

# Anticipating runoff from farm fields

Wesley Brooks

## 1 Executive Summary

In this report, we discuss a model to predict runoff due to rain falling on farm fields. The analysis is based on measurements of runoff during and after rain events at five Wisconsin farms.

First, we discuss a model that makes predictions based only on the soil moisture, which is easily measured with an inexpensive instrument. We conclude that a single model performs well at all five farms, though we do not know how well it would generalize to new sites. A more complex model that, in addition to soil moisture, uses the amount of precipitation as a predictor is discussed next. That model uses linear discriminants to predict whether or not a rainfall event will produce runoff. The linear discriminants are generated by weighted regression.

The more complicated model improves on the predictive performance of the simpler model, mostly by reducing the number of false positives; there is also some improvement in the number of true positives. In this case, there is some evidence that we should have a separate model for each farm, but the evidence is not conclusive. This model might not generalize as well as the simpler model, but should still outperform that model in predicting when runoff will occur.

## 2 Report

### 2.1 Goal

Runoff from farm fields pollutes Wisconsin's lakes and rivers by washing soil, fertilizer, and chemicals into the waterways that drain the fields. The damage is worst when the runoff occurs immediately after manure, fertilizer, or chemicals have been applied to the field. It's also costly for the farmer to treat a field and then have that effort washed away. So it is a win-win scenario if farmers can avoid treating their fields when they know that an impending storm will cause rain water to run off their fields.

Runoff will occur when too much rain falls too quickly to be absorbed by the soil. This is a function of the storm's intensity and of the initial condition of the soil. This project seeks to produce a predictive tool to advise farmers when to expect runoff, based on the weather report and on the reading from an inexpensive soil-moisture probe. In order to make sure it gets used, the tool must be as simple as possible.

Each field is different - for instance some fields are "tiled", which means that a porous material is installed below the soil to increase the rate of drainage. Ideally, the local factors like tiling that promote or impede drainage will work through their effect on soil moisture, which would allow one tool to predict runoff from fields that vary in their hydrologic structure. In order to decide whether to recommend a single global tool or several locally-adapted tools, we need to test whether the best predictive tool differs significantly between farms.

## 2.2 Data

Five Wisconsin farms (Pagel, Saxon, Koepke, Riecher, and Pioneer) participated in this study. The streams that drain each farm were identified, and a flow gauge was installed in each to measure the surface water runoff from the farm. Care was taken to identify all possible drainage streams (each farm is drained by between one and three streams), so no runoff escaped detection. The gauges were installed in channels that are dry except when they carry runoff, so we are confident that any flow measured represents runoff from the fields and not normal stream flow, with exceptions at Koepke Farm noted below.



Figure 1: The dots mark the farms in this study.

Rainfall was measured at a single bucket-type rain gauge on each farm (at least at Koepke and Riecher Farms). These gauges work by counting the number of times that the weight of accumulated rain tips a bucket whose volume is known. Monitoring equipment recorded the time of each bucket tip, so we know when the rain began and ended and can calculate the intensity of rainfall by the time between tips.

At each farm, each rain event is an entry in the data set. A rain event begins when the bucket is first tipped and ends when the bucket goes two hours without being tipped (the two hours are not included in the rain event, but if the rain resumes after a delay of, say, 90 minutes, then it counts as a continuous event). During this study, the longest rain event lasted more than 36 hours, while the shortest lasted about seven minutes. Events were eliminated if less than 0.1 inches of rain fell, or if the ground was frozen at the time (frozen ground makes it inevitable that any rain will cause runoff, so those events are not an interesting prediction problem.)

Land surveys were used to find the total area drained by the gauged streams. Multiplying the number of acres drained by the streams by the number of inches of rain that fell at the rain gauge produces an estimate of the volume of rain that fell on each farm during each event. Total flow at the stream gauges divided by the total volume estimates the proportion of rain that ended up as runoff. In a few cases at Koepke Farm, this proportion was surprisingly high - including one event where the proportion was greater than unity. The likely explanation is that there is a spring at Koepke Farm that flows only under very wet conditions and that this spring was activated by the biggest rain events, so more water flowed than fell.

