



## Article

# Segmentation of PMSE Data Using Random Forests

Dorota Jozwicki <sup>1,\*</sup> , Puneet Sharma <sup>2</sup> , Ingrid Mann <sup>1</sup> and Ulf-Peter Hoppe <sup>1</sup>

<sup>1</sup> Department of Physics and Technology, UiT the Arctic University of Norway, 9019 Tromsø, Norway; ingrid.b.mann@uit.no (I.M.); ulf-peter.hoppe@uit.no (U.-P.H.)

<sup>2</sup> Department of Automation and Process Engineering, UiT the Arctic University of Norway, 9019 Tromsø, Norway; puneet.sharma@uit.no

\* Correspondence: dorota.s.jozwicki@uit.no

**Abstract:** EISCAT VHF radar data are used for observing, monitoring, and understanding Earth's upper atmosphere. This paper presents an approach to segment Polar Mesospheric Summer Echoes (PMSE) from datasets obtained from EISCAT VHF radar data. The data consist of 30 observations days, corresponding to 56,250 data samples. We manually labeled the data into three different categories: PMSE, Ionospheric background, and Background noise. For segmentation, we employed random forests on a set of simple features. These features include: altitude derivative, time derivative, mean, median, standard deviation, minimum, and maximum values corresponding to neighborhood sizes ranging from 3 by 3 to 11 by 11 pixels. Next, in order to reduce the model bias and variance, we employed a method that decreases the weight applied to pixel labels with large uncertainty. Our results indicate that, first, it is possible to segment PMSE from the data using random forests. Second, the weighted-down labels technique improves the performance of the random forests method.

**Keywords:** space physics; upper atmosphere; random forests; segmentation



**Citation:** Jozwicki, D.; Sharma, P.; Mann, I.; Hoppe U.-P. Segmentation of PMSE Data Using Random Forests. *Remote Sens.* **2022**, *14*, 2976. <https://doi.org/10.3390/rs14132976>

Academic Editor: Michael E. Gorbunov

Received: 26 May 2022

Accepted: 16 June 2022

Published: 22 June 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Polar Mesospheric Summer Echoes (PMSE) are radar echoes that form at about 75 to 95 km altitude at polar latitudes during the summer months. A recent long-term study of observations at 53.5 MHz made over two decades at 69 N (and many others) showed that they appear between mid-May and the end of August and are most likely to appear in June and July with an average occurrence of 95 percent [1].

Formation of PMSE requires the presence of turbulence, free electrons and charged aerosols. The charged aerosols themselves contain water ice and require the presence of very low temperature, the adequate water vapor [1–3], and nucleation centers to facilitate heterogeneous condensation. Meteor Smoke Particles (MSP) have been identified as the likely condensation nuclei. They result from meteor ablation and recondensation. In addition to them, the water vapor and the cold temperature at mid and high latitudes at the mesopause during the summer months allow the ice particles to form [4]. The combination of neutral air turbulence and the effect of negatively charged ice particles result in irregularities in the electron density distribution which generates the observed radar echoes, see, e.g., [1].

PMSE and Noctilucent Clouds (NLC) are observed during a similar time of the year and at similar heights and observations showed that the NLC tended to appear at the bottom of PMSE [2]. PMSE and NLC have the potential to reveal details about the atmosphere, including many changes during recent decades. An increase of NLC occurrence over the years has already been noticed in observations from 1964 to 1988 [5] and one could argue that climate change may have reached the edge of space. To better understand this, systematic studies of PMSE over time can be helpful because they reveal the existence of water ice particles at the height where they are observed.

We aim to develop a method to investigate the thickness of PMSE, their shape, and the variation of PMSE height with time over the years. This requires the classification

and segmentation of the radar images that contain PMSE. For our analysis, we employ random forests as a segmentation method. Next, we use a technique already developed by others [6] to select labeled data points. In this paper, we focus on segmenting PMSE signal in images of VHF (Very High Frequency) radar data from EISCAT (where EISCAT is the name of the research organization that operates the radar). The EISCAT VHF radar is located in Tromsø, Northern Norway and operates in the 224 MHz band. These images are time-height color plots of the radar data. For simplicity, we will employ the term “image” throughout this paper when referring to those plots. The parameter shown in all these images is the equivalent electron density, i.e., the number density of free electrons which would create an equivalent radar signal if it were due to incoherent scatter [7]. The equivalent electron density is very nearly proportional to the signal-to-noise ratio employed by many publications with PMSE observations [e.g., [8,9]]. For the purpose of this study, we have no use for any absolute value of PMSE strength. Therefore, we do not calibrate the data with all the steps which would be necessary to obtain an absolute radar reflectivity [10].

The paper is structured as follows: In Section 2, we outline the theory associated with the method used for segmentation of the PMSE signal and the metrics we used to evaluate its performance. In Section 3, we describe the procedure used for the segmentation of the PMSE signal. Section 4 highlights the results obtained from our analysis. In Section 5, we describe the results in detail and discuss key challenges. Finally, in Section 6, we outline the conclusions based on this study.

## 2. Theory

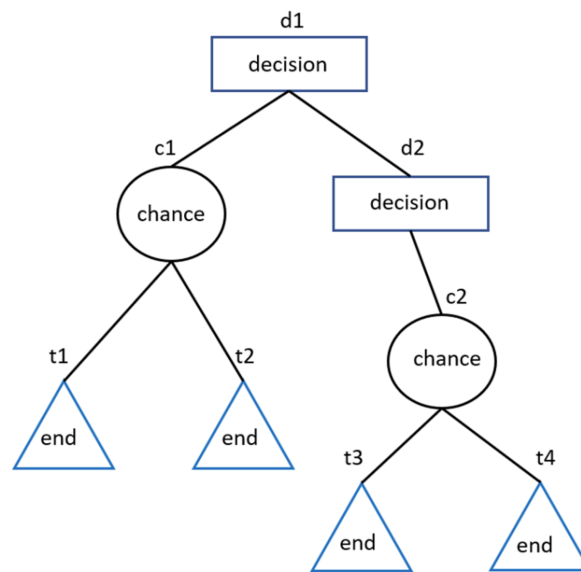
### 2.1. Random Forests

A way to characterize and segment data is to use a decision tree. A decision tree is represented as a directed graph  $G = (V, E)$ ,  $E \subset V^2$ , where  $V$  is a finite set of nodes split into three disjoint sets  $V = D \cup C \cup T$ , where  $D$  is decision node,  $C$  is chance node, and  $T$  is terminal or end node [11]. The different nodes represent different phases of a decision problem sequence [11]. In a decision node, based on observations about an item, we select an action. In Figure 1, there are two edges  $(d_1, c_1)$  and  $(d_1, d_2)$  originating from decision  $d_1$  and one of edges lead to another decision node  $d_2$ . In a chance node which represents the probability of an outcome, we again select an edge randomly. In Figure 1, there are two edges  $(c_1, t_1)$  and  $(c_1, t_2)$  originating from chance node  $c_1$  and two edges  $(c_2, t_3)$  and  $(c_2, t_4)$  originating from chance node  $c_2$ . Terminal or end nodes  $(t_1, t_2, t_3, \text{ and } t_4)$  represent the outcome of a sequence of actions. For instance, it could be an items target value (regression) or its category (classification).

Several decision tree algorithms exist in the literature. Algorithms such as: ID3 [12], C4.5 [13] and C5.0 [14] employ information gain (which uses the concept of entropy) for deciding on the features to use for split at each step in building the tree, whereas another decision tree algorithm, e.g., CART [15] uses Gini impurity for the splitting criterion.

Decision tree-based methods are easy to understand and interpret; however, they are not robust. For instance, a small change in data or noise in the features can lead to large change in the tree and their associated outputs [16]. This implies that decision tree might not generalize well for unseen data.

Random forests is a decision tree-based ensemble learning method that has several advantages, such as having a built-in estimate of generalization error, depending only on one or two tuning parameters, and providing a measure of importance of different features of data [17].



**Figure 1.** Example of a decision tree with decision, chance, and terminal or end nodes.

Random forests use Breiman's bootstrap aggregation or bagging technique in which several individual decision trees are trained on different subsets of the training dataset, also known as random sampling with replacement [17]. Furthermore, random forests use random subsets of available features for building the individual trees, also known as feature bagging. In the study by Probst et al. [18], it is suggested that the number of features to be randomly selected (*mtry*) for classification tasks usually have a default value of  $\sqrt{p}$ , where  $p$  is the total number of features. However, *mtry* can be increased from its default value to improve the probability that at least one of randomly selected features is a strong predictor [18,19].

For regression, generalization error is measured using out-of-bag mean square error as:

$$MSE_{oob} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{f}_{oob}(x_i))^2, \quad (1)$$

where  $\hat{f}_{oob}(x_i)$  is out-of-bag prediction for bootstrap sample  $i$ ,  $y_i$  is its actual outcome [17], and  $N$  is the number of samples. For classification, generalization error is measured using the out-of-bag error rate as:

$$E_{oob} = \frac{1}{N} \sum_{i=1}^N (I(y_i \neq \hat{f}_{oob}(x_i))), \quad (2)$$

i.e., it assigns 0 to error for a correct classification and 1 for an incorrect classification [17]. For example, if a PMSE sample is misclassified as a Noise sample, 1 is added to the sum in the out-of-bag error rate. If the PMSE sample is correctly classified as PMSE, 0 is added to the sum in the out-of-bag error rate. When we apply these equations to our data,  $x_i$  corresponds to a given sample  $i$  of our dataset. Then,  $f(x_i)$  is the predicted label by the model for this given sample  $i$ , while  $y_i$  is the actual label of this sample.  $\hat{f}_{oob}(x_i)$  means that we consider the predicted label by the model for the sample  $x_i$  in the out-of-bag dataset. Finally,  $I(y_i \neq \hat{f}_{oob}(x_i))$  is a measure of how close  $y_i$  is to  $\hat{f}_{oob}(x_i)$ . It is a loss function defined to minimize the expected value of the loss. Bootstrapping ensures that individual decision trees are unique, which reduces the overall variance of the random forests method [20]. Finally, the prediction is obtained by aggregating the decisions of

individual trees in the case of regression, or by taking the majority vote in the case of classification.

