DREE-RF: A Radar-Based Rainfall Energy Estimation Model Using Random Forest

Jingxuan Zhu, Qiang Dai, Yuanyuan Xiao, Jun Zhang, Lu Zhuo, and Dawei Han

***Abstract*—Current radar techniques focus on rainfall observations, leaving a research gap in rainfall energy (*E*) involving the interaction of raindrops and land surface processes. *E* is defined as the accumulated kinetic energy per unit rainfall and is a key parameter in the understanding process of the rainfall impact on the land surface. Utilizing the capability of dual-polarization radar to detect the rainfall microphysics characteristics, this study proposes the first computational model for estimating *E* from radar signals. The model investigates the mechanistic correlation between the radar dual-polarization parameters and *E*, and finds that specific differential phase (K_{DP}) and horizontal reflectivity (Z_H) have the strongest correlation with *E*. Therefore, the study develops radar-based empirical regression and random forest (RF) models for *E* estimation, where RF models consider whether the sensitive K_{DP} is available. The results show that the RF models improve the accuracy of estimating *E* and has a Pearson coefficient greater than or equal to 0.97 with station measured *E*, and their spatially extensive capability of the models is further validated. In addition, the Pearson values of daily *E* estimated from radar data based on TRM and RF are 0.85 and 0.92, respectively, indicating that RF is better than TRM when using real radar. This study contributes to enhancing the understanding of rainfall processes in the context of climate change and have great potential for applications in hydrological modeling, flood forecasting, and agricultural planning.**

***Index Terms*—dual polarization; radar; random forest (RF); raindrop size distribution (DSD); rainfall energy**

I. INTRODUCTION

RADAR is essential for quantifying rainfall with high spatial and temporal resolution by measuring reflectivity, providing crucial support for estimating water exchanges between the atmosphere and the land surface [1], [2]. However, the interaction between rainfall and land surface hydrology processes involves more than just the amount of rainfall. It also encompasses the impact of raindrops on surface elements, which is determined by the energy carried by raindrops upon reaching the ground. For instance, raindrops impact the surface with kinetic energy, breaking up soil aggregates and displacing soil particles [3], [4], [5]. These droplets also generate turbulence on water surfaces, changing the gas exchange rates between the interfaces and affecting the carbon and oxygen concentrations and pH levels of the water surface [6], [7]. Moreover, drop energy has a significant impact on the spread of diseases and pests in plants, as well as the cycling of soil nutrients, which are essential for the ecosystem [8], [9]. These energy-driven impacts are challenging to accurately capture using traditional radar estimates alone.

The most precise estimates of rainfall energy (*E*) come from disdrometers that can directly measure drop size distribution (DSD) and speed of raindrops. However, deploying such instruments on a large scale is often cost prohibitive and logistically challenging. To address this, previous research established empirical relationships between rainfall kinetic energy per depth (*KE*) and rainfall intensity (*I*) for specific locations and climatic conditions, including power-law [2], [10], linear [11], polynomial [12], logarithmic [13], [14], and exponential [15], [16] relationships. The exponential form is the most commonly used today because it has ability to yield a positive value at zero intensity and an asymptotic limit at higher intensities. In particular, the exponential *KE–I* relationship recommended in the RUSLE2 manual [5] has been widely used in various studies related to rainfall erosivity and water-induced soil erosion [17], [18], [19], [20]. However, these empirical methods derived from site observations suffer from the difficulty of limited spatial representation [1], [21].

With the advancement of remote sensing technology, ground radars offer the possibility of large-scale *E* estimation. Ground-based weather radar, with its high measurement accuracy and wide coverage, has become a valuable tool for large-scale rainfall measurements, playing a crucial role in in applications such as precipitation forecasting and meteorological research [22]. Decades of research have significantly improved the quality control techniques applied to radar data, enhancing their accuracy and reliability [23], [24], [25], [26]. One of the main advantages of ground-based radar is its ability to capture the intricate details of precipitation processes at the regional scale, allowing for comprehensive monitoring of rainfall pattern distribution. Dual-polarization radar can transmit and receive two polarized signals (typically horizontal and vertical polarization), and the differential information can provide additional characteristic parameters of observed objects. For

This work was supported by the National Natural Science Foundation of China under Grant 42371409 and 42201020. *(Corresponding author: Qiang Dai.)*

Jingxuan Zhu, Qiang Dai, Yuanyuan Xiao and Jun Zhang are with the Key Laboratory of Virtual Geographic Environment, Ministry of Education, Nanjing Normal University, Nanjing 210023, China (e-mail: jingxuanzhu@njnu.edu.cn; q.dai@njnu.edu.cn; xiaoyuanyuan_nnu@njnu.edu.cn; jun.zhang@njnu.edu.cn).

Lu Zhuo is with School of Earth and Environmental Sciences, Cardiff University, Cardiff CF10 3AT, UK (ZhuoL@cardiff.ac.uk).

Dawei Han is with Department of Civil Engineering, University of Bristol, Bristol BS8 1TR, UK (d.han@bristol.ac.uk).

instance, commonly used dual-polarization parameters include horizontal reflectivity (Z_H), differential reflectivity (Z_{DR}) and specific differential phase (K_{DP}). These parameters are closely associated with DSD [27], [28], indicating that radar data can provide an opportunity to calculate E directly. K_{DP} is characterized by stability and reliability, being impervious to radar attenuation and beam blocking. However, there are uncertainties in using K_{DP} to retrieve DSD parameters, as K_{DP} is calculated based on the differential phase shift (φ_{DP}), which may contain noise, especially in cases of low rainfall intensities. To mitigate the effects of noise, φ_{DP} measurements are typically filtered within a certain range, often exceeding several kilometers [29]. Research often impose restrictions on K_{DP}, for instance, Bringi et al. [30] only applies scenarios where K_{DP} > 0.3 deg km^{-1}, while Rico-Ramirez [31] sets the threshold at K_{DP} > 0.1 deg km^{-1}, and Huang et al. [32] defines the usable range of K_{DP} as [0.3, 1].

Although radar-based rainfall estimation has made significant progress, there is still a research gap in radar E estimation. For instance, the relationship between horizontal reflectivity and rainfall intensity (Z_H–I) proposed by Marshall and Palmer [33] has become the most widely used method to estimate rainfall intensity. Various quantitative precipitation estimation (QPE) techniques have also continued to evolve [34], [35], [36]. However, the direct relationship between Z_H and E does not perform adequately due to the nonlinearity between raindrop size and velocity, offering a challenge in accurately estimating E using regression methods. Recognizing the limitations of fixed regression relationships, the study turns to the random forest (RF) algorithm to take full advantage of the large range and high spatial and temporal resolution of radar. Known for its adaptability and robustness in developing regression models [37], [38], RF has the potential to overcome spatial limitations by effortlessly handling high-dimensional datasets, making it possible in scenarios like estimating radar-based E where dual-polarization variables interact in a complex manner. Furthermore, RF provides the benefit of mitigating overfitting, a frequently encountered issue in regression modeling, by merging the outputs of several decision trees, leading to an improved generalization of the model [39].

In order to improve E estimation more accurately and avoid reliance of intensity, we propose two kinds of large-scale Dual-polarization radar-based Rainfall Energy Estimation models using RF (DREE-RF), considering whether K_{DP} is available. The study utilized the distinction between different polarization radar electromagnetic echoes to retrieve E at large scales. This is the first comprehensive method for estimating E from radar signals, addressing the limitations of the current study and has important implications for our knowledge regarding the evolution of rainfall processes in the context of climate change.

II. Study Area and Data

A. Study area

The study area for our study is centered around the Hameldon Hill radar, which is located in the eastern part of Hameldon Hill in Lancashire, England, at a latitude of 53.75°N and a longitude of 2.29°W. This hill is part of the Pennines, a range of hills and mountains that stretches across northern England. Hameldon hill itself has a notable elevation, rising to approximately 409 meters above sea level, providing an advantageous position for various observational purposes, including meteorological research. Due to its geographical location in the UK, Hameldon Hill experiences a typical maritime climate characterized by relatively mild temperatures, high humidity, and frequent precipitation year-round. These weather conditions make it an ideal location for studying rainfall patterns, storm systems, and atmospheric phenomena. Figure 1 identifies the location of weather radar station and its 50- and 80-km buffer ranges, as well as the locations of the two disdrometers around the radar.

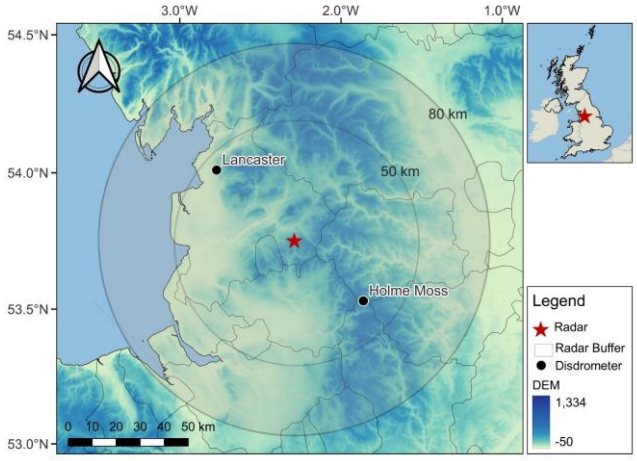


Fig. 1. Map of disdrometers, Hameldon Hill radar and its 50- and 80-km buffer ranges, with terrain elevation in the background.