Farm	Begin	End	No. Events	No. With Runoff
Pagel	6/2006	10/2009	148	15
Saxon	5/2007	9/2010	158	27
Koepke	5/2005	9/2008	181	27
Riecher	3/2004	9/2010	354	52
Pioneer	4/2003	9/2010	426	83

The data is plotted in Figure 2, which shows that runoff tends to occur when the ground is already wet, or when a lot of rain falls.

After the data had been collected, the soil moisture probe at Pioneer was found to be miscalibrated, and the measurements of soil moisture from that probe were decreased by five percent (e.g. from 30% to 25%) to account for the error.

Since snow accumulates on the gauge during the winter, some of what looks like rain is actually the accumulated snow melting in spring. Tips of the bucket that seemed to be caused by melting snow were removed from the data set.

The intensity of each event was measured by the greatest amount of rain that fell in a five

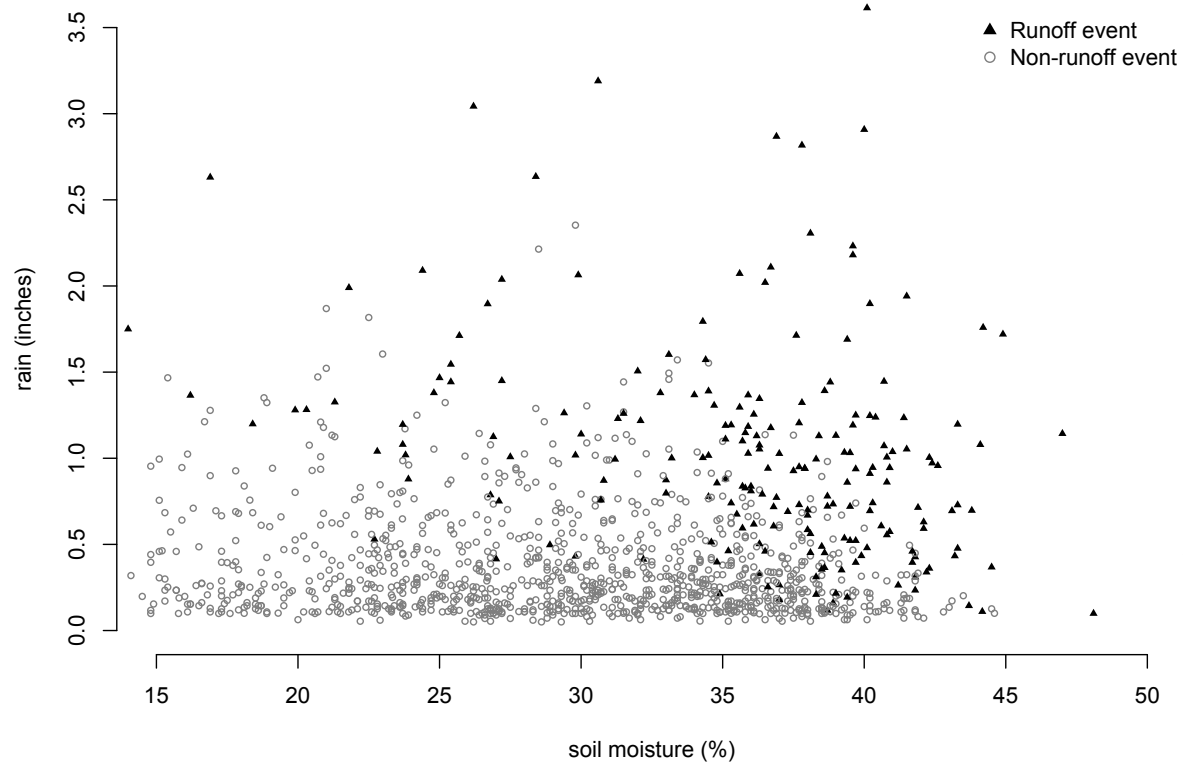


Figure 2: The soil moisture and precipitation data for all five farms, coded by whether or not the storm resulted in runoff.

minute, ten minute, fifteen minute, thirty minute, and one hour period during the event. While intensity might be a strong predictor of runoff, it cannot be used in the predictive tool because it is unknown until the storm happens, which is too late for use in the decision-making process.

### 2.3 Analysis based on soil moisture

In the interest of making the runoff-predicting tool as simple to use as possible, we focus on predicting whether or not runoff will occur, rather than on predicting the amount of runoff. Intuitively, one expects that runoff is more likely when more rain falls, and when the soil moisture is greater to begin with. The plots in figure ?? tell us that this is indeed the case at each of the farms.

