



Technical Note HydroMet: A New Code for Automated Objective Optimization of Hydrometeorological Thresholds for Landslide Initiation

Jacob L. Conrad ^{1,2}, Michael D. Morphew ^{1,2}, Rex L. Baum ¹ and Benjamin B. Mirus ^{1,*}

¹ U.S. Geological Survey, Geologic Hazards Science Center, Golden, CO 80401, USA;

jlconrad@usgs.gov (J.L.C.); mdmorphew@gmail.com (M.D.M.); baum@usgs.gov (R.L.B.)

² Colorado School of Mines, Golden, CO 80401, USA

Correspondence: bbmirus@usgs.gov

Abstract: Landslide detection and warning systems are important tools for mitigation of potential hazards in landslide prone areas. Traditionally, warning systems for shallow landslides have been informed by rainfall intensity-duration thresholds. More recent advances have introduced the concept of hydrometeorological thresholds that are informed not only by rainfall, but also by subsurface hydrological measurements. Previously, hydrometeorological thresholds have been shown to improve capabilities for forecasting shallow landslides, and they may ultimately be adapted to more generalized landslide forecasting. We present HydroMet, a code developed in Python by the U.S. Geological Survey, which allows users to guide the automated estimation of hydrometeorological thresholds for a site or area of interest, with the flexibility to select preferred threshold variables for the antecedent hydrologic conditions and the triggering meteorological conditions. Users can import hydrologic time-series data, including rainfall, soil-water content, and pore-water pressure, along with the times of known landslide occurrences, and then conduct objective optimization of warning thresholds using receiver operating characteristics. HydroMet presents many additional options, including selecting the threshold formula, the timescale of possible threshold variables, and the skill statistics used for optimization. Users can develop dual-stage thresholds for watch and warning alerts, with a lower, risk-averse threshold to avoid missed alarms and a less conservative threshold to minimize false alarms. Users may also choose to split their inventory data into calibration and evaluation subsets to independently evaluate the performance of optimized thresholds. We present output and applications of HydroMet using monitoring data from landslide-prone areas in the U.S. to demonstrate its utility and ability to produce thresholds with limited missed and false alarms for informing the next generation of reliable landslide warning systems.

Keywords: software; landslide early warning systems; hydrometeorological thresholds; automated optimization

1. Introduction

Landslides are a common geologic hazard worldwide that are often destructive and deadly. Landslide early warning systems (LEWS) offer a promising approach for risk reduction in steep, soil-mantled terrain by providing alerts in advance of the potential conditions for landsliding. Typically, LEWS rely on thresholds to distinguish between conditions that are conducive to shallow landsliding from those that are not. The most common and established landslide threshold approach relies on the rainfall intensity-duration (ID) formulation [1–4], such that higher intensity shorter duration storms above the threshold, and lower intensity longer duration storms above the threshold, are both considered equally likely to trigger landsliding events. However, in many settings the antecedent soil wetness conditions influence the variability in rainfall triggering amounts, which has supported an emerging approach that relies on soil hydrology and rainfall to define hydrometeorological thresholds [5–11]. In general, these studies show that explicit



Citation: Conrad, J.L.; Morphew, M.D.; Baum, R.L.; Mirus, B.B. HydroMet: A New Code for Automated Objective Optimization of Hydrometeorological Thresholds for Landslide Initiation. *Water* **2021**, *13*, 1752. https://doi.org/10.3390/ w13131752

Academic Editor: Samuele Segoni

Received: 21 May 2021 Accepted: 22 June 2021 Published: 25 June 2021

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



Copyright: © 2021 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/). consideration of soil antecedent moisture conditions can reduce the number of missed alarms as well as the number of false alarms relative to rainfall-only thresholds. This follows logically, since a smaller storm occurring on soils that are already very wet may be more likely to trigger landslides than a larger storm on a drier soil, which is a nuance that cannot be fully captured by a simple threshold that considers only rainfall conditions. Additionally, while ID thresholds are typically only applicable to forecasting shallow landslides and debris flows, hydrometeorological thresholds could potentially account for the gradual, steady subsurface wetting associated with deeper and seasonally active landslides. Various approaches and scientific software have been developed to optimize rainfall ID thresholds using statistical methods to analyze rainfall conditions during past landslide events [12–15], but no standardized approaches have been established for developing and testing hydrometeorological thresholds. Various formulations have been explored, most of which consider a hydrological variable indicative of the antecedent wetness conditions and a meteorological variable indicative of the storm triggering conditions, which is consistent with the cause-trigger framework proposed by Bogaard and Greco [16]. For example, the hydrological variable could be antecedent soil moisture and the meteorological variable the cumulative storm total at the time of landsliding.

Here, we present the development of a new scientific software called HydroMet, which can be used to optimize thresholds using rainfall and hydrologic monitoring data. The software builds upon previous research that demonstrates the utility of hydrometeorological thresholds [5] and allows end users considerable flexibility in guiding the objective optimization thresholds for their area of interest. HydroMet is a statistical analysis tool to develop objective empirical thresholds for landslide initiation, and thus it requires some knowledge of landslide timing, as well as the local conditions contributing to their cause and triggering. HydroMet was developed by the U.S. Geological Survey (USGS) Landslide Hazards Program using the Python programming language. This paper outlines the theoretical basis and implementation, including output file interpretation, as well as an example file for new users.

