



Palestine Polytechnic University

Deanship of Graduate Studies and Scientific Research

Master of Civil Engineering

**An Integrated Machine Learning and GIS Approach for Optimizing  
Hydrological SCS-CN Parameters and Runoff Prediction in the West  
Bank**

**Rafiq Mazen Fareed Al-Rajabi**

Supervisor

Dr Mutaz Qafisheh

*Thesis submitted in partial fulfillment of the requirements for the degree  
Master of Civil Engineering*

September 2026

The undersigned hereby certify that they have read, examined, and recommended to the Deanship of Graduate Studies and Scientific Research at Palestine Polytechnic University:

An Integrated Machine Learning and GIS Approach for Optimizing Hydrological SCS-CN  
Parameters and Runoff Prediction in the West Bank

**Rafiq Mazen Fareed Al-Rajabi**

in partial fulfillment of the requirements for the degree of Master in Civil Engineering.

**Graduate Advisory Committee:**

Prof./Dr. ....,

(Supervisor), University (typed)

Signature: \_\_\_\_\_

Date: \_\_\_\_\_

Prof./ Dr. ....,

(Co-supervisor), University (typed)

Signature: \_\_\_\_\_

Date: \_\_\_\_\_

Prof./Dr.....,

(Internal committee member), University ( typed).

Signature: \_\_\_\_\_

Date: \_\_\_\_\_

Prof./Dr. ....,

(External committee member), University (typed).

Signature: \_\_\_\_\_

Date: \_\_\_\_\_

Thesis Approved by:

Name: \_\_\_\_\_

Dean of Graduate Studies & Scientific Research  
Palestine Polytechnic University

Signature:.....

Date:.....

رقم الإيداع عن كتاب \_\_\_\_\_

لمؤلفه \_\_\_\_\_

رقم الإيداع لدى المكتبة الوطنية الفلسطينية ( \_\_\_\_\_ )

الوصفات \_\_\_\_\_

o يتحمل المؤلف/الناشر كامل المسؤولية القانونية عن مُنتجه.

# An Integrated Machine Learning and GIS Approach for Optimizing Hydrological SCS-CN Parameters and Runoff Prediction in the West Bank

**Rafiq Mazen Fareed Al-Rajabi**

## ABSTRACT

Estimating runoff is a practical necessity in semi-arid regions, where a single intense storm can cause flooding in a season that is otherwise short of water. This study looks at the West Bank, where sparse gauging makes rainfall-runoff modelling difficult, and combines the SCS-CN method, GIS and machine learning to get more reliable runoff estimates.

We processed rainfall records from 41 stations covering 2020–2024 and used them to build a hydrological database of nine variables, the most important being maximum precipitation (Pmax) and the curve number (CN). Before trusting the data-driven results, we calibrated a HEC-HMS model against observed discharge from the Alexander Basin. Adjusting CN for antecedent moisture condition (AMC) was what allowed the simulated peak discharge to match the observed one.

Six machine learning models were then compared: Support Vector Regression (SVR), Random Forest, XGBoost, K-Nearest Neighbours (KNN), Decision Tree and Linear Regression. SVR came out on top with  $R^2 = 0.981$  and an RMSE of 0.79 mm. Random Forest was second ( $R^2 = 0.974$ ), followed by XGBoost ( $R^2 = 0.958$ ). Across the models, Pmax turned out to be the strongest driver of surface runoff.

These results suggest that machine learning can be a dependable tool for estimating runoff in ungauged basins, and the same approach could be applied to other semi-arid regions for flood risk assessment and water resource management.

Keywords: SCS-CN Method, Machine Learning, GIS, Runoff Prediction, West Bank, Antecedent Moisture Condition, Curve Number Optimization.

نهج متكامل للتعلم الآلي ونظم المعلومات الجغرافية لتحسين معايير SCS-CN الهيدرولوجية والتنبؤ بالجريان السطحي في الضفة الغربية

رفيق مازن الرجبي

### الخلاصة

يعدّ تقدير جريان المياه ضرورة عملية في المناطق شبه القاحلة، حيث يمكن لعاصفة واحدة شديدة أن تُسبب فيضانات في موسم يعاني من نقص المياه. تتناول هذه الدراسة الضفة الغربية، حيث يُصعّب نقص محطات القياس نمذجة العلاقة بين هطول الأمطار وجريان المياه، وتجمع بين طريقة SCS-CN ونظم المعلومات الجغرافية والتعلم الآلي للحصول على تقديرات أكثر دقة لجريان المياه.

قمنا بمعالجة سجلات هطول الأمطار من 41 محطة تغطي الفترة من 2020 إلى 2024، واستخدمناها لبناء قاعدة بيانات هيدرولوجية تضم تسعة متغيرات، أهمها الحد الأقصى لهطول الأمطار ( $P_{max}$ ) ورقم المنحنى (CN). قبل اعتماد النتائج المستندة إلى البيانات، قمنا بمعايرة نموذج HEC-HMS باستخدام بيانات التصريف المرصودة من حوض الإسكندر. وقد مكّن تعديل رقم المنحنى (CN) وفقاً لحالة الرطوبة السابقة (AMC) من مطابقة ذروة التصريف المُحاكاة مع ذروة التصريف المرصودة.

