Reply RC1

Dear Referee,

Thank you for your time to review our manuscript and for all your constructive suggestions considering our study. It helped to improve the quality of the manuscript. We reply to your comments below. Our response to the comments appears in bold and revised text as *italic*.

Main comment:

All methods and comparisons are presented together, which overwhelms the narrative. I strongly suggest reorganizing the content into a more hierarchical or modular format. For instance, the manuscript could be structured as follows: a) Evaluation of the noise correction methods for C_{nn}, b) Performance assessment using remote sensing vs eddy covariance constraints, and c) Sensitivity to turbulence scaling assumptions (free vs full MOST).

This is just a suggested outline, and the author is welcome to be creative. However, a clearer and more modular structure would greatly improve the readability and impact of the paper.

We understand the reasoning of the reviewer. We would like to point out that point a) is actually covered in our previous paper: van der Valk et al. (2025), which is referred to in multiple places of the manuscript. Regarding the structure of the results and discussion, we have decided to add subsection headers to improve readability, also based on the comments of the other reviewers. In the results section, we added the following subsections: *4.1.1 Energy-balance method versus two-wavelength method, 4.1.2 C_{nn} Correction methods, 4.1.3 Free Convection scaling, 4.1.4 Comparison with alternative L_vE methods*

In the discussion section, we added the following subsections: *5.1 Energy-balance method versus two-wavelength method, 5.2 MOST L_vE estimates versus free-convection* and *5.3 Potential of CMLs to estimate L_vE*

Specific comments

- In my opinion, Figure 1 does not represent the core contribution of the paper and mainly provides background or contextual information. Therefore, I suggest moving it to the supplementary materials.
  As we think it is important for the reader to understand what are the implications of applying the free-convection scaling instead of the complete MOST scaling, we decided to explain this in the main text. Deriving this from the equations would be difficult.

- The manuscript states that the observation period spans from April 1 to October 1. However, the full time series (TS) over this period is not shown anywhere. While it is reasonable to highlight selected days to illustrate the diurnal cycle, an overview of the entire time series is important to assess the consistency and overall performance of the method.
  We agree with the reviewer that these two example time series are not illustrative of the performance over the entire data period. Therefore, we changed these time series to median diurnal cycles per month for all methods, including the uncertainty for the CML methods (25th and 75th percentiles). As a result, the text is changed as follows:

A comparison of the reference methods for the full data period shows an overall comparable behaviour (Fig. 3), indicating negligible influence of differences in footprints between the two references. On average, however, the MWS-2λ method shows an overestimation in comparison to EC. For the monthly median diurnal cycles, all CML methods with the spectral noise correction overestimate L_vE compared to the references (Fig. 4). This overestimation is roughly constant throughout our data period, with the exception of June, during which the overestimation in comparison to the reference methods is larger. The diurnal cycle is well captured by all CML methods, even when considering the interquartile range. Noteworthy, is the similarity between the CML-2λ and CML-EBM-OBS methods in the diurnal cycles, both in median values and uncertainty. The clearest differences between these methods occur before sunset, when the CML-2λ method overestimates L_vE, while the CML-EBM-OBS estimates are constrained by the measured net radiation.