B. Disdrometer data

This study collected data from two Laser Precipitation Monitor (LPM) disdrometer stations including Holme Moss station (53.53°N, 1.86°W) and Lancaster station (54.01°N, 2.77°W), both located within a 50 km radius of the radar site. The data source was the Disdrometer Verification Network (DiVeN) [40], which is a disdrometer dataset consists of 14 Thies Clima LPMs deployed across the entire UK from 2017 to 2019. Holme Moss station and Lancaster station are situated at distances of 42.94 km and 37.96 km from the radar center, respectively. The LPM operates by recording laser signal attenuation to determine DSD and can estimate particle fall velocity by considering the particle's duration within the laser beam. Each LPM covers a measurement area of 45.6 cm^2 with dimensions of 228 mm in length and 20 mm in width. It is capable of recording the quantity of raindrops passing through in each minute, encompassing 22 different diameter categories ranging from 0.125 mm to 8 mm, as well as 20 different velocity categories ranging from 0 to 10 m·s^{-1}.

The DSD data used in this study underwent quality control processes based on [41], which involved the removal of outliers and non-rainfall information. For the Holme Moss station, the DSD measurements were taken from March 10, 2017, to September 30, 2019, resulting in 52,105 usable records after processing. At the same time, for the Lancaster station,

measurements were recorded from February 22, 2017, to September 30, 2019, yielding 66,450 usable records after data processing.

C. C-band radar data

Nestled in the landscapes of Lancashire, UK, the Hameldon Hill radar serves as an essential component of the radar network of the UK Meteorological Office [42]. Capable of simultaneously transmitting and receiving horizontally and vertically polarized waves, this C-band radar uses a sampling scheme with a range bin length of 600 m at five elevations. Parameters including horizontal reflectivity (Z_H), differential reflectivity (Z_{DR}), differential phase shift (φ_{DP}), correlation coefficient (ρ_{HV}), and radial velocity (RV) are measured at all elevations. The radar has a maximum range of 255 km and a nominal beam width of 1 degree. It conducts multiple scans at various angles every 5 minutes. In this study, scan data at the lowest radar elevation angle (0.5°) were used to analyze the spatial distribution of E.

The estimation of K_{DP} involves three primary steps [29], [43]: 1) φ_{DP} unwrapping, where, due to the wrapped nature of the measured φ_{DP}, a one-dimensional phase unwrapping process is employed; 2) φ_{DP} wavelet analysis filtering, used to reduce fluctuations induced by clutter, which tend to cause significant oscillations; and 3) K_{DP} calculation, wherein the least squares method is utilized with a variable-distance approach to fit the unwrapped φ_{DP} data and derive the K_{DP} values [44]. Then, the raw radar data were subjected to attenuation corrections [45], [46] and nonweather clutter removal [47] to achieve corrected Z_H and Z_{DR}.

III. Methodology

A. Framework

The study framework encompasses several key steps aimed at constructing and evaluating a model for estimating E based on dual polarization parameters (Fig. 2). Initially, disdrometer data is utilized to extract minute-by-minute parameters of DSD parameters, which include the mass-weighted mean drop diameter (D_m), normalised intercept parameter (N_w), and shape parameter (μ). These parameters are subsequently used to compute the corresponding accumulated E for each minute. Following the DSD parameter derivation, the T-matrix method is utilized to calculate C-band dual-polarization radar parameters, which encompass Z_H, Z_{DR} and K_{DP}.

This study proposes two random forest (RF) models, denoted as RF1 and RF2, and compares them with the traditional regression model (TRM) and DSD-intermediated regression model (DIRM). RF1 incorporates Z_H, Z_{DR}, and K_{DP}, while RF2 omits K_{DP} to cope with the possible unreliability of K_{DP} data that does not fall within the interval [0.3,1]. All models are calibrated (or trained) and validated using an 80%/20% split, referred to as the datasets TRM-C, TRM-V, DIRM-C, DIRM-V, RF-T, and RF-V, respectively. The E estimation results, E_{TRM}, E_{DIRM}, E_{RF1}, and E_{RF2}, are obtained and subsequently

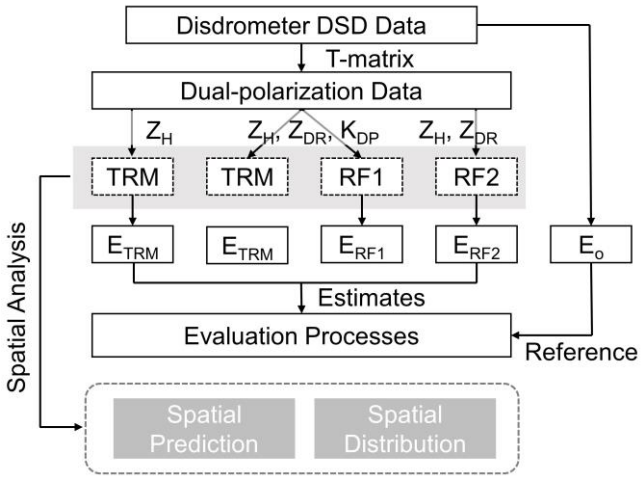


Fig. 2. The framework of this study for estimating E from dual-polarization data using TRM, DIRM and RF models.

evaluated against the reference E_o derived from disdrometer DSD observations using various statistical metrics. Furthermore, this study undertakes an evaluation of the spatial extension capabilities of the TRM, DIRM and RF models by utilizing data from a different disdrometer location. This assessment is pivotal in comprehending the reliability and adaptability of these models across diverse geographic regions. Finally, the study calculates the spatial distribution of precipitation energy using C-band radar data, providing insight into the microphysical mechanisms of rainfall.

B. DSD model

DSD model is conventionally described using a normalized three-parameter gamma model, a widely accepted approach in meteorology (Dai and Han, 2014; Ulbrich, 1983). The mathematical expression for this model is

$$N(D) = N_w f(\mu)\left(\frac{D}{D_m}\right)^{\mu} \exp\left[-\frac{(4+\mu)D}{D_m}\right] \quad (1)$$

$$f(\mu) = \frac{6(4+\mu)^{\mu+4}}{4^4\Gamma(\mu+4)}, \quad (2)$$

where $N(D)$ represents the count of particles within each unit diameter interval, D_m (mm) signifies the mass-weighted mean drop diameter of the DSD model, N_w ($\mathrm{mm^{-1}\cdot m^{-3}}$) is the normalised intercept parameter, $\Gamma(n)$ denotes the gamma function, and $f(\mu)$ is a function associated with the shape parameter μ.

Then, by summing the number of raindrops in each diameter and velocity range, $N(D)$ can be calculated from disdrometer observations as shown below:

$$N(D_i) = \sum_{i=1,j=1}^{I,J} \frac{n_{i,j}}{A\Delta t v_j \Delta D_i}. \quad (3)$$

In the provided equations, I and J correspond to the bin indices for diameter and velocity, respectively. D_i (mm) signifies the mean diameter within the ith size bin, v_j ($\mathrm{m\cdot s^{-1}}$) represents the mean particle velocity associated with the jth velocity bin. The term n_{ij} denotes the number of particles found

within the specific diameter bin i and velocity bin j, A_i (m^2) designates the sampling area, while Δt and ΔD refer to the sampling time and diameter interval, respectively.

C. Parameter calculation from disdrometer observations

Rainfall kinetic energy is the main factor that causes soil particles to be dislodged and splashed, and the physical structure of soil to be decomposed and eroded. Assuming that the drop is a uniform sphere, according to the generalized kinetic energy formula, its kinetic energy (e, unit: J) is related to the diameter (D, unit: mm), the terminal velocity (v, unit: $m \cdot s^{-1}$) and the water density (ρ, 1 $g \cdot cm^{-3}$):

$$e = \frac{1}{2} m v^2 = \frac{1}{12} 10^{-6} \pi \rho v_i^2 D_i^3 . \quad (4)$$

Then, the total kinetic energy of raindrops falling in the unit volume, or KE ($J \cdot m^{-2} \cdot mm^{-1}$), is generally used to express E, defined as:

$$KE = \frac{e}{AP} = \frac{1}{AP} \sum n \cdot e \quad (5)$$

$$P = I \cdot t . \quad (6)$$

$$I = 6\pi \times 10^{-4} \int N(D) D^3 v(D) dD , \quad (7)$$

where e (J) and P (mm) are the total raindrop kinetic energy and rainfall depth in a specific time step t (h), and I is the rainfall intensity ($mm \cdot h^{-1}$) calculated from DSD data and drop velocity. Since the sampling interval of the disdrometer is 1 minute, the t value is set to 1/60.