The importance of features is calculated based on the permutation importance method proposed by Breiman [21]. It is calculated as follows: first, we use out-of-bag samples for estimating the predictions from each tree with a selected feature  $f$  [17]. Second, the feature  $f$  is randomly permuted in out-of-bag samples and their predictions are calculated. Third, we calculate the difference between the prediction scores for the permuted and the original. Fourth, the average of differences over all trees within the random forest is an estimate of importance of the feature  $f$ .

Random forests methods are fast, simple, and easy to interpret via permutation importance [22]. They have been used in several applications such as: pattern recognition [23], object detection [24], remote sensing [25], and image segmentation [26].

## 2.2. Weighted-Down Technique

In their paper [6], Almeida et al. propose a novel method that can reduce both model bias and model variance. In their method, first, estimation of the pixel-wise label uncertainty of training data is performed. For instance, given a sample  $x_i$  with label  $y_i$ , its neighborhood uncertainty score is calculated as:

$$b_{x_i} = \frac{-C * \left(\frac{k_{y_i}}{k} \ln\left(\frac{k_{y_i}}{k} * \frac{k_{y_i}}{\sum d_{x_i}}\right)\right)}{-\sum_{j=1}^C \left(\frac{k_j}{k} \ln\left(\frac{k_j}{k} * \frac{k_j}{\sum d_j}\right)\right)}, \quad (3)$$

where  $C$  is the number of classes,  $k$  is the number of neighbors we consider for each sample (we employed  $k = 11$  for our experiments),  $k_{y_i}$  is the number of neighbors with same label as  $x_i$ ,  $k_j$  is the number of neighbors with class label  $y_j$ , and  $d_{x_i}$  represents a vector with normalized distances to the  $k_{y_i}$  neighbors with the same class label as  $x_i$ . For more details on the significance of the terms in the numerator and denominator of Equation (3), we refer the reader to the original paper [6]. Next, the training sample weights are adjusted such that the samples with high uncertainty are weighted-down and those with low uncertainty are weighted-up.

## 2.3. Metrics for Evaluation of Performance

In this section, we briefly discuss the different metrics used for evaluating the performance of methods used for data segmentation.

### 2.3.1. Classification Error

The classification error  $E$  is defined as the ratio of the number of misclassified samples, i.e., sum of the False Positives and False Negatives to that of the total number of samples. The values of classification error are in the range  $[0,1]$ , where values closer to zero indicate fewer misclassifications, hence better performance.

### 2.3.2. Logarithmic Loss

Logarithmic loss  $L$  is based on the predicted class probabilities and is considered to be a more refined metric than classification error [18,27]. Logarithmic loss is defined as:

$$L = \sum_{j=1}^n w_j \log(1 + e^{-m_j}), \quad (4)$$

where  $n$  is number of samples, the weight for observation  $j$  is  $w_j$  and the weights are normalized to sum to 1, and  $m_j$  is scalar classification score [28].

### 2.3.3. Area Under ROC Curve (AUC)

A receiver operating characteristic (ROC) curve is a simple and visual way to summarize the performance of a classifier [29]. Assuming a two class prediction problem where

the output is either positive or negative, an ROC curve is created by plotting true positive rate against false positive rate.

For the test samples, true positive rate is defined as the ratio of number of correct positive outcomes to that of all the positive samples [29]. The false positive rate is defined as the ratio of number of incorrect positive outcomes to that of all negative samples [29]. Finally, the area under the ROC Curve (AUC) gives us a scalar value in the range [0,1], which is used to measure the performance of a classifier. A random guessing classifier can give an AUC of 0.5; hence, any realistic classifier should have an AUC value more than 0.5 [29].

### 3. Method

In this section, we discuss the dataset used for our experiments. Next, we briefly explain the labeling procedure and the weighted-down technique employed in this study. After that, we explain the options used for random forests and finally the feature extracted from the data.

#### 3.1. Dataset

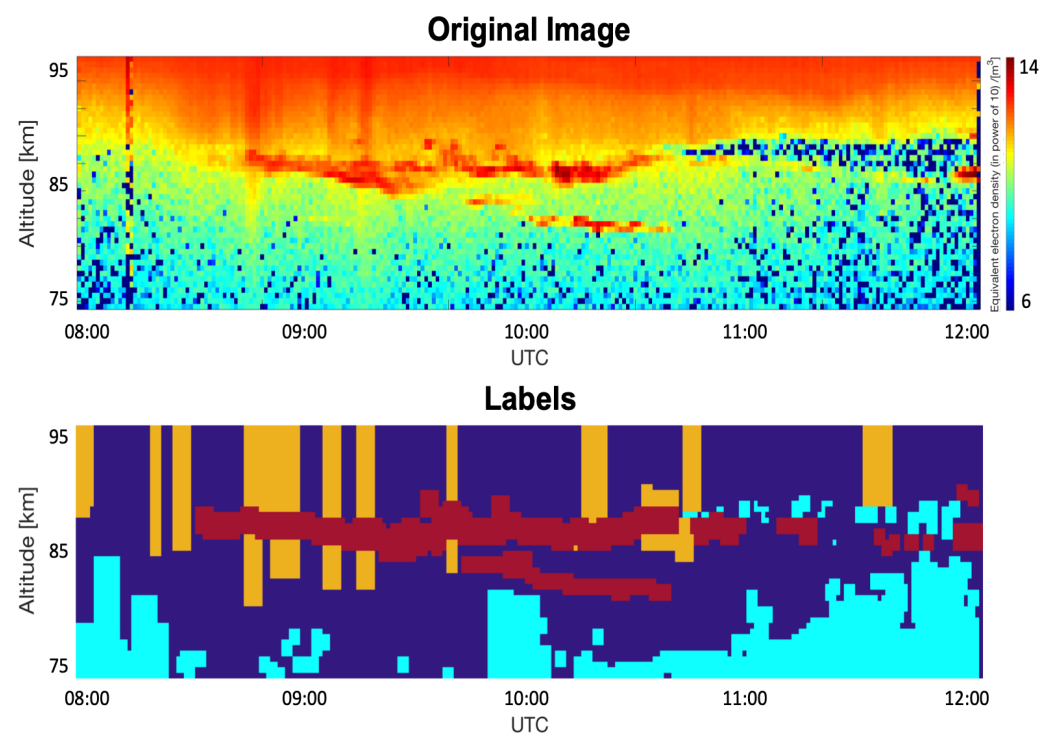
The data we used for the analysis comes from EISCAT VHF radar located near Tromsø, Norway. The images contain measured backscattered power as a function of altitude and time. We use a height range of 75 to 95 km for our analysis, and observations typically last several hours. The height resolution is 0.30 to 0.45 km and the time resolution is of approximately one minute. We downloaded the data written in ASCII format from the Madrigal website. For further information about the 30 observation days used in this study, their dates and times are listed in Table 1.

**Table 1.** This table lists the observation dates, given in (day. month. year), that are used for this study. The observation time is given in Universal time (UTC) in the format (hour: minutes: seconds).

Dates	Start Time in UTC	End Time in UTC
(dd:mm:yyyy)	(hh:mm:ss)	(hh:mm:ss)
28 June 2008	07:58:33	08:36:18
30 June 2008	07:59:38	12:07:30
02 July 2008	10:24:30	11:59:02
10 June 2009	09:03:42	11:56:09
14 July 2009	08:19:33	11:33:15
16 July 2009	08:47:30	10:06:26
17 July 2009	07:49:44	11:59:30
30 July 2009	12:15:29	15:59:08
06 July 2010	07:00:30	23:59:30
07 July 2010	00:00:30	21:59:27
08 July 2010	09:00:42	12:59:03
09 July 2010	09:00:24	12:59:09
01 June 2011	08:34:31	10:02:07
08 June 2011	07:23:50	13:01:07
09 June 2011	08:01:45	12:59:26
12 June 2012	07:13:31	23:59:28
29 June 2012	10:21:57	10:30:04
11 July 2012	07:54:57	13:09:40
13 June 2013	07:12:33	08:59:26
28 June 2013	07:02:43	12:58:28
12 July 2013	00:00:28	21:58:28
27 July 2013	08:56:36	13:05:14
27 June 2014	09:03:48	12:59:38
01 July 2014	09:00:36	13:00:24
22 July 2014	22:26:33	23:59:28
23 July 2014	00:00:28	09:26:28
10 August 2015	09:14:40	16:12:28
12 August 2015	20:04:40	23:59:28
13 August 2015	00:00:28	01:59:26
20 August 2015	00:00:28	01:59:26