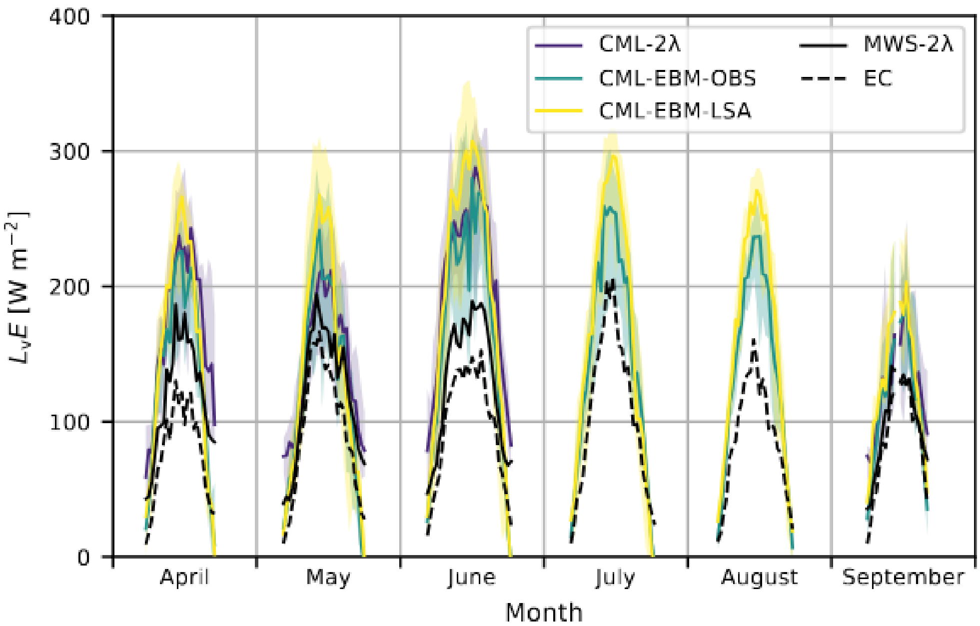

Figure 4. Median diurnal cycles and interquartile ranges (shading) of 30-min L_vE estimates of the methods for the Nokia CML in comparison with the reference instruments. CML-2λ uses the two-wavelength scintillation method with the CML and LAS, CML-EBM-OBS uses the measured energy balance method as constraint to infer the turbulent heat fluxes and CML-EBM-LSA uses the estimated net radiation by LSA SAF instead of the measured net radiation. Shown estimates are obtained using the spectral noise method for the CML estimates and complete Monin-Obukhov scaling. We removed timestamps with less than 10 available days in order to obtain representative diurnal cycles. We refer the reader to Appendix A for a complete overview of the used abbreviations.

- For the Monin–Obukhov Similarity Theory (MOST) scaling, two additional parameters— z0m (roughness length for momentum) and d (displacement height)—are required. These parameters depend on the characteristics of the roughness elements, vegetation height, and atmospheric conditions. However, the manuscript does not provide any information on how these values were determined or set.
  We indeed did not mention these in the methodology section. Our apologies for this omission. For the roughness length at the location of our experimental setup (Cabauw, the

Netherlands), we use the values reported in the MSc thesis from Robert Moonen (2021), who determined the roughness lengths based on EC data, depending on wind direction:
40° – 100°: 0.03 m
100° – 160°: 0.01 m
160° – 280°: 0.005 m
280° – 40°: 0.01 m
These values are in the same order as found by Verkaik and Holtslag (2007).

We relate the displacement height to the roughness length based on Brutsaert (1982):
$d_0 = 16/3\ z_0$

We added this into Section 2 as follows:
Thus, next to the scintillometer measurements, the two-wavelength method requires temperature, humidity and horizontal wind speed measurements. Additionally, it requires more complex measurements of the correlation between temperature and humidity fluctuations, the roughness length for momentum and the displacement height. For the roughness length z_0, we make use of values reported by Moonen (2021), who determined the roughness length for various wind sectors using EC data (Table 1) and found similar values to Verkaik and Holtslag (2007). For the displacement height d, we use the relationship d = 16/3 z_0 from Brutsaert (1982).
In order to obtain the turbulent heat fluxes, Eqs. (4 - 7) have to be solved iteratively. In figures, we refer to the two-wavelength method using the suffix 2λ.

Table 1. Lower and upper bound of wind direction classes and roughness length z_0 values as determined by Moonen (2021).

Lower bound [°]	Upper bound [°]	z_0 [m]
40	100	0.03
100	160	0.01
160	280	0.005
280	40	0.01

- The current version of Figure 6 is not very clear or easily readable, especially regarding the size of the symbols in relation to the scale of the graph.
  We agree that identifying individual symbols for some methods is hard, given the selected limits of the y-axis. However, we selected these scales on purpose, to emphasize the largest differences between methods, and not focus on minor differences in statistical metrics between methods.

- Line 241-242: The statement "When considering the H estimates as well, the overall performance of both EBM versions reduces" is unclear. Does this mean that when metrics are computed jointly for both H (sensible heat flux) and LvE (latent heat flux), the overall performance of the Energy Balance Methods (EBM) decreases? If so, how was the joint evaluation done?
  We refer here to the fact that for the sensible heat flux, the two-wavelength method performs better than the EBM. We agree that this was not clearly formulated. We replaced this sentence with:

The H estimates of both EBM versions perform less well than the estimates of the two-wavelength method. This is in line with our expectations, because the LAS signal dominates these estimates.

- Line 245: "For the two-wavelength method, this overestimation is fully attributed to the LvE, since H is constrained by the LAS, while for the EBM, this overestimation can be distributed among LvE and H." while for the EBM, this overestimation can be distributed among LvE and H. The reasoning that overestimation is "fully attributed to the LvE" in the two-wavelength method, due to LAS constraining H, is understood. However, this relies on the implicit assumption that H is accurately captured by the LAS. Could the authors clarify whether this assumption holds true in this context?
  We are not sure what the reviewer means here. The overestimation we are referring to here is in comparison to the two-wavelength method with the MWS and LAS (and EC). So, the LAS estimates for the reference and the CML are exactly the same. This implies that all differences between the reference and the CML can be attributed to the L_vE estimates. For clarification, we added as follows:
  For the two-wavelength method, this overestimation is fully attributed to the L_vE, since H is constrained by the same LAS estimates for the CML-2λ and the MWS-2λ methods, while for the EBM methods this overestimation can be distributed among L_vE and H.

  Moreover, as can be seen in Fig. C1, the differences between the *H* estimates of the EC and the two-wavelength method with the MWS and LAS show roughly the same statistical metrics as the L_vE comparison.

- The discussion and conclusion sections feel somewhat verbose. Please streamline the content to avoid redundancy.
  Based on your comments and the comments of other reviewers, we have adapted the discussion and conclusion sections, and we think the content in these sections is now less redundant.

