Global Surface Runoff estimation with ECMWF Daily Precipitation data through Google Earth Engine

Background

Surface runoff (also known as overland flow) is the flow of water occurring on the ground surface when excess rainwater can no longer sufficiently rapidly infiltrate in the soil. This can occur when the soil is saturated by water to its full capacity, and the rain arrives more quickly than the soil can absorb it, which may result in increased flood risk.

Generation of runoff data at a global scale has until recently presented many complexities due to the amount of processing and data storage necessary. With the advent of Google Earth Engine this type of analysis becomes possible as the necessary datasets on precipitation, soil, land cover and topography are available through this cloud service which also offers the possibility to process the data, converting it into maps and analysis ready data.

Elements to implement a Runoff Map

The amount of daily runoff on any given area depends on the total precipitation for that day, the previous condition of the soil (if it is already saturated from previous rains), the Hydrologic Soil Group (HSG), land cover and slope.

Generation of a Global Curve Number map

The combination HSG, land cover and slope are used to generate an index called **Curve Number (CN).** The curve number method was developed by the USDA Natural Resources Conservation Service, which was formerly called the *Soil Conservation Service* or SCS. The curve number is an empirical parameter used in hydrology for predicting direct runoff or infiltration from rainfall excess.

The runoff equation is :

$$Q = \begin{cases} 0 & \text{for } P \leq I_a \\ \frac{(P-I_a)^2}{P-I_a+S} & \text{for } P > I_a \end{cases}$$

where

- Q is runoff (m^3/m^2)
- P is rainfall (m)
- S is the potential maximum soil moisture retention after runoff begins (m)
- I_a is the initial abstraction (m), or the amount of water before runoff, such as infiltration, or rainfall interception by vegetation.

- $I_a = 0.05S$

The runoff curve number, CN, is then related

$$S = \frac{1000}{CN} - 10$$

CN has a range from 30 to 100. The lower the curve number, the more permeable the soil is, the lower runoff is generated. Runoff cannot begin until the initial abstraction has been met, the curve number aids on calculating what the Potential Soil Moisture might be.

In order to generate a Global Curve Number map we divide the work into the following steps :

1. Creating a Hydrologic Soil Group (HSG) map
2. Obtaining a land cover map (simplified from CCI Land Cover 2019 – 300 m sp)
3. Generating a slope map (flat/hilly) from USGS SRTM – 30-m sp
4. Merge the classification of the previous maps and remapping (assigning values per class) the classification into a Curve Number map.

1. Generate Global Hydrologic Soil Group map

The Hydrologic Soul Group is a classification of the soil into four classes HSG. A, HSG. B, HSG. C and HSG. D, from higher to lower infiltration rates. Sandy soils will have a higher infiltration rate than clay soils.

The clay content in the soil can be used to determine the Hydrologic Soil Group it belongs to.

The availability in GEE of a global dataset (ISRIC – SoilGrids , GEE script) containing clay content in g/kg at 6 standard depths (0, 5, 15, 30, 60 and 100 cm) allows us to derive the HSG for pixels at 250m resolution.

We have generated a script in GEE that calculates the average content of clay in the first 100 centimeters of depths (using a trapezoidal calculation as a way to calculate a weighted average of clay content for the 0-100 cm depth) and then reclassified those pixels into one of the 4 HSG classes (A, B ,C and D) using this logic.

HSG A → Clay content % < 10% → pixel value 100

HSG B → 10% >= Clay content % < 20% → pixel value 200

HSG C → 20% >= Clay content % < 40% → pixel value 300

HSG D → Clay content % >= 40% → pixel value 400

In the resulting image the class values for *A, B, C, D* will be *100, 200, 300 and 400*. These values are chosen so that we are able to generate unique class combinations with the values of Land Cover (possible values 10, 20, 30, 40, 50 and 60) and slope (0 for flat and 1 for steep) when the Curve Number map is generated.

Source of the GEE script with the algorithm:
https://code.earthengine.google.com/?scriptPath=users%2Fopenforisinitiative%2FEarthMap%3ACFI%2FHydrologicSoilGroup

Resulting GEE Asset (HSG classification):
https://code.earthengine.google.com/?asset=users/openforisearthmap/Hydrology_EarthMap/HydrologySoilGroup_SoilGrids250m

This is the visualization of the map in Earth Map LINK TO VISUALIZE

2. Simplify CCI Land Cover map into 6 Land Cover classes

The most common Land Cover product being used currently is the European Space Agency´s Climate Change Initiative Land Cover (CCI) dataset (http://www.esa-landcover-cci.org/). This dataset has been added to GEE and we have taken the last global image available from 2018 to create a "simplified" version of the land cover where the 40 original classes have been grouped into 6 Land Cover (LC) classes.