### 3.2. Labeling

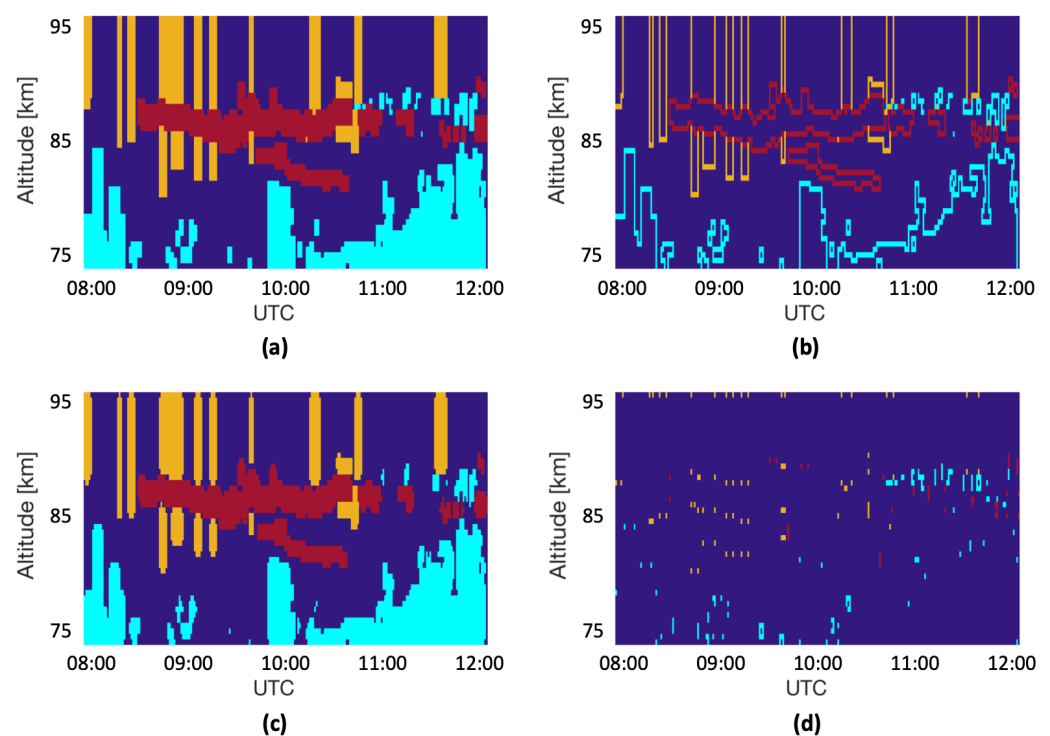
We labeled the data manually and pixel by pixel, using the built-in Matlab Image Labeler App. In this way, a given labeled pixel belongs to one of three classes of interest, namely, PMSE, Ionospheric background, and Background noise. The regions of interest that are considered in this paper are discussed in more detail in our previous work [30]. The PMSE is characterized by a region where coherent scattering occurs, whereas the Ionosphere is a region where incoherent scattering occurs. The Noise region also displays incoherent scatter but because the signal is low the region has a lot of missing values (NaNs). As a result, it makes this region look different compared to the Ionosphere. The labeling was performed by visually recognizing a PMSE pattern. We based this on the fact that the amplitude of the PMSE looks greater than its surroundings, and it has a particular wavy structure that makes it different from the background. Figure 2 shows an example of the manual labeling process for a given image. We represented the original image as a heatmap, where the blue pixels represent the minimum values, the red pixels represent the maximum values, and the other colored pixels represent the values in between minimum and maximum. We use this same color code also later in this paper. This refers to the equivalent electron density from the standard GUISDAP analysis [31]. As for the labels part of Figure 2, the cyan colored pixels represent the Background noise, the yellow colored pixels belong to the Ionospheric background, and the dark red colored pixels represent the PMSE. Finally, the dark blue colored pixels represent unlabeled data. We partially labeled 18 images out of the 30, that contain a total of 56,250 samples (pixels). We used 60 percent of the labeled data for training (33,750 samples) and 40 percent for quantitative testing (22,500 samples). For qualitative testing, we used all the images. In addition, qualitative analysis was made by visual inspection of the segmented images by a domain expert.



**Figure 2.** An example for manual labeling of the observations from 30 June 2008. The altitude range varies from 75 km to 95 km. The observation starts at 08:00 UTC and finishes at 12:00 UTC. The color scale in the original image represents the equivalent electron density to the power of 10, per cubic meter. In the labels part, the cyan, yellow and dark red color represent, respectively, the Background noise, Ionospheric background, and PMSE classes. The dark blue color represents unlabeled data.

### 3.3. Labels with Reduced Weighting

Next, we use the weighted-down technique described in Section 2.2 that aims to reduce both model bias and variance by reducing the weighting for pixel labels with large uncertainty and increasing it for labels with small uncertainty. For this, as a first step on the manually labeled data, for instance as shown in Figure 3a, we apply edge erosion to obtain a set of pixel labels that should be given lesser weight. As the labels we used in our experiment do not overlap, the edge erosion step generates a set of pixel labels along the label boundaries, as shown in Figure 3b. Finally, we calculate uncertainty scores based on Equation (3) in Section 2.2 for these pixel labels, and any pixel labels with non zero uncertainty scores are not kept for further analysis. Figure 3c shows the pixel labels after removing pixel labels with uncertainty, and Figure 3d shows the pixel labels that are removed from further analysis.



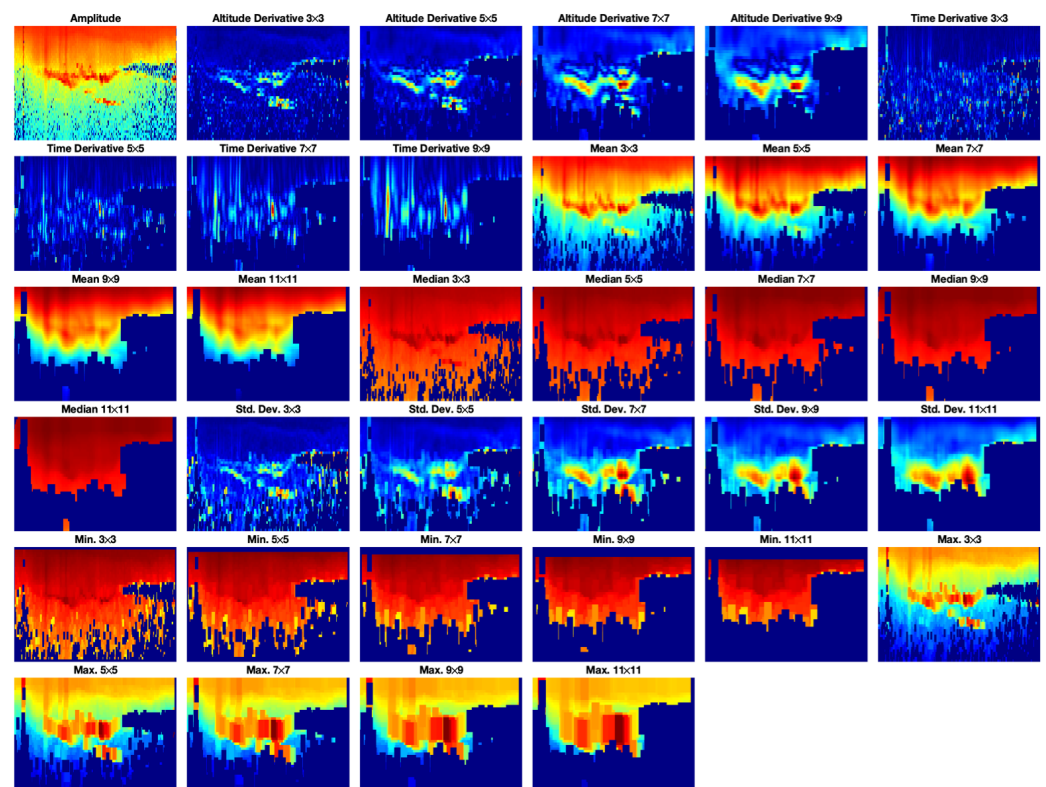
**Figure 3.** Preprocessing step with the labels: (a) shows the original labels, (b) shows the contours of the labels, (c) shows the image after removing the labeled pixels using weighted-down labels technique, and (d) shows the removed labeled pixels. For all four images, the red colored pixels belong to the PMSE class, the yellow pixels represent the Ionospheric background, and the cyan pixels belong to the Background noise class. Finally, the dark blue pixels represent unlabeled data. All of the plots have the same axes: the horizontal axis represents the time which starts at 8:00 UTC and finishes at 12:00 UTC. The vertical axis represents the altitude which ranges from 75 km to 95 km. The chosen observation day is 30 June 2008.

### 3.4. Random Forests Application

For our experiments, we employ random forests for training and evaluating the performance of data segmentation. For this we use MATLAB. The training is performed on an ensemble of bagged classification trees where number of trees in the forest is 500. In line with the study by Probst et al. [18] the number of trees is kept high, i.e., 500, samples are drawn with replacement, and  $p$ -value is used as the splitting rule. In addition, we enable surrogate decision split in order to allow the random forests to make a decision in case of missing data. This is done to accommodate for instances where we obtain missing amplitude values, i.e., NaNs in the data.

### 3.5. Feature Extraction

For each pixel, we extracted a dataset of features which is used as input together with its label. For each pixel we calculate features such as: mean, standard deviation, median, minimum and maximum values associated with neighborhoods ranging in sizes  $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$ ,  $9 \times 9$ , and  $11 \times 11$  pixels where the pixel is at the center. In addition, we compute vertical and horizontal Gradient magnitudes using Sobel kernels for filter sizes  $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$ , and  $9 \times 9$ , see [32]. Horizontal gradient operators calculate the time derivatives, and vertical gradient operators calculate the altitude derivatives. Furthermore, for each pixel its altitude and amplitude are included as features. This generates a feature vector with 35 dimensions. We plotted the different features extracted from the data in Figure 4 for the given observation day, 30 June 2008, except for altitude which is not illustrated in this figure. In the figure, the image in the first row and column represents the normalized amplitude. Then, from left to right and from top to bottom, the four next images represent the vertical gradient magnitudes for filter sizes of  $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$ , and  $9 \times 9$  pixels. The following four images represent the horizontal gradient magnitudes for the same filter sizes. The next five images represent the mean values for filters filter sizes of  $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$ ,  $9 \times 9$ , and  $11 \times 11$  pixels. In a similar way, we represent the median values, the standard deviation, the minimum values, and finally the maximum values for the same filter sizes ranging from  $3 \times 3$  to  $11 \times 11$  pixels.



**Figure 4.** Visualization of the different features extracted from the data (with the exception of altitude) in the form of heat maps where red pixels correspond to highest values, blue pixels correspond to lowest values, and yellow pixels represent the values in between. The rows in the figure represent the different features, and the columns represent the different filter sizes that were used. All 34 plots have the same axes: the horizontal axis represents the time which starts at 8:00 UTC and finishes at 12:00 UTC. The vertical axis represents the altitude which ranges from 75 km to 95 km. The observation day chosen in this figure is 30 June 2008.