- The manuscript lacks a discussion on how differences in footprints, sampling frequencies, and spatial/temporal representativeness among the various systems, such as eddy covariance (EC), microwave scintillometer (MWS-LAS), and commercial microwave links (CMLs), might affect the flux estimates. While a detailed intercomparison may be beyond the scope of this study, a short paragraph acknowledging these differences and citing relevant literature would greatly help readers interpret discrepancies in the time series and understand the strengths and limitations of each measurement system.
  We understand the comment of the reviewer. However, originally the site of Cabauw was selected given its relatively homogeneous terrain, especially in the dominant south-westerly wind direction. Also, a comparison by our reference instruments illustrates the similar behaviour between the EC and the MWS-LAS (Fig. 3). Moreover, the footprints of the CML and MWS are nearly identical, given their collocated installation.
  To emphasize this, we added in Section 3.2 the following:
  As reference instruments, we use an optical-microwave scintillometer setup and an eddy-covariance system (EC). The scintillometer setup consists of a Radiometer Physics RPG-MWSC-160 microwave scintillometer, transmitting at 160.8 GHz (i.e., a wavelength of 1.86 mm), and a Kipp & Zonen LAS Mk-II, transmitting at 352.7 THz (i.e., a wavelength of 850 nm), which are both sampled at 1 kHz. Hereafter we refer to this system as MWS-2λ. The path of

this scintillometer setup is nearly identical to the path of the CML (e.g., Fig. 2), so that differences in footprints can be neglected. It should be noted that this setup has a major data gap between 10 July and 6 September 2024. The EC system consists of a sonic anemometer (Gill-R50) and an open-path H_2O/CO_2 sensor (LICOR-7500), sampled at 10 Hz, and installed at 3 meter above the ground (Bosveld et al., 2020). The fluxes obtained with the EC can be directly obtained from the KNMI Data Platform, next to other more common meteorological measurements. The EC data is available during the entire study period. During the dominant south-westerly wind direction, the footprints of the EC, scintillometer setup and CML all predominantly consist of grass fields.

And in Section 4.1:

A comparison of the reference methods for the full data period shows an overall comparable behaviour (Fig. 3), indicating negligible influence of differences in footprints between the two references. Even though on average, the MWS-2λ shows an overestimation in comparison to the EC. In general, the uncertainty of the CML methods observed in two example timeseries does not deviate largely from the intercomparison between reference methods (Fig. 4).

- Additional minor thought: While comparing to the EC tower's available energy (R_n – G), it would be very helpful if the authors could plot the sum of H + LvE derived from the CML data vs EC tower H+LvE. Although the footprints and instruments are not perfectly collocated, such a comparison would provide valuable insight into the total energy captured by the CML-based approach. This could also be benchmarked against typical energy closure rates observed at flux towers (approximately 80% on clear days), offering an overview of the method's overall performance.

  We understand the reasoning of the reviewer, and agree that it is indeed relevant to know for the reader how well the energy balance closes, to get an idea of the performance of the EBMs. However, we do think that the proposed analysis is out of scope for this paper, also because there are multiple studies performed to this (e.g., Kroon, 2004, which is in Dutch, but the main findings are reported also in this triennial report by KNMI in English). We therefore added as follows in Section 2.2 (also based on other reviewer's comments):

  For the in-situ method, we use the measured ground heat flux, while for the remotely sensed data product we assume 10 % of the net radiation is used for the ground heat flux. Note that the measured energy balance hardly ever closes, especially in more complex measurement environments, e.g., forests or cities (Mauder et al., 2020). For the field site used in this study, Cabauw in the Netherlands, typically an imbalance during day-time is found between 10% (afternoons) to 40% (mornings) (Kroon, 2004). For our data period we find similar values using EC data (not shown).

Reply RC2

Dear Referee,

Thank you for your time to review our manuscript and for all your constructive suggestions considering our study. It helped to improve the quality of the manuscript. We reply to your comments below. Our response to the comments appears in bold and revised text as *italic*.

Main comments:

- General figure comment: The way the figures are organized makes them take up a lot of space on screen/paper resulting in a lot of unnecessary white space. The white space makes the figures break up the text significantly more than normal. I'm not sure if the journal editors will organize the figures better but I recommend some reorienting of the plot frames to minimize white space.
  We understand the reviewer. Indeed during the publication phase, also copy-editing will take place, so that figures will nicely fit onto pages in combination with the text. Also single-column figures 1, 3, 6 and 8 will be fit into a single column.

