (LERES), Makkink model (MAK24) [1], Penman-Monteith model (LEPMJS) [7] and lysimeter (LELYS). Note LE behaviour on and after the 13$^{\text{th}}$ when the lysimeter grass was mowed.

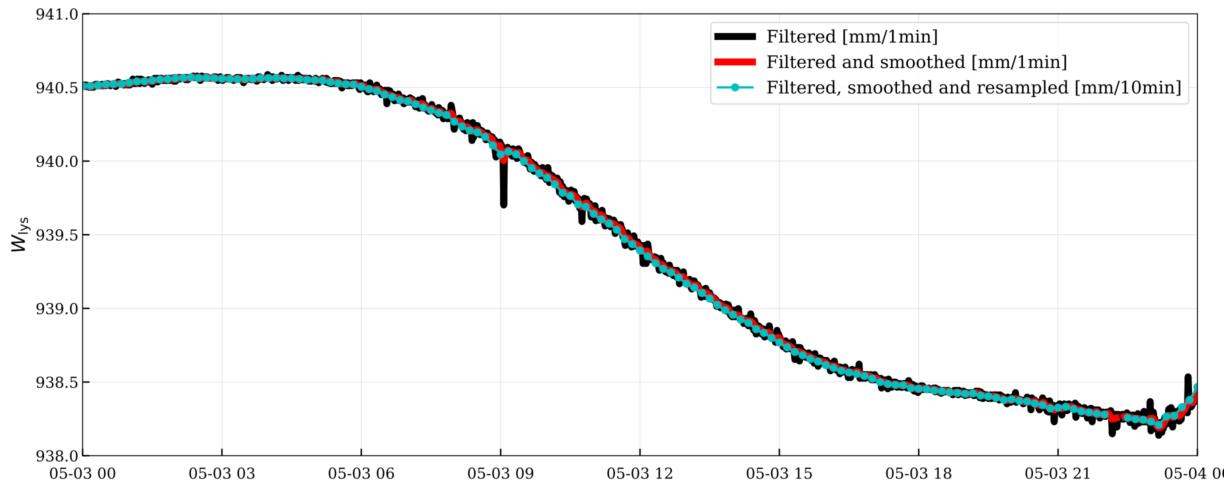

Figure 11: Lysimeter weight on May 3rd, 2021, showcasing the effects of smoothing using a Triangle Moving Average and resampling to ten-minute (10min) averages where a minimum number of seven one-minute samples (1min) must be available.

The optional precipitation filter is applied after resampling because it is based on the 10-minute averaged precipitation intensity observations of the nearby KNMI precipitation gauge. The precipitation filter threshold can be adjusted, but is currently set to 0.0 mm/10min when applied. Therefore, when the gauge reports non-zero precipitation, the lysimeter samples are omitted. The effectiveness of higher threshold values should be investigated. Furthermore, our observations showed that due to the fact that the precipitation gauge is knowingly less sensitive to small precipitation events and approximately 300 m south of the lysimeter, filtering data points before and after the gauge registers the event is more effective than removing only samples during the precipitation gauge-reported event. Therefore, a window of ten minutes before and after the gauge reported non-zero precipitation is flagged for removal when the precipitation filter is applied. Note, that this window can also be adjusted and further investigation for optimization is recommended as well. Both the precipitation flag array and gauge precipitation intensity are provided in the post-processed data file. However, the lysimeter data therein is not filtered based on the precipitation flag because this flag is optional. Nonetheless, an example of the precipitation filter application is shown in Figure 12 (without smoothing for distinction). The effectiveness of this filter is showcased in the following section and applying this filter is highly recommended for studying ET.