#### 4. Results

The performance of random forests based on segmentation methods using different combinations of features, i.e., filter sizes, number of randomly selected features  $mtry$ , can be seen in Tables 2 and 3. While the results from using original labels are outlined in Table 2, the results we obtained using weighted-down labels (discussed in Section 3.3) are shown in Table 3. We evaluate the performance in terms of classification error for the test data, classification error for the out-of-bag (OOB) data samples, logarithmic error for the test data, logarithmic error for the out-of-bag (OOB) samples, Area under the ROC Curves (AUCs) for PMSE, Ionospheric background, and Noise. Note that for each filter size in the table, we use the following features: altitude derivative, time derivative, mean, median, standard deviation, minimum, and maximum. Additionally, we also use other features such as altitude and amplitude. This means that for each filter size, the feature vector has nine dimensions. At the filter size  $11 \times 11$ , the gradient filter is  $9 \times 9$ . For one filter size, e.g.,  $3 \times 3$  we use  $mtry = 3, 6, 9$ . After that, we use all filter sizes and select  $mtry = 5, 10, 15, 20, 25, 30$  and  $35$ , where  $35$  is the total length of the feature vector obtained from using all filter sizes.

The results for original labels can be seen in in Table 2 and the results for weighted-down labels can be found in Table 3. In both cases, the logarithmic error and logarithmic error OOB have the best performance for a filter size of  $7 \times 7$  pixels with  $mtry = 9$ , and the worst one for all filter sizes and an  $mtry = 5$ . The classification error and the classification error OOB have the best results for the filter size  $3 \times 3$  associated with  $mtry = 3$  for original labels, or the filter size  $7 \times 7$  associated with  $mtry = 9$  and the filter size  $11 \times 11$  with  $mtry = 6$  for the weighted-down labels. The worst performance on the other hand was obtained in the case using all filter sizes and  $mtry = 5$ , and the filter size  $9 \times 9$  in combination with  $mtry = 3$ . Almost all the AUC curves had the best performance for the combination of the filter size  $6 \times 6$  with an  $mtry = 6$ . The only exception is the AUC Ion. Back. metric with original labels, for which the best performance was obtained for a filter size of  $5 \times 5$  pixels with  $mtry = 6$ . We see that for all filter sizes, i.e., 35-dimensional feature vector and  $mtry = 5, 10, 15, 20$ , there are slight improvements in the performance for the scores associated with different evaluation metrics. This can possibly indicate that the performance of the random forests algorithm benefits from multi-resolution features extracted using the different filter sizes. However, the performance decreases when all the filter sizes are used and an  $mtry$  equal to 35.

Based on the results obtained in Tables 2 and 3, we choose the filter size  $7 \times 7$  and  $mtry = 9$  for qualitative analysis. We used this classification model on 30 images, out of which 12 were new data for the model. Figures 5–8 show the predicted labels for the classification model (with a filter size  $7 \times 7$  and  $mtry = 9$ ). In all four cases, the prediction of PMSE labels by the model looks poor. Figure 9 shows the predictor importance for both original (a) and weighted-down (b) labels. We can see that in both cases, the altitude is clearly dominating over the other features. The importance value in the vertical axis is using an arbitrary scale, and the results were averaged over 10 iterations. The error bars represent one standard deviation from the average.

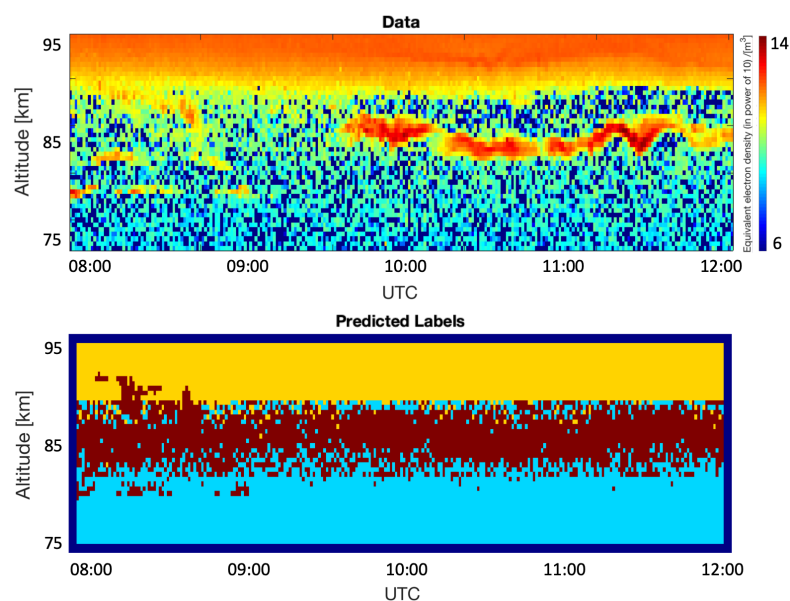
For qualitative testing, the predicted labels were generated for all the 30 test images and for all the cases (all filter sizes and all  $mtry$  values) shown in both Tables 2 and 3. Although in Tables 2 and 3, the values of classification error, classification error OOB, logarithmic error and logarithmic error OOB are worse for all filter sizes and  $mtry = 5$ , these parameters gave us the best predicted labels. To illustrate our qualitative analysis, we use four examples. The predicted labels that we observed were the best (all filter sizes and  $mtry = 5$ ) are shown in Figures 10–13. These figures show the predicted labels for the same time and dates as Figures 5–8. Figures 14 and 15 show the corresponding predictors importance for, respectively, original labels and weighted-down labels in the case of using all filter sizes together with  $mtry = 5$ . The results were averaged over 10 iterations, and the error bars represent one standard deviation from the average.

**Table 2.** Results of the classification using original labels. These values are obtained after five iterations of each experiment. Each field contains the mean over these five iterations, followed by one standard deviation.

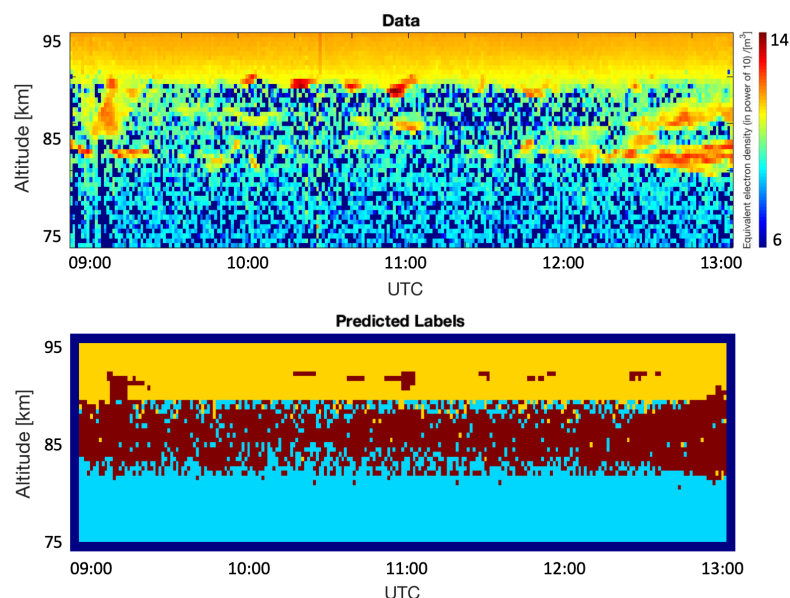
Filter Size	<i>mtry</i>	Logarithmic Error		Logarithmic Error OOB		Classification Error		Classification Error OOB		AUC Ion. Back.		AUC Noise		AUC PMSE	
		Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
3 × 3	3	0.39468	0.00115	0.39455	0.00112	0.10498	0.00018	0.10926	0.00058	0.98396	0.00041	0.98103	0.00039	0.97001	0.00035
3 × 3	6	0.37179	0.00093	0.37093	0.00095	0.11519	0.00096	0.11285	0.00223	0.98883	0.00020	0.98663	0.00013	0.97768	0.00017
3 × 3	9	0.37161	0.00004	0.37035	0.00003	0.13725	0.00035	0.13305	0.00016	0.98667	0.00005	0.96972	0.00005	0.96404	0.00009
5 × 5	3	0.41333	0.00176	0.41349	0.00171	0.12525	0.00424	0.12593	0.00332	0.98593	0.00033	0.97600	0.00035	0.96480	0.00079
5 × 5	6	0.38747	0.00062	0.38795	0.00059	0.11493	0.00062	0.11739	0.00042	0.98933	0.00009	0.98159	0.00057	0.97461	0.00010
5 × 5	9	0.36193	0.00011	0.36262	0.00013	0.11379	0.00026	0.11519	0.00011	0.98901	0.00012	0.98134	0.00018	0.97612	0.00033
7 × 7	3	0.42243	0.00114	0.42247	0.00114	0.12452	0.00366	0.12899	0.00226	0.98280	0.00025	0.97158	0.00029	0.96459	0.00121
7 × 7	6	0.39271	0.00126	0.39313	0.00128	0.11445	0.00099	0.11632	0.00052	0.98836	0.00014	0.98084	0.00022	0.97499	0.00013
7 × 7	9	0.36027	0.00005	0.36120	0.00006	0.10833	0.00076	0.11041	0.00042	0.98652	0.00008	0.98213	0.00005	0.97406	0.00008
9 × 9	3	0.42842	0.00069	0.42829	0.00070	0.14273	0.00432	0.13902	0.00356	0.98083	0.00033	0.96983	0.00027	0.95969	0.00123
9 × 9	6	0.38914	0.00147	0.38861	0.00150	0.12105	0.00097	0.11740	0.00070	0.98594	0.00020	0.98062	0.00057	0.97007	0.00010
9 × 9	9	0.36235	0.00004	0.36162	0.00005	0.11938	0.00031	0.11632	0.00026	0.98276	0.00022	0.97944	0.00012	0.96780	0.00010
11 × 11	3	0.42290	0.00061	0.42310	0.00064	0.12776	0.00066	0.12760	0.00057	0.98121	0.00029	0.97027	0.00042	0.96548	0.00061
11 × 11	6	0.37492	0.00113	0.37509	0.00118	0.11085	0.00025	0.10928	0.00042	0.98337	0.00053	0.98262	0.00028	0.97105	0.00051
11 × 11	9	0.36253	0.00003	0.36265	0.00005	0.11887	0.00065	0.11831	0.00046	0.98089	0.00016	0.97877	0.00008	0.96882	0.00009
All Sizes	5	0.43433	0.00164	0.43477	0.00165	0.13891	0.01048	0.13772	0.00575	0.98039	0.00056	0.97096	0.00076	0.95898	0.00072
All Sizes	10	0.41768	0.00102	0.41817	0.00101	0.10884	0.00035	0.11096	0.00079	0.98535	0.00008	0.97749	0.00077	0.96591	0.00058
All Sizes	15	0.40385	0.00124	0.40429	0.00122	0.11200	0.00479	0.10980	0.00438	0.98752	0.00014	0.98091	0.00016	0.96857	0.00042
All Sizes	20	0.39397	0.00142	0.39433	0.00142	0.12100	0.00198	0.11975	0.00159	0.98850	0.00021	0.98234	0.00017	0.97017	0.00019
All Sizes	25	0.38476	0.00107	0.38505	0.00108	0.12963	0.00031	0.12798	0.00038	0.98852	0.00008	0.98334	0.00012	0.97092	0.00005
All Sizes	30	0.37788	0.00076	0.37810	0.00077	0.12929	0.00015	0.12834	0.00004	0.98817	0.00011	0.98329	0.00014	0.97074	0.00042
All Sizes	35	0.37096	0.00004	0.37113	0.00007	0.12963	0.00018	0.12967	0.00055	0.98651	0.00014	0.97192	0.00016	0.96238	0.00096