Therefore, the cumulative E at the mth minute, denoted as E_m, is represented by the following equation, which is derived from the unit E for that minute, KE_m, and the rainfall intensity, I_m:

$$E_m = KE_m \cdot P_m = \frac{KE_m \cdot I_m}{60} . \quad (8)$$

On an event scale, the accumulated E represents the sum of the storm energy over all time steps, which is expressed as follows:

$$E = \sum KE \cdot P . \quad (9)$$

As for radar parameters at the C-band frequency, including Z_H expressed in units of $mm^6 \cdot m^{-3}$, Z_{DR} in dB, and K_{DP} in $deg \cdot km^{-1}$, can be computed utilizing the T-matrix scattering technique [49], [50] from disdrometer data:

$$Z_{H,V} = \frac{4\lambda^4}{\pi^4 \left| K_w^4 \right|} \sum \left| f_{H,V}(D_i) \right|^2 N(D_i) D_i^3 \Delta D_i \quad (10)$$

$$Z_{DR} = 10 \log_{10} \frac{Z_H}{Z_V} \quad (11)$$

$$K_{DP} = \frac{180 l}{\pi} \sum \mathrm{Re}\left[f_{HH}(0, D_i) - f_{VV}(0, D_i) \right] N(D_i) \Delta D_i , \quad (12)$$

where f_H and f_V (D_i) represents the backscatter amplitude of a raindrop for horizontal and vertical polarizations, while f_H (0, D_i) and f_V (0, D_i) correspond to the standard forward scattering amplitudes. K_w (0.96) denotes the dielectric factor of water, and l (53 mm) signifies the radar wavelength.

D. TRM and DIRM using empirical relationships

The TRM of estimating E using radar signals can be divided into three main steps. Firstly, the rainfall intensity is estimated from empirical relationships using the Z_H parameter and the Z_H–I relationship. Subsequently, the E per unit depth (KE) is calculated by combining the KE–I relationship. Finally, the KE per time step is multiplied with the rainfall depth to determine the accumulative E for the specified time period.

In the large-scale area, the exponential KE–I empirical relationship is widely used to estimate rainfall KE, recommended by (USDA-Agricultural Research Service [5] in RUSLE2:

$$KE = 29[1 - 0.72 \exp(-0.082 I)] . \quad (13)$$

The radar reflectivity factor for Z_H and I relationship is a fundamental concept in meteorology. This relationship provides a crucial tool for estimating precipitation rates based on radar observations. The typical form of the Z_H–I relationship is expressed as:

$$Z_H = a I^b , \quad (14)$$

where a and b are empirical constants that vary depending on factors such as the type of precipitation, radar wavelength, and geographic location. The Z_H–I relationship helps meteorologists convert radar data into valuable rainfall information, enabling accurate rainfall estimation and weather forecasting.

The DIRM employs dual-polarization radar parameters to regressively derive the DSD parameters, which are then used to calculate E. The following form is a widely used tool for regressing DSD parameters (including the drop median diameter D_0, the generalized intercept parameter N_w and the shape parameter μ) on dual-polarization parameters [38]:

$$D_0 = a + b Z_{DR} + c Z_{DR}^2 + d Z_{DR}^3 , \quad (15)$$

$$LWC = a Z_H 10^{(b Z_{DR} + c Z_{DR}^2 + d Z_{DR}^3 + e Z_{DR}^4)} , \quad (16)$$

$$N_w = \frac{3.67^4}{\pi \rho} \frac{LWC}{D_0^4} , \quad (17)$$

$$\mu = a + b\lambda + c\lambda^2 , \quad (18)$$

$$\lambda = m Z_{DR}^n , \quad (19)$$

where the liquid water content (LWC) is related to Z_H and Z_{DR}.

E. RF-based E retrieval model

RF is an ensemble learning method renowned for its versatility and accuracy. It operates by constructing multiple decision trees during RF-T and outputs predictions based on the ensemble average or mode. The distinctive feature of RF lies in its ability to mitigate overfitting and enhance generalization performance, making it well-suited for complex relationships between input variables. To initialize the RF model, essential parameters need to be configured. The initial parameters include the number of trees in the forest, the maximum depth of each tree, and the minimum number of samples required to split a node. These parameters play a pivotal role in determining the model's performance, as determined by analyzing the out-of-bag mean squared error (OOB MSE) of MATLAB's TreeBagger tool. In a Random Forest, each decision tree is trained on a bootstrap sample of the original dataset, and some

data points are left out (out-of-bag) during the training process. The OOB MSE is calculated by comparing the predicted values for the out-of-bag data points to their actual values, providing a measure of how well the model generalizes to unseen data. A lower OOB MSE indicates better predictive performance.

The construction of the RF model involves establishing relationships between dual-polarization parameters (Z_H, Z_{DR}, K_{DP}) and E. The dataset, comprising 80% of the available observations, is utilized for training the RF model. Each tree in the RF ensemble captures distinct patterns and dependencies within the data. The model undergoes iterative refinement during the RF-T process, enhancing its ability to generalize across diverse datasets. After the RF-T phase, the remaining 20% of the dataset serves as the TRM-V set for evaluating the model's predictive accuracy.

F. Evaluation methods

To delineate the relationship between radar dual-polarization parameters and E, Pearson correlation coefficient is employed to calculate the respective correlation measures. It assesses the linear association between the dual-polarization parameters and E, providing information on the strength and direction of this relationship. In addition, five additional metrics—mean absolute error (MAE), mean bias error (MBE), root mean square error (RMSE), relative bias (BIAS), and Nash–Sutcliffe model efficiency coefficient (NSE)—are employed to evaluate the results of estimating E using the TRM, DIRM and RF models. MAE quantifies the average magnitude of errors, giving a measure of the overall accuracy. MBE highlights the bias or tendency of the model to overestimate or underestimate the E. RMSE provides information about the magnitude of errors on a root mean square scale, emphasizing larger errors. BIAS, as a relative measure, indicates the systematic tendency of the model to consistently overestimate or underestimate. Finally, NSE evaluates the efficiency of the model by comparing the predicted values with the observed ones, considering both the mean and variability. All of the metrics mentioned above are defined in Table I.

TABLE I
STATISTICAL EVALUATION METRICS IN THIS STUDY

Evaluation Metrics	Formulas	Perfect value
Pearson correlation coefficient	$Pearson = \frac{\sum (E_i - \overline{E})(O_i - \overline{O})}{\sqrt{\sum (E_i - \overline{E})}\sqrt{\sum (O_i - \overline{O})}}$	1
Mean absolute error (MAE)	$MAE = \frac{\sum \lvert O_i - E_i \rvert}{n}$	0
Mean bias error (MBE)	$MBE = \frac{\sum (O_i - E_i)}{n}$	0
Root mean square error (RMSE)	$RMSE = \sqrt{\frac{1}{n}\sum (O_i - E_i)^2}$	0
Relative bias (BIAS)	$BIAS = \frac{\sum (E_i - O_i)}{\sum O_i} \times 100\%$	0%
Nash–Sutcliffe model efficiency coefficient (NSE)	$NSE = 1 - \frac{\sum (O_i - E_i)^2}{\sum (O_i - \overline{O})^2}$	1

IV. RESULTS

A. Responses of dual-polarized radar parameters to E

Using the T-matrix method, C-band dual-polarization parameters, including Z_H, Z_{DR}, and K_{DP}, were calculated from the DSD parameters obtained from the Holme Moss disdrometer. Fig. 3 displays an occurrence plot illustrating the relationships between the three radar parameters and the three DSD parameters. Previous studies have found notable relationships between the dual-polarization parameters and the DSD parameters [27], [28]. For example, D_0 and Z_{DR} have an apparent polynomial regression relationship (Fig. 3b), which is widely used to estimate D_0 using radar signals, while N_w can be derived using Z_H and Z_{DR} co-inversion.

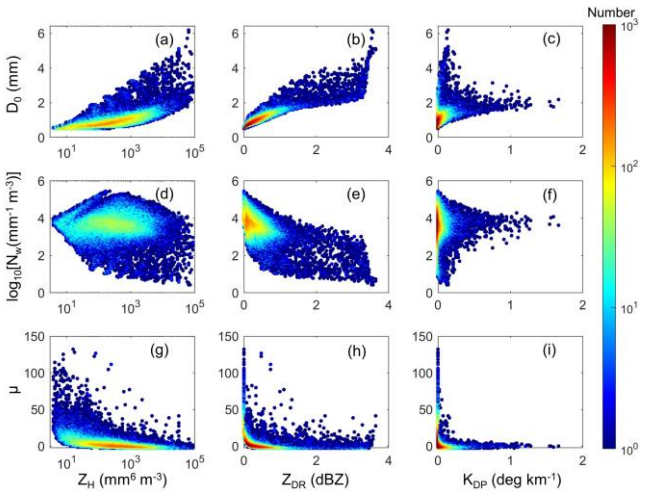


Fig. 3. Occurrences of relationships between DSD parameters and dual polarization parameters.

In order to more visually analyze the relationship between the radar dual polarization parameters and the E, the sampling occurrence distribution of the three relationships, which are obtained from the Holmes Moss disdrometer, are shown in Fig. 4. It is noteworthy that the relationship between K_{DP} and E is more centered on a straight line. In contrast, the dispersion of data points characterizes the connection between E and Z_{DR}, reflecting a weaker correlation between the variables. The Pearson coefficient was used to compare the correlation between the radar bipolarization parameters and E. The strongest Pearson correlation was found between the K_{DP} and

E, which was an astonishing 0.94. The Z_H and Z_{DR} had relatively low Pearson values of 0.70 and 0.36, respectively, in comparison with *E*.