- The way the results are presented and discussed – i.e., a multitude of acronyms and organization on plots – makes interpreting the results and following the discussion a bit cumbersome. At the very least, sub-section headings would be helpful that remind the reader what the plethora of non-standard acronyms mean. I leave it up to the authors to decide how best to achieve this clarity, but due to the myriad of techniques being compared, some more nuanced form of organization is necessary.
  We agree with the reviewer that the organization might be difficult to understand without any headers. Therefore, also based on the other reviewers' comments, we have added headers in the results and discussion section. In the results section, we added the following subsections: *4.1.1 Energy-balance method versus two-wavelength method, 4.1.2 C_{nn} Correction methods, 4.1.3 Free Convection scaling, 4.1.4 Comparison with alternative L_vE methods*
  In the discussion section, we added the following subsections: *5.1 Energy-balance method versus two-wavelength method, 5.2 MOST L_vE estimates versus free-convection* and *5.3 Potential of CMLs to estimate L_vE*

- The authors seem to focus their scintillometry references to EU-based sources (Specifically in the paragraph starting on Line 337). There are many studies performed in the U.S. and China post Ward et al. 2015 that help to paint a more thorough picture of how well scintillometry works over generally heterogenous terrain. I won't provide specific examples, but I strongly encourage the authors to include references outside of their realm of influence – we are no longer in the preinternet era where it was nearly impossible to know if someone on the opposite side of the world is performing similar research. Science should not be limited by political borders, and more global citations will lead to better dissemination of knowledge and more efficient scientific progress.
  We agree with the reviewer that in some cases our sources are EU-based. Therefore, we add some recent studies scintillometry has been applied in heterogeneous or complex terrain. Moreover, many other studies on heterogeneity or over cities have mostly focussed on obtaining a sensible heat flux by using a LAS, whereas here we originally

aimed to illustrate the potential of microwave scintillometry for latent heat fluxes. We do realise that it would be valid to also mention these studies here. We added as follows:
Generally, microwave scintillometry has proven itself as reliable method to estimate LvE over different landscapes, such as heterogeneous farmlands (e.g., Meijninger et al., 2002, 2006; Beyrich et al., 2005; Xu et al., 2023), cities (Ward et al., 2015a). Also, in areas with a more complex topography, scintillometry has shown its value for estimating the turbulent heat fluxes, especially on larger-scales, such as over vineyards (Perelet et al., 2022) or hilly forests (Isabelle et al., 2019). For H, several studies, some only using an optical scintillometer, have shown the potential over heterogeneous farmland (e.g., Beyrich et al., 2002; Ezzahar et al., 2007), arid regions with sparse vegetation (e.g., Asanuma and Iemoto, 2007; Kleissl et al., 2009), and cities (e.g., Lagouarde et al., 2006; Lee et al., 2015). Yee et al. (2015) compares several scintillometers…

Specific comments:

- Line 57: "…i.e. the spectral noise correction method" – a reference here would be useful or at the very least a description on what the correction entails. Right now the study has very minimal data QC notes
  It is unclear what reviewer is referring to, because the preceding text explains the nature of the corrections. Moreover, this paragraph starts with stating the reference. Therefore, we leave it as is.

- Line 189: "the Nokia CML vibrates during higher wind speeds" – Do you have any information on the beamwidth of the Nokia link? You mention vibration and link to the other paper in review but their handling of the vibrations is the same single sentence. I have no problem with the cutoff above 8 m s-1 . I am more interested in whether these commercial icrowave links are less sensitive to vibrations than the RPG system.
  The half-power beamwidth of the Nokia is 1.6 degrees. However, this does not mean that the signal intensity is constant within this main lobe.
  The Nokia CML is more sensitive vibrations than the MWS, mostly due to relatively basic mounting system of the Nokia. Therefore, we added as follows:
  … higher wind speeds (van der Valk et al., 2025). This is caused by the relatively weak mounting system of the Nokia, as no vibrations are found in MWS even though both are mounted in the same mast…

- Line 190: "Additionally, we remove rainy intervals or those following a rain event within an hour…" – How do you determine the 1-hour cutoff? Radomes do get wet and require a drying off time. In my experience, this is strongly dependent on precipitation amount and can take upwards of 6 hours to dry following strong precipitation. Perhaps a better metric would be when the average signal has returned to XX% of the pre-rainfall signal strength.
  We fully agree with the reviewer that a more sophisticated method would be to base this on the average signal strength in comparison to the pre-rainfall signal strength. Typically, it takes around 5-10 minutes for the Nokia antenna to dry and return to pre-rainfall signal strength, as is shown in Fig. 13 in van Leth et al. (2018), where they use the same Nokia link. Setting the cutoff to 30 minutes might have sufficed as well; however, we selected here the 1-hour cutoff to be certain that the antenna covers would be definitely dry.

- Section 3.3 – Acronyms should not be used in section headings unless they are known ubiquitously, like TKE and MOST.
  We changed into:
  Remotely-sensed net radiation estimates

- Figure 6 – Description of the frames should be in the caption. It is not unheard of for readers to skim through figures first and with the large number of accronyms used in this manuscript the definitions of the acronyms used in this, and other, figures gets lost without a reminder for the reader in the caption. This comment applies to all figures.
  We understand the comment of the reviewer. In combination with this comment and comments of other reviewers, we have decided to include an abbreviation list in the appendix. To refer the reader to this, we add in the theory section, experiment section and in the captions:
  We refer the reader to Appendix A for a complete overview of the used abbreviations.