**Table 3.** Results of the classification using weighted-down labels. These values are obtained after five iterations of each experiment. Each field contains the mean over these five iterations, followed by one standard deviation.

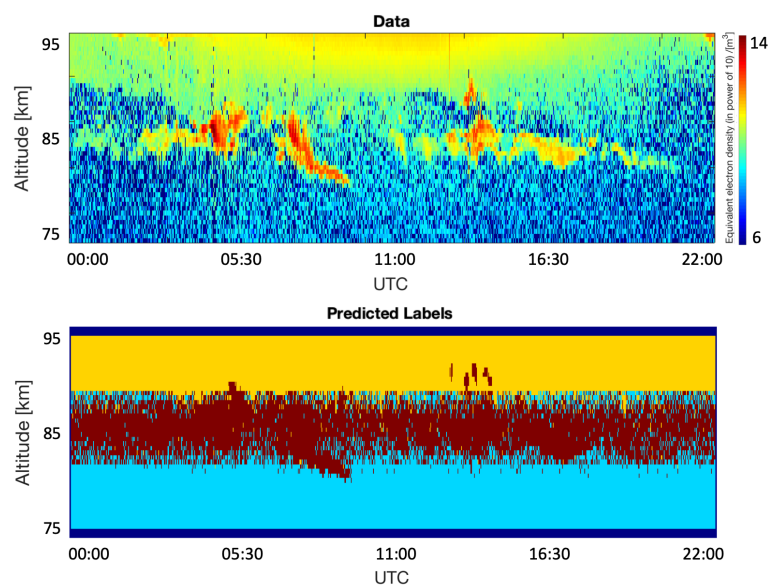
Filter Size	<i>mtry</i>	Logarithmic Error		Logarithmic Error OOB		Classification Error		Classification Error OOB		AUC Ion. Back.		AUC Noise		AUC PMSE	
		Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std
3 × 3	3	0.38950	0.00162	0.38927	0.00162	0.09828	0.00050	0.09697	0.00054	0.98806	0.00039	0.98300	0.00093	0.97536	0.00047
3 × 3	6	0.36683	0.00040	0.36659	0.00039	0.10360	0.00182	0.10437	0.00153	0.99176	0.00016	0.98826	0.00015	0.98212	0.00009
3 × 3	9	0.36459	0.00004	0.36427	0.00007	0.11532	0.00022	0.11578	0.00041	0.99063	0.00006	0.97384	0.00008	0.97110	0.00024
5 × 5	3	0.41009	0.00097	0.40947	0.00101	0.10974	0.00501	0.10625	0.00234	0.98778	0.00034	0.97695	0.00019	0.97437	0.00089
5 × 5	6	0.38508	0.00056	0.38407	0.00060	0.10504	0.00023	0.10131	0.00043	0.99152	0.00004	0.98119	0.00087	0.98128	0.00058
5 × 5	9	0.36253	0.00018	0.36111	0.00016	0.10598	0.00023	0.10088	0.00035	0.99095	0.00009	0.97723	0.00026	0.97715	0.00036
7 × 7	3	0.41946	0.00203	0.42020	0.00200	0.11242	0.00602	0.11725	0.00443	0.98662	0.00047	0.97677	0.00038	0.97086	0.00129
7 × 7	6	0.38950	0.00128	0.39053	0.00126	0.10166	0.00063	0.10692	0.00086	0.99168	0.00010	0.98475	0.00013	0.97987	0.00014
7 × 7	9	0.35469	0.00007	0.35640	0.00008	0.09460	0.00039	0.10087	0.00040	0.98998	0.00002	0.98615	0.00006	0.97921	0.00068
9 × 9	3	0.42364	0.00314	0.42302	0.00314	0.11726	0.00428	0.11962	0.00382	0.98395	0.00041	0.97341	0.00063	0.96720	0.00155
9 × 9	6	0.38314	0.00176	0.38282	0.00173	0.10587	0.00037	0.10529	0.00046	0.98849	0.00033	0.98271	0.00028	0.97598	0.00030
9 × 9	9	0.35714	0.00002	0.35719	0.00002	0.10516	0.00026	0.10440	0.00031	0.98648	0.00013	0.98298	0.00012	0.97465	0.00010
11 × 11	3	0.41998	0.00261	0.41864	0.00259	0.11798	0.00039	0.11496	0.00037	0.98394	0.00048	0.97410	0.00019	0.96995	0.00030
11 × 11	6	0.37212	0.00042	0.37113	0.00042	0.09978	0.00030	0.09628	0.00007	0.98557	0.00026	0.98462	0.00010	0.97410	0.00010
11 × 11	9	0.35753	0.00004	0.35661	0.00002	0.10448	0.00025	0.10290	0.00031	0.98402	0.00008	0.98029	0.00005	0.97274	0.00007
All Sizes	5	0.43022	0.00128	0.43013	0.00129	0.12580	0.01500	0.12236	0.00691	0.98521	0.00048	0.97142	0.00144	0.96726	0.00061
All Sizes	10	0.41479	0.00193	0.41452	0.00195	0.10435	0.00043	0.10037	0.00059	0.98886	0.00012	0.97768	0.00084	0.97209	0.00045
All Sizes	15	0.40186	0.00152	0.40145	0.00156	0.10167	0.00621	0.09960	0.00439	0.99041	0.00025	0.98078	0.00013	0.97444	0.00032
All Sizes	20	0.39105	0.00107	0.39054	0.00106	0.11213	0.00276	0.10939	0.00199	0.99138	0.00015	0.98215	0.00007	0.97609	0.00015
All Sizes	25	0.38112	0.00227	0.38047	0.00232	0.11830	0.00013	0.11614	0.00013	0.99157	0.00007	0.98331	0.00009	0.97686	0.00021
All Sizes	30	0.37277	0.00038	0.37199	0.00041	0.11913	0.00015	0.11693	0.00028	0.99121	0.00007	0.98287	0.00022	0.97694	0.00041
All Sizes	35	0.36581	0.00003	0.36490	0.00003	0.11985	0.00008	0.11747	0.00011	0.98922	0.00016	0.97259	0.00018	0.96867	0.00028



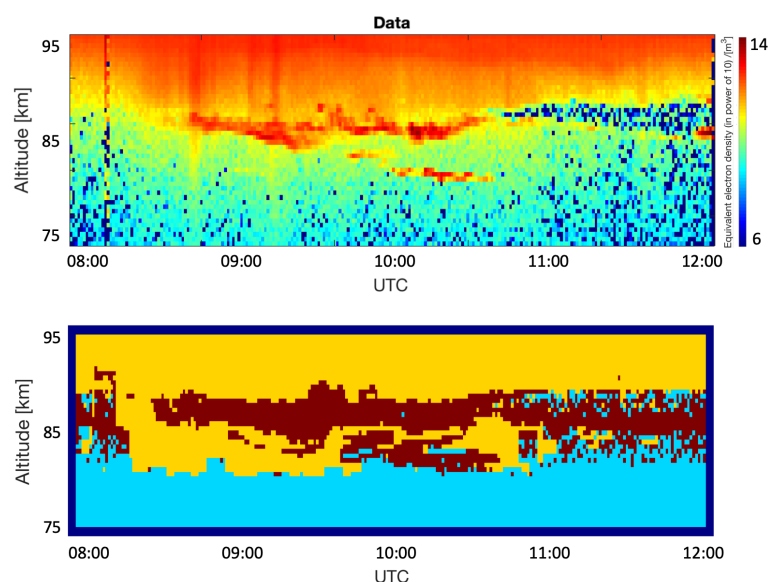
**Figure 5.** Results of segmentation using the random forests method with the  $7 \times 7$  filter size, weighted-down labels, and  $mtry = 9$ . The weighted-down labels technique is taken from the study by [6]. The data are obtained from the observation day, 17 July 2009. The image on the top illustrates the original image where the color scale represents the equivalent electron density to the power of 10, per cubic meter. The image at the bottom represents the predicted labels. Yellow, cyan and dark red represents, respectively, the region of the image labeled as Ionospheric background, Background noise, and PMSE. The horizontal axis on both images represents the time which starts at 8:00 UTC and finishes at 12:00 UTC. The vertical axis represents the altitude which ranges from 75 km to 95 km.