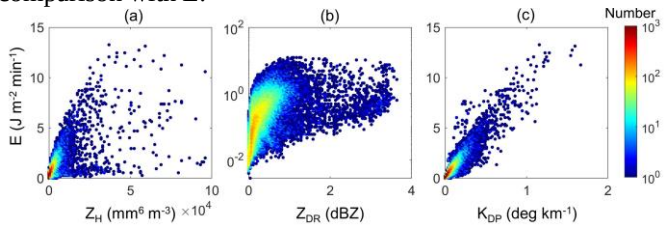


Fig. 4. Occurrences of relationships between radar dual-polarization parameters Z_H (a), Z_{DR} (b), and K_{DP} (c) with *E*.

B. E estimation using TRM and DIRM

At the Holmes Moss station, using the minute rainfall intensity data obtained from the disdrometer, and matching it with the disdrometer-derived C-band Z_H results. The first 80% of the data points were selected as the TRM-C dataset and the last 20% of the records were selected as the validation dataset, with data amounts of 35,258 and 8,815, respectively. The TRM-C dataset is used to fit a local Z_H–I empirical formula, which results in a power-form relationship Z_H=320.6$I^{1.56}$, shown as the black line in Fig. 5. Then, taking disdrometer-derived Z_H as the independent variable and combining it with the Z_H–I regression relationship, the estimated rainfall intensity (I_{TRM}) can be obtained based on TRM. Following this, using the *KE*–*I* empirical relationship recommended by RUSLE2 (Eq. 13), the corresponding unit *E* (KE_{TRM}) is determined. Ultimately, the product of I_{TRM} and KE_{TRM} yields the accumulative *E* for each minute (E_{TRM}).

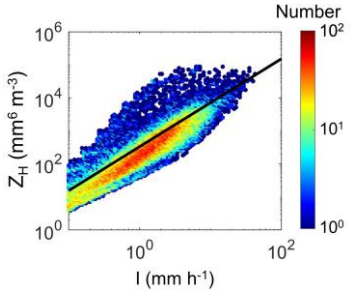


Fig. 5. Occurrences of the C-band Z_H–I relationship in colored dots and the fitting formula in the black line within the TRM-C dataset.

Fig. 6 shows the comparison of the three DSD parameters (including D_0, $\log_{10}N_w$ and μ) obtained using the disdrometer-derived DSD regression relationship with the measured DSD parameters of the disdrometer. D_0 is best fitted for small raindrops, $\log_{10}N_w$ is not exactly symmetrically distributed along the diagonal, and the regression results for μ are very different from the measured values. In fact, it is difficult to estimate μ accurately using conventional regression methods, especially in the presence of measurement errors [51].

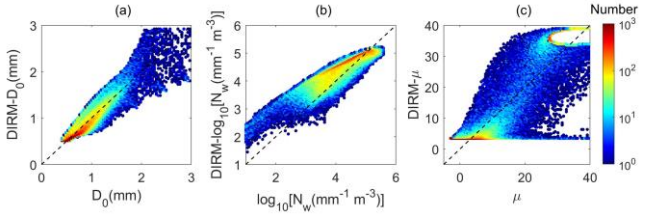


Fig. 6. Comparison of D_0 (a), $\log_{10}N_w$ (b) and μ (c) using DIRM with the DSD-derived true value (black dashed line is diagonal).

Fig. 7 (a, b) displays a scatterplot comparison between the minute accumulative *E* (E_{TRM}) calculated by the TRM model and the accumulative *E* (E_o) computed from disdrometer observed DSD information. Considering E_o as the true value, the black dashed line in the graph represents the diagonal of the plot, which then indicates that points closer to this line represent better estimation results. It is evident from the graph that both the TRM-C and TRM-V datasets exhibit some degree of dispersion in TRM results. There are instances of overestimation in *E* using TRM-C set at the Holmes Moss station, with some E_{TRM} values exceeding 15 J m^{-2} min^{-1}, and even approaching 30 J m^{-2} min^{-1}, while DSD-derived E_o shows almost no values exceeding 15 J m^{-2} min^{-1}.

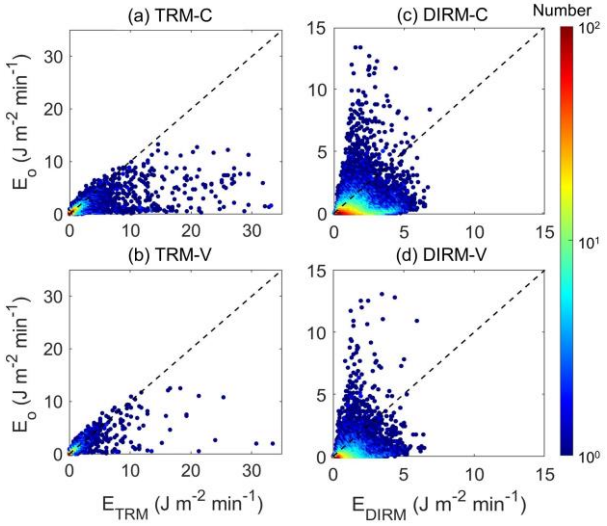


Fig. 7. Comparison of minute accumulative *E* obtained from TRM-C (a), TRM-V (b), DIRM-C (c), DIRM-V (d) dataset using TRM with the DSD-derived true value (black dashed line is diagonal).

Similarly, Fig. 7 (c, d) presents a scatterplot comparing the cumulative *E* in minutes (E_{DIRM}) calculated by the DIRM model with the E_o from the disdrometer. As can be seen from the figure, the DIRM results are all very scattered, with a large number of samples far from and below the diagonal line, indicating that the E_{DIRM} values are large compared to the true values, although most of the E_{DIRM}s do not exceed 5 J m^{-2} min^{-1}.

Table II presents several evaluation metrics for the TRM and DIRM results for both the calibrated and validated datasets. It is noticeable that all datasets exhibit relatively low Pearson correlation coefficients with observed values, for the DIRM results. In fact, the Pearson coefficient for the TRM-C dataset

is as low as 0.69, which is even lower than the TRM-V dataset. In terms of errors, the TRM-V results display higher MAE, MBE, and RMSE compared to the TRM-C results. The MBE, in particular, exceeds the TRM-C results by a factor of two. The fact that the TRM-V has better results may be due to the insufficient amount of data and the different time coverage of the two datasets, indicating large uncertainties of the TRM. Additionally, the NSE for the TRM-C dataset is -1.85, falling below zero, indicating a poor simulation accuracy for the TRM model. The TRM-C set for TRM displays a bias higher than 12, whereas the TRM-V set shows a bias result of less than -1. The contrast between the two sets is significant, and with opposing signs, reveals that TRM carries a level of uncertainty. The Pearson coefficient of DIRM, on the other hand, is much smaller than that of TRM, which is 0.39 for both datasets, and the MAE is almost twice as large as that of TRM, and the absolute value of BIAS is also very high, suggesting that DIRM has a much larger uncertainty than TRM. However, its RMSE is lower, which may be due to the small range of values in E_{DIRM}.

TABLE II
EVALUATION METRICS FOR THE TRM RESULTS FOR THE TRM-C, TRM-V, DIRM-C AND DIRM-V DATASETS

	Pearson	MAE	MBE	RMSE	BIAS	NSE
TRM-C	0.69	0.35	-0.08	1.41	12.24	-1.85
TRM-V	0.75	0.31	0.01	1.12	-1.61	-0.23
DIRM-C	0.39	0.61	0.54	0.96	-69.53	-0.41
DIRM-V	0.39	0.62	0.49	0.99	-59.79	-0.51

C. E estimation using RF models

During practical radar observations, K_{DP} values below 0.1 deg km^{-1} are vulnerable to the radar inherent noise and other factors, which leads to low data reliability. Consequently, this study created two RF models to address available and unavailable K_{DP} scenarios. RF1, the first model, includes RF[Z_H, Z_{DR}, K_{DP}], while RF2, the second model, uses RF[Z_H, Z_{DR}]. The data used in this section is the same as section 4.2, again divided into a RF-T dataset and a RF-V dataset. Before constructing the RF models, it was essential to determine the number of leaf nodes and trees for the model using the RF-T dataset. In this study, the MATLAB TreeBagger tool was employed to build the RF. Initially, different values for the number of leaf nodes (5, 10, 20, and 50) were considered. Both models with 5 leaf nodes exhibited the lowest mean squared error, and all curves almost reached a plateau after 100 trees. Therefore, we set the number of trees in the model to 100 and obtain each RF model with dual-polarization parameters to retrieve E.

After establishing the RF models, the input parameters from both the RF-T and RF-V datasets were fed into the RF models to obtain the respective minute E E_{RF1} and E_{RF1}, as shown in Fig. 8. In Fig. 8 (a, b), it is evident that the scatterplot of E_{RF1} versus E_o closely aligns with the diagonal line, indicating that the E_{RF1} value of the RF model closely matches the E_o value with a small difference. Likewise, the comparison of E_{RF2} and E_o displayed in Fig. 8 (c, d), as well as the outcome of RF1 presented in Fig. 8 (a, b), exhibit considerable similarity.