  Additionally, we extended the caption in each figure, so that the abbreviations are easier to interpret. The caption for Fig. 6 becomes for example:
  ***Figure 6**. Statistical metrics per method and scaling to obtain L_vE estimates using the Nokia CML for both correction methods (shape) versus both reference instruments (color). The solid line indicates the statistical metrics of the reference instruments versus each other (Fig. 3). CML-2λ method uses the two-wavelength scintillation method with the CML and LAS, CML-EBM-OBS uses the measured energy balance method as constraint to infer the turbulent heat fluxes and CML-EBM-LSA uses the estimated net radiation by LSA SAF instead of the measured net radiation. The "FC"-suffix refers to the free-convection scaling. The dotted line shows the statistical metrics of a comparison between the L_vE estimates directly obtained from LSA SAF versus the MWS-2λ method. The dashed line represents L_vE estimates based on the measured available energy (R_{net} −G) and the Bowen ratio β obtained from the EC-system, i.e., (R_{net} −G)/(1+β) versus the MWS-2λ method. The used Bowen ratio is a median value for the full data period (excluding nighttime intervals), as a means to obtain an objectively selected, representative Bowen ratio value to estimate L_vE from only net radiation measurements. We refer the reader to Appendix A for a complete overview of the used abbreviations.*

Reply RC3

Dear Referee,

Thank you for your time to review our manuscript and for all your constructive suggestions considering our study. It helped to improve the quality of the manuscript. We reply to your comments below. Our response to the comments appears in bold and revised text as *italic*.

- Line 26: "Bastiaanssen et.al. 1998, Mu et.al., 2007" more recent references, as well as overview papers, could have been used here to illustrate the point the authors make.
  The references used here are two algorithms to determine evaporation, which is currently not clear. We remove these references and replace with a more appropriate review:
  In comparison, satellite remote sensing estimates of evaporation have a better spatial coverage, but have a low temporal and limited spatial resolution (Zhang et al., 2016).

- Line 52: "so that the Cnn values are overestimated"; better to replace this with: "resulting in an overestimation of Cnn values".
  We changed accordingly.

- Lines 54-57: Please split up this sentence to make it more clear.
  We changed as follows:
  In order to correct for that, we proposed two methods. The first method applies a high-pass filter and subtracts a low quantile of the resulting variances of the Nokia CML, called the constant noise correction method. The second method corrects for the noise in the Nokia CML by comparing with an MWS and selecting parts of the power spectra where the Nokia Flexihopper behaves in correspondence with scintillation theory, called the spectral noise correction method. The latter method also considers different crosswind conditions and corrects for the omitted scintillations using scintillation theory.

- Line 91: "to be 0.8"; maybe add that this is for daytime conditions.
  We added as follows:
  Similar to Ward et al. (2015b), we assume r_{Tq} to be 0.8 for daytime conditions, who find this to be a reasonable value, consistent with values obtained from fast-response sensors (e.g., Kohsiek, 1982; Meijninger et al., 2002).

- Line 127: "To do so, closure of the measured energy balance is assumed" It might be good to mention at this point that closure of the energy balance is almost never accomplished, especially not in "complex measurement environments (e.g. forests or cities)".
  We agree with the reviewer and rephrased as follows (also based on other reviewer's comments):
  …is used for the ground heat flux. Note that the measured energy balance hardly ever closes, especially in more complex measurement environments, e.g., forests or cities (Mauder et al., 2020). For the field site used in this study, Cabauw in the Netherlands, typically an imbalance during day-time is found between 10% (afternoons) to 40% (mornings) (Kroon, 2004). For our data period we find similar values using EC data (not shown).

  As alternative constraint, we considered prescribing a Bowen ratio instead of net radiation, however that did not yield promising results…

- Line 162: "from 1 April 2024 to 1 October 2024"; please mention that only daytime data is used.
  We followed the suggestion of the reviewer:
  We use daytime data from from 1 April 2024 to 1 October 2024, which corresponds to a growing season in the Netherlands.

- Lines 164-165: "The footprints … included." It might be considered to show the typical footprint in Figure 2 to illustrate this statement.
  The dominant wind direction at Cabauw is south-westerly, for which the footprint is almost homogeneous grass fields with small ditches in between. To emphasize that in most cases the footprints are homogeneous, we rephrased as follows:
  The dominant wind direction at Cabauw is south-westerly, so that the footprints of the scintillometers and EC mostly consist of grass fields. Only for northerly wind directions built-up area may partly be included.

- Line 169: Replace "we show" by "it is shown", since the authors are not (exactly) the same.
  We followed the suggestion of the reviewer.

- Line 171: add "a" between "is" and "more".
  We followed the suggestion of the reviewer.

- Lines 184-186: "In order … time intervals". Please rephrase this to make clear what exactly has been done here.
  We rephrased as follows:
  In order to be able to compare daily E estimates, we only aggregate the 30-min time intervals per day that are available for both the Nokia CML and the reference instrument.