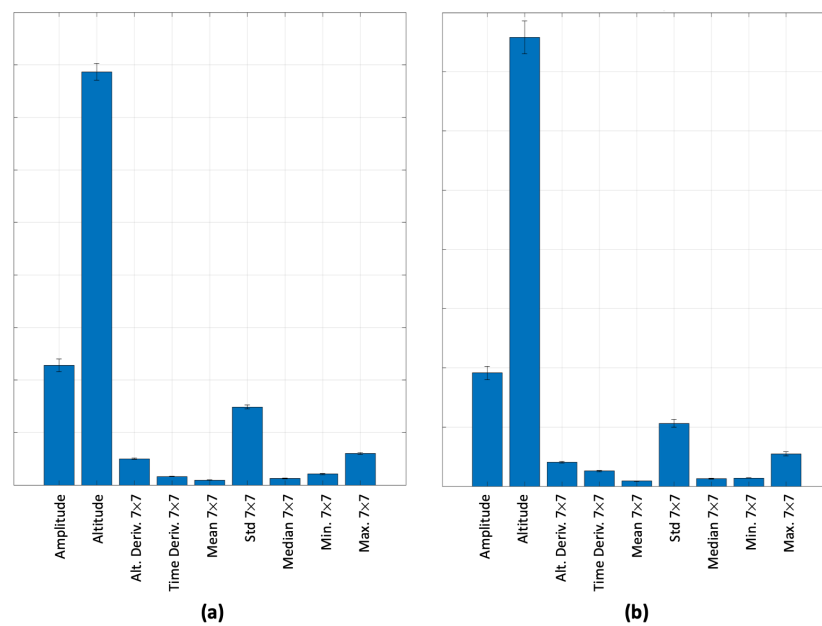
**Figure 6.** Results of segmentation using the random forests method with the  $7 \times 7$  filter size, weighted-down labels, and  $mtry = 9$ . The weighted-down labels technique is taken from the study by [6]. The data are obtained from the observation day, 8 July 2010. The image on the top illustrates the original image where the color scale represents the equivalent electron density to the power of 10, per cubic meter. The image at the bottom represents the predicted labels. Yellow, cyan and dark red represents, respectively, the region of the image labeled as Ionospheric background, Background noise, and PMSE. The horizontal axis on both images represents the time which starts at 9:00 UTC and finishes at 13:00 UTC. The vertical axis represents the altitude which ranges from 75 km to 95 km.



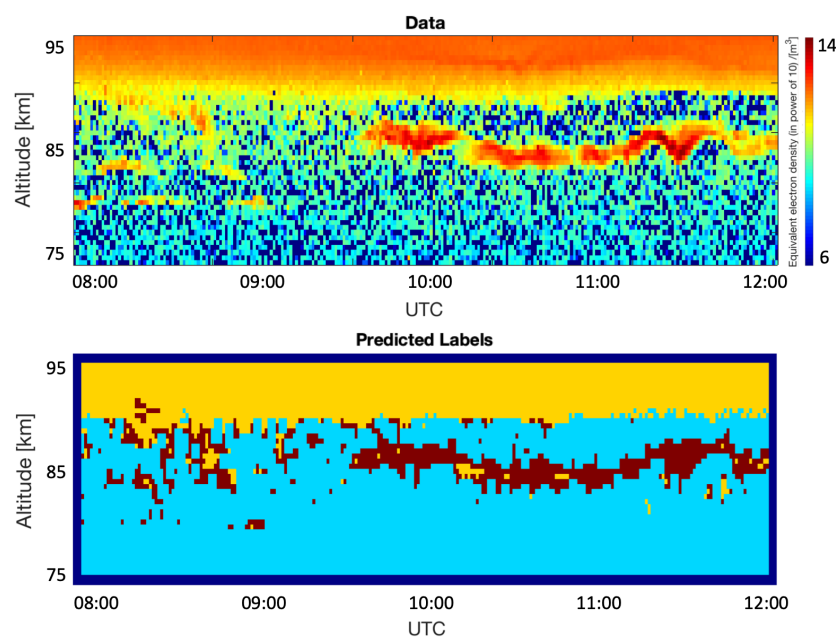
**Figure 7.** Results of segmentation using the random forests method with the  $7 \times 7$  filter size, weighted-down labels, and  $mtry = 9$ . The weighted-down labels technique is taken from the study by [6]. The data are obtained from the observation day, 7 July 2010. The image on the top illustrates the original image where the color scale represents the equivalent electron density to the power of 10, per cubic meter. The image at the bottom represents the predicted labels. Yellow, cyan and dark red represents, respectively, the region of the image labeled as Ionospheric background, Background noise, and PMSE. The horizontal axis on both images represents the time which starts at 00:00 UTC and finishes at 22:00 UTC. The vertical axis represents the altitude which ranges from 75 km to 95 km.



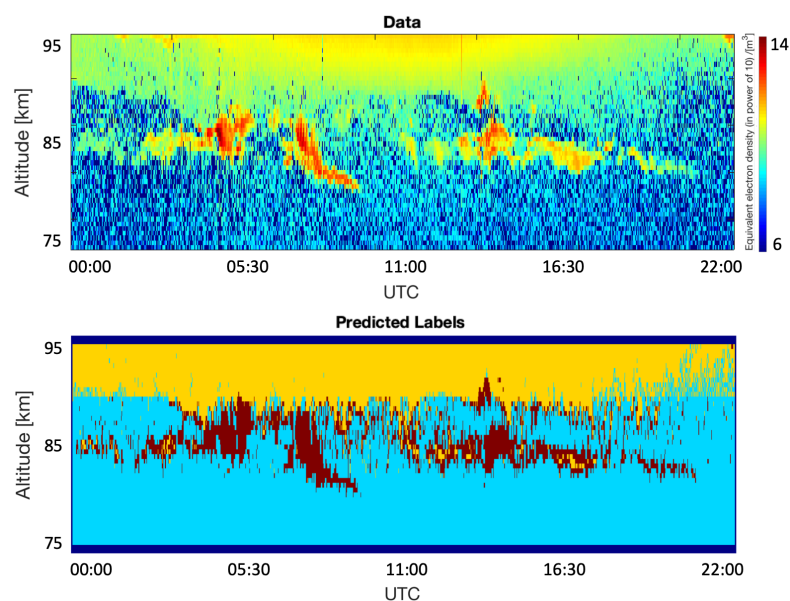
**Figure 8.** Results of segmentation using the random forests method with the  $7 \times 7$  filter size, weighted-down labels, and  $mtry = 9$ . The weighted-down labels technique is taken from the study by [6]. The data are obtained from the observation day, 30 June 2008. The image on the top illustrates the original image where the color scale represents the equivalent electron density to the power of 10, per cubic meter. The image at the bottom represents the predicted labels. Yellow, cyan and dark red represents, respectively, the region of the image labeled as Ionospheric background, Background noise, and PMSE. The horizontal axis on both images represents the time which starts at 8:00 UTC and finishes at 12:00 UTC. The vertical axis represents the altitude which ranges from 75 km to 95 km.



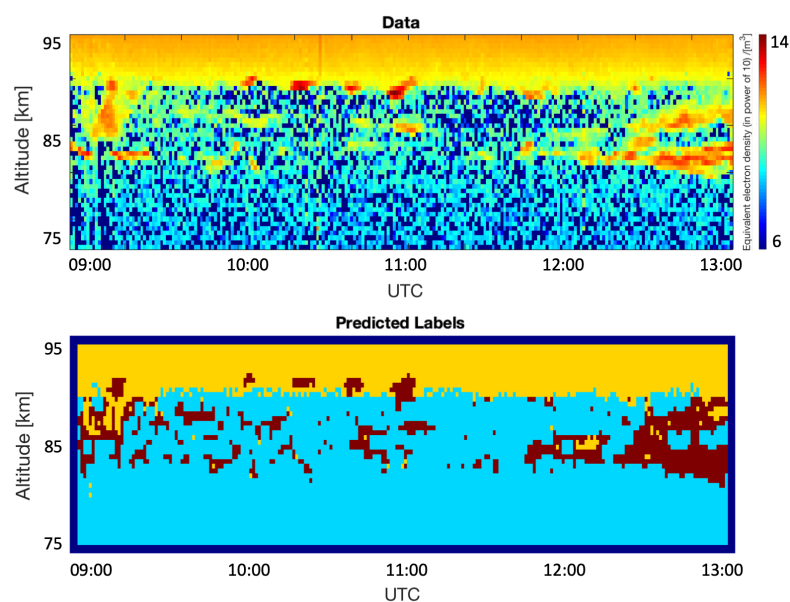
**Figure 9.** Predictor importance for random forests used on original (a) and weighted-down (b) labels with a  $7 \times 7$  filter size and  $mtry = 9$ . The horizontal axis lists all the predictors, and the vertical axis shows their importance using an arbitrary scale. Higher values mean that the algorithm assigned to them a higher importance to classify the data efficiently. These values are averaged over 10 iterations, and the error bars represent one standard deviation from the average.