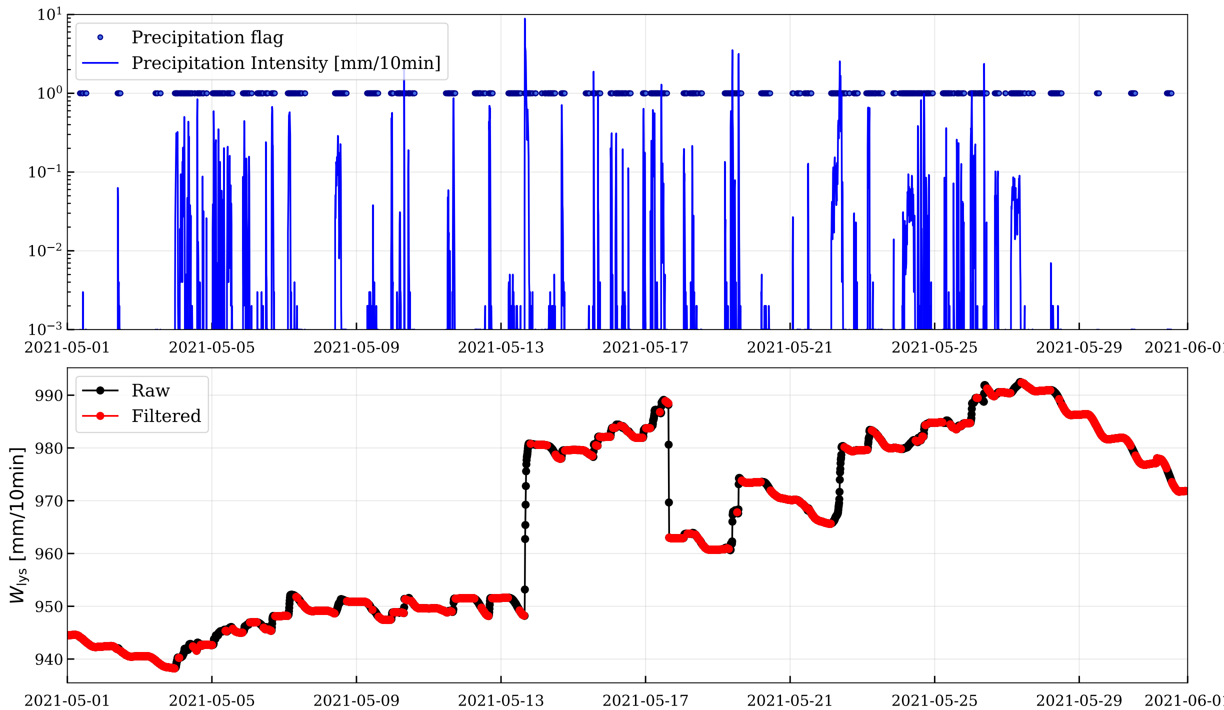

Figure 12: Example of precipitation flagged data points (1.0) in May, 2021, alongside the precipitation intensity (PI) reported by the nearby KNMI precipitation gauge. Below, the lysimeter weight is shown comparing 10-minute averaged raw unfiltered values and filtered values. Filtered values have been subjected to all flags up to and including the precipitation filter.

5 Calculating Latent Energy

While ET and precipitation are calculated as described in Section 2.3, determining the Latent Energy (LE) requires additional input. Specifically, if precipitation is known, deviating changes in lysimeter weight can be attributed to latent processes such as ET and condensation. For instance, if precipitation is not known, an increase in lysimeter weight due to condensation would be attributed to precipitation based on Equation 1. We can determine LE from the difference between the lysimeter weight change and the nearby precipitation intensity (PI) measurements:

$$LE_{\text{lys}}\ [\text{mm}/10\ \text{min}] = \Delta W_{\text{lys}}\ [\text{mm}/10\ \text{min}] - PI\ [\text{mm}/10\ \text{min}] \tag{3}$$

Figure 13 shows an example of the LE_{lys} calculated from the filtered and smoothed lysimeter data compared to the Makkink model which represents the potential ET based on temperature and incoming solar radiation observations. For distinction, the first days are characterized by little to no precipitation until May 4^{th}. Without precipitation, we observe good agreement between the 10-minute averaged LE_{lys} and the Makkink model which assumes unlimited water availability (and other simplifications). Accordingly, we observe that the potential ET often exceeds the lysimeter measurements as expected, but with relatively good agreement as both display diurnal behaviour, increasing throughout the day and near zero at night. However, the lysimeter is capable of capturing negative values which represent condensation. This agreement breaks down on May 4^{th}, which is characterized by frequent precipitation, significantly impacting the LE_{lys} measurements.