### 2.3.1 Breakpoints

Mr. Wunderlin's first request was that I decide whether some existing work on the project was valid. In order to decide what value of soil moisture should form the decision boundary between predicting runoff or no runoff, he had fit a piecewise linear regression model (Figure 3), where the decision boundary was the value of soil moisture that minimized the mean of the standard error of the two regression lines, with the two standard errors weighted by the number of observations covered by the respective regression line:

$$\text{Optimization criterion} = n_{\text{lower}} \sqrt{\frac{\text{SSR}_{\text{lower}}}{n_{\text{lower}} - 2}} + n_{\text{upper}} \sqrt{\frac{\text{SSR}_{\text{upper}}}{n_{\text{upper}} - 2}}$$

Where SSR is the sum of squared residuals,  $n$  is the number of observations covered by the regression line, and upper or lower refers to which soil moisture regime is covered by the regression line. Note that within each soil moisture regime the residuals are squared then summed, but the square root is taken before averaging between the two regimes. Eliminating the square-root step results in a procedure to minimize the sum of squared errors, which would be more in line with standard practice but would lead to somewhat different breakpoints.

My recommendation is to change the minimizing criterion to be the sum of absolute errors over the training set. The change from the current method is minor, but the procedure is more standard and is more easily communicated. For the rest of this section, breakpoints are chosen to minimize the sum of absolute errors. The splits under the new method are almost the same as under the original method:

**Deciding whether to split by farm** There are two ways to decide whether each farm needs a unique model, or whether a single model can fit the data across all the farms. We can use the binomial deviance to decide whether the farm term explains enough of the observed variation

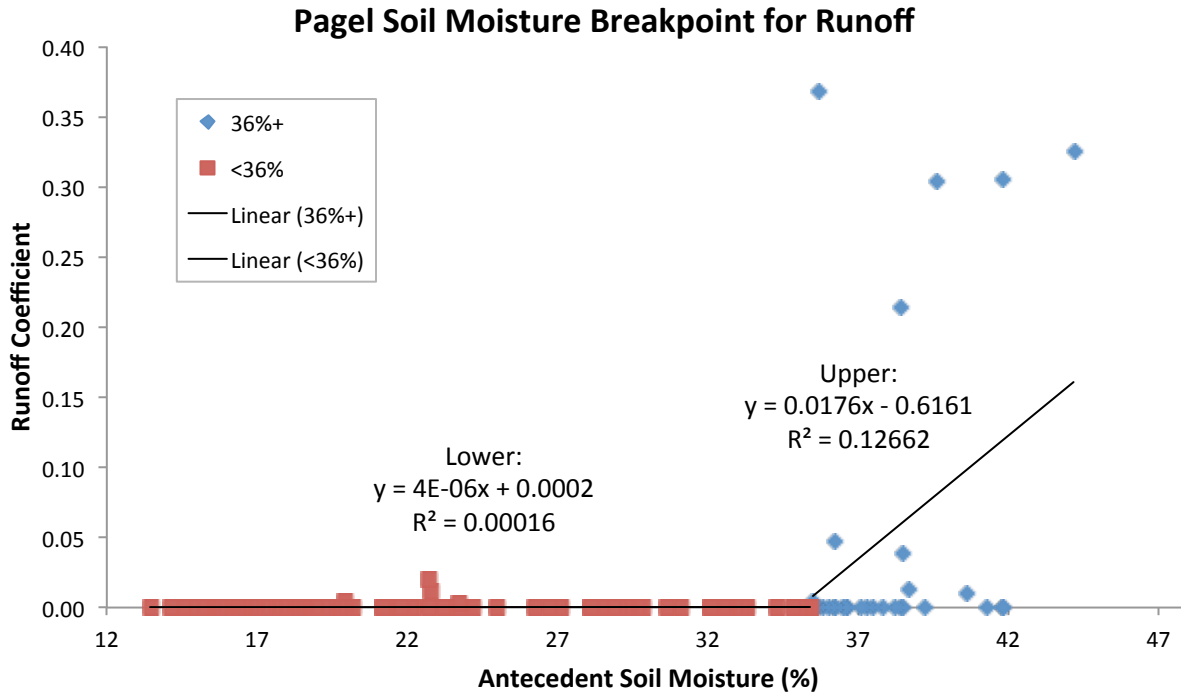


Figure 3: The piecewise-linear regression model used at the Pagel Farm site.

to include it in the model, or we can use cross-validation to test whether there is a meaningful difference between the predictive performance of a model with the farm term and a model without it.