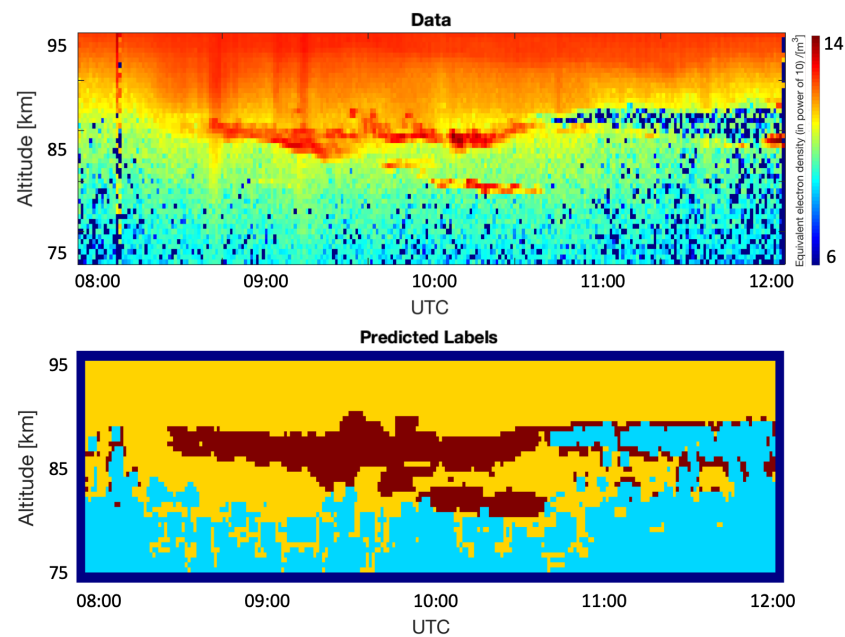
**Figure 10.** Results of segmentation using the random forests method ( $mtry = 5$ ) and weighted-down labels technique from the study by [6]. The data are obtained from the observation day on the 17 July 2009. The image on the top illustrates the original image where the color scale represents the equivalent electron density to the power of 10, per cubic meter. The image at the bottom represents the predicted labels. Yellow, cyan and dark red represents, respectively, the region of the image labeled as Ionospheric background, Background noise, and PMSE. The horizontal axis on both images represents the time which starts at 7:50 UTC finishes at 12:00 UTC. The vertical axis represents the altitude which ranges from 75 km to 95 km.



**Figure 11.** Results of segmentation using the random forests method ( $mtry = 5$ ) and weighted-down labels technique from the study by [6]. The data are obtained from the observation day, 7 July 2010. The image on the top illustrates the original image where the color scale represents the equivalent electron density to the power of 10, per cubic meter. The image at the bottom represents the predicted labels. Yellow, cyan and dark red represent, respectively, the region of the image labeled as Ionospheric background, Background noise, and PMSE. The horizontal axis on both images represents the time which starts at 00:00 UTC and finishes at 22:00 UTC. The vertical axis represents the altitude which ranges from 75 km to 95 km.



**Figure 12.** Results of segmentation using the random forests method ( $mtry = 5$ ) and weighted-down labels technique from the study by [6]. The data are obtained from the observation day of 8 July 2010. The image on the top illustrates the original image where the color scale represents the equivalent electron density to the power of 10, per cubic meter. The image at the bottom represents the predicted labels. Yellow, cyan and dark red represents, respectively, the region of the image labeled as Ionospheric background, Background noise, and PMSE. The horizontal axis on both images represents the time which starts at 09:00 UTC and finishes at 13:00 UTC. The vertical axis represents the altitude which ranges from 75 km to 95 km.

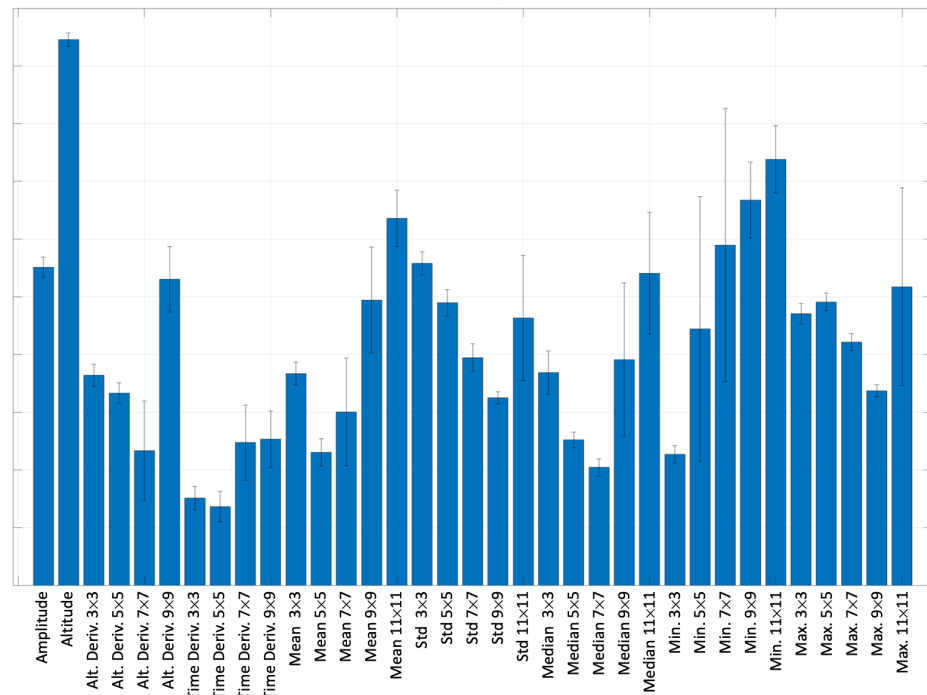


**Figure 13.** Results of segmentation using the random forests method ( $mtry = 5$ ) and weighted-down labels technique from the study by [6]. The data are obtained from the observation day, 30 June 2008. The image on the top illustrates the original image where the color scale represents the equivalent electron density to the power of 10, per cubic meter. The image at the bottom represents the predicted labels. Yellow, cyan and dark red represents, respectively, the region of the image labeled as Ionospheric background, Background noise, and PMSE. The horizontal axis on both images represents the time which starts at 8:00 UTC and finishes at 12:00 UTC. The vertical axis represents the altitude which ranges from 75 km to 95 km.

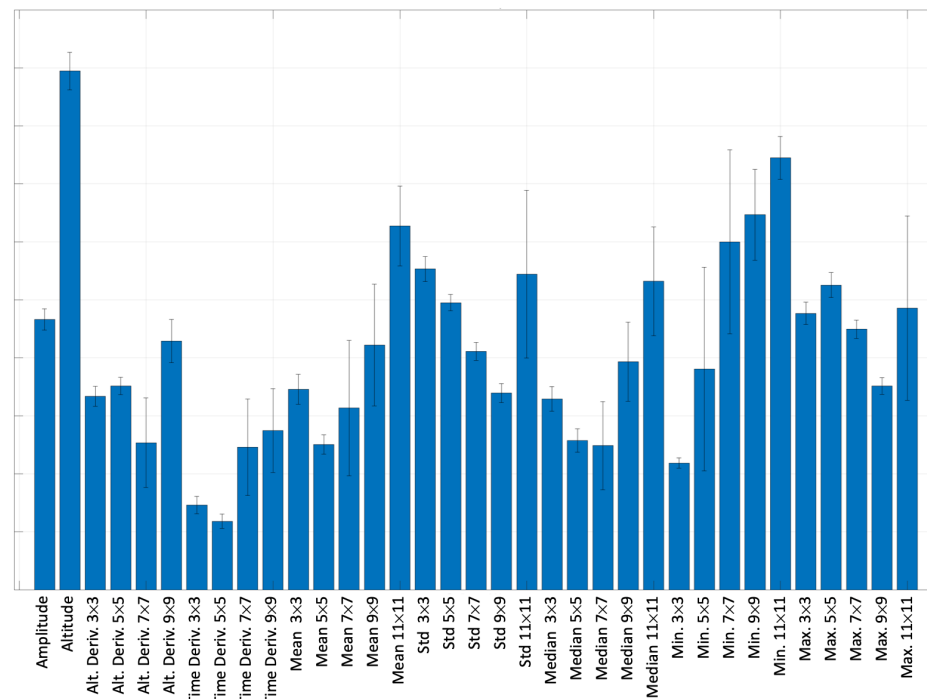
In Table 3, the evaluation scores for all combinations of filter sizes are slightly better than the evaluation scores in Table 2. This can imply that by employing weighted-down labels (discussed in Section 3.3), we can achieve a reduction in both model bias and variance, hence, leading to an improved performance. Although the performance gain achieved using weighted-down labels technique from the study by [6] is marginal, for further studies involving data labels beyond the three categories used in this paper, the gains achieved using the weighted-down labels technique could be significant.

When all 35 features are used as predictors, i.e.,  $mtry = 35$ , one can see that their importance is varying with different filter sizes. We can note that when original labels are used for the random forests algorithm, the predictor importance is different (as shown in Figure 14) as compared to when we use the weighted-down labels (as shown in Figure 15). The importance of all the different predictors is plotted on an arbitrary scale which is linear and relative. One can note that in Figure 14, the most important feature is altitude which is something that is also used in practice to determine if a signal is PMSE. After that, the second most important feature is the  $11 \times 11$  minimum value, followed by the  $9 \times 9$  minimum value, and then the  $11 \times 11$  mean value, and so on. The importance of features is similar for the weighted-down labels case as shown in Figure 15, where the first 6 predictors are the same. After that, the order is slightly different. This implies that features extracted across multiple scales, i.e., ranging from  $3 \times 3$  to  $11 \times 11$  can play an important role in improving the prediction of the random forests. Finally, one can see that in Figure 15, the error bars are slightly smaller using weighted-down labels compared to Figure 14, where original labels are used.





**Figure 14.** Predictor importance for random forests used on original labels with all filter sizes and  $mtry = 5$ . The horizontal axis lists all the predictors, and the vertical axis shows their importance using an arbitrary scale. Higher values mean that the algorithm assigned to them a higher importance to classify the data efficiently. These values are averaged over 10 iterations, and the error bars represent one standard deviation from the average.



**Figure 15.** Predictor importance for random forests used on weighted-down labels with all filter sizes and  $mtry = 5$ . The horizontal axis lists all the predictors, and the vertical axis shows their importance using an arbitrary scale. Higher values mean that the algorithm assigned to them a higher importance to classify the data efficiently. These values are averaged over 10 iterations, and the error bars represent one standard deviation from the average.