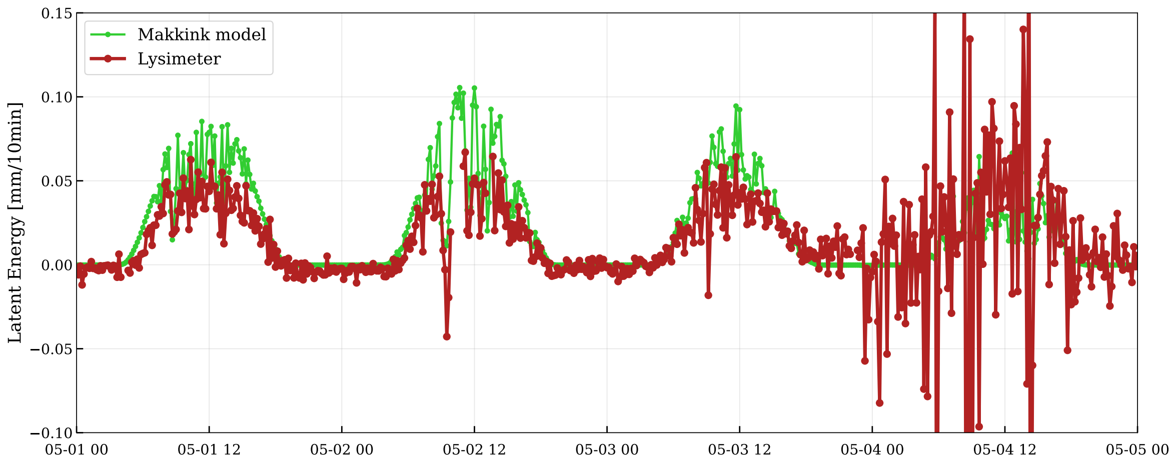

Figure 13: Example of lysimeter latent energy compared to the Makkink model for three clear days followed by a day with precipitation [MM-DD HH].

Inherent differences between the two instruments cause issues when calculating LE. Adding the known precipitation intensity to the difference in lysimeter weight should leave only latent processes; however, in this case, the precipitation gauge measurements do not precisely represent this. As mentioned, the instruments are roughly 300 m apart and the gauge is not as responsive as the lysimeter. Differences in detecting precipitation events compromise the LE calculation. Therefore, the optional precipitation filter (described in Section 4) can be beneficial by omitting precipitation events and the samples immediately before and after. As an example, Figure 14a compares the unfiltered and filtered lysimeter LE for an entire month, showing the influence when also including the precipitation filter. Visible noise and unrealistic outliers in the unfiltered data are reduced after automatic and manual filtering of the lysimeter weight. Although reduced, LE calculated using the filtered data still exhibits some unrealistic outliers, many of which are removed when the precipitation filter is also applied. When applied, it is not possible to study latent processes during precipitation events, but this is when ET is minimal. Interesting phenomena, such as the dew formation on the 30^{th}, is still observed, which would be automatically reported as precipitation by the lysimeter without determining LE.

To confirm the data removed by the precipitation filter is physically unrealistic, we compare to the processed data with and without the precipitation filter to the Makkink model in Figure 14b as a scatter plot. Overall, we observe that the precipitation filter excludes many high magnitude LE values when the Makkink model estimates little to no LE. While applying the precipitation filter appears to improve the dataset by reducing unrealistic values and showing better agreement with the model, data availability is significantly reduced. Therefore, future studies should consider the gap-filling and associated bias to further improve this dataset. Furthermore, a higher precipitation threshold and different window sizes for the filter should be investigated to retain more meaningful data. Operationally, a more sensitive, closer precipitation sensor could improve the quality of these calculations and would therefore benefit the SEB observations array at Cabauw.