تمت مقارنة ستة نماذج للتعليم الآلي: انحدار متجه الدعم (SVR)، والغابة العشوائية، وXGBoost، وخوارزمية أقرب الجيران (KNN)، وشجرة القرار، والانحدار الخطي. تصدرّ نموذج SVR القائمة بمعامل تحديد ( $R^2$ ) قدره 0.981 وخطأ متوسط تربيعي (RMSE) قدره 0.79 ملم. وجاءت الغابة العشوائية في المرتبة الثانية ( $R^2 = 0.974$ )، تليها خوارزمية XGBoost ( $R^2 = 0.958$ ). وخلصت الدراسة إلى أن قيمة  $P_{max}$  هي العامل الأقوى المؤثر على جريان المياه السطحية.

تشير هذه النتائج إلى أن التعلم الآلي أداة موثوقة لتقدير جريان المياه في الأحواض غير المقاسة، ويمكن تطبيق هذا النهج نفسه على مناطق شبه قاحلة أخرى لتقييم مخاطر الفيضانات وإدارة موارد المياه.

### الكلمات المفتاحية

طريقة رقم المنحنى (SCS-CN)، التعلم الآلي، نظم المعلومات الجغرافية (GIS)، التنبؤ بالجريان السطحي، الضفة الغربية، حالة الرطوبة السابقة للمطر (AMC)، تحسين رقم المنحنى

## DECLARATION

I declare that the Master's Thesis entitled "An Integrated AI and GIS Approach for Optimizing SCS-CN Parameters and Runoff Prediction in the West Bank" is my own original work, and hereby certify that unless stated otherwise, all work contained within this thesis is my own independent research and has not been submitted for the award of any other degree at any institution, except where due acknowledgement is made in the text.

Student Name: Rafiq Mazen Al-Rajabi

Signature: \_\_\_\_\_

Date: \_\_\_\_\_

## DEDICATION

To my beloved parents, for their endless prayers, sacrifices, and unconditional support.

To my dear wife, for her boundless patience, encouragement, and love throughout this journey.

To my family and friends, who stood by my side.

## ACKNOWLEDGEMENT

I would like to thank Allah, first and foremost, for granting me the strength and capability to complete this research.

I would also like to express my sincere gratitude and deep appreciation to my supervisor, Dr. Mutaz Qafisheh, for his invaluable guidance, constructive feedback, and continuous encouragement throughout this study. His expertise and mentorship were instrumental in shaping this research.

Furthermore, I extend my sincere gratitude to Palestine Polytechnic University and all its faculty members for their educational support and for providing a conducive academic environment throughout my studies.

Finally, my deepest love and gratitude belong to my family, especially my mother, my father, and my wife, as well as all my friends who supported me. Thank you for your endless patience, unconditional love, and moral support during the entirety of my academic journey. Your prayers and encouragement kept me going, and this milestone is a reflection of your continuous support.

## LIST OF ABBREVIATIONS

| <b>Abbreviation</b> | <b>Definition</b>                              |
|---------------------|------------------------------------------------|
| AI                  | Artificial Intelligence                        |
| AMC                 | Antecedent Moisture Condition                  |
| ANN                 | Artificial Neural Network                      |
| CN                  | Curve Number                                   |
| CORINE              | Coordination of Information on the Environment |
| DEM                 | Digital Elevation Model                        |
| DT                  | Decision Tree                                  |
| GIS                 | Geographic Information Systems                 |
| HSG                 | Hydrologic Soil Group                          |
| Ia                  | Initial Abstraction (mm)                       |
| IDW                 | Inverse Distance Weighting                     |
| LMS                 | Local Meteorological Service                   |
| KNN                 | K-Nearest Neighbors                            |
| LiDAR               | Light Detection and Ranging                    |
| LSMA                | Linear Spectral Mixture Analysis               |
| LULC                | Land Use and Land Cover                        |
| MAE                 | Mean Absolute Error                            |
| ML                  | Machine Learning                               |
| MSE                 | Mean Squared Error                             |

|                  |                                               |
|------------------|-----------------------------------------------|
| NRCS             | Natural Resources Conservation Service        |
| P <sub>max</sub> | Maximum 24-hour Precipitation Depth (mm)      |
| PPU              | Palestine Polytechnic University              |
| PWA              | Palestinian Water Authority                   |
| Q                | Runoff Depth (mm)                             |
| R <sup>2</sup>   | Coefficient of Determination                  |
| RBF              | Radial Basis Function                         |
| RF               | Random Forest                                 |
| RMSE             | Root Mean Square Error                        |
| S                | Potential Maximum Retention (mm)              |
| SCS              | Soil Conservation Service                     |
| SCS-CN           | Soil Conservation Service Curve Number Method |
| SRTM             | Shuttle Radar Topography Mission              |
| SVR              | Support Vector Regression                     |
| TC               | Time of Concentration (hours)                 |
| TL               | Lag Time (minutes)                            |
| USDA             | United States Department of Agriculture       |
| WMS              | Watershed Modelling System                    |
| XGBoost          | Extreme Gradient Boosting                     |