- Line 206: "comparable"; unclear to what; please add clarification.
  This is indeed unclear. We refer here to the statistical metrics from our comparison of LSA SAF estimates with measured net radiation for our data period. We rephrased as follows:
  For Cabauw during our data period, the net radiation obtained with LSA SAF compared to the measured net radiation on average shows similar error estimates, though with an overestimation (Fig. A1).

- Line 222: "the expected differences"; please explain what these expected differences are.
  We mean here the higher L_vE values for the sunny day than for the cloudy day. We rephrased as follows:
  Moreover, the L_vE estimates on the sunny day are higher than on the cloudy day for all methods, as would be expected.

- Lines 231: "In this…versus the MWS-2λ." Please use shorter sentence to make this more clear.
  We changed as follows:
  …In this plot, we also show the intercomparison between the reference methods and two comparisons with alternative methods to derive L_vE. These alternatives are the L_vE estimates obtained directly from LSA SAF versus those from the MWS-2λ method and L_vE estimates

based on R_{net} − G and a Bowen ratio versus those from the MWS-2λ method. We show these methods to illustrate how a readily available (former) and a basic experimental method from only net radiation estimates (latter) perform in comparison to the reference instruments. The used Bowen ratio in the latter method is obtained from the EC and is the median ratio for the full data period (excluding nighttime intervals). We use a median value as a means to obtain an objectively selected, representative Bowen ration value to estimate L_vE from only net radiation measurements.

- Lines 233-234: “This Bowen … intervals).” Please explain/justify why this was done (instead of using the actual Bowen ratio).
  See our response to your previous comment.

- Line 237-238, figure 5, caption: remove “together”
  We changed accordingly.

- Line 237-238, figure 6, caption “The used Bowen ….intervals).” Please explain why this constant ratio is used instead of the (more appropriate?) instantaneous/daily value.
  We added as follows (also in the other figures):
  The used Bowen ratio is a median value for the full data period (excluding nighttime intervals), as a means to obtain an objectively selected, representative Bowen ratio value to estimate L_vE from only net radiation measurements. We refer the reader to Appendix A for a complete overview of the used abbreviations.
  Also see our changed text in our to your comment on L231.

- Line 238 and further; please use the same abbreviation throughout the paper (for example not EBM without addition versus later on EBM-LSA and/or EBM-$ß_{EC}$). Maybe a good idea would be to add an abbreviation list at the start/end of the paper for more easy reference.
  The EBM used here is the energy balance method making use of the measured net radiation. We realise that this can be unclear. Therefore, we change EBM with observations to EBM-OBS. Moreover, in Sect. 2, we also added for the EBMs a clarification per method on their abbreviation at the end of that specific subsection. Lastly, we also added an abbreviation list in the appendix. We think that this will help the reader more easily understand the structure of the abbreviations.
  So the end of Sect. 2.2 becomes:
  …of the prescribed Bowen ratio. In the remainder of this article, we refer to the EBM using in-situ radiation data with EBM-OBS, and for the method using the LSA data products, we use EBM-LSA.
  To refer the reader to the abbreviation list, we added in the theory section, experiment section and in the captions:
  We refer the reader to Appendix A for a complete overview of the used abbreviations.

- Line 241: add “it” between “but” and “has”.
  We changed accordingly.

- Lines 243-245. This sentence is insufficient explanation at this point. Only mentioning the Van der Valk 2025 paper is not enough; one or two additional explanatory lines would be required here.

We refer here to the differences in the performance of the turbulent heat fluxes and refer to van der Valk et al. (2025), to guide the reader to the overestimation of C_{nn} (and not specifically to the correction methods). In the sentence following these lines, we explain what causes this behaviour. To clarify, we rephrased as follows:
…two turbulent heat fluxes at Cabauw. These differences in performance between the turbulent heat fluxes are mostly a consequence of the nature of these methods in combination with the overestimation of C_{nn} by the CML (see van der Valk et al., 2025).

- Line 246-248. "If desired… and H." Doing so would be rather arbitrary and considered finetuning, which is not a good approach to solve this particular phenomenon.
  We agree with the reviewer. We removed this part, as it does not fit in this context.

- Line 256: "two alternative methods"; unclear what is meant with these two alternative methods. Please explain.
  We explain these two methods in the following sentences. To emphasize this, we change here as follows:
  A comparison with two alternative methods to retrieve L_vE estimates shows that estimating L_vE using CMLs can be beneficial, especially regarding the spread. One of these alternative methods is based on the measured energy balance and prescribing a median Bowen ratio based on the EC data, i.e., the best possible estimation of the Bowen ratio, results in higher IQR in comparison to the MWS-2λ than any of the methods using the CML. The other alternative is the L_vE estimates directly obtained from LSA SAF data product. A comparison between these LSA SAF L_vE estimates versus the MWS-2λ is also outperformed on the IQR by the majority of the methods using the CML. In comparison to this method, it should be noted that the EBM using LSA SAF radiation data only shows a minor improvement.

- Line 259: "between between"; remove one of these.
  Agreed.