**Binomial deviance** Our model predicts whether or not runoff will occur. This is a binary decision, and the decrease in binomial deviance measures the improvement in model fit if new explanatory variables are included. The deviance is the difference in the log-likelihood between two models. Since the deviance asymptotically follows a chi-squared distribution, it can be used to test the null hypothesis that including the farm as a parameter in the model does not improve the model's fit. The log-likelihood of a binomial model is calculated as follows:

$$\text{Log-likelihood} = x_0 \log \frac{x_0}{n_0} + (n_0 - x_0) \log \frac{n_0 - x_0}{n_0} + x_1 \log \frac{x_1}{n_1} + (n_1 - x_1) \log \frac{n_1 - x_1}{n_1}$$

Farm	Std. error	Abs. errors
Koepke	41.5	41.3
Pagel	35.5	36.9
Pioneer	35.5	34.4
Riecher	35.5	35.5
Saxon	35.5	33.0
Overall	35.5	35.3

Table 1: Soil moisture breakpoints computed using the weighted average standard error (as used by Mr. Wunderlin in the original work), and the sum of absolute errors (as recommended here).

where  $x_0$  is the number of runoff events when the soil moisture is below the breakpoint,  $n_0$  is the total number of observations from below the breakpoint,  $x_1$  is the number of runoff events when the soil moisture is above the breakpoint, and  $n_1$  is the total number of observations above the breakpoint. Then the deviance is:

$$D = -2 \times (\text{Log-likelihood}_{\text{model with farm terms}} - \text{Log-likelihood}_{\text{model without farm terms}})$$

Since there are five farms in the study, there are four degrees of freedom to the farm terms, and the test for a farm effect is based on the chi-squared distribution with four degrees of freedom. This test is summarized in Table 2, and we conclude that there is no need to include the farm as a parameter in the model.

Model	Log-lik	d.f.	Deviance	p-value
No farm terms	974.15			
With farm terms	972.01	4	4.28	0.369

Table 2: The chi-squared test for farm effect based on the binomial deviance has a p-value of 0.369, which indicates that there is no evidence that each farm needs a unique model.



**Cross-validation** Another way of looking at the model’s accuracy is to set aside some observations before the model-building process, then using the model to predict the outcome of those observations. This process is called *cross validation*, and it is a powerful way to measure a model’s predictive performance. In this section we compare the performance of models built using farm as a predictor to that of those built without terms for farm. Ten-fold cross validation (splitting the data into ten sections and predicting each one by cross-validation). Since the split into ten sections is random, the process was repeated 200 times to see whether the results are sensitive to the way that the data is split into sections. The results are in Table 3.

Model	False positives	False negatives	True positives	True negatives
No farm terms	225.5 (6.9)	77.2 (3.5)	126.6 (3.5)	836.7 (6.8)
With farm terms	239.8 (0.4)	65.0 (0.2)	138.9 (0.3)	802.3 (0.5)

Table 3: Mean and (standard deviation) of statistics summarizing the predictive performance when using one model per farm and when using a single model for all farms. Results were obtained by 200 repetitions of 10-fold cross-validation.

There are about the same number of errors overall between the two methods, so adding the farm term does not improve the model fit, and it is reasonable to use just one model to predict runoff at these five farms.

## 2.4 Analysis based on precipitation and soil moisture

Intuitively, the more rain that falls, the more likely runoff becomes. We can therefore use the amount of rain as a predictor of when runoff will occur. Since the measured rain depth is unknown until after the storm is over, in practice this method will rely on the weather forecast. However we don’t have the historical weather forecasts from the study period, so the measured rainfall was used to build the models and evaluate the method. We are assuming that the forecast is an unbiased estimate of the actual amount of rain (meaning that, even though the forecast won’t get the amount

of rain exactly right, it is just as likely to miss low as to miss high.)