2. Theoretical Basis

2.1. Optimizing with Receiver Operating Characteristics

The objective optimization of thresholds in HydroMet uses receiver operating characteristics (ROC), which employ the confusion matrix (Figure 1) to quantify the pairing between positive and negative events, and true and false predictions [17,18]. As in other LEWS studies using ROC analysis, we define positive events as the occurrence of a landslide within the specified time interval, and we define negative events as the lack of a reported landslide during the interval. True predictions correctly identify the occurrence of positive and negative events, whereas false predictions incorrectly identify those events. As seen in Figure 1, true positives (TP) correctly predict the occurrence of landsliding; true negatives (TN) correctly predict the lack of landsliding; false positives (FP) incorrectly predict landsliding when no known landslides occurred (type I errors); and false negatives (FN) incorrectly predict no landslides when in fact one or more landslides occurred (type II errors). Thus, a perfect threshold would include only TP and TN, though in practice some number of FP and FN are expected due to the spatially variable nature of precipitation and the subsurface hydromechanical processes that trigger landslides across a given landscape. The value of ROC analysis is that it provides a method to quantitatively evaluate the TP, TN, FP, and FN to objectively optimize a threshold to maximize correct predictions.

	Landslide Occurred	Landslide Did Not
		Occur
Predicted Landslide	True Positive (TP)	False Positive (FP)
		(False Alarm or Type I
		Error)
No Landslide	False Negative (FN)	True Negative (TN)
Predicted	(Missed Alarm or Type II	_
	Error)	

Figure 1. Confusion matrix describing the relationship between the occurrence of landslides and the prediction of landslides.

Landslides are relatively episodic hazards, so datasets used for ROC analysis typically include relatively few positive events (i.e., landslides) and many negative events (i.e., no landslides); when developing thresholds for accurate forecasting, there are often only a small number of true positives and numerous true negatives. The real challenge in minimizing errors lies in achieving the desired balance between the false positives and false negatives, which may depend on the LEWS objectives. Other scientific fields might strive for high accuracy as a measure of success, but in landslide forecasting, the large number of false negatives will render the accuracy measurement useless for determining the success of a landslide threshold. Therefore, we turn towards other ROC metrics to evaluate the effectiveness of the calculated threshold. For example, an optimistic threshold would seek to minimize the false negatives (type II errors), such that all landslide events were correctly predicted in advance without any missed alarms, but this would come at the expense of more numerous false positives (type I errors). In contrast, a more pessimistic threshold would seek to minimize those false alarms at the expense of a few additional missed alarms when landslides do occur (type II errors). To this end, the following skill statistics are often used to objectively optimize thresholds with those differing priorities in mind:

True Positive Rate (TPR) =
$$\frac{TP}{TP+FN}$$

False Positive Rate (FPR) = $\frac{FP}{FP+TN}$
Threat Score (TS) = $\frac{TP}{TP+FN+FP}$
Precision = $\frac{TP}{TP+FP}$
True Skill Statistic (TSS) = $TPR - FPR$
Optimal Point (OP) = $\sqrt{FPR^2 + (TPR-1)^2}$

All these skill statistics vary between zero (0.0) and one (1.0); for the TPR, threat score, precision, and true skill statistic, a larger value indicates a better score with a perfect score of one, whereas for the FPR and optimal point, a smaller value indicates a better score with a perfect score of zero. For a given dataset, HydroMet will display a range of scores for a threshold formulation (e.g., linear threshold for antecedent 3-day rainfall and recent 1-day rainfall) depending on the actual threshold equation, whereas each possible threshold equation will exhibit a unique value for each skill statistic. The threshold equation represents the mathematical formulation of the line separating conditions that predict a landslide from those that predict no landslide. This range of scores for a given threshold formulation using the ROC curve approach, which is further helpful in optimizing hydrometeorological thresholds and rainfall ID thresholds alike.

Each possible threshold equation is plotted in terms of the FPR (x-axis) and TPR (yaxis) where the connected points produce the ROC curve; the area under the curve (AUC) statistic indicates how well the general formulation can perform. Threshold formulations with AUC = 0.5 are considered random models in that they predict correctly as often as they predict incorrectly, whereas values between 0.5 and 1.0 indicate models with some predictive power such that the greater the AUC the better the threshold formulation. The ROC curve can thereby illustrate how thresholds with the same variables, but different formulae, can produce LEWS with very different priorities in terms of false alarms versus missed alarms. The example in Figure 2 uses ROC values given from multiple HydroMet analyses to show the ROC curve of a suite of thresholds developed for Portland, OR, whereby the four different skill statistics are used to optimize their respective thresholds using 3-days recent cumulative precipitation (P3) and 1-day antecedent saturation (S1) for each case. The ROC curve does not show the suite of all possible thresholds tested but only the most optimal threshold for a given formulation and set of variables, as determined by the highest TPR and subsequently the lowest FPR. It is possible to have multiple combinations of TPR and FPR but only the higher TPR is selected for each statistic as the optimal rate. The optimal daily threshold for each skill statistic is shown (Table 1).



Figure 2. HydroMet output plot showing optimal ROC curve for Portland, OR, using 3-days recent cumulative precipitation (P3) and 1-day antecedent saturation (S1). Each skill statistic was optimized on a range of threshold values. The skill statistics were tested for cumulative precipitation and antecedent saturation between 1 and 15 days. The resulting optimal daily threshold is shown. Optimal point and true skill statistic had the same optimal true positive rate and false positive rate, shown in orange.