## List of Figures

|                                                                                                    |           |
|----------------------------------------------------------------------------------------------------|-----------|
| <b>Figure 2.1 : West Bank area.....</b>                                                            | <b>17</b> |
| Figure 2.2: West Bank Digital Elevation Model STRM(DEM).....                                       | 19        |
| Figure 2.3: West Bank Annual Rainfall.....                                                         | 21        |
| Figure 2.5 West Bank Major Hydraulic Zone .....                                                    | 23        |
| Figure 2.6 West Bank watershed and sub-watershed Classification .....                              | 24        |
| Figure2.7: West Bank Soil Classification.....                                                      | 26        |
| Figure2.8 :West Bank land use Landsat (CORINE).....                                                | 28        |
| Figure 2. 9 :West Bank land use Sentinel-2 (USGS) .....                                            | 29        |
| Figure7.10. Distribution of key hydrological and physiographic variables .....                     | 79        |
| Figure7. 11. Correlation matrix of physiographic features and three candidate runoff targets. .... | 80        |
| Figure 7.12. Model comparison across all three metrics. ....                                       | 81        |
| Figure 7.13. Actual vs. predicted runoff depth for all six models on the test set. ....            | 82        |
| Figure 7.14 Five-fold cross-validation $R^2$ scores for the SVR model. ....                        | 83        |
| Figure 7.15. Feature importance from Random Forest, XGBoost, and Decision Tree. ....               | 84        |
| Figure 7.16 Residuals vs. predicted values (left) and residual distribution.....                   | 84        |
| Figure 7.17 ML-predicted peak flow vs. HEC-HMS software output. ....                               | 85        |
| Figure 7.18. ML-predicted peak flow vs. the GIS raster-based method. ....                          | 86        |
| Figure7. 19. Distribution of relative error against both benchmarks. ....                          | 86        |
| Figure 7.20. Side-by-side comparison: the actual HEC-HMS runoff depth pushed through .....         | 87        |
| Figure 21:Nahal Alexander HEC-HMS Calibration Diagnostic .....                                     | 89        |
| Figure 7.22: Multi-Method Peak Discharge Comparison by Sub-Basin.....                              | 90        |
| Figure 7.23: Nahal Alexander Simulated Hydrograph, 25–26 November 2020 .....                       | 92        |
| Figure 7.24: Nahal Alexander Simulated Hydrograph, 18–19 February 2021.....                        | 92        |
| Figure 7.25: Nahal Alexander Simulated Hydrograph, 23–24 January 2022.....                         | 93        |
| Figure C- 26: Rainfall Station Location.....                                                       | 113       |
| Figure C-27:AMC 2020, West Bank .....                                                              | 114       |
| Figure C-28:Pmax 2020, West Bank .....                                                             | 115       |
| Figure C-29 :CN 2020, West Bank.....                                                               | 116       |
| Figure C- 30 qmm 2020, West Bank .....                                                             | 117       |
| Figure C- 31 :AMC 2021, West Bank .....                                                            | 118       |

|                                          |     |
|------------------------------------------|-----|
| Figure C- 32: Pmax 2021, West Bank ..... | 119 |
| Figure C-33 CN 2021, West Bank.....      | 120 |
| Figure C-34 qmm 2021, West Bank .....    | 121 |
| Figure C-35:AMC 2022, West Bank .....    | 122 |
| Figure C-36: Pmax 2022, West Bank .....  | 123 |
| Figure C-37 CN 2022, West Bank.....      | 124 |
| Figure C-38 : qmm 2022, West Bank .....  | 125 |
| Figure C-39 :AMC 2023, West Bank .....   | 126 |
| Figure C-40:Pmax 2023, West Bank .....   | 127 |
| Figure C-41 CN 2023, West Bank.....      | 128 |
| Figure C-42 qmm 2023, West Bank .....    | 129 |
| Figure C-43:AMC 2024, West Bank .....    | 130 |
| Figure C-44:Pmax 2024, West Bank .....   | 131 |
| Figure C-45:CN 2024, West Bank.....      | 132 |
| Figure C-46:qmm 2024, West Bank .....    | 133 |

## List of Tables

|                                                                                                                                                  |     |
|--------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Table 2.1: Key water resources statistics for the West Bank .....                                                                                | 31  |
| Table 3.2: Curve Numbers for European CORINE Land Use Classification (AMC-II).....                                                               | 37  |
| Table 3.3 :Curve Numbers for USDA Land Use Classification (AMC-II). ....                                                                         | 38  |
| Table 3.4 SCS Rainfall Distributions (24-Hour Storm).....                                                                                        | 42  |
| Table 4. 5: Soil Types of the West Bank: Texture Classification and Hydrologic Soil Group ...                                                    | 52  |
| Table 6: PBIAS Performance Criteria .....                                                                                                        | 63  |
| Table 7.6: Israeli and Palestinian Meteorological Service Rainfall Monitoring Stations Used in This Study .....                                  | 67  |
| Table 6. 8 : Input Features Selected for Machine Learning Modelling .....                                                                        | 70  |
| Table 7.9 : Sub-watershed Areas and Weighted Composite Curve Numbers (European CORINE and USDA Classification, AMC-I, AMC-II, and AMC-III) ..... | 73  |
| Table 7.10 : SCS-CN Runoff Results for 2020: Q, S, and Ia per Sub-watershed.....                                                                 | 75  |
| Table 7.11 Model comparison across all three metrics.....                                                                                        | 80  |
| Table 7.12 Runoff depth (mm) modeling target .....                                                                                               | 84  |
| Table 7.13. : Muskingum routing stability diagnosis .....                                                                                        | 88  |
| Table 7. 14: Annual peak flow and runoff summary — ).....                                                                                        | 88  |
| Table 7.15: HEC-HMS calibration results — Nahal Alexander: .....                                                                                 | 90  |
| Table 7. 16: Calibrated Curve Numbers for Nahal Alexander sub-basins (AMC-I, II, III and global calibration) .....                               | 91  |
| Table 17 Comparative Analysis of Previous Studies and the Current Study .....                                                                    | 96  |
| Table 18 Distinguishing Features of the Present Study .....                                                                                      | 101 |
| Table A. 19: Station AMC classification and Pmax for 2020 (selected stations).....                                                               | 110 |