### 2.4.1 Linear Discriminants

A linear discriminant is a straight line that separates the runoff events from the non-runoff events (Figure 4). It is produced by linear regression, where the outcome is coded as +1 for observations where runoff occurred and as -1 for observations where no runoff occurred. Weighted regression was used here because there are an unequal number of runoff events and non-runoff events. Non-runoff events were given a weight of one, while runoff events were given a weight of three.

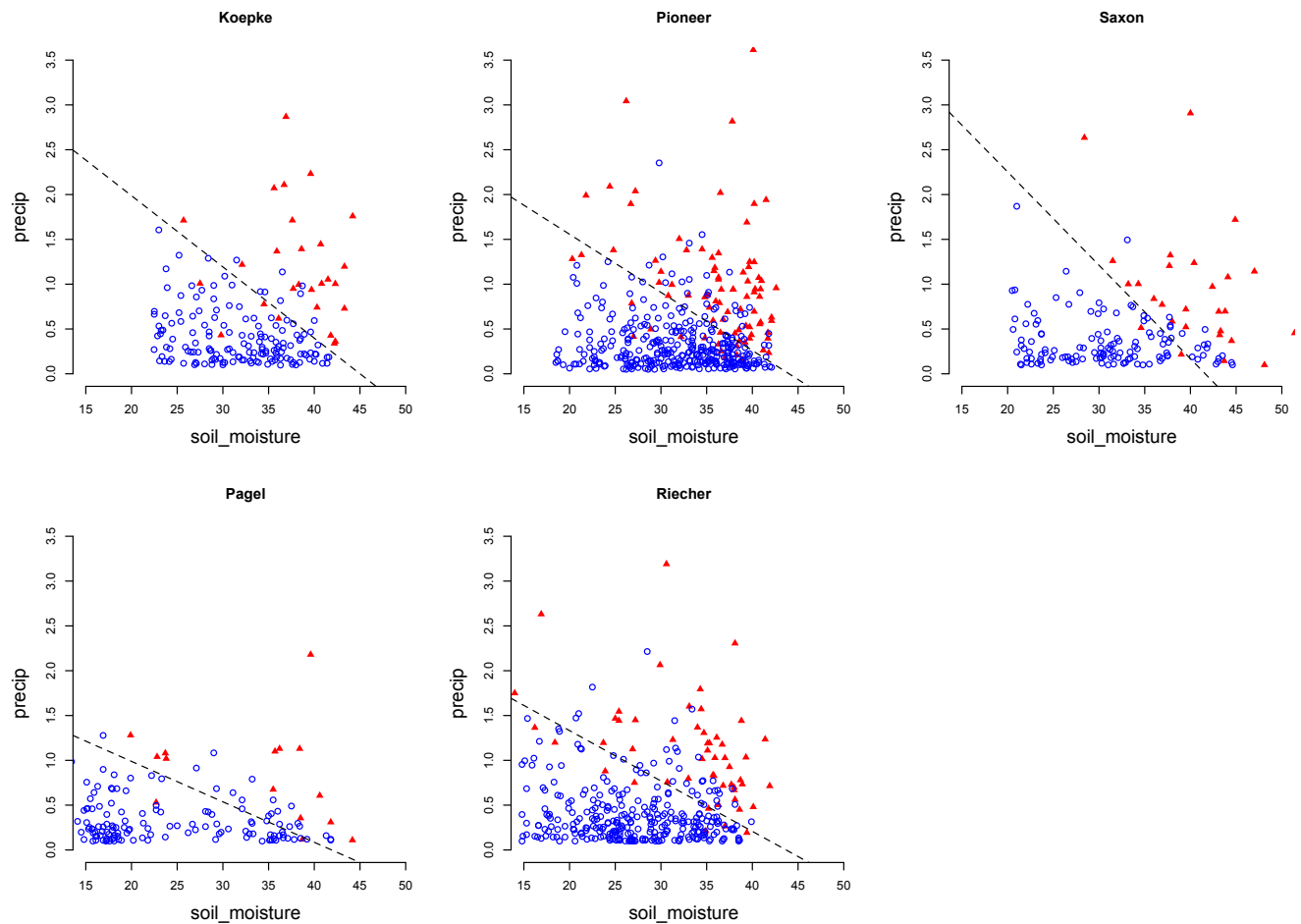


Figure 4: These are the linear decision rules at each farm individually. Triangles are runoff events, circles are non-runoff observations.