Optimized Skill Statistic	Optimal Threshold TPR	Optimal Threshold FPR	Formula P (mm), S (unitless)
threat score	0.5333	0.0155	P3 > 54, S1 > 0.70
optimal point	0.9000	0.0918	P3 > 24, S1 > 0.70
precision	0.3222	0.0009	P3 > 100, S1 > 0.70
true skill statistic	0.9000	0.0918	P3 > 24, S1 > 0.70

Table 1. Summary of skill statistics calculated for four contrasting thresholds with the same variables and daily timescales of 3 days of cumulative precipitation (P3) and 1-day average antecedent saturation (S1) as shown in Figure 2.

As seen in Figure 3, the selected skill statistic can affect the optimal threshold, even for the same rainfall and soil saturation time duration. The optimal threshold using threat score as the optimizing metric can be seen in Figure 3a, whereas Figure 3b used the same variable durations (P3:S1) but optimized for the optimal point statistic. Due to the conditions present for the landslides in this area, both thresholds happen to converge on an optimal saturation of above 0.7, whereas the precipitation levels differ from 54 mm for threat score and 24 mm for optimal point. Because of this difference, the threat score threshold has fewer false alarms (purple triangles) but more missed/failed alarms (red squares). The threat score is considered to be a more pessimistic (or conservative) threshold compared to a more optimistic (or balanced) optimal point threshold, which reduces the number of missed landslides (red squares) but increases the number of false alarms. This is again a trade-off that needs to be discussed when choosing an optimal threshold for a landslideprone area. Users can choose to be more proactive with warnings, but this increases the percentage of false alarms, thus reducing credibility of those who must alert the landslideprone areas of a landslide occurrence. On the other hand, a more conservative approach might reduce false alarms, but this increases the chance of missing an event and elevating the risk of damage or injury to those in the area due to lack of situational awareness. To address these contrasting needs by different stakeholders, the concept of dual thresholds has been introduced to allow both a more conservative and a more balanced threshold to be displayed within a single threshold space [5–7].



Figure 3. HydroMet output plot showing optimized threshold using a P3:S1 variables for two different skill statistics: threat score and optimal point, showing respectively a more conservative and a more balanced threshold. (**a**) An optimized threshold using the threat score statistic, which resulted in a threshold of precipitation greater than 54 mm and a saturation greater than 0.7. (**b**) The optimized threshold using the optimal point statistic, which resulted in a threshold of precipitation greater than 0.7.

Benchmarks for defining good, bad, or acceptable threshold performance scores or AUC values for LEWS are not clearly established, but given the considerable heterogeneity within any area of interest and the uncertainty associated with spatial variability in rainfall and antecedent conditions, a perfect model is not expected. In this paper, we present a real-world example to provide users with a concrete frame of reference for evaluating their own HydroMet results. Additionally, HydroMet provides the option of split sampling one's data to allow for calibration and independent evaluation of optimized thresholds.

2.2. Threshold Formulation

In theory, it is possible to employ HydroMet to develop thresholds with any two time-series variables as inputs, but we emphasize the use of an antecedent hydrologic variable and a triggering precipitation. Here, we emphasize soil saturation (calculated using volumetric water content), pore-water pressures, and rainfall. However, in principle, any combination of variables could be used, such as stream discharge, snowmelt, water-table fluctuations, excess rainfall to account for interception, volumetric water content (VWC), and even infiltration time series. To calculate the soil saturation in previous studies [5–10], we normalized (VWC) using the maximum observed VWC during saturated conditions, which has allowed us to develop thresholds using hydrologic time-series data without detailed knowledge of the soil parameters. We used both in situ monitoring data [5,6] and satellite derived estimates of rainfall and subsurface hydrologic conditions [8], and there is no specific requirement in HydroMet for how users collect or pre-process these data.

To avoid redundancy between the cause and triggering variables proposed in the framework of Bogaard and Greco [16], HydroMet automatically assigns the antecedent conditions to begin directly before the triggering variable, without any overlap. To allow users to develop thresholds that can leverage rainfall forecasts to provide advance warning for LEWS, HydroMet can use measured rainfall during the interval of a landslide event as a proxy for a completely accurate rainfall forecast, though uncertainty in rainfall conditions would then need to be either assumed negligible or considered manually.

For threshold definition, rainfall and landslide occurrence data are typically discretized into regular intervals, such as hourly or daily timesteps. HydroMet can accommodate data with any temporal discretization for the threshold variables and landslide timing, which must be defined by the user at the outset.

Alternative formulas have been considered for hydrometeorological thresholds, including linear [6] and bilinear functions [5,9], interpolated line segments without a mathematical function [7], cutoff values for integration of antecedent conditions with traditional rainfall thresholds [4,10], and more complex logical operators using soil moisture data [11]. HydroMet allows two options, linear or bilinear thresholds, the user must select one option at the outset of the threshold development exercise but can perform repeated analysis to compare the two possible formulations for threshold equations. Previous research and LEWS development have introduced the concept of dual-stage thresholds for watch and warning alerts [5–7,19,20]. These typically include a lower, risk-averse threshold that minimizes missed alarms to indicate the conditions for potential landsliding conditions, in combination with a higher, less-conservative (more balanced) threshold that minimizes false alarms to indicate the conditions when landsliding is very likely. HydroMet allows the user to define such dual thresholds using the threat score for a higher, more conservative threshold and the optimal point for a lower, more balanced threshold. Otherwise, a single threshold can be optimized using any single desired skill statistics (see equations in Section 2.1).

3. Implementation

3.1. System Requirements

HydroMet was developed in Python and the source code can be downloaded from the USGS code repository https://code.usgs.gov/ghsc/lhp/hydromet (accessed on 23 June 2021).