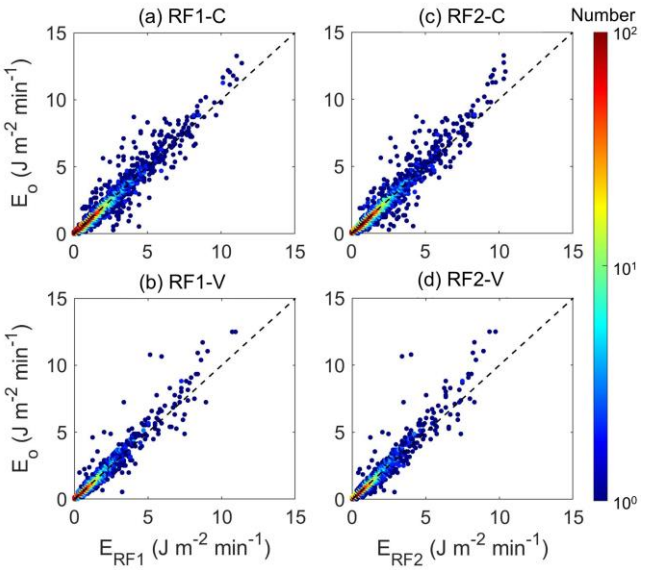


Fig. 8. Comparison of minute accumulative E obtained from RF1-T (a), RF1-V (b), RF2-T (c) and RF2-V (d) dataset with the DSD-derived true value (black dashed line represents the diagonal).

To quantify the performance of the two RF models, Table III presents a total of six statistical metrics. The superiority of the RF1 model, which incorporates K_{DP} data, over the RF2 model, which does not, is worth noting. Both RF models demonstrate remarkable Pearson correlation coefficients of over 0.97, with RF1 achieving 0.98 for both the RF-T and RF-V sets. The NSE values for RF1 are marginally higher than those for RF2, even reaching 0.97 in the RF-T set. In terms of error metrics, the MBE values for all the data set are close to zero, except for the RF1 RF-V data. Despite a slightly lower performance on the RF1 validation dataset compared others, the RF1-V results are still acceptable. Furthermore, the validation set of RF1 and RF2 demonstrates significantly enhanced performance in comparison to the TRM model, with an RMSE that is only 20.54% and 23.21% of the TRM model, respectively.

TABLE III
EVALUATION METRICS FOR THE RF RESULTS FOR THE RF-T AND RF-V DATASETS

	Pearson	MAE	MBE	RMSE	BIAS	NSE
RF1-T	0.98	0.06	0.00	0.16	-0.23	0.96
RF1-V	0.98	0.09	0.03	0.23	-3.67	0.95
RF2-T	0.98	0.06	0.00	0.17	-0.30	0.96
RF2-V	0.97	0.09	0.03	0.26	-4.15	0.93

D. Spatial prediction capabilities of TRM and RF model

In order to assess the spatial predictive capability of the models, data from the Lancaster station (located within the scanning range of the radar), were fed into the TRM, DIRM and RF models built at the Holmes Moss station to assess their spatial simulation performance (Fig. 9). In Fig. 9a, the comparison results between TRM and the ground truth remain wide scattered, mostly concentrated below the diagonal line, indicating a clear tendency to overestimate E. E_{DIRM}, on the other hand, is lower than the reference value, and the red high-

density region is concentrated in the lower left part of the plot with no distribution along the diagonal (Fig. 9b). In contrast, the RF models perform much better (Fig. 9c, d), with most of the scatterplot points still closely match the diagonal.

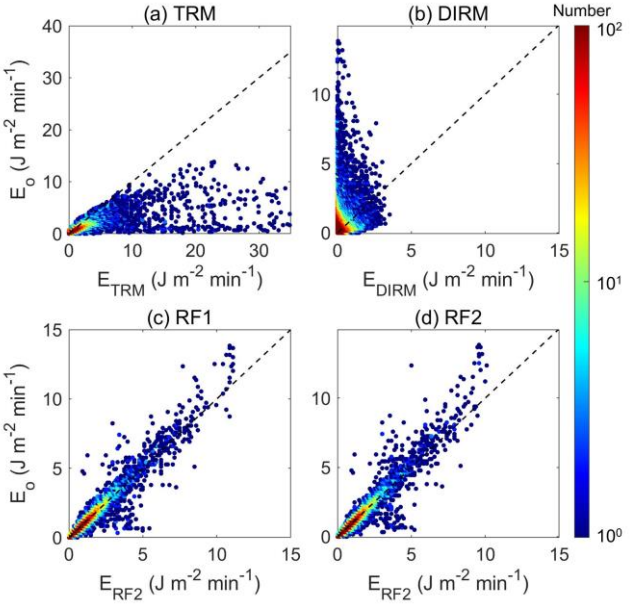


Fig. 9. Comparison of minute accumulative *E* based on TRM (a), DIRM (b), RF1 (c) and RF2 (d) with the DSD-derived true value (black dashed line is diagonal) at Lancaster station.

Table IV shows that the Pearson of E_{TRM} and E_{DIRM} compared to the reference values are only 0.73 and 0.32 respectively, and their NSEs are even lower than -2. In contrast, Pearson of the RF model is 0.97, and the NSE is 0.95, which is slightly lower than that of its performance at the Holmes Moss station, but it is still significantly better than that of the TRM model. In terms of error metrics, the MAE, MBE and RMSE of the RF model results are significantly better than those of the TRM and DIRM models. In addition, the very high positive BIAS values of the TRM model results suggest that there is a greater tendency to overestimate *E* values at Lancaster station. Meanwhile, the very poor Pearson and NSE of DIRM indicate that this method of introducing DSD as an intermediate variable introduces more errors and is not suitable for *E* estimation, especially cross-station estimation.

TABLE IV
EVALUATION METRICS FOR THE RF RESULTS FOR THE TRM, DIRM AND RF DATASETS IN LANCASTER STATION

	Pearson	MAE	MBE	RMSE	BIAS	NSE
TRM	0.73	0.34	-0.21	1.70	35.59	-2.96
DIRM	0.32	0.30	-0.10	0.66	32.61	-3.48
RF1	0.97	0.07	0.00	0.21	-0.83	0.94
RF2	0.97	0.08	0.01	0.23	-1.24	0.93

E. Spatial prediction capabilities of TRM and RF

Within the 50 km scanning range of the Hameldon Hill radar, two disdrometers are located, as highlighted in black squares in Fig. 10. The radar data underwent preprocessing steps, including attenuation correction and clutter removal. Non-rainfall information with ρ_{hv} less than 0.98 was eliminated. The TRM and RF models, developed using data from the Holme Moss station, were applied to estimate *E* within the radar scanning scope. Fig. 10 depicts the Plan Position Indicator (PPI) scan of estimated E_{TRM} and E_{RF} at 1139 UTC on August 26, 2018 and 1605 UTC on September 20, 2018. In the RF computations, two models, RF1 and RF2, are used based on whether K_{DP} falls within the [0.1,3] range. The DIRM results are not shown in this section due to their poor performance.

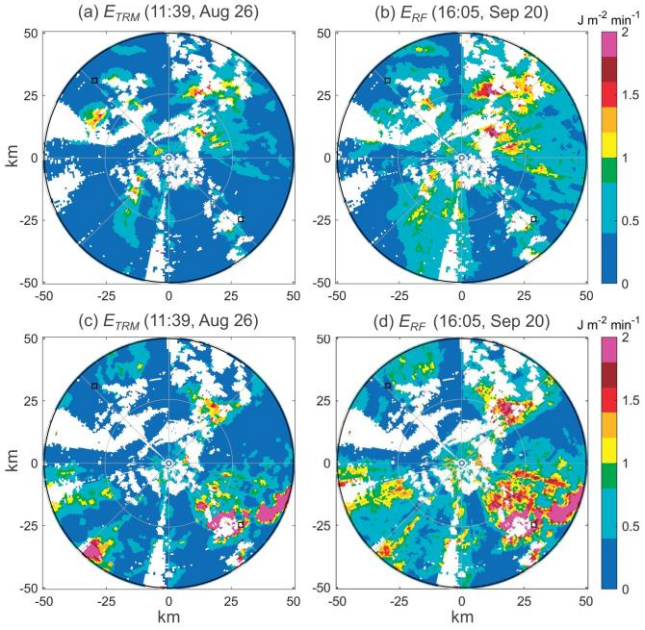


Fig. 10. Estimated E_{TRM} (a) and E_{RF} (b) for 11:39 UTC 26 Aug 2018 and E_{TRM} (c) and E_{RF} (d) for 1605 UTC 20 Sep 2018. The black squares show the locations of the two disdrometers in this study.

In Fig. 10, It can be seen that the E_{RF} is overall higher than the E_{TRM}, which is in line with the pattern found in Fig. 9 that TRM has a tendency to be underestimated. At this time, most of the intense rainfall (Z_H>40 dBZ) was concentrated northeast of Holme Moss station, with E_{TRM}s and E_{RF}s exceeding 200 J m^{-2} min^{-1} and 400 J m^{-2} min^{-1}, while the center of the intense rainfall was excluded from the *E* calculations due to the mixing of hailstones that created gaps in the graph. It is noteworthy that the RF-estimated *E* is overall higher than the TRM-estimated *E* in Fig. 10, which is in difference to the TRM-estimated bias derived from the two disdrometers. However, the BIAS results with opposite signs for the RF-T and RF-V sets and the Pearson values consistently below 0.80 in Table II also illustrate the unreliability of the TRM. In addition, due to radar sensitivity limitations, low intensity rainfall is less frequently identified, again likely to elicit a different pattern than the disdrometer.