The final classes and values are:

- No Vegetation (bareland): 10
- Croplands : 20
- Herbaceous Cover : 30
- Shrubland : 40
- Tree Cover : 50
- Water Cover : 60

The mapping (conversion from a CCI class into a simplified LC class) of the classes has been performed using this logic:

ORIGINAL CCI CLASS	GROUPED LC CLASS
0 (No data)	No Data 70
10,11,12 (Rainfed cropland)	Cropland 20
30 (Mosaic cropland (>50%) / natural vegetation (tree, shrub, herbaceous cover) (<50%)	Cropland 20
40 (Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland (< 50%))	Herbaceous 30
50 (Tree cover, broadleaved, evergreen, closed to open (>15%))	Tree Cover 50
60,61,62 (Tree cover, broadleaved, deciduous, closed to open (> 15%))	Tree Cover 50
70,71,72 (Tree cover, needleleaved, evergreen, closed to open (> 15%))	Tree Cover 50
80,81,82 (Tree cover, needleleaved, deciduous, closed to open (> 15%))	Tree Cover 50
90 (Tree cover, mixed leaf type (broadleaved and	Tree Cover 50

ORIGINAL CCI CLASS	GROUPED LC CLASS
needleleaved))	
100 (Mosaic tree and shrub (>50%) / herbaceous cover (< 50%))	Herbaceous 30
110 (Mosaic herbaceous cover (>50%) / tree and shrub (<50%))	Shrubland 40
120,121,122 (Shrubland)	Shrubland 40
130 (Grassland)	Herbaceous 30
140 (Lichens and mosses)	No Vegetation 10
150,151,152,153 (Sparse vegetation (tree, shrub, herbaceous cover))	No Vegetation 10
160 (Tree cover, flooded, fresh or brakish water)	Water cover 60
170 (Tree cover, flooded, saline water)	Water cover 60
180 (Shrub or herbaceous cover, flooded, fresh-saline or brakish water)	Water cover 60
190 (Urban areas)	No Vegetation 10
200,201,202 (Bare areas)	No Vegetation 10
210 (Water Bodies)	Water cover 60
220 (permanent snow and ice)	No Vegetation 10

GEE script to generate and visualize the asset :
https://code.earthengine.google.com/?scriptPath=users%2Fopenforisinitiative%2FEarthMap%3ACFI%2FCCILC_Remapping

GEE Asset with the 2018 CCI LC map reclassified (to accelerate processing of Curve Number map) :
https://code.earthengine.google.com/?asset=users/openforisearthmap/Hydrology_EarthMap/CCILC2019_MappedToSimplifiedLC

3. CN - Generate Slope map

The slope map we use is derived from the Shuttle Radar Topography Mission (SRTM, see Farr et al. 2007) digital elevation data model (30 m / pixel). A slope calculation algorithm has been applied to the elevation model (standard GEE slope algorithm) and then we have used a very simple formula to classify the pixels into hilly or flat.
If pixel slope gradient > 3% → Hilly (value 1)
Else if pixel slope <= 3% → Flat (value 0)

The slope classification into Hilly/Flat is done in the Curve Number script (method **getSteepSlopeMap**) :
https://code.earthengine.google.com/?scriptPath=users%2Fopenforisinitiative%2FEarthMap%3ACFI%2FCurveNumber

4. CN - Merging maps and remapping to Curve Number

Once the Hydrologic Soil Group (HSG), Simplified Land Cover (LC) and slope images are generated it is possible to combine the values of each pixel to generate new classes of CN.

HSG values can be:

A-100 / B-200 / C-300 / D-400

Simplified Land Cover (LC) values can be:

10 -No Veg / 20-Crops / 30–Herbaceous / 40-Shrubs / 50–Trees / 60-Water

Slope values can be:

0-Flat / 1-Hilly

Using these values (hundreds + tens + units) we can generate unique combinations without fear of misidentifying pixels, i.e. the combination of HSG, Land Cover and slope can never be equal for two different combinations. Examples :

HSG. A (100) + Crops (20) + Flat (0) → Class 120

HSG. B (200) + Trees (50) + Hilly (1) → Class 151

This approach to generate unique combinations works if there are less than 10 classes in each different map.

Now we take the values of the pixels to apply a mapping table that assigns Curve Number value for each combination (CN in bold, followed by the combine class number):

	No Vegetation (10)		**Cropland (20)**		**Harbaceous (30)**		**Shrubland (40)**		**Tree cover (50)**		**Water cover (60)**	
HSG	**Flat (0)**	**Hilly (1)**	**Flat (0)**	**Hilly (1)**	**Flat (0)**	**Hilly (1)**	**Flat (0)**	**Hilly (1)**	**Flat (0)**	**Hilly (1)**	**Flat (0)**	**Hilly (1)**
A (100)	**92** 110	**95** 111	**57** 120	**61** 121	**41** 130	50 131	**17** 140	**39** 141	**33** 150	**39** 151	**100** 160	**100** 161
B (200)	**92** 210	**95** 211	**70** 220	**73** 221	**61** 230	**69** 231	**33** 240	**55** 241	**54** 250	**60** 251	**100** 260	**100** 261
C (300)	**92** 310	**95** 311	**77** 320	**81** 321	**75** 330	**79** 331	**67** 340	**69** 341	**69** 350	**73** 351	**100** 360	**100** 361
B (400)	**92** 410	**95** 411	**81** 420	**85** 421	**86** 430	**88** 431	**76** 440	**77** 441	**77** 450	**79** 451	**100** 460	**100** 461