The software requires Python 3 and the Python packages NumPy, matplotlib, pandas, sklearn, math, and os. Further details are available at the code repository.

3.2. Input Requirements

Input data should be placed into a folder titled 'Data.' Within the 'Data' folder, each location should have a labeled subfolder, for example 'Portland' for the city of Portland. For each location, there should be two text files: one with the time series hydrologic data and the other with the times of landslide occurrences, unless only the positive pore-water pressures are used as a proxy for landslide occurrence. The hydrologic data would be labeled '*location*RSP.txt' and the landslide occurrences would be '*location*LandslideTimes.txt,' i.e., for Portland this would be 'portlandRSP.txt' and 'portlandLandslideTimes.txt.' Capitalization of the city or location is not important. The specific units of rainfall measurement, pressure, and time of data recording is not predetermined, but users must take care to ensure that all variables use consistent units within the dataset. Text files should be organized as follows:

- Rainfall, saturation, pressure (RSP) files are four-column, header-less files
 - Column 1: time of data recording, as a date number, (e.g., number of days since January 1, 0001)
 - O Column 2: rainfall measurement, in length (e.g., mm)
 - Column 3: saturation measurement, unitless
 - \bigcirc Column 4: pressure measurement, in ML⁻¹T⁻² (e.g., kPa)
- LandslideTimes files are single column, header-less files
 - Column 1: time of landslide recording, as a date number

It is important to note that column 4 on the RSP data is only needed when pore pressure is being used as a proxy for landslide occurrences. The data must be formatted according to these requirements and file paths organized correctly into these folders. When the program is run, the user will first indicate the location used. HydroMet will open and read the two data files as needed. As noted earlier, users can apply the HydroMet software to use any preferred time-series variable (e.g., infiltration instead of soil saturation), but here we emphasize the use of rainfall, soil saturation, and pore-water pressures in the file names and output plots since we have used these variables successfully in previous analyses [5–7].

3.3. Additional Options

The user has many options to specify and inputs to alter to obtain the desired threshold analysis. The code provides places for user input throughout the analysis, which are discussed below. Note that the inputs are not case sensitive, so capitalization of inputs is not important. Documentation in this paper uses capital letters for consistency.

Variable duration—The user specifies the time duration, in number of days for cumulative rainfall and antecedent soil saturation. HydroMet asks four separate input questions, noted as the minimum days of cumulative rainfall, maximum days of cumulative rainfall, minimum days of antecedent saturation, and maximum days of saturation. Valid input values are integers greater than or equal to 1. These durations will then be used as a time component of the threshold on which the hydrologic impacts are evaluated.

Threshold formulation (linear vs. bilinear)—The user can then denote linear or bilinear threshold formulation. Linear is denoted by the letter 'L' and bilinear is denoted by the letter 'B'.

Skill statistic—The threshold is optimized based on the skill statistic chosen. The four options are threat score ('T'), optimal point ('O'), true skill statistic ('S'), and precision ('P'). Descriptions and equations used to calculate these skill statistics are provided in Section 2.1.

Discretization of Threshold Optimization Search—The user is prompted to specify the minimum, maximum, and spacing between values to be tested at saturation thresholds. These values are given by three numbers separated by a space. The saturation is a percentage given as a decimal so only values from 0–1 for the minimum and maximum are allowed. For bilinear thresholds, a recommended value for this input is (0, 1, 0.1) to request HydroMet to step through all values starting at 0 in intervals of 0.1 until 1. However, for linear thresholds, it may be necessary to test x-intercept values of saturation that are above unity, in which case a recommended value is (0, 1.3, 0.1). These values can be refined, and ranges beyond this initial recommendation should be explored by users.

The user is then prompted to specify the minimum, maximum, and spacing of precipitation values. These values must be provided by the user as three integers separated by a space. The unit of rainfall is measured length, so the minimum is 0 and the maximum is not restricted. A recommended initial range of values for rainfall in millimeters is (0, 200, 2) which can and should also be adjusted by users depending on the rainfall conditions for the site of interest.

Split sample—The user can choose to use the entire dataset for optimization or split the data into subsets for training (calibration) and testing (evaluation). This latter option is useful if the user has enough representative data to independently test the thresholds developed. If the entire dataset is used for optimization, the user will input 'All'. If the user would like to split the data, they must decide whether they want to split the RSP data into two sections based on one of two methods: (a) as a percentage of the duration of record ('RSP'), where the landslide data are automatically assigned to the two samples accordingly; or (b) a percentage of the recorded landslides ('LS'), where the RSP data are then assigned accordingly. If one of these two data splitting approaches was selected, the user must then designate the percentage used for training (calibration) set and the remainder will be reserved as independent testing (evaluation) data. This percentage is an integer from 1 to 100.

Rainfall forecast uncertainty—By default, HydroMet uses a 24-h "error-free" forecast in the precipitation variable for optimizing all thresholds. Use of rainfall forecasts was found to add to the utility of hydrometeorological thresholds for informing LEWS [6]. For example, if a 3-day precipitation is chosen, the program will use 48 h of rainfall prior to the time of landslide occurrence and 24 h of rainfall data recorded during the time of landslide occurrence to act as a prediction. For optimization, reliance on a 24 h rainfall forecast would require using the previously recorded rainfall data that do not include the influence of forecast uncertainty. However, for real-time application of an optimized threshold, uncertainty is likely in that 24-hour-precipitation forecast, which could influence threshold performance. To crudely simulate the potential influence of uncertainty in rainfall forecasts and evaluate how the thresholds respond to a variable rainfall prediction, the user can choose to add a percentage of rainfall uncertainty to the prediction. For example, if the user wants to add a 10% uncertainty to the precipitation forecast, the rainfall prediction in the future 24-h period will have a random error of \pm 5%. The percentage is entered as an integer from 1 to 100.