Fig. 11 compares the cumulative daily *E* estimated from radar data and disdrometer data at Lancaster station. The study selected the minute-scale times when the two types of data overlap and accumulated to the daily scale. Using the disdrometer-derived daily *E* as the reference, it can be seen that all the daily *E* values estimated using radar data are lower. Among them, the RF-based radar samples are closer to the diagonal, showing that the values are taken closer to the real values. The Pearson values of 0.85 and 0.92 for the TRM and

RF, respectively, indicate that the RF outperforms the TRM in the case of using real radar.

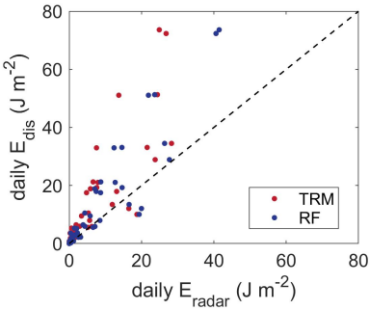


Fig. 11. Comparison of daily cumulative values of *E* using radar data (based on TRM and RF) and disdrometer data (black dashed line is diagonal) at Lancaster station in 2018.

V. Discussion

The significance of *E* cannot be underestimated, especially in the study of raindrop interaction with land surface processes. Disdrometers provide a method to observe DSD parameters, allowing for precise estimation of *E*. However, the station-based disdrometer makes it challenging to apply this method on a large scale. On the other hand, ground-based radar systems offer broad scanning coverage, and previous research [27], [28] have revealed a strong correlation between dual-polarization radar parameters and rainfall microphysical characteristics, presenting a potential solution for large-scale *E* estimation. While current radar-based QPE methods have made significant strides in providing accurate measurements of rainfall amounts, there remains a research gap in estimating *E*, which reflects the impact of raindrops on surface elements.

Traditional methods for calculating *E* often rely on empirical functions of rainfall intensity. These methods are widely used and appreciated for their simplicity. However, they come with certain limitations. They tend to oversimplify the complex relationship between rainfall microphysical characteristics and rainfall intensity, overlooking the diversity and intricacies of precipitation [1]. In many cases, these methods calculate *E* primarily based on rainfall intensity, which may not accurately reflect the true impact of the rainfall on the surface. Another issue is that these methods are typically based on meteorological conditions specific to particular regions, limiting their applicability to different geographical areas. They often disregard regional differences and fail to fully account for the geographical and meteorological variations in rainfall characteristics. Using the same empirical formula in different geographical environments may lead to inaccurate estimations, especially in applications where higher precision is required. Furthermore, researchers have paid close attention to the choice of the forms of the empirical relationship, such as exponential, logarithmic and power functions, which have a significant effect on the estimation results. Even though empirical methods of estimating *E* are useful in certain situations, they cannot comprehensively meet the requirements of different regions and various meteorological conditions. Therefore, there is a necessity to explore more accurate and widely applicable estimation methods to better understand and harness the potential of *E*.

The DIRM approach, which employs the DSD parameters as intermediary variables, is theoretically capable of achieving a high degree of accuracy in the *E* estimation. However, the relatively straightforward inversion of the aforementioned DSD parameters with the dual-polarization parameters often yields unsatisfactory results, introducing a considerable uncertainty that significantly impairs the *E* estimation. This is particularly evident when compared to the TRM method, which is based on the intensity regression relation and offers a more robust estimation of *E*. It should be noted that there have been some studies that have proposed methods for estimating DSD with dual-polarization parameters with high accuracy [e.g., 52-56]. However, these methods are more complicated and are not considered in this study. Future studies will also further consider these methods to improve the estimation of DSD by dual-polarization radar, thus optimizing the estimation capability of *E* by dual-polarization radar.

The DREE-RF models, which utilize radar dual-polarization parameters, offer an alternative method for estimating *E*. This model capture the interactions between radar signals and microphysical information of *E*, which could yield more precise and adaptable results. This approach is particularly advantageous in large-scale regions that exhibit diverse rainfall patterns and microphysical properties. The utilization of radar dual-polarization parameters has the potential to enhance comprehension of subtle variations in rainfall microphysics and their impact on *E*. This is due to the association of these parameters with microphysical information, as represented by DSD information, which drives *E* through raindrop microphysical processes. In addition, radar scans cover a wider range of areas compared to station devices, enabling more accurate rainfall microphysical information to be extended to larger geographic regions. Moreover, this research proposes that the RF model, which concentrates on microphysical mechanisms, provides a marked benefit for spatial forecasts. The DREE-RF results shows potential for improving large-scale spatial estimations of *E* accuracy when compared to traditional empirical formula methods. This emphasizes the possibility of significantly enhancing our comprehension and application of *E*, particularly in areas with diverse rainfall patterns and microphysical properties.

Despite the promising performance of DREE-RF models in estimating *E*, it is vital to consider certain uncertainties. First, the disdrometer data itself carries inherent uncertainties, despite undergoing quality control. It is challenging to eliminate all recording errors entirely, and the instrument itself may introduce systematic errors. Second, the amount of disdrometer data utilized to construct the model may be limited. This study is limited by the use of data from a single disdrometer over a period of no more than three years, which may hinder the ability of the model to capture variations in rainfall microphysical characteristics across diverse regions and over extended time periods. Third, due to data constraints, terrain and environmental factors that may impact the microphysical processes of raindrops were not considered in this research. Limitations of the modeling process should be acknowledged as important parameters such as elevation, temperature,

pressure, and aerosol conditions, among others were not taken into account. Additionally, radar data can also introduce uncertainty into the *E* estimation, especially in the spatial analysis, where the calculated *E* still has many gap regions in space due to the sensitivity of radar scanning, clutter and beam blockage interference, especially over complex terrain. These limitations could impact the accuracy and applicability of the model under different conditions. Further research need broaden the dataset (including radar and disdrometer data), integrate supplementary environmental factors, and account for extended time scales, in order to optimize the functionality of models, and further, to eliminate the gap region of radar using suitable means of spatial interpolation. Additionally, future studies will consider incorporating more existing regression methods and other machine learning techniques to enhance the performance and robustness of the model.

VI. Conclusion

E plays a vital role in comprehending the interaction between rainfall and various surface elements. This study presents the first comprehensive method for estimating large-scale *E* using dual-polarization parameters. Results were compared with empirical regressional relationship and applied to actual radar data. The primary objectives of this research were to resolve two key issues. First, empirical relationships rely on localized observations from disdrometers, hindering estimation of *E* on a larger scale. Second, empirical relationships that are based on rainfall intensity demonstrate geographic limitations, making them less useful in regions with unique terrain or meteorological conditions. The following outlines the specific conclusions.

1. The relationship between dual-polarization parameters and *E* was explored in this study, with Pearson coefficients of 0.70 for Z_H, 0.36 for Z_{DR}, and 0.94 for K_{DP}.
2. The results of the TRM based on regression formulas at the Holmes Moss station show instances of overestimated *E*, with Pearson correlation coefficients falling below 0.80 compared to disdrometer results.
3. The results based on the RF models outperform the TRM in terms of Pearson and error indices. All the Pearson values with disdrometer results for RF are not lower than 0.97, and RF1 using the K_{DP} performing slightly better than RF2.
4. RF exhibits superior spatial extension capabilities compared to TRM, with Pearson values of 0.73 and 0.97, respectively.
5. In the validation of real radar, the Pearson values of daily *E* estimated from radar data based on TRM and RF are 0.85 and 0.92, respectively, indicating that RF is better than TRM in the case of using real radar.

By revealing a strong association between dual-polarization parameters and *E*, the research has advanced our understanding of the intricate dynamics governing rainfall processes. Furthermore, the demonstrated superior performance of the DREE-RF over the traditional approaches underscores the potential for improved *E* estimation. This model has great potential for use in hydrological modeling, flood forecasting, and agricultural planning, and significantly advance the understanding of precipitation processes in the context of climate change.