- Line 259: "estimates directly obtained from LSA SAF"; unclear what is meant here. Please explain.
  We mean the L_vE estimates that can be obtained from LSA SAF data product. See our reply to your comment on L256 on our addition.

- Lines 218-261: In this section too many methods are intercompared to each other, where the assessment is based on three different parameters (MBE, IQR, r). Apart from the fact that it is difficult to read a piece of text with numerous abbreviations also the comparison is described by means of jumping between approaches and also between assessment parameters. This needs to be described in a much more systematic manner. It is advised to use consistent acronyms, split the section into subsections where performance versus the reference method's (only) are described w.r.t. MBE, IQR and r. Preferably only one reference method (which is the EC observation) should be used.
  We understand the comment of the reviewer, also in line with the comments of the other reviewers. In the results section, we added the following subsections to guide the reader:
  4.1.1 Energy-balance method versus two-wavelength method, 4.1.2 C_{nn} Correction methods, 4.1.3 Free Convection scaling, 4.1.4 Comparison with alternative L_vE methods

Moreover, we changed the EBM with observations to EBM-OBS, which we think is an intuitive abbreviation.

Lastly, we changed the structure and sentences of (now) section 4.1.1, so that the comparison between the two-wavelength method and EBM-OBS is emphasized, and the EBM-LSA is only mentioned at the end of this section:
Overall, the EBM-OBS outperforms the other two methods for the L_vE estimates. It has a lower MBE and IQR than the two-wavelength method and the EBM-LSA. All methods have a comparable r in comparison to the MWS-2λ. We would have expected the two-wavelength method to perform best, as this is closest to the traditional two-wavelength setup, but it has a higher MBE than both the EBM versions. The H estimates of both EBM versions perform less well than the estimates of the two-wavelength method, as was also expected because the LAS signal dominates these estimates. (Fig. B1). The H estimates of EBM-OBS have an MBE and IQR similar to the L_vE estimates, even though L_vE is the highest of the two turbulent heat fluxes at Cabauw. These differences in performance between the turbulent heat fluxes are mostly a consequence of the nature of these methods in combination with the overestimation of C_{nn} by the CML (see van der Valk et al., 2025). For the two-wavelength method, this overestimation is fully attributed to the L_vE, since H is constrained by the LAS, while for the EBM methods this overestimation can be distributed among L_vE and H. Note that the EBM-LSA has a higher MBE and IQR than the EBM-OBS, most likely due to the overestimation and uncertainty of R_{net} by LSA SAF (Fig. A1).

Lastly, we disagree with the reviewer that only the EC observations should be used as reference. Scintillometers have proven their value over the past years during dedicated field campaigns. Moreover, the MWS-LAS setup measures along the same path as the CML, so that this reference is most representative of the turbulent flux estimates.

- Lines 277-278: "the performance…is roughly comparable": This is an incorrect statement. The r are lower in all four cases and the MBE and IQR are up to a factor 2.6 higher!
  We agree that this is not correctly phrased. We meant here that when using the two-wavelength method, the performance of the *L_vE* estimates for the spectral noise method and free-convection scaling is roughly comparable to the references. We should have also indicated that this is not the case for the complete scaling. We rephrase as follows:
  When using the two-wavelength method with the spectral noise method and free-convection scaling, the performance of the L_vE estimates is roughly comparable to the intercomparison between our reference instruments, the MWS-2λ and EC systems. Using the complete scaling instead, the performance of the L_vE estimates unexpectedly decreases, due to the overestimation of C_{nn} after correction (van der Valk et al., 2025).

- Line 282: Add "an" before "CML"
  We changed accordingly.

- Lines 282-289: This section is mainly a repetition of what is described already in section 4.
  We agree with the reviewer that this is a repetition of what is described in Section 4; however, we want to emphasize this as we think this is one of the more important results. It illustrates that there is potential in using an EBM, although this depends largely on the quality of the net radiation estimates. Therefore, we leave this as is.

- Lines 293-294: “This underlines …the MBE” This statement is correctly mentioning that the method is suffering from an in-accurate product (Rn) that provides the upper limit of the turbulent fluxes. Because of this, a discussion on the accuracy of the relative contribution of H and LE to Rn would be more beneficial here. This would also better fit the topic of the paper, namely determination of turbulent LE flux.
  Based on your comment here, line 127, lines 349-356, and other reviewer’s comments, we have added a paragraph discussing the energy-balance closure in this Sect. 5.1 and elaborated on the energy-balance closure in Sect. 2.2 (see our reply to your comment on line 127). In the discussion, we moved the paragraph from L349-356 to this location and adapted the paragraph more towards energy-balance closure:
  However, application of the EBM is not trivial in general, let alone over complex terrain, such as cities or forests. Typically, the observed energy balance does not close (Mauder et al., 2020), which is also the case for our field site (Kroon, 2004). For cities the addition of the anthropogenic heat flux as heat source for the turbulent heat fluxes (e.g., Oke et al., 2017) and the significant amount of heat storage (e.g., Sun et al., 2017) can complicate the application of the EBM (e.g., Harman and Belcher, 2006; Miao et al., 2012). These fluxes are added to the energy balance, so that the assumed energy balance for the EBM versions is not valid for every CML in cities. Moreover, the validity of the EBM also depends on the location of the CML, for example the mounting height, which affects the footprint of the CML to be on local scales, i.e., streets for cities, or more regional scales, i.e., city (or neighbourhood) scales.