Positive pore pressure threshold—The final option a user can elect to perform is a similar threshold analysis as described above, but instead of known landslide occurrences, positive pore pressure is used as a proxy in the LandslideTimes input file. This approach may be desirable in cases where positive pore-water pressures are a known precursor to shallow landsliding, but the timing of landslides is unknown or poorly constrained [7,9]. Positive pore pressure threshold analysis provides the variable duration, threshold format, skill statistic, and discretization of threshold optimization search options. Split sampling and uncertainty are not implemented directly in HydroMet for positive pore pressure analysis.

3.4. Operational Instructions

All user input is completed within the configuration script, called ConfigHydroMey.py. A flowchart of user decisions can be seen in Figure 4 below. Once all inputs are put into the configuration script, the file is run to make sure there are no errors. If an error message



is generated, the user must correct any errors before continuing to run the HydroMet.py program.

Figure 4. A flowchart of the user options and inputs with the configuration script ConfiHydroMet.py. The configuration script starts with the green circle and moves through the flow chart. If a specific analysis is desired, the user will then input the corresponding values denoted by the yellow ellipses. Once all inputs are selected, the code can be run, and results are shown.

The various threshold options provided by HydroMet allow the user to specify all possible threshold options within their desired analysis and produce an optimal threshold with a wide array of end goals in mind. Since all study areas of interest are different and the purpose of each threshold can vary depending on user requirements, HydroMet's utility is maximized with its fluidity of options. Detailed operational instruction can be found within the repository titled operational_instructions.md.

4. Output Files

The output text and figures will be placed in the same user-specified directory as the input files. After execution of the program, subfolders can be seen for each location. For example, the Portland output text and figures will be contained within a subfolder called 'Portland'.

Depending on the completed analysis, the subfolder will contain three sub-subfolders and a figure. The first subfolder is titled 'ROC Metric' and contains the results of the Threshold Analysis function. 'ROC Metric' will contain subfolders for each ROC metric calculated, which will be labeled as 'Threat Score', 'True Skill Statistic', 'Precision', and 'Optimal Point'. Inside these will be subfolders for each split of the data. They will be labeled as 'All Data' (for the full data set), 'Training Data Split' (for the training set split), and 'Testing Data Split' (for the test set split). Within this folder are the results. There are four items per iteration number. 'Round{value}.txt', where value refers to the iteration number, contains labeled columns and rows of the calculated values for Y-int, X-int, FPR, TPR, Threat Score, Precision, True Skill Statistic, Optimal Point, Percent Exceeded, and AUC for each daily threshold option provided. 'Round{value}.png' is the figure for the optimal threshold. The best daily threshold is found from the above .txt file. 'ROCRound{value}.txt' contains the false positive rate (FPR) in one column and the true positive rate (TPR) in the next. This is used to plot the ROC curve. 'BestThresholdRound{value}.txt' contains a labeled row of the best threshold for the round. All user input values are listed plus the calculated values seen in 'Round{value}.txt'. These text file outputs are summarized in Table 2.

Table 2. The three output text files and contents are shown. The information contained has some intentional overlap. The main difference from Round{value} and BestThresholdRound{value} is that Round{value} shows the optimal threshold for each precipitation: saturation daily combination whereas BestThresholdRound{value} gives just the optimal threshold for the desired ROC metric and all possible variables used in the threshold analysis, not just the optimal threshold.

	Round{Value}	ROCRound{Value}	BestThresholdRound{Value}
Range of precipitation days to evaluate			•
Range of saturation days to evaluate			•
Threshold curve ROC Metric used to optimize threshold			•
X intercept range			•
Y intercept range			•
Optimal cumulative rainfall days	•		•
Optimal antecedent saturation days	•		•
Optimal threshold y intercept	•		•
Optimal threshold x intercept	•		•
FPR	•	•	•
TPR	•	•	•
Threat score	•		•
Precision score	•		•
True skill statistic	•		•
Optimal Point score	•		•
Percent of landslides exceeding threshold	•		•
AUC	•		•
Best threshold from each P:S day combination	•		
Single Optimal Threshold			•

The second folder is titled 'Combined' and contains the results of the Combined Optimal Point/ Threat Score Threshold calculation. 'Combined' will contain the .png figures of the combined optimal point/threat score plots. The plot files will be labeled as 'Rainfall-{value}_Saturation-{value}, where value refers to the variable duration.

The third folder is titled 'PorePressure' and contains the results of the threshold analysis using pore pressure as a proxy for landslide occurrences. This folder is organized similarly as the 'ROCMetric' folder, with subfolders of each of the four ROC metrics used in evaluation. Within those are the text files and figures as described above for the 'ROCMetric' folder. Since this analysis does not allow for a train/test splitting of the data, no subfolders are needed to distinguish between all data, training data, or testing data.

One standalone figure contains the Pore Water Pressure and Landslide Occurrence Correlation. This output file will be titled 'PositivePressureCorrelation.png'.