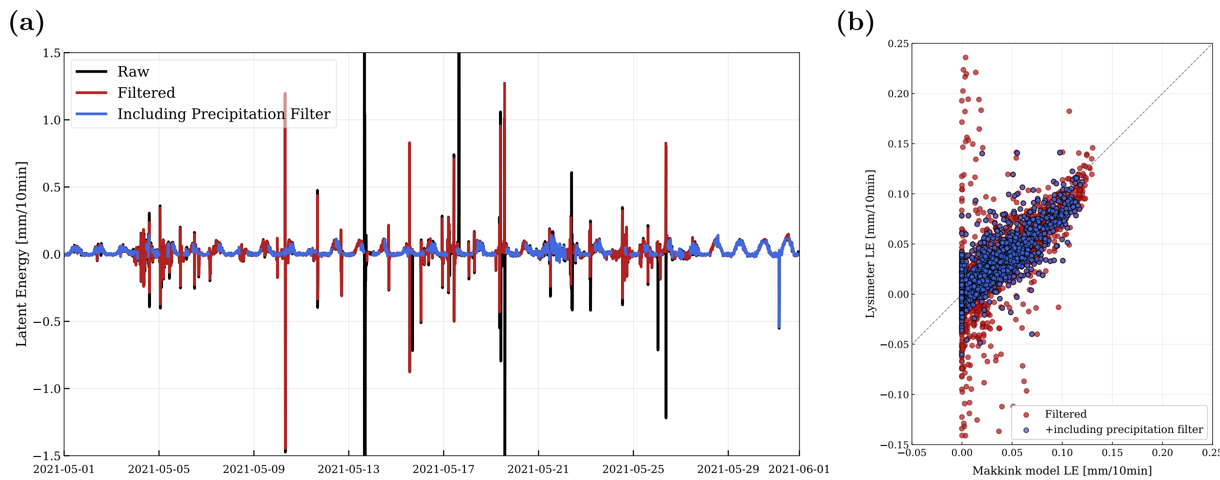

Figure 14: Lysimeter latent energy throughout May, 2021, at different stages of filtering. **(a)** Time series comparing the raw unfiltered and filtered (and smoothed) data alongside the same series when the precipitation filter is also applied. **(b)** The same data before and after the precipitation filter is applied shown as a scatter plot comparing to the Makkink model values.

Furthermore, while the KNMI precipitation gauge and Makkink model are useful references for data validation, other comparisons may be more constructive. For instance, more detailed and therefore possibly more representative models are available to estimate the potential ET [8]. A variety of observations are available at this site for intercomparison. Therefore, ultimately, the lysimeter LE should be compared not only to models but to the other established LE methods to evaluate its performance in future studies.

6 Summary

This report outlines the quality control procedure applied to data generated by the Eijkelkamp smart lysimeter installed at KNMI's Cabauw observational site for atmospheric research. Therefore, the focus of this report is the necessary post-processing of the lysimeter data, not evaluating this instrument's performance which is recommended for future study. In short, the raw 1-minute lysimeter data provided by Eijkelkamp as .csv files is processed by a python code developed to perform quality control, generating a new .csv file containing the lysimeter weight, flag arrays, and most importantly the validated 10-minute averaged precipitation, evapotranspiration (ET), and latent energy (LE).

Both mechanical and manual flags are applied to the data, and these flagged data is subsequently filtered out. Firstly, some erroneous data is identified based on the mechanical limitations of the instrument. Specifically, changes in lysimeter weight that exceed machine limits or samples during draining events are automatically removed. Naturally, any data with errors reported by the instrument are also removed. Secondly, manual flagging is performed based on visual inspection which addresses physically unrealistic data due to external interference or in-operability. A core remaining issue is that manual interference is relatively unknown and undocumented for most years, such as irregular grass maintenance and drought. Once erroneous data are removed, smoothing is performed using a triangle moving average and the data is resampled to 10-minute averages with a condition that at least seven one-minute samples are available. An adjustable precipitation flag is created based on input from nearby KNMI precipitation gauge observations. However, this flag is not applied directly but provided as optional to improve the data quality. The lysimeter is not recommended for investigating ET during precipitation events. Filtering out anomalous or unrealistic data is complex and some compromised samples are unavoidably undetected or in-distinguishable.

While the lysimeter reports hourly precipitation and ET values, the 10-minute differences in lysimeter weight are used to calculate these observations and LE with a better time resolution. The 10-minute averaged values showed good agreement with the KNMI precipitation gauge and Makkink model for potential ET. It is important to note then that the LE is calculated using two instruments with individual uncertainties, limitations and potentially undetected interference. As expected, the precipitation gauge is less responsive than the lysimeter and the setup could benefit from a closer, and more sensitive, precipitation sensor. Also, as anticipated, the lysimeter ET measurements tended to be lower than the modelled potential ET but exhibited similar behaviour. These data should also be compared to more realistic models. Ultimately, the LE is calculated by subtracting the known precipitation (reported by the KNMI precipitation gauge) from the total lysimeter weight change, isolating the contribution of latent processes. We observed that LE data quality is degraded during precipitation events which requires further investigation and improvements using the precipitation flag should also be explored further. Moreover, when studying ET, the precipitation filter is highly recommended. Nonetheless, these direct LE values can provide further insight into the surface energy balance and have potential as a reference for less direct LE observation methods, particularly regarding dew and transpiration.

We recommend continued development towards further improving and automating the lysimeter data quality control procedure. For example, we are exploring automatic outlier detection to inspect unknown interference that causes sudden drastic changes. However, this code requires further development to ensure interesting physical phenomenon are retained. While eliminating all interference is impossible, such as the influence of run-off, freezing, or irregular grass condition etc., the lysimeter should be compared to the neighboring indirect methods for measuring LE (refer to Section 2.2) to investigate and evaluate its performance. Furthermore, lysimeter data quality may also be improved by applying more suitable smoothing methods such as an adaptive window adaptive threshold smoothing algorithm [15] which should be explored. Additionally, the dataset can be further enhanced by filling the gaps created by filtering out erroneous data. As is, the validated, high-resolution lysimeter datasets provide a good foundation to investigate latent processes towards better understanding of the surface energy and hydrological balance.