The Curve Number in water cover areas is assumed to be 100 (all precipitation equals to run-off)

Curve Number GEE Script can be accessed here :
https://code.earthengine.google.com/?scriptPath=users%2Fopenforisinitiative%2FEarthMap%3ACFI%2FCurveNumber

Resulting GEE Asset is here :
https://code.earthengine.google.com/?asset=users/openforisearthmap/Hydrology_EarthMap/CurveNumber_vSoilGrids250m

Extracting Run-Off maps and data

Once we have generated a Global Curve Number (CN) map at 250m resolution (as described previously) we can proceed onto using the data to estimate run-off derived from ECMWF ERA5 LAND Precipitation (P) daily data.

Before calculating the Run-Off we produce these two maps:

- Using the Curve-Number (CN) global map, generate a **Maximum Potential Soil Moisture (S)** map. This is a map showing what is the potential of absorption of water in the soil. The Soil Moisture (S) Map is calculated with this expression: $S_{inches} = (1000/CN) - 10$
  The Soil Moisture is translated from inches to mm by using $S_{mm} = S_{inches} * 25.4$

- Using the Maximum Potential Soil Moisture map, we can derive the **Initial Abstraction (Ia)** map. This basically estimates how much it should rain before the runoff starts.
  Under average conditions the Ia is calculated by as 20% of the Maximum Potential Soil Moisture. $Ia = 0.2 * S$

To calculate the Run-Off (Q), we use hourly precipitation data from the ECMWF ERA5 Land dataset. The first hour (0) of each day contains the precipitation of the previous 24 hours in the *total_precipitation* band.

We want to consider the previous conditions of the soil. The previous rain events could make the Initial Abstraction lower than the standard Ia. To do this, we calculate **the total precipitation of the 3 previous days**. For the pixels where the precipitation of the previous days exceed the **1.5 times** de Initial Abstraction, the Initial Abstraction is reduced to just **30% of the standard Ia to account that the soil might be near capacity**.(i.e. in a day where there has been no rain in the previous days, Ia might be 100 mm, while for a day where the previous days it has rained 200mm (above the 1.5*Ia threshold) then the Ia for the single day is reduced to 30mm, after which we consider that runoff starts.

Once this Initial Abstraction ($Ia_{prevcon}$) is calculated we can proceed to calculate run-off.
Run-Off (Q) is supposed to start only if the Precipitation (P) exceeds Initial Abstraction considering previous condition ($Ia_{prevcon}$). Once the $P > Ia_{prevcon}$ then the runoff is calculated with this formula:

$$Q = \begin{cases} 0 & \text{for } P \le I_a \\ \frac{(P-I_a)^2}{P-I_a+S} & \text{for } P > I_a \end{cases}$$

where

- Q is runoff (m)
- P is rainfall (m)
- S is the potential maximum soil moisture retention after runoff begins (m)
- I_a is the initial abstraction (m) - which has been adapted to precious conditions- or the amount of water before runoff, such as infiltration, or rainfall interception by vegetation.

In Earth Map the Run-Off is shown aggregated by year or month. Earth Map aggregates the data by summing up the run-off for the days within the month or year.

The code to the GEE script for RunOff calculation :
https://code.earthengine.google.com/?scriptPath=users%2Fopenforisinitiative%2FEarthMap%3ACFI%2FRunOff

Notes for the user

This runoff map is generated using a simplified version of the CN method, which itself presents several limitations that the users should consider before using the dataset: because of its empirical origins, the CN method is restricted to certain geographic regions and land use types. Moreover, it was not designed to describe the spatial variability of runoff, it only computes the lumped surface runoff for a given rainfall event from small agricultural, forest, and urban watersheds. We extended the CN method to cover the globe, but it should be considered that this CN runoff map is not to be used in naturally bare areas: sandy bare areas, e.g. the Sahara Desert, result in high runoff which is obviously wrong.

Model finetuning

We suggest implementing the following improvements to get a higher accuracy dataset :

- Increase the land cover classes to include natural and artificial bare land surfaces.
- Integrate Land Cover maps per year instead of a static **asset.**
- Consider the total 5 days antecedent rainfall (mm) for the Antecedent Moisture Conditions (AMC) I, II and III over different seasons (growing, dry).
- Increase the slope classes (from 2 to 6):
    - Little or no slope: 0 - 3 % gradient
    - Gentle slopes: 3 - 9 % gradient
    - Moderate slopes: 9 - 15 % gradient
    - Steep slopes: 15 - 30 % gradient
    - Extremely steep slopes: 30 - 60% gradient
    - Excessively steep slopes: > 60% gradient