## Table of Contents

|                                                                             |      |
|-----------------------------------------------------------------------------|------|
| Abstract.....                                                               | III  |
| الخلاصه .....                                                               | IV   |
| Declaration.....                                                            | V    |
| Dedication .....                                                            | VI   |
| Acknowledgement .....                                                       | VII  |
| List of abbreviations .....                                                 | VIII |
| List of Figures .....                                                       | X    |
| List of Tables .....                                                        | XII  |
| Chapter 1: Introduction .....                                               | 1    |
| 1.1 General Introduction .....                                              | 1    |
| 1.2 Limitation of the Soil Conservation Service Curve Number (SCS-CN) ..... | 2    |
| 1.2.1 Lack of Temporal Dynamics and Intensity.....                          | 3    |
| 1.2.2 Fixed Initial Abstraction Ratio (Ia).....                             | 4    |
| 1.2.3. Spatial Homogeneity vs. Distributed Modeling.....                    | 4    |
| 1.2.4 Discrete Antecedent Moisture Condition (AMC) Jumps.....               | 4    |
| 1.2.5 Topographical and Slope Limitations.....                              | 5    |
| 1.2.6 Problem Statement.....                                                | 6    |
| 1.3 Research Aim, Objectives, and Hypotheses .....                          | 7    |
| 1.4 Research Hypotheses .....                                               | 8    |
| 1.5 Research Questions.....                                                 | 8    |
| 1.6 Significance and Contribution of the Research .....                     | 9    |
| 1.7 Study Area Overview.....                                                | 10   |
| 1.8 Thesis Organisation .....                                               | 10   |
| CHAPTER 2 HYDROLOGICAL BACKGROUND AND STUDY AREA .....                      | 12   |
| 2.1 Introduction.....                                                       | 12   |

|                                                                      |    |
|----------------------------------------------------------------------|----|
| 2.2 The Hydrological Cycle and Runoff Generation .....               | 12 |
| 2.3 Rainfall-Runoff Processes .....                                  | 13 |
| 2.4 Factors Affecting Runoff .....                                   | 13 |
| 2.5 Hydrology of Arid and Semi-Arid Regions.....                     | 15 |
| 2.6 The West Bank: Geographic and Physical Setting .....             | 15 |
| 2.6.1 Location .....                                                 | 15 |
| 2.6.2 West Bank Digital Elevation Model (DEM) .....                  | 18 |
| 2.6.3 West Bank Climate .....                                        | 20 |
| 2.7.4 West Bank Geology and Hydrogeology .....                       | 22 |
| 2.7.5 West Bank Soils.....                                           | 25 |
| 2.7.6 Land Use and Land Cover .....                                  | 27 |
| 2.8 Water Resources and Management Challenges in the West Bank ..... | 30 |
| 2.9 Previous Hydrological Studies in the Region .....                | 31 |
| 2.10 Summary .....                                                   | 32 |
| Chapter 3: THE SCS-CN METHOD.....                                    | 34 |
| 3.1 Introduction.....                                                | 34 |
| 3.2 Origin and Development of the Method .....                       | 34 |
| 3.3 Theoretical Basis and Governing Equations.....                   | 34 |
| 3.4 Curve Number Determination.....                                  | 36 |
| 3.4.1 Hydrologic Soil Groups .....                                   | 36 |
| 3.4.2 Land Use and Land Cover .....                                  | 37 |
| 3.4.3 Antecedent Moisture Condition.....                             | 39 |
| 3.5 The Initial Abstraction Ratio .....                              | 39 |
| 3.6 Composite (Weighted) Curve Number .....                          | 40 |
| 3.7 Modifications and Improvements of the SCS-CN Method.....         | 40 |
| 3.8 Definition of SCS Rainfall Distributions.....                    | 41 |
| 3.8.1 Origin and Structure of the Type 1A Distribution .....         | 42 |
| 3.9 Muskingum Channel Routing: Theory and Stability Criteria.....    | 43 |
| 3.9.1 Storage Equation.....                                          | 43 |