References

[1] Q. Dai *et al.*, "Radar remote sensing reveals potential underestimation of rainfall erosivity at the global scale," *Science Advances*, vol. 9, no. 32, p. eadg5551, Aug. 2023, doi: 10.1126/sciadv.adg5551.
[2] D. T. Meshesha, A. Tsunekawa, M. Tsubo, N. Haregeweyn, and F. Tegegne, "Evaluation of kinetic energy and erosivity potential of simulated rainfall using Laser Precipitation Monitor," *CATENA*, vol. 137, pp. 237–243, Feb. 2016, doi: 10.1016/j.catena.2015.09.017.
[3] Q. Dai, J. Zhu, S. Zhang, S. Zhu, D. Han, and G. Lv, "Estimation of rainfall erosivity based on WRF-derived raindrop size distributions," *Hydrology and Earth System Sciences*, vol. 24, no. 11, pp. 5407–5422, Nov. 2020, doi: 10.5194/hess-24-5407-2020.
[4] K. G. Renard, *Predicting Soil Erosion by Water: A Guide to Conservation Planning with the Revised Universal Soil Loss Equation (RUSLE)*. U.S. Department of Agriculture, Agricultural Research Service, 1997.
[5] USDA-Agricultural Research Service, *RUSLE2 Science Documentation*. Washington, D.C., 2013.
[6] L. Gutiérrez-Loza *et al.*, "Air–sea CO2 exchange in the Baltic Sea—A sensitivity analysis of the gas transfer velocity," *Journal of Marine Systems*, vol. 222, p. 103603, Oct. 2021, doi: 10.1016/j.jmarsys.2021.103603.
[7] D. T. Ho, L. F. Bliven, R. Wanninkhof, and P. Schlosser, "The effect of rain on air-water gas exchange," *Tellus B*, vol. 49, no. 2, pp. 149–158, 1997, doi: 10.1034/j.1600-0889.49.issue2.3.x.
[8] S. Alt and A. Kollar, "Hydrodynamics of raindrop impact stimulate ascospore discharge of Venturia inaequalis," *Fungal Biology*, vol. 114, no. 4, pp. 320–324, Apr. 2010, doi: 10.1016/j.funbio.2010.01.009.
[9] T. Gilet and L. Bourouiba, "Fluid fragmentation shapes rain-induced foliar disease transmission," *Journal of The Royal Society Interface*, vol. 12, no. 104, p. 20141092, Mar. 2015, doi: 10.1098/rsif.2014.1092.
[10] S. W. Park, J. K. Mitchell, and G. D. Bubenzer, "Splash erosion modeling: physical analyses," *Transactions of the ASAE*, vol. 25, no. 2, pp. 357–361, 1980.
[11] J. Nyssen *et al.*, "Rainfall erosivity and variability in the Northern Ethiopian Highlands," *Journal of Hydrology*, vol. 311, no. 1, pp. 172–187, Sep. 2005, doi: 10.1016/j.jhydrol.2004.12.016.
[12] C. E. Carter, J. D. Greer, H. J. Braud, and J. M. Floyd, "Raindrop Characteristics in South Central United States," *Transactions of the ASAE*, vol. 17, no. 6, pp. 1033–1037, 1974.
[13] D. T. Meshesha, A. Tsunekawa, M. Tsubo, N. Haregeweyn, and E. Adgo, "Drop size distribution and kinetic energy load of rainfall events in the highlands of the Central Rift Valley, Ethiopia," *Hydrological Sciences Journal*, vol. 59, no. 12, pp. 2203–2215, Dec. 2014, doi: 10.1080/02626667.2013.865030.
[14] W. H. Wischmeier and D. D. Smith, "Rainfall energy and its relationship to soil loss," *Eos, Transactions American Geophysical Union*, vol. 39, no. 2, pp. 285–291, 1958, doi: 10.1029/TR039i002p00285.
[15] L. C. Brown and G. R. Foster, "Storm erosivity using idealized intensity distributions," *Transactions of the ASAE*, vol. 30, no. 2, pp. 379–386, 1987.
[16] C. J. Rosewell, "Rainfall Kinetic Energy in Eastern Australia," *Journal of Climate & Applied Meteorology*, vol. 25, no. 11, pp. 1695–1701, Nov. 1986, doi: 10.1175/1520-0450(1986)025<1695:RKEIEA>2.0.CO;2.
[17] A. A. Fenta *et al.*, "Improving satellite-based global rainfall erosivity estimates through merging with gauge data," *Journal of Hydrology*, vol. 620, p. 129555, May 2023, doi: 10.1016/j.jhydrol.2023.129555.
[18] P. Panagos *et al.*, "Projections of soil loss by water erosion in Europe by 2050," *Environmental Science & Policy*, vol. 124, pp. 380–392, Oct. 2021, doi: 10.1016/j.envsci.2021.07.012.
[19] P. Panagos *et al.*, "Global rainfall erosivity assessment based on high-temporal resolution rainfall records," *Sci Rep*, vol. 7, no. 1, Art. no. 1, Jun. 2017, doi: 10.1038/s41598-017-04282-8.
[20] Y. Xie, S. Yin, B. Liu, M. A. Nearing, and Y. Zhao, "Models for estimating daily rainfall erosivity in China," *Journal of Hydrology*, vol. 535, pp. 547–558, Apr. 2016, doi: 10.1016/j.jhydrol.2016.02.020.
[21] J. Zhu, S. Zhang, Q. Yang, Q. Shen, L. Zhuo, and Q. Dai, "Comparison of rainfall microphysics characteristics derived by numerical weather prediction modelling and dual-frequency precipitation radar," *Meteorological Applications*, vol. 28, no. 3, p. e2000, 2021, doi: 10.1002/met.2000.
[22] J. Wang, L. Zhuo, D. Han, Y. Liu, and M. A. Rico-Ramirez, "Hydrological Model Adaptability to Rainfall Inputs of Varied Quality,"