5. Example Applications

The current Python scripts for HydroMet and ConfigHydroMet can be found in a Git repository located at https://code.usgs.gov/ghsc/lhp/hydromet (accessed on 23 June 2021) The repository contains the two Python scripts, a short ReadMe file describing an overview of the software and application of the software, operational instructions, example RSP and landslide data, and an example demonstrating the file structure of the input data, and output folders. These examples rely on in situ monitoring data from Mukilteo, Washington, presented previously in Mirus et al. [5,6]. However, as noted earlier, it is possible to apply HydroMet using input data derived from any source, including numerical models [7,10] or satellite derived hydrometeorological data [8].

When using a desktop Python application, the user can download HydroMet.py and ConfigHydroMet.py and the example data and place them in the appropriate file configuration. The example configuration script contains inputs to perform a simple threshold analysis. The only necessary modification is to update the line labeled 'file_location', highlighted in Figure 5. The file location should match the file directory path specific to the user's computer. The configuration script serves as a template for the user to substitute valid parameters for their own study area and analysis. Following the prompts provided in the configuration script, the user can enter respective inputs. Values must be entered exactly as seen in the configuration script example. Any response with parentheses, brackets, or commas separating values needs to be replicated. If this is done incorrectly, the configuration script will produce an error message preventing proper execution of HydroMet.

```
1 # -*- coding: utf-8 -*-
    Created on Fri Oct 23 11:08:36 2020
    '''This is the configuration file for the HydroMet script. '''
8
    import os
    ""Enter the location of data to be used and the file path directory for where the file and data is located.""
    location = 'Mukilteo
12 file_location = 'C:/Users/{username}/Documents/HydroMet/'
   ""Would you like to calculate the landslide threshold based on a range of precipitation
14
   and saturation days? [Yes or No]. ''
   XYThresh = 'Yes'
    rainfallDurationMin = 1
18 rainfallDurationMax = 2
    satDurationMin = 1
20
   satDurationMax = 2
    curveType = 'b
    rocMetric = 't'
    xParameters = [0, 1, 0.1]
24 yParameters = [0, 200, 2]
    '''Percentage' is only needed if 'LS' or 'RSP" is selected for 'dataSplit.'''
   dataSplit = 'All
    percentage = 80
```

Figure 5. First 29 lines of the configuration script ConfigHydroMet.py. The file location highlighted above in line 12 will need modification to reflect the user's file directory structure. The use of quotation marks around inputs can be seen in lines 11,12, 16, 21 22, and 27. The use of brackets and commas separating values is seen in lines 23 and 24.

For the simple threshold analysis, many variables are required by user input and the final results can vary greatly when different parameters are selected. First, a minimum and maximum range is needed for both the cumulative precipitation and antecedent saturation days. Next, the user chooses between a linear or a bilinear threshold formula. Next, the ROC metric used for optimization is selected from the options of threat score, optimal point, true skill statistic, and precision. Next, the user selects the minimum, maximum, and spacing values to use as potential threshold values. These values can also be thought of as the range of values on the x and y axis that HydroMet analyzes during the optimization. Then, the user either uses all the data for optimization or chooses to split the data to some percentage as a training (calibration) dataset and the remainder as a test (evaluation) dataset. The data can be split in two ways, either the landslide occurrences are split by a specified percentage, and the hydrologic data are organized to the correct time or the hydrologic data are split by some percentage and the landslides are organized accordingly. Finally, the user inputs an iteration number for ease of documentation of the results, since it is presumed that users may conduct several iterations of optimizations before settling on a threshold to use for the intended LEWS. Once ConfigHydroMet.py and HydroMet.py are executed, the elapsed time for calculation of the four daily threshold combinations from the example configuration script will be displayed (Figure 6).

Figure 6. Example of the elapsed time of each daily threshold calculation from the repository. Once all four are calculated, the optimal threshold based on the desired ROC metric is shown, as seen above. For this analysis, the optimal threshold was identified with a 2-day cumulative rainfall and a 2-day antecedent saturation.

The example from the repository will produce a bilinear 2-day precipitation (P2) and 2-day antecedent saturation (S2) threshold optimized for Threat Score, shown with yellow lines (Figure 7). The output text files described in Table 2 will populate in the respective folders (Figure 8).

The multi-threshold analysis is based on a dual optimization of optimal point and threat score. The thresholds are calculated using the same ROC metric equations defined above for the optimal point and threat score. Since the threat score metric will favor more missed alarms and the optimal point will favor more false alarms, the multi-threshold allows for a range of landslide-possible and landslide-probable thresholds. When selecting the dual threshold option, the discretization of the optimization search is predefined, so the user does not need to further specify the range of search values for this step. The example problem from Mukilteo produces dual P3:S1 thresholds (Figure 9).

Pore pressure analysis calculates the percentage of landslides that occur with a positive pore pressure. This analysis can be used in areas where landslide timing is known in order to determine the correlation between a landslide occurring with positive pore pressure. The percentage of landslides occurring with a positive pore pressure is then printed to the screen. Figure 10 shows an example of the pore-pressure time series plots for Mukilteo.



Landslide Threshold Analysis for Mukilteo using Threat Score-Full Data

Figure 7. HydroMet output plot showing the optimal threshold for the example in Mukilteo, Washington. The yellow square shows the optimal rainfall and saturation thresholds. The blue Xs are events where there was no landslide and the threshold would have predicted no landslide. The green circles are where a landslide occurred, and the threshold would have correctly predicted a landslide based on the precipitation and saturation values at that time. The purple triangles and red squares are false alarms and missed alarms, respectively, where the threshold incorrectly assumed or missed a landslide event.

```
You have selected "No" to a Multi Threshold Analysis. If this is incorrect,
please try again and input "Yes" exactly as written here.
You have selected "No" to a Pore Pressure Analysis. If this is incorrect, please
try again and input "Yes" exactly as written here.
You have selected "No" to a Proxy Threshold Analysis. If this is incorrect,
please try again and input "Yes" exactly as written here.
```

Figure 8. After the threshold analysis is shown in the users' Python interface, a print statement will appear. This statement will tell the user that they have selected 'No' on performing any of the other analyses available. If these other options are selected as 'Yes,' their respective results will appear and the print statement will not, indicating that proper analysis was performed.