|                                                                  |    |
|------------------------------------------------------------------|----|
| 3.9.2 Routing Coefficients .....                                 | 44 |
| 3.10 Tools and Software for SCS-CN Application.....              | 44 |
| 3.11 Limitations of the SCS-CN Method .....                      | 46 |
| 3.12 Summary .....                                               | 47 |
| CHAPTER 4 GIS AND REMOTE SENSING IN HYDROLOGICAL MODELLING ..... | 48 |
| 4.1 Introduction.....                                            | 48 |
| 4.2 The Role of GIS in Hydrology.....                            | 48 |
| 4.3 Digital Elevation Models .....                               | 48 |
| 4.4 Watershed and Stream Network Delineation.....                | 49 |
| 4.5 Remote Sensing Data Sources .....                            | 50 |
| 4.6 Classification accuracy assessment.....                      | 51 |
| 4.7 Soil Data and Hydrologic Soil Groups .....                   | 52 |
| 4.8 Review of GIS-Based Runoff Estimation Studies .....          | 54 |
| 4.9 Summary .....                                                | 55 |
| CHAPTER 5 ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING.....      | 56 |
| 5.1 Introduction.....                                            | 56 |
| 5.2 Machine Learning Categories .....                            | 56 |
| 5.2.1 Supervised Learning .....                                  | 56 |
| 5.2.2 Unsupervised Learning .....                                | 56 |
| 5.2.3 Reinforcement Learning .....                               | 57 |
| 5.3 Regression Models for Runoff Prediction .....                | 57 |
| 5.3.1 Multiple Linear Regression.....                            | 57 |
| 5.3.2 Decision Tree .....                                        | 57 |
| 5.3.3 Random Forest .....                                        | 58 |
| 5.3.4 Extreme Gradient Boosting (XGBoost).....                   | 58 |
| 5.3.5 Support Vector Regression .....                            | 58 |
| 5.3.6 K-Nearest Neighbors .....                                  | 59 |
| 5.3.7 Linear regression.....                                     | 59 |

|                                                                                |    |
|--------------------------------------------------------------------------------|----|
| 5.4 Metaheuristic Optimization Algorithms for Parameter Calibration .....      | 60 |
| 5.5 Feature Selection and Feature Importance.....                              | 60 |
| 5.6 Model Development Workflow .....                                           | 61 |
| 5.7 Model Performance Evaluation Metrics .....                                 | 61 |
| 5.8 Model Evaluation Metrics.....                                              | 61 |
| 5.9 Python Libraries and Tools.....                                            | 63 |
| 5.10 Applications of Artificial Intelligence in Rainfall-Runoff Modelling..... | 64 |
| 5.11 Summary .....                                                             | 65 |
| CHAPTER 6 RESEARCH METHODOLOGY .....                                           | 66 |
| 6.1 Overview of the Methodological Framework.....                              | 66 |
| 6.2 Study Area and Data Sources .....                                          | 66 |
| 6.3 Computation of the SCS-CN Parameters.....                                  | 69 |
| 6.4 Construction of the Research Dataset.....                                  | 69 |
| 6.5 Data Pre-processing .....                                                  | 70 |
| 6.6 Machine-Learning Models and Hyperparameter Tuning .....                    | 71 |
| 6.7 Summary .....                                                              | 72 |
| CHAPTER 7 RESULTS AND DISCUSSION.....                                          | 73 |
| 7.1 Introduction.....                                                          | 73 |
| 7.2 Curve Number Results.....                                                  | 73 |
| 7.3 Annual SCS-CN Runoff Results (2020-2024).....                              | 75 |
| 7.4 Machine Learning Model Performance .....                                   | 78 |
| 7.4.1 Overview: Six-Target, six-Algorithm Evaluation.....                      | 78 |
| 7.4.2 Cross-Validation Stability.....                                          | 83 |
| 7.4.3 Feature Importance .....                                                 | 83 |
| 7.4.4 Residual Diagnostics.....                                                | 84 |
| 7.4.5 From Runoff Depth to Peak Discharge: SCS Conversion and Validation ..... | 84 |
| 7.4.6 Diagnostic: Isolating the Source of Error .....                          | 87 |
| 7.5 Nahal Alexander: Observed Streamflow and HEC-HMS Calibration .....         | 87 |
| 7.5.1 Nahal Alexander Muskingum Stability Diagnosis .....                      | 87 |

|                                                           |     |
|-----------------------------------------------------------|-----|
| 7.5.2 Observed Annual Flow Statistics.....                | 88  |
| 7.5.3 Calibration Performance Before Adjustment.....      | 89  |
| 7.5.4 AMC-Based CN Calibration Results .....              | 91  |
| CHAPTER 8 CONCLUSIONS AND RECOMMENDATIONS .....           | 94  |
| 8.1 Summary of Principal Findings .....                   | 94  |
| 8.2 Conclusions.....                                      | 94  |
| 8.3 Recommendations for Future Research .....             | 95  |
| 8.4 Distinctions from Previous Studies.....               | 96  |
| REFERENCES .....                                          | 103 |
| APPENDIX.....                                             | 110 |
| Appendix A: Annual Station Rainfall Data (2020–2024)..... | 110 |
| Appendix B : West Bank Map .....                          | 112 |

## Chapter 1: INTRODUCTION

### 1.1 General Introduction

The Mediterranean basin is a distinctive hydro-meteorological and ecological system of great complexity, characterized by long dry summers governed by subtropical high-pressure systems and winter precipitation controlled by westerly cyclonic fronts. For decades, civil and water-resources engineering in the region has assumed climatic stationarity, designing infrastructure based on historical Intensity–Duration–Frequency (IDF) curves. This assumption is no longer valid: climate change is altering the spatial and temporal precipitation patterns of the basin, which is now recognised as a major climate-change hotspot (IPCC, 2023). Regional climate models project a systematic decrease in total annual rainfall under the Representative Concentration Pathways RCP 4.5 and RCP 8.5, together with a concentration of the residual precipitation into high-intensity, short-duration storms that impose unprecedented hydraulic loading on natural catchments and engineered drainage networks (Stosic et al., 2024)

Topographically, the West Bank is a complex environment for hydraulic engineers. Steep gradients and a rapid descent from the central mountain ridge to the low Jordan Valley produce fast catchment response times. Prolonged, severe summer droughts bake the soil and greatly reduce the antecedent moisture condition, lowering the initial infiltration capacity (Shadeed & Almasri, 2010a). During high-intensity winter storms, the rainfall rate quickly exceeds the infiltration rate, and saturation-excess overland flow converts the majority of the rainfall volume into high-velocity surface runoff with high peak discharges downstream (Shadeed & Almasri, 2010b).