Once again, we use both the binomial deviance and cross-validation to assess whether it makes

sense to use one model at all five farms, or whether we should have one model per farm. Cross-validation is based on precipitation rounded to the nearest half inch of rainfall in order to better reflect the conditions under which the model will be used.

Model	Log-lik	d.f.	Deviance	p-value
No farm terms	699.1			
With farm terms	689.7	4	19.0	0.0008

Table 4: The chi-squared test for farm effect based on the binomial deviance has a p-value of 0.0008, so including the farm in the model significantly improves the model fit.

Model	False positives	False negatives	True positives	True negatives
No farm terms	64.2 (1.2)	58.6 (1.4)	145.2 (1.4)	998.0 (1.3)
With farm terms	74.1 (2.5)	58.9 (2.2)	145.0 (2.2)	988.1 (2.5)

Table 5: Mean and (standard deviation) of statistics summarizing the predictive performance when using one model per farm and when using a single model for all farms. Results were obtained by 200 repetitions of 10-fold cross-validation. Predictions were based on rainfall rounded to the nearest half-inch.

This time, on the question of whether we should separate the models by farm, the results are more equivocal. The binomial deviance indicates that a separate model for each farm better fits the data than does a single model for all five farms. But no real improvement in predictive power is apparent from the cross-validation. In this case, it is probably OK to use just one model for all five farms, but we should not assume that it will generalize well to new farms beyond the five in this study.

### 3 Conclusion

Finally, we combine Table 3 and Table 5 to compare the predictive power of a model that uses only the soil moisture as a predictor to a model that uses both soil moisture and precipitation (rounded to the nearest half-inch) as predictors. In both cases, we use a single global model at all five farms.

Model	False positives	False negatives	True positives	True negatives
Soil moisture only	225.5 (6.9)	77.2 (3.5)	126.6 (3.5)	836.7 (6.8)
Soil moisture and precip	64.2 (1.2)	58.6 (1.4)	145.2 (1.4)	998.0 (1.3)

Table 6: The model using both the soil moisture and precipitation to make predictions improves on the model that only uses soil moisture mainly by causing fewer false positives (while at the same time identifying a greater number number of true positives).

Based on this comparison, the best model to use for predicting runoff is the model that uses rainfall and antecedent soil moisture. That model is summarized in Table 7, where each entry indicates the soil moisture breakpoint when a certain quantity of rain (given by the row) falls on a certain farm (given by the column). The first column is the aggregate model; it can be used at all the farms and should be used when applying this method to any new farms.

We have developed models to predict runoff of rainwater from five Wisconsin farms and shown that they can accurately predict when runoff will occur based on the antecedent soil moisture and the quantity amount of rainfall. When only soil moisture is used as a predictor there seems to be no need to find a different breakpoint for each farm, but when we consider the amount of rainfall as a predictor, it begins to seem that different models perform best at each farm.

Using the amount of rainfall as a predictor improves performance mostly by reducing the number of false positives, but we also get an increase in the number of true positives. Since we rounded rainfall to the nearest half-inch before using it in prediction, the results should generalize well to

the sort of prediction that would be possible based on a weather forecast. Where we could run into problems is if the weather forecast consistently over- or under-estimates the expected rainfall.

Forecast rainfall (in)	Aggregate	Koepke	Pagel	Pioneer	Riecher	Saxon
0	49%	49%	55%	47%	49%	44%
0.5	39%	42%	39%	38%	39%	39%
1.0	30%	34%	23%	30%	28%	33%
1.5	20%	26%	8%	21%	18%	27%
2.0	10%	19%	*	12%	8%	21%
2.5	*	11%	*	4%	*	16%
3	*	3%	*	*	*	10%
3.5+	*	*	*	*	*	4%

Table 7: Decision thresholds for the model incorporating soil moisture and rainfall - runoff is predicted when the soil moisture exceeds the threshold in the table (found at the intersection of the farm column and the row that matches the inches of rain that are forecast). The thresholds are generated under the assumption that rainfall is known only to the nearest half-inch. Aggregate refers to the global model; the others are specific to one farm. An (\*) means that this amount of rain will produce runoff regardless of the antecedent soil moisture.