## 5. Discussion

In the case with the filter size  $7 \times 7$  and  $mtry = 9$  which quantitatively gives the best results, the segmentation results as shown in Figures 5–8 are worse. This could be due to its poor generalization to new unseen data. Based on the segmentation results obtained in Figures 10–13, we can note that using the random forests for all filter sizes with  $mtry = 5$  and weighted-down labels technique from the study by Almeida et al. [6], it is possible to segment the data into the three different categories of interest. Furthermore,  $mtry = 5$  is in line with the study by Probst et al. [18], which suggests that the recommended number of predictors is approximately equal to the square root value of the total number of features, which is 35 in our case.

In one of the images used for qualitative testing, for instance Figure 13, we notice an unusual case where part of the PMSE signal is not accurately segmented. We observe that some pixels were classified as PMSE at the border between the Ionospheric background and the Background noise. This happens around 11:00 to 12:00 UTC, while we can clearly see that there is no PMSE at that time on the original image above. We think this is the result of the fact that PMSE are usually having a pattern elongated horizontally. Because of this, the model might make the horizontal patterns stand out more, and therefore give more importance to the vertical gradients, hence to the altitude derivatives. This is something that Figures 14 and 15 confirm, where altitude derivatives were given more importance than the time derivatives. In the future, we aim to use this segmentation approach to extract the PMSE signal from the vast dataset of EISCAT observations in order to analyze in detail the structures of the PMSE signals and compare the PMSE signals from different time periods in the solar cycle.

## 6. Conclusions

This study outlines a framework to segment PMSE from the Ionospheric background and the Background noise in images obtained from EISCAT VHF radar data. We manually labeled the data into three different categories: PMSE, Ionospheric background, and Background noise, representing in total a dataset of 56,250 labeled samples. For segmentation, we employed random forests on a set of simple features. These features include: altitude derivative, time derivative, mean, median, standard deviation, minimum, and maximum values corresponding to neighborhood sizes ranging from 3 by 3 to 11 by 11. We also used the amplitude and altitude additionally as features. Next, we used a weighting-down technique on the data labels to reduce the model bias and variance.

First, our results show that it is possible to extract PMSE signal from the data, when using all sizes for feature extraction and  $mtry = 5$ . Second, by employing the weighted-down labels technique, we note an improvement in the performance of random forests.

For future studies, PMSE could be investigated over a broader dataset comprising several years of observations for one complete solar cycle. Information such as the thickness or shape of PMSE over the years could also be analyzed to gain further understanding of its origin and evolution.

**Author Contributions:** Conceptualization, D.J., P.S., I.M. and U.-P.H.; Data curation, D.J.; Funding acquisition, I.M.; Investigation, D.J. and P.S.; Project administration, I.M.; Software, D.J. and P.S.; Supervision, P.S., I.M. and U.-P.H.; Validation, P.S., I.M. and U.-P.H.; Writing—original draft, D.J.; Writing—review & editing, D.J., P.S., I.M. and U.-P.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was carried out within a project funded by Research Council of Norway, NFR 275503. The Norwegian participation in EISCAT and EISCAT3D is funded by Research Council of Norway, through research infrastructure grant 245683. The EISCAT International Association is supported by research organizations in Norway (NFR), Sweden (VR), Finland (SA), Japan (NIPR and STEL), China (CRIPR), and the United Kingdom (NERC).

**Data Availability Statement:** EISCAT VHF and UHF data are available under <http://www.eiscat.se/madrigal/> (accessed on 15 June 2022).

**Acknowledgments:** The authors thank the Research Council of Norway for the funding and the EISCAT International Association which is supported by research organizations in Norway (NFR), Sweden (VR), Finland (SA), Japan (NIPR and STEL), China (CRIPR), and the United Kingdom (NERC) for making the data easily accessible.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Latteck, R.; Renkowitz, T.; Chau, J.L. Two decades of long-term observations of polar mesospheric echoes at 69°N. *J. Atmos. Sol.-Terr. Phys.* **2021**, *216*, 105576. [[CrossRef](#)]
2. Cho, J.Y.N.; Röttger, J. An updated review of polar mesosphere summer echoes: Observation, theory, and their relationship to noctilucent clouds and subvisible aerosols. *J. Geophys. Res. Atmos.* **1997**, *102*, 2001–2020. [[CrossRef](#)]
3. Rapp, M.; Lübken, F.J. Polar mesosphere summer echoes (PMSE): Review of observations and current understanding. *Atmos. Chem. Phys.* **2004**, *4*, 2601–2633. [[CrossRef](#)]
4. Avaste, O. Noctilucent clouds. *J. Atmos. Terr. Phys.* **1993**, *55*, 133–143. [[CrossRef](#)]
5. Gadsden, M. A secular change in noctilucent cloud occurrence. *J. Atmos. Terr. Phys.* **1990**, *52*, 247–251. [[CrossRef](#)]
6. Almeida, M.; Zhuang, Y.; Ding, W.; Crouter, S.E.; Chen, P. Mitigating Class-Boundary Label Uncertainty to Reduce Both Model Bias and Variance. *ACM Trans. Knowl. Discov. Data* **2021**, *15*, 27. [[CrossRef](#)]
7. Mann, I.; Häggström, I.; Tjulin, A.; Rostami, S.; Anyairo, C.C.; Dalin, P. First wind shear observation in PMSE with the tristatic EISCAT VHF radar. *J. Geophys. Res. Space Phys.* **2016**, *121*, 11271–11281. [[CrossRef](#)]
8. Ecklund, W.L.; Balsley, B.B. Long-term observations of the Arctic mesosphere with the MST radar at Poker Flat, Alaska. *J. Geophys. Res. Space Phys.* **1981**, *86*, 7775–7780. [[CrossRef](#)]
9. Hoppe, U.P.; Hall, C.; Röttger, J. First observations of summer polar mesospheric backscatter with a 224 MHz radar. *Geophys. Res. Lett.* **1988**, *15*, 28–31. [[CrossRef](#)]
10. Hocking, W.; Rüster, R.; Czechowsky, P. Absolute reflectivities and aspect sensitivities of VHF radio wave scatterers measured with the SOUSY radar. *J. Atmos. Terr. Phys.* **1986**, *48*, 131–144. [[CrossRef](#)]
11. Kaminski, B.; Jakubczyk, M.; Przemyslaw, S. A framework for sensitivity analysis of decision trees. *Cent. Eur. J. Oper. Res.* **2018**, *26*, 135–159. [[CrossRef](#)] [[PubMed](#)]
12. Quinlan, J.R. Induction of Decision Trees. *Mach. Learn.* **1986**, *1*, 81–106. [[CrossRef](#)]
13. Quinlan, J.R. *C4.5: Programs for Machine Learning*; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 1993.
14. Kuhn, M.; Johnson, K. *Applied Predictive Modeling*; Springer: New York, NY, USA, 2013.
15. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Chapman and Hall/CRC: Monterey, CA, USA, 1984.
16. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning: With Applications in R*; Springer Publishing Company, Incorporated: New York, NY, USA, 2014.
17. Cutler, A.; Cutler, D.R.; Stevens, J.R. Random Forests. In *Ensemble Machine Learning: Methods and Applications*; Zhang, C., Ma, Y., Eds.; Springer: Boston, MA, USA, 2012; pp. 157–175. [[CrossRef](#)]
18. Probst, P.; Wright, M.N.; Boulesteix, A.L. Hyperparameters and tuning strategies for random forest. *WIREs Data Min. Knowl. Discov.* **2019**, *9*, e1301. [[CrossRef](#)]
19. Grömping, U. Variable Importance Assessment in Regression: Linear Regression versus Random Forest. *Am. Stat.* **2009**, *63*, 308–319. [[CrossRef](#)]
20. Karpievitch, Y.V.; Hill, E.G.; Leclerc, A.P.; Dabney, A.R.; Almeida, J.S. An Introspective Comparison of Random Forest-Based Classifiers for the Analysis of Cluster-Correlated Data by Way of RF++. *PLoS ONE* **2009**, *4*, e7087. [[CrossRef](#)]
21. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
22. Tyrallis, H.; Papacharalampous, G.; Langousis, A. A Brief Review of Random Forests for Water Scientists and Practitioners and Their Recent History in Water Resources. *Water* **2019**, *11*, 910. [[CrossRef](#)]
23. Ristin, M.; Guillaumin, M.; Gall, J.; Van Gool, L. Incremental Learning of Random Forests for Large-Scale Image Classification. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 490–503. [[CrossRef](#)]
24. Gall, J.; Razavi, N.; Van Gool, L. An Introduction to Random Forests for Multi-class Object Detection. In *Outdoor and Large-Scale Real-World Scene Analysis*; Dellaert, F., Frahm, J.M., Pollefeys, M., Leal-Taixé, L., Rosenhahn, B., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 243–263.
25. Belgiu, M.; Drăguț, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [[CrossRef](#)]
26. Mahapatra, D. Analyzing Training Information From Random Forests for Improved Image Segmentation. *IEEE Trans. Image Process.* **2014**, *23*, 1504–1512. [[CrossRef](#)]
27. Probst, P.; Boulesteix, A.L. To Tune or Not to Tune the Number of Trees in Random Forest. *J. Mach. Learn. Res.* **2017**, *18*, 6673–6690.

28. MathWorks. Available online: <https://se.mathworks.com/help/stats/compactclassificationensemble.loss.html> (accessed on 8 July 2021).
29. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [[CrossRef](#)]
30. Jozwicki, D.; Sharma, P.; Mann, I. Investigation of Polar Mesospheric Summer Echoes Using Linear Discriminant Analysis. *Remote Sens.* **2021**, *13*, 522. [[CrossRef](#)]
31. Lehtinen, M.S.; Huuskonen, A. General incoherent scatter analysis and GUISDAP. *J. Atmos. Terr. Phys.* **1996**, *58*, 435–452. [[CrossRef](#)]
32. Doug Morris, R.R. Sobel gradient kernels ( $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$ ,  $9 \times 9$ ). Available online: <https://imagej.nih.gov/nih-image/download/user-macros/slowsobel.macro> (accessed on 11 May 2021).