This runoff potential is estimated by the empirical Soil Conservation Service Curve Number (SCS-CN) model developed by the United States Department of Agriculture (USDA, 1986), which merges the physical parameters of a catchment into a single quantifiable measure, the Curve Number. Unplanned urban expansion and rapid land-use change have converted large areas of pervious agricultural and natural land into impervious concrete and asphalt, driving composite Curve Numbers toward their maximum and shortening the time of concentration, with the result

that municipal stormwater networks designed for lower peak flows are increasingly overloaded (Abushandi & Merkel, 2013)

In such heterogeneous terrain, the SCS-CN model requires sophisticated spatial discretization, which is achieved by integrating a Geographic Information System (GIS) with a high-resolution Digital Elevation Model (DEM) to derive slope, flow direction, flow accumulation, and automated watershed delineation (Tarboton, 1997). By superimposing topography with hydrologic-soil-group and land-use layers, a high-resolution Curve Number grid is generated, enabling precise spatial mapping of runoff-generation areas. Coupling this framework with Python-based automation and, more recently, with machine learning, transforms runoff estimation into a rapid, data-driven engineering process suitable for predictive flood early-warning and water-harvesting design under a volatile climate (Alfieri et al., 2012).

### 1.2 Limitation of the Soil Conservation Service Curve Number (SCS-CN)

The Soil Conservation Service Curve Number (SCS-CN) method, originally developed by the United States Department of Agriculture from empirical rainfall-runoff data collected on small agricultural watersheds across the United States, remains the most widely used model for estimating direct surface runoff from storm rainfall. Its appeal lies in a parsimonious structure that reduces watershed response to a single lumped parameter, the curve number, derived from tabulated combinations of hydrologic soil group and land cover. This same simplicity, however, limits the method's reliability once it is transferred outside the conditions under which it was originally calibrated, and several structural weaknesses become particularly consequential in complex, semi-arid terrain such as the West Bank.

The fixed initial abstraction ratio ( $I_a = 0.2S$ ) embedded in the standard formulation is a documented source of error under Mediterranean-type rainfall regimes dominated by short-duration, high-intensity convective storms, and regional studies have repeatedly favored lower ratios to better reproduce observed hydrographs. The model's three-tier antecedent moisture condition (AMC) classification offers only a coarse discretization of what is, physically, a continuous variable governing infiltration capacity. Treating CN as a static, spatially lumped parameter further obscures the fine-scale heterogeneity of soil, slope, and land cover typical of dissected mountainous watersheds. In data-scarce, largely ungauged basins, CN values are

commonly assigned from standard lookup tables developed for U.S. conditions with little or no local calibration against observed rainfall-runoff response, a practice that introduces systematic bias rather than reducing it.

These constraints have motivated a growing body of research that couples Geographic Information Systems (GIS) with Artificial Intelligence (AI) and machine learning to spatially resolve and empirically optimize SCS-CN parameters, offering a data-driven alternative to the conventional lookup-table approach (Li & Liu, 2017) .

#### 1.2.1 Lack of Temporal Dynamics and Intensity

An additional limitation, distinct in kind from those discussed above, concerns the mathematical form of the method itself. The SCS-CN equation is a purely volumetric transformation: it converts a single value of total storm rainfall depth into a single value of total runoff depth through the retention parameter  $S$ , without reference to how that depth accumulated over time. Two storms of identical total depth can differ substantially in temporal distribution, as when a short, high-intensity convective cell is compared with a longer event of comparatively low intensity. This distinction carries direct physical consequences: runoff generation depends on the relationship between rainfall intensity and the soil's infiltration capacity at each instant, not on cumulative depth alone. The short, intense storm typically produces a higher peak discharge and a different total runoff volume than the low-intensity storm of equal depth, yet the SCS-CN equation returns an identical estimate for both, since neither intensity nor duration enters the calculation.

This volumetric framing produces a structural inconsistency within the broader rainfall-runoff modelling sequence. Once a runoff depth has been estimated, its conversion into a discharge hydrograph typically proceeds through unit hydrograph or routing procedures that depend explicitly on the Time of Concentration ( $T_c$ ), usually computed in minutes to represent a watershed's characteristic response time. The runoff-generation component of the model is therefore insensitive to storm timing, while the routing component applied immediately afterward treats timing as its central variable. In small, steep, fast-responding catchments, where storm duration relative to  $T_c$  often governs peak discharge more strongly than total rainfall depth alone, this mismatch between a time-blind input and a time-dependent transformation constrains the physical realism of the resulting hydrograph. (Kujawiak et al., 2020).

### 1.2.2 Fixed Initial Abstraction Ratio ( $I_a$ )

The classic equation assumes a fixed initial abstraction ratio of  $I_a = 0.2S$ . Decades of field calibrations have demonstrated that this ratio is often a severe overestimation, particularly in semi-arid environments like the West Bank, where soil crusting and distinct land covers dominate. Research indicates a ratio of 0.05 or lower is more representative. Applying a static 0.2 ratio without localized calibration leads to under-predicting early-stage runoff (Mishra & Singh, 2003a).