- Lines 304-306: “Moreover…estimates”. These lines are very confusing, please rephrase.
  We agree that these lines are confusing. Based on your suggestions at line 238 and 259, we consistently use the abbreviations EBM-OBS and EBM-LSA for describing the two used EBMs. Due to these changes, the sentence has become:
  Moreover, in comparison to L_vE estimates obtained from LSA SAF data product, the EBM-OBS shows an improvement, while EBM-LSA does not show any improvement in comparison to these LSA SAF L_vE estimates.

- Line 310: “These two”, unclear what is meant with “these two”, please add an explanation (e.g. variables/parameters?).
  We refer here to the eliminated variables described in the preceding sentence, i.e., horizontal wind speed and roughness length. We rephrased as follows:
  These eliminated variables have a relatively large influence….

- Lines 316-319. Please describe/demonstrate why this (i.e. increase in LE and reduction of H due to free convection assumption) occurs instead of saying that this occurs.
  We agree that it is also important to describe why this increase in L_vE occurs, while H decreases. We added as follows:
  However, it must be noted that for the EBM versions, the use of free-convection causes an increase in the LvE estimates and a reduction in H in comparison to the complete scaling. H reduces due to the strong relation with the free-convection wind scaling variable u_{fc} (Eq. 12), which decreases compared to the friction velocity u_*, so that L_vE has to increase as a consequence of the prescribed available energy that needs to be distributed among the two turbulent heat fluxes.

- Line 323: “The former”; unclear what is meant with “The former”, please explain.
  We mean the bichromatic method. We changed as follows:

The bichromatic method is not applicable to the Nokia CML…

- Line 346: please remove "so that Cnn estimates….et.al. 2015b)."
We followed the suggestion of the reviewer.

- Line 350: "shows" should be replaced by "show"
We changed accordingly.

- Lines 349-356: Indeed as mentioned also previously w.r.t. line 127; the energy balance or closure does not hold everywhere, in fact almost nowhere, and especially not over complex terrain (cities and forests). Agani, this is more a discussion on the energy balance approach than on the proper functioning of CMLs for LE estimates. It would be nice/better if the discussion section would focus more on the CML method itself.
We moved and rearranged this paragraph to (now) Sect. 5.1, where we use it to discuss the energy-balance closure problem. See our reply to your comments on line 293-294.

- Lines 361-364: Not clear what is meant with this section, please add clarification, or remove.
We agree with the reviewer that this is not clear and also does not add much to the discussion. Therefore, we removed it accordingly.

- Lines 370-371: Performance is roughly comparable to the reference; this is not true, see also remarks made w.r.t. lines 277-278.
Similar to your previous comment, we meant to refer here to the performance of the free-convection method, which we did not specify. We changed as follows:
Using the two-wavelength method, the performance of the turbulent heat flux estimates obtained with the free-convection scaling and spectral noise correction method is comparable to the comparison between the reference MWS-2λ method and EC system, while application of the two-wavelength method combined with the complete scaling performs less well.

- Lines 380-382: "Yet, for … of the CML": It is unclear why this statement would illustrate the added value of the CML. Please explain.
We agree with the reviewer that this is unclear. Moreover, after reconsideration we think that this does not add much to the story, therefore we removed this sentence.

- Line 397: "In general, …scintillometers" In fact this was already illustrated in the Van der Valk et.al. 2025 paper. To my opinion, in the current paper it is illustrated that operational scintillometry for LE estimates is not (yet?) possible with CMLs with a sufficient accuracy and that attempts to increase the accuracy are either hampered by inaccurate restriction/limitation determination or result in deteriorating sensible heat flux estimates. This in itself is a useful story to tell, though maybe less satisfying.
We agree with the reviewer that this is also a main message of our research, which is currently not well enough reflected in the conclusions. We changed this paragraph as follows:
In general, our results illustrate the possibility to use CMLs as scintillometers. Also after aggregation of the 30-min L_vE estimates to daily E estimates, the performance remains comparable for almost all methods and days. This aggregation might be particularly

interesting for hydrological applications, for example on spatial scales of catchments, which could be combined with the possibility to monitor rainfall using the same CMLs. However, our results also illustrate that the accuracy of the L_vE estimates using CML networks will largely depend on the quality of the radiation estimates. The quality of widely available radiation data, such as LSA SAF, seems too low for our purposes, and needs to be addressed in future research. Additionally, attempts to estimate evaporation using different CML types and employed sampling strategies of networks would be required, while also the performance of the proposed methods and scalings need to be tested in different climatic settings. If these issues would be addressed, CMLs could show a large potential to be used to estimate evaporation, especially considering the existing infrastructure which is also present on locations where other observations are lacking.