Water Resources Research, vol. 59, no. 2, p. e2022WR032484, 2023, doi: 10.1029/2022WR032484.
[23] V. N. Bringi, M. A. Rico-Ramirez, and M. Thurai, "Rainfall Estimation with an Operational Polarimetric C-Band Radar in the United Kingdom: Comparison with a Gauge Network and Error Analysis," *Journal of Hydrometeorology*, vol. 12, no. 5, pp. 935–954, Oct. 2011, doi: 10.1175/JHM-D-10-05013.1.
[24] Q. Dai, Q. Yang, D. Han, M. A. Rico-Ramirez, and S. Zhang, "Adjustment of Radar-Gauge Rainfall Discrepancy Due to Raindrop Drift and Evaporation Using the Weather Research and Forecasting Model and Dual-Polarization Radar," *Water Resources Research*, vol. 55, no. 11, pp. 9211–9233, 2019, doi: 10.1029/2019WR025517.
[25] Q. Dai and D. Han, "Exploration of discrepancy between radar and gauge rainfall estimates driven by wind fields," *Water Resources Research*, vol. 50, no. 11, pp. 8571–8588, 2014, doi: 10.1002/2014WR015794.
[26] Y. Song, D. Han, and J. Zhang, "Radar and rain gauge rainfall discrepancies driven by changes in atmospheric conditions," *Geophysical Research Letters*, vol. 44, no. 14, pp. 7303–7309, 2017, doi: 10.1002/2017GL074493.
[27] V. N. Bringi, V. Chandrasekar, J. Hubbert, E. Gorgucci, W. L. Randeu, and M. Schoenhuber, "Raindrop Size Distribution in Different Climatic Regimes from Disdrometer and Dual-Polarized Radar Analysis," *Journal of the Atmospheric Sciences*, vol. 60, no. 2, pp. 354–365, Jan. 2003, doi: 10.1175/1520-0469(2003)060<0354:RSDIDC>2.0.CO;2.
[28] M. Le and V. Chandrasekar, "Raindrop Size Distribution Retrieval From Dual-Frequency and Dual-Polarization Radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 5, pp. 1748–1758, May 2012, doi: 10.1109/TGRS.2011.2167683.
[29] Y. Wang, J. Zhang, P.-L. Chang, C. Langston, B. Kaney, and L. Tang, "Operational C-Band Dual-Polarization Radar QPE for the Subtropical Complex Terrain of Taiwan," *Advances in Meteorology*, vol. 2016, p. e4294271, Dec. 2015, doi: 10.1155/2016/4294271.
[30] V. N. Bringi, G.-J. Huang, V. Chandrasekar, and E. Gorgucci, "A Methodology for Estimating the Parameters of a Gamma Raindrop Size Distribution Model from Polarimetric Radar Data: Application to a Squall-Line Event from the TRMM/Brazil Campaign," *Journal of Atmospheric and Oceanic Technology*, vol. 19, no. 5, pp. 633–645, May 2002, doi: 10.1175/1520-0426(2002)019<0633:AMFETP>2.0.CO;2.
[31] M. A. Rico-Ramirez, "Adaptive Attenuation Correction Techniques for C-Band Polarimetric Weather Radars," *IEEE Trans. Geosci. Remote Sensing*, vol. 50, no. 12, pp. 5061–5071, Dec. 2012, doi: 10.1109/TGRS.2012.2195228.
[32] X. Huang *et al.*, "Attenuation correction for C-band dual-polarization radar reflectivityfactor and the accuracy improvement of precipitation estimation," *Journal of the Meteorological Sciences*, vol. 38, no. 2, pp. 237–246, Apr. 2018.
[33] J. S. Marshall and W. M. K. Palmer, "THE DISTRIBUTION OF RAINDROPS WITH SIZE," *Journal of the Atmospheric Sciences*, vol. 5, no. 4, pp. 165–166, Aug. 1948, doi: 10.1175/1520-0469(1948)005<0165:TDORWS>2.0.CO;2.
[34] M. Liao and A. P. Barros, "Toward Optimal Rainfall for Flood Prediction in Headwater Basins—Orographic QPE Error Modeling Using Machine Learning," *Water Resources Research*, vol. 59, no. 11, p. e2023WR034456, 2023, doi: 10.1029/2023WR034456.
[35] T.-S. Yo, S.-H. Su, J.-L. Chu, C.-W. Chang, and H.-C. Kuo, "A Deep Learning Approach to Radar-Based QPE," *Earth and Space Science*, vol. 8, no. 3, p. e2020EA001340, 2021, doi: 10.1029/2020EA001340.
[36] J. Zhang *et al.*, "A Dual-Polarization Radar Synthetic QPE for Operations," *Journal of Hydrometeorology*, vol. 21, no. 11, pp. 2507–2521, Oct. 2020, doi: 10.1175/JHM-D-19-0194.1.
[37] M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 114, pp. 24–31, Apr. 2016, doi: 10.1016/j.isprsjprs.2016.01.011.
[38] Q. Yang, Q. Dai, S. Zhang, K. Zhu, and L. Zhang, "Raindrop Size Distribution Retrieval Model for X-Band Dual-Polarization Radar in China Incorporating Various Climatic and Geographical Elements," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–17, 2022, doi: 10.1109/TGRS.2022.3168586.
[39] S. Zhou and L. Mentch, "Trees, forests, chickens, and eggs: when and why to prune trees in a random forest," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 16, no. 1, pp. 45–64, 2023, doi: 10.1002/sam.11594.
[40] B. S. Pickering, R. R. Neely III, and D. Harrison, "The Disdrometer Verification Network (DiVeN): a UK network of laser precipitation instruments," *Atmospheric Measurement Techniques*, vol. 12, no. 11, pp. 5845–5861, Nov. 2019, doi: 10.5194/amt-12-5845-2019.
[41] Q. Yang, Q. Dai, D. Han, Y. Chen, and S. Zhang, "Sensitivity analysis of raindrop size distribution parameterizations in WRF rainfall simulation," *Atmospheric Research*, vol. 228, pp. 1–13, Nov. 2019, doi: 10.1016/j.atmosres.2019.05.019.
[42] D. Harrison, S. Georgiou, N. Gaussiat, and A. Curtis, "Long-term diagnostics of precipitation estimates and the development of radar hardware monitoring within a radar product data quality management system," *Hydrological Sciences Journal*, vol. 59, no. 7, pp. 1277–1292, Jul. 2014, doi: 10.1080/02626667.2013.841316.
[43] Y. Wang and V. Chandrasekar, "Algorithm for Estimation of the Specific Differential Phase," *Journal of Atmospheric and Oceanic Technology*, vol. 26, no. 12, pp. 2565–2578, Dec. 2009, doi: 10.1175/2009JTECHA1358.1.
[44] H. Huang, G. Zhang, K. Zhao, and S. E. Giangrande, "A Hybrid Method to Estimate Specific Differential Phase and Rainfall With Linear Programming and Physics Constraints," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 1, pp. 96–111, Jan. 2017, doi: 10.1109/TGRS.2016.2596295.
[45] V. N. Bringi, V. Chandrasekar, N. Balakrishnan, and D. S. Zrnić, "An Examination of Propagation Effects in Rainfall on Radar Measurements at Microwave Frequencies," *Journal of Atmospheric and Oceanic Technology*, vol. 7, no. 6, pp. 829–840, Dec. 1990, doi: 10.1175/1520-0426(1990)007<0829:AEOPEI>2.0.CO;2.
[46] T. Islam, M. A. Rico-Ramirez, D. Han, and P. K. Srivastava, "Sensitivity associated with bright band/melting layer location on radar reflectivity correction for attenuation at C-band using differential propagation phase measurements," *Atmospheric Research*, vol. 135–136, pp. 143–158, Jan. 2014, doi: 10.1016/j.atmosres.2013.09.003.
[47] W. Hall, M. A. Rico-Ramirez, and S. Krämer, "Offshore wind turbine clutter characteristics and identification in operational C-band weather radar measurements," *Quarterly Journal of the Royal Meteorological Society*, vol. 143, no. 703, pp. 720–730, 2017, doi: 10.1002/qj.2959.
[48] C. W. Ulbrich, "Natural Variations in the Analytical Form of the Raindrop Size Distribution," *Journal of Climate and Applied Meteorology*, vol. 22, no. 10, pp. 1764–1775, 1983.
[49] M. N. Anagnostou, E. N. Anagnostou, J. Vivekanandan, and F. L. Ogden, "Comparison of Two Raindrop Size Distribution Retrieval Algorithms for X-Band Dual Polarization Observations," *Journal of Hydrometeorology*, vol. 9, no. 3, pp. 589–600, Jun. 2008, doi: 10.1175/2007JHM904.1.
[50] P. Barber and C. Yeh, "Scattering of electromagnetic waves by arbitrarily shaped dielectric bodies," *Appl. Opt., AO*, vol. 14, no. 12, pp. 2864–2872, Dec. 1975, doi: 10.1364/AO.14.002864.
[51] E. Gorgucci, V. Chandrasekar, V. N. Bringi, and G. Scarchilli, "Estimation of Raindrop Size Distribution Parameters from Polarimetric Radar Measurements," *Journal of the Atmospheric Sciences*, vol. 59, no. 15, Art. no. 15, Aug. 2002, doi: 10.1175/1520-0469(2002)059<2373:EORSDP>2.0.CO;2.
[52] E. Yoshikawa, V. Chandrasekar, T. Ushio, and Z. Kawasaki, "Raindrop size distribution (DSD) retrieval for X-band dual-polarization radar," *2012 IEEE International Geoscience and Remote Sensing Symposium*, Munich, Germany, 2012, pp. 2411-2414, doi: 10.1109/IGARSS.2012.6351005.
[53] A. Adachi, K. Takahisa, and Y. Hiroshi, "Estimation of raindrop size distribution and rainfall rate from polarimetric radar measurements at attenuating frequency based on the self-consistency principle," *Journal of the Meteorological Society of Japan. Ser. II* 93.3, pp. 359-388, 2015, doi: 10.2151/jmsj.2015-020.
[54] V.Mahale, G. Zhang, M. Xue, J. Gao, and H. Reeves, "Variational Retrieval of Rain Microphysics and Related Parameters from Polarimetric Radar Data with a Parameterized Operator," *Journal of Atmospheric and Oceanic Technology*, vol. 36, no. 12, pp. 2483-2500, 2015, doi: 10.1175/JTECH-D-18-0212.1 .
[55] H. Huang, K. Zhao, G. Zhang, D. Hu, and Z. Yang, "Optimized raindrop size distribution retrieval and quantitative rainfall estimation from polarimetric radar," *Journal of Hydrology*, vol. 580, pp. 124248, 2020, doi: 10.1016/j.jhydrol.2019.124248.
[56] Y. Sun, H. Xiao, H. Yang, L. Feng, H. Chen and L. Luo, "An Inverse Mapping Table Method for Raindrop Size Distribution Parameters Retrieval Using X-band Dual-Polarization Radar Observations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 11, pp. 7611-7632, Nov. 2020, doi: 10.1109/TGRS.2020.2982687.

Jingxuan Zhu received the B.S. degree from Nanjing Normal University, Nanjing, China, in 2015, where she is currently pursuing the Ph.D. degree in cartography and geographical information system.

Her research interests include rainfall estimation and prediction based on radar remote sensing and rainfall kinetic energy and rainfall erosivity estimation on large scale.

Qiang Dai (Member, IEEE) received the Ph.D. degree in civil engineering from the University of Bristol, Bristol, U.K., in 2014.

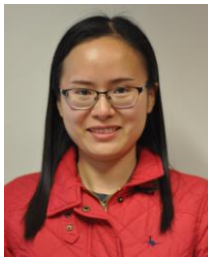
He is a currently Professor with the School of Geography, Nanjing Normal University, Nanjing, China. His research interests are meteorology, hazards risk assessment, radar hydrology, and numerical weather prediction model. His key expertise within the center exists in meteorology, hazards, radar hydrology, and numerical weather prediction model.

Yuanyuan Xiao received the B.S. degree from Jiangxi Normal University, Nanchang, China, in 2022.

She is currently pursuing the Ph.D. degree in cartography and geographical information system from Nanjing Normal University, Nanjing, China. Her research interests include the relationship between rainfall and aerosols.

Jun Zhang received the Ph.D. degree in civil engineering from the University of Bristol, Bristol, U.K., in 2017.

She is currently an Associate Professor with the School of Geography, Nanjing Normal University, Nanjing, China. Her research interests are the simulation of water cycle in multiple scales and resolutions, groundwater simulations, urban flood.



Lu Zhuo received the M.Eng. and Ph.D. degrees in civil engineering from the University of Bristol (UoB), Bristol, U.K., in 2011 and 2016, respectively.

She is currently working as a Lecturer with Cardiff University, Cardiff, UK. Her research interests include multiple natural hazards modelling and monitoring (e.g., floods, landslides, and earthquake), remote sensing of environment, and disaster risk management.

Dawei Han was born in Tianjin City, China, in 1961. He received the B.Eng. and M.Sc. degrees in water conservancy from the North China University of Water Conservancy and Electric Power, Zhengzhou, China, in 1982 and 1984, respectively, and the Ph.D. degree in radar hydrology from the University of Salford, Salford, U.K., in 1991.

He is currently a Professor of hydroinformatics with the Department of Civil Engineering, Uniersity of Bristol, Bristol, U.K. His research interests include hydroinformatics, real-time flood forecasting, flood risk management, remote sensing and geographic information system, natural hazards, and water resources management.