### 1.2.3. Spatial Homogeneity vs. Distributed Modeling

Originally developed as a lumped model, the SCS-CN method assumes a uniform watershed. Using a composite Curve Number for large, heterogeneous basins obscures critical spatial variations. To resolve this, modern applications leverage robust GIS environments (e.g., utilizing ArcPy for automated spatial processing) to transition from lumped to distributed modeling, calculating pixel-by-pixel CN values based on high-resolution Digital Elevation Models (DEM) and land-use rasters (Mishra & Singh, 2003a).

### 1.2.4 Discrete Antecedent Moisture Condition (AMC) Jumps

The standard SCS-CN procedure represents antecedent soil moisture, a continuously varying state governed by recent rainfall, evapotranspiration, and drainage, through three discrete classes: AMC I (dry), AMC II (average), and AMC III (wet). Each class carries its own curve number, derived from the reference CN (II) value through a fixed conversion relationship, and the watershed's runoff response on a given day is computed using whichever class the antecedent rainfall record assigns to that day. The classification is therefore a stepped approximation of an underlying continuum: soil moisture does not, in physical reality, occupy one of three discrete states, but moves continuously between them as a function of the catchment's wetting and drying history. This mismatch between a discrete classification and a continuous physical process produces a threshold effect at the boundaries between classes. Because the governing variable, commonly the five-day antecedent rainfall total, is evaluated against fixed numerical cutoffs, two days with nearly identical antecedent moisture can fall into different AMC classes if their totals lie on opposite sides of a threshold, even by a narrow margin. The resulting curve numbers, and consequently the computed runoff depths, can then differ sharply between such days despite the

near absence of any real difference in soil-moisture state. Days whose antecedent conditions lie close to a class boundary, often described as transitional days, are therefore particularly susceptible to this effect, and the resulting estimation error is systematic: it originates in the discretization scheme itself rather than in random variability of the underlying data. A continuous or empirically fitted moisture-adjustment function would instead allow curve number to vary smoothly with antecedent rainfall, avoiding the abrupt discontinuities that a fixed three-class scheme imposes on an inherently gradual process.(Kujawiak et al. 2020)

#### 1.2.5 Topographical and Slope Limitations

The curve number values tabulated in the standard SCS-CN reference tables were derived from empirical rainfall-runoff records collected primarily on small agricultural watersheds with slopes of approximately five per cent, conditions representative of gently sloping to rolling terrain rather than the steep gradients found in mountainous catchments. The baseline equation carries no explicit slope term: curve number, and by extension the retention parameter  $S$ , is treated as invariant with respect to gradient, so that two otherwise identical land segments, one on a five percent slope and one on a considerably steeper incline, are assigned the same CN value provided their soil group and land cover match.

This omission runs counter to established infiltration theory. As slope increases, the velocity of overland flow increases correspondingly, which reduces the residence time available for water to infiltrate before it is conveyed downslope. The opportunity time for infiltration is therefore inversely related to gradient, and steep terrain systematically directs a larger share of rainfall into surface runoff than a CN value calibrated on gentler reference conditions would predict. In mountainous, semi-arid catchments where slopes substantially exceed the standard five percent baseline, dissected terrain such as the West Bank highlands being a case in point, this unadjusted framework tends to underestimate both peak discharge and total runoff volume. Slope-adjusted curve number formulations addressing this gap have been proposed in the literature, typically applying a multiplicative correction to  $CN(II)$  that scales with slope steepness beyond the reference gradient, in order to better reproduce observed runoff response in steep terrain.(Mishra & Singh, 2003a)

### 1.2.6 Problem Statement

Water scarcity is one of the major environmental and socio-economic challenges to Palestinian development (World Bank, 2024). Palestinians have less than 70 cubic meters of water available per person a year, well below the internationally accepted threshold of water poverty – 500 cubic meters a year. This is one of the lowest in the world (“West Bank and Gaza Country Climate and Development Report,” 2023). In 1995, the Oslo II Accords assigned most of the shared transboundary groundwater resources to Israel, and this was compounded by the absence of surface water infrastructure on the Palestinian side (Alawna & Garcia, 2025)

Most of the wadi runoff in the West Bank goes, untapped, into the Jordan River, the Dead Sea, or the Mediterranean. This shows that stormwater harvesting is an opportunity waiting to be tapped. Reliable estimates of runoff depth and peak discharge are important for the design of retention reservoirs, recharge facilities, and harvesting systems in these spatially diverse catchments. This study proposes estimates from a GIS-based SCS-CN parameterization framework constrained by LSTM neural networks with limited observational data.

Beyond the limitations inherent to the SCS-CN formulation itself, the application of rainfall-runoff models in the West Bank is further constrained by three interrelated technical difficulties concerning data availability. First, observed streamflow records suitable for routine model calibration are unavailable for many Palestinian wadis, owing to the absence of a continuous, permanently instrumented hydrometric network. Second, rainfall-monitoring infrastructure is disproportionately concentrated within networks operated in Israel, a dependency that produces uneven spatial coverage across Palestinian catchments and constrains local institutions' access to the meteorological data required for hydrological analysis. Third, even where data are available, the pronounced spatial heterogeneity of West Bank terrain complicates their application: Shadeed and Almasri (2010a) concluded that a lumped-parameter modeling approach is inappropriate for catchments in which soil, topographic, land-use conditions vary markedly over short distances, and that spatially distributed models are required instead to adequately represent runoff-generation processes.