Figure 9. HydroMet output plot showing dual-threshold analysis for Mukilteo, Washington, using a P3:S1 daily threshold. For both skill statistics, threat score and optimal point, a range of threshold values were tested to find the optimal threshold, shown above. Only landslide events and non-landslide events are shown since the determination of a missed alarm is dependent upon which skill statistic is used.



Figure 10. HydroMet output plot showing correlation between landslide occurrences and positive pore water pressure. The total length of time for the data set is shown with the occurrence of a landslide as a dot. The red line signifies where pore pressure is equal to zero, therefore any value above the line would be a positive pore-water pressure. The top plot shows the full range of pore-water pressure values and the bottom is a zoomed in look at the zero-pressure line, highlighting positive and negative values.

Finally, proxy threshold analysis uses positive pore pressure as a proxy for landslide occurrence where timing of landslides is unknown. This proxy allows a hydrologic threshold to be calculated. This option is very similar to the standard application of HydroMet using landside data, where user input is used for many of the same variables in the calculation. The minimum and maximum for both the rainfall days and saturation days is acquired followed by the choice of a linear or bilinear threshold and the choice of ROC metric. The range and spacing of threshold values is given and an iteration number is given for documentation. Unlike the landslide threshold based on known landslide data, this section uses the occurrence of the positive pore pressure as a proxy for a landslide. From this, the same calculations are performed to find the optimal threshold. Figure 11 shows a comparison of a true landslide analysis and a proxy analysis using pore water pressure. Both analyses use a range of daily thresholds of a range of 1–2 days cumulative rainfall and 1–2 days of antecedent saturation (P1–2: S1–2) optimized with threat score. The optimal threshold is calculated from the four daily combinations with the highest performing threshold presented.



Figure 11. Two HydroMet output plots showing a comparison of using observed landslide timing versus positive pore-water pressure as a proxy for landslide occurrence. Threat score was the statistic used for optimization and both analyses tested a range of 1–2 days precipitation and 1–2 days of saturation to give four daily threshold options to optimize. (a) Using actual timing of landslide occurrence, a daily threshold of P2:S2 is optimal with a threshold of P2 > 32 mm and S2 > 0.8. (b) When positive pore water pressures are a proxy for landslide occurrence, the optimal threshold is P1:S2 with P1 > 10 mm and S2 > 0.8.

6. Summary

The HydroMet program was developed to determine landslide thresholds based on cumulative precipitation and antecedent saturation data. Traditional thresholds were only informed with rainfall intensity-duration, but the inclusion of soil hydrology, as implemented in HydroMet, is effective for increasing the accuracy of landslide forecasting thresholds. Along with the use of precipitation and saturation, HydroMet incorporates many variable inputs that allow for flexibility in the desired parameters and resulting thresholds. Examples from the repository are shown to provide useful insight into the results provided from the code, but these should not be considered definitive in terms of the possible scope of applications for HydroMet. In the basic threshold analysis, the inputs include a variable duration for antecedent data, a discretization of threshold search parameters, the choice of skill statistic used in optimization, and the choice between linear or bilinear threshold formulas. In addition to these input decisions previously used,

HydroMet allows for a split-sampling of data to use as a train and test subcategory and the option to add uncertainty to rainfall prediction. Along with a single prediction, HydroMet allows users to develop a dual threshold, explore the relation between pore water pressure and landslide occurrences, and finally to use these positive pore water pressures as a proxy for landslide occurrence for threshold optimization when landslides events are unknown.

The Git repository currently contains the HydroMet and ConfigHydroMet Python files, along with detailed operational instructions on the required input data format, variable input and descriptions and output folder contents. The original source data [21] and current configuration script provided allows the user to quickly set up the program and run a basic analysis to gain familiarity with how the program works. The user can then adjust the threshold variables, formulations, and other assumptions to see how the resulting outputs change or use their own hydrometeorological and landslide data to perform threshold analyses for their area of focus. HydroMet can help inform crucial landslide early warning systems for areas of high landslide potential. The software provides tools to dynamically visualize optimal thresholds that minimize the number of missed and false alarms.

Author Contributions: Conceptualization, B.B.M. and M.D.M.; methodology, J.L.C. and M.D.M.; software, J.L.C. and M.D.M.; validation, J.L.C.; formal analysis, J.L.C. and M.D.M.; investigation, B.B.M. and M.D.M.; resources, B.B.M. and R.L.B.; data curation, B.B.M.; writing—original draft preparation, J.L.C. and B.B.M.; writing—review and editing, J.L.C. and B.B.M.; visualization, J.L.C.; supervision, B.B.M. and R.L.B.; project administration, R.L.B. and B.B.M.; funding acquisition, R.L.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data used in the Mukilteo example application of HydroMet can be found in the paper by Smith et al. and is available at: https://doi.org/10.5066/F7NZ85WX (accessed on 23 June 2021).