Acknowledgements: The author would like to acknowledge the work of Shasha Li for her initial exploration of lysimeter data quality within the Cabauw In-situ team. Appreciation is also extended to Cisco (Evert I.F.) de Bruijn, Judith Jongen-Boekee and the entire Cabauw-Insitu team for their continued input and support which help shape this work. Additionally, thank you to Fred C. Bosveld who contributed valuable preliminary work acquiring and implementing the lysimeter and complimentary measurements, as well as the infrastructure for Figure 10.

References

[1] G. F. Makkink. Testing the Penman formula by means of lysimeters. *Journal of the Institution of Water Engineers*, 11:277–288, 1957.

[2] F. C. Bosveld, P. Baas, A. C. M. Beljaars, A. A. M. Holtslag, J. Vilà Guerau de Arellano, and B. J. H. van de Wiel. Fifty Years of Atmospheric Boundary-Layer Research at Cabauw Serving Weather, Air Quality and Climate. *Boundary-Layer Meteorology*, 177(2-3):583–612, 12 2020.

[3] Royal Eijkelkamp. Eijkelkamp Smart Lysimeters: User manual (original instructions). Technical Report M-1680/1681E, Giesbeek, the Netherlands, 2023.

[4] H. R. A. Wessels. Cabauw meteorological data tapes 1973–1984; Description of instrumentation and data processing for the continuous measurements. Technical Report WR 84-6, Royal Netherlands Meteorogical Institute, De Bilt, 1984.

[5] J. G. van der Vliet. Elf jaar Cabauw-metingen. Technical report, Royal Netherlands Meteorogical Institute, De Bilt, 1988.

[6] F. C. Bosveld. The Cabauw In-situ Observational Program 2000-Present: Instruments, Calibrations and Set-up. Technical report, Royal Netherlands Meteorological Institute (KNMI), De Bilt, 2020.

[7] J. L. Monteith. Evaporation and environment. In *The State and Movement of Water in Living Organisms*, volume 19 of *Symposia of the Society for Experimental Biology*, pages 205–234. Cambridge University Press, 1965.

[8] H. A. R. De Bruin and A. A. M. Holtslag. A simple parameterization of the surface fluxes of sensible and latent heat during daytime compared with the penman-monteith concept. *Journal of Applied Meteorology and Climatology*, 21(11):1610 – 1621, 1982.

[9] S. R. de Roode, F. C. Bosveld, and P. S. Kroon. Dew formation, eddy-correlation latent heat fluxes, and the surface energy imbalance at cabauw during stable conditions. *Boundary-Layer Meteorol*, 135, 2010.

[10] A. C. M. Beljaars and F. C. Bosveld. Cabauw Data for the Validation of Land Surface Parameterization Schemes. *Journal of Climate*, 10(6):1172–1193, 6 1997.

[11] G. C. Topp, J. L. Davis, and A. P. Annan. Electromagnetic determination of soil water content: Measurements in coaxial transmission lines. *Water Resources Research*, 16(3):574–582, 1980.

[12] J. Bijma. Onderzoek Problemen en Tekortkomingen Operationele Neerslagmeter. Technical Report Internal Report IR2008-04, Royal Netherlands Meteorogical Institute, De Bilt, 2008.

[13] H. A. R. de Bruin. From penman to makkink. In *Evaporation and Weather*, volume 39 of *Proceedings and Informations*, page 505. DLO Winand Staring Centre, Wageningen, Netherlands, 1997.

[14] L. Zhang, X. Li, Y. Wang, X. Li, and X. Zhang. A new approach to estimating potential evapotranspiration using the penman–monteith equation. *Journal of Hydrology*, 552:1–9, 2017.

[15] A. Peters, J. Groh, F. Schrader, W. Durner, H. Vereecken, and T. Pütz. Towards an unbiased filter routine to determine precipitation and evapotranspiration from high precision lysimeter measurements. *Journal of Hydrology*, 549:731–740, 2017.

Royal Netherlands Meteorological Institute

PO Box 201 | NL-3730 AE De Bilt
Netherlands | www.knmi.nl