Acknowledgments: We are grateful to Leo De Vita and Claudia Meisina for organizing this special issue on rainfall-induced shallow landslides: modeling and warning. We are grateful to Dianne Brien for review of the HydroMet software and an earlier version of this paper. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Caine, N. The rainfall intensity-duration control of shallow landslides and debris flows. Geogr. Ann A Phys. Geogr. 1980, 62, 23–27.
- Keefer, D.K.; Wilson, R.C.; Mark, R.K.; Brabb, E.E.; Brown, W.M.; Ellen, S.D.; Harp, E.L.; Wieczorek, G.F.; Alger, C.S.; Zatkin, R.S. Real-time landslide warning during heavy rainfall. *Science* 1987, 238, 921–925. [CrossRef] [PubMed]
- 3. Guzzetti, F.; Peruccacci, S.; Rossi, M.; Stark, C.P. The rainfall intensity–duration control of shallow landslides and debris flows: An update. *Landslides* **2008**, *5*, 3–17. [CrossRef]
- 4. Segoni, S.; Piciullo, L.; Gariano, S.L. A review of the recent literature on rainfall thresholds for landslide occurrence. *Landslides* **2018**, *15*, 1483–1501. [CrossRef]
- Mirus, B.B.; Morphew, M.D.; Smith, J.B. Developing hydro-meteorological thresholds for shallow landslide initiation and early warning. *Water* 2018, 10, 1274. [CrossRef]
- 6. Mirus, B.B.; Becker, R.; Baum, R.L.; Smith, J.B. Integrating real-time subsurface hydrologic monitoring with empirical rainfall thresholds to improve landslide early warning. *Landslides* **2018**, *15*, 1909. [CrossRef]
- Thomas, M.A.; Mirus, B.B.; Collins, B.D. A physics-based approach to identify thresholds for rainfall-induced shallow landsliding. *Geophys. Res. Lett.* 2018, 45, 9651–9661. [CrossRef]
- 8. Thomas, M.A.; Collins, B.D.; Mirus, B.B. Assessing the feasibility of satellite-based thresholds for hydrologically driven landsliding. *Water Resour. Res.* 2019, 55, 9006–9023. [CrossRef]
- 9. Thomas, M.A.; Mirus, B.B.; Smith, J.B. Hillslopes in humid-tropical climates aren't always wet: Implications for hydrologic response and landslide initiation in Puerto Rico. *Hydrol. Processes* **2020**, *34*, 4307–4318. [CrossRef]

- Segoni, S.; Rosi, A.; Lagomarsino, D.; Fanti, R.; Casagli, N. Brief communication: Using averaged soil moisture estimates to improve the performances of a regional-scale landslide early warning system. *Nat. Hazards Earth Syst. Sci.* 2018, 18, 807–812. [CrossRef]
- 11. Wicki, A.; Lehmann, P.; Hauck, C.; Seneviratne, S.I.; Waldner, P.; Stähli, M. Assessing the potential of soil moisture measurements for regional landslide early warning. *Landslides* **2020**, *17*, 1881–1896. [CrossRef]
- 12. Segoni, S.; Rosi, A.; Rossi, G.; Catani, F.; Casagli, N. Analysing the relationship between rainfalls and landslides to define a mosaic of triggering thresholds for regional-scale warning systems. *Nat. Hazards Earth Syst. Sci.* 2014, 14, 2637–2648. [CrossRef]
- 13. Piciullo, L.; Gariano, S.L.; Melillo, M.; Brunetti, M.; Peruccacci, S.; Guzzetti, F. Definition and performance of a threshold-based regional early warning model for rainfall-induced landslides. *Landslides* **2017**, *14*, 995–1008. [CrossRef]
- Baum, R.L.; Fischer, S.J.; Vigil, J.C. THRESH—Software for Tracking Rainfall Thresholds for Landslide and Debris-Flow Occurrence, User Manual: U.S. Geological Survey Techniques and Methods, Book 14. 2018. Available online: https://doi.org/10.3133/ tm14A2 (accessed on 23 June 2021).
- 15. Melillo, M.; Brunetti, M.T.; Peruccacci, S.; Gariano, S.L.; Roccati, A.; Guzzetti, F. A tool for the automatic calculation of rainfall thresholds for landslide occurrence. *Environ. Modell. Softw.* **2018**, *105*, 230–243. [CrossRef]
- 16. Bogaard, T.; Greco, R. Invited perspectives: Hydrological perspectives on precipitation intensity-duration thresholds for landslide initiation: Proposing hydrometeorological thresholds. *Nat. Hazards Earth Syst. Sci.* **2018**, *18*, 31–39. [CrossRef]
- 17. Swets, J.A. Measuring the accuracy of diagnostic systems. *Science* **1988**, 240, 1285–1293. [CrossRef] [PubMed]
- 18. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, 27, 861–874. [CrossRef]
- 19. Chleborad, A.F.; Baum, R.L.; Godt, J.W.; Powers, P.S. A prototype system for forecasting landslides in the Seattle, Washington, area. *Rev. Eng. Geol.* 2008, *20*, 103–120.
- 20. Baum, R.L.; Godt, J.W. Early warning of rainfall-induced shallow landslides and debris flows in the USA. *Landslides* **2010**, *7*, 259–272. [CrossRef]
- Smith, J.B.; Baum, R.L.; Mirus, B.B.; Michel, A.R.; Stark, B. Results of Hydrologic Monitoring on Landslide-Prone Coastal Bluffs near Mukilteo, Washington, U.S. Geological Survey Data Release. 2017. Available online: https://doi.org/10.5066/F7NZ85WX (accessed on 23 June 2021).