Addressing these compounding constraints calls for approaches to data acquisition that extend beyond conventional gauge-based monitoring. Collaboration with local research institutions and water authorities, combined with the incorporation of remote sensing techniques, offers a

practicable means of supplementing sparse ground-based records and thereby improving the calibration and predictive accuracy of future hydrological models applied to Palestinian catchments.

### 1.3 Research Aim, Objectives, and Hypotheses

The current study aims to establish and evaluate an integrated methodological pipeline that combines the Soil Conservation Service Curve Number (SCS-CN) approach, GIS-based spatial analysis, and supervised machine learning algorithms to better estimate Curve Number parameters and improve runoff prediction accuracy and value metric discharge prediction accuracy in the West Bank watershed system.

To achieve the aforementioned aim, the following objectives have been established

- To compile and spatially analyses a comprehensive hydrological dataset in the West Bank, including: rainfall records from the Palestinian and Israeli Meteorological Service monitoring network; digital soil and land cover classifications; and topographic data derived from SRTM DEMs, for the period 2020 to 2024.
- A comparison between the European and American systems and calculating the numbers of the composite curve for 33 sub-watersheds in the West Bank using both the European Land Cover Classification (CORINE) and the USDA land use system under AMC-AMC-II, AMC-III conditions, and I. Since CORINE coverage is not available for parts of the study area located within Israeli-controlled territories, the USDA classification was adopted for these areas because it provides continuous coverage for all thirty-three targeted sub-watersheds, where the hydrological calculations were done using the American system.
- Examine several supervised machine learning regression models such as – Random Forest, Decision Tree, XGBoost, SVR, KNN, and Linear Regression – on a dataset of 160 watershed-year observations defined by nine physiographic and meteorological features and to determine the model configuration that offers the highest predictive accuracy on an independent test set.

- Development of a physically based post-processing pipeline, which transforms model-predicted runoff depths to volumetric discharge, rates to allow comparison with benchmarks from traditional hydrological software.
- Identifying key hydrological factors governing surface runoff generation in the Western Agreement—through an analysis of task significance, objectives for guiding and utilizing future data collection, and strategies for simplifying correlations.

#### 1.4 Research Hypotheses

The present study is structured around the following primary hypotheses:

H1-The integration of GIS-derived spatial parameters within the SCS-CN computational framework produces runoff estimates of greater spatial representativeness than those obtained through conventional, non-spatially distributed implementations of the method.

H2-The application of advanced ensemble and kernel-based machine learning algorithms to physically meaningful SCS-CN input features yields a statistically significant improvement in runoff prediction accuracy relative to conventional linear regression approaches.

H3-Maximum 24-hour precipitation depth ( $P_{max}$ ) and the Curve Number (CN) constitute the dominant predictors of direct runoff depth across the West Bank study area, a finding anticipated to be consistent with the theoretical underpinnings of the SCS-CN methodology

#### 1.5 Research Questions

The following key research questions guide the study:

- How do composite Curve Numbers from European CORINE and USDA land use classifications differ across West Bank sub-watersheds, and what are the differences in estimated runoff depth?

- Which machine learning algorithms are best suited for the prediction of SCS-CN runoff depths for the study watershed set, and what is their improvement over traditional linear regression?
- Which hydrological features are most important for the West Bank content for runoff generation in the West Bank according to the ML feature importance analysis?
- Can we translate ML-predicted runoff depths into volumetric discharge rates that are comparable with the results of well-known hydrological software?
- How does inter-annual variability of rainfall and AMC affect runoff generation patterns across the West Bank watershed system during the 2020-2024 study period?

### 1.6 Significance and Contribution of the Research

This study advance the SCS-CN method by treating the Curve Number as a dynamic parameter that varies with antecedent moisture conditions, integrating this variability within a geographic information system, and implementing the computation through algorithms applied across a complete set of watersheds in the West Bank.

The applied contribution is more concrete. The study produces spatially complete runoff estimates for 33 West Bank sub-watersheds across a five-year observation period, among the first datasets of this kind for the region. That gives planners a quantitative basis for stormwater design, flood risk assessment, and water-harvesting feasibility work. Support vector regression performed as a strong surrogate model for runoff prediction in this analysis that result matters practically: it gives Palestinian water resource engineers a fast, low-cost way to estimate runoff in catchments that have no gauge at all. The models and the scripts behind them are released as open-source code (Kumar et al., 2024), intended for local institutions, NGOs, and municipalities that want to use them directly rather than build something similar from scratch.

This also adds to a growing body of evidence on how well machine learning performs in data-scarce hydrological settings, a relevance that extends to other parts of the developing world facing a similar combination of occupation and political fragmentation in their own water management.

## 1.7 Study Area Overview

The West Bank constitutes the westernmost part of the Jordan River watershed and is the upland recharge zone of the Mountain Aquifer system. It has a total land area of some 5,860 km<sup>2</sup>. It is geographically located between latitude 31.3°N and 32.6°N and longitude 34.9°E and 35.6°E. To the west and north is the occupation of Palestine 1948, to the east is Jordan, and to the southeast is the Dead Sea basin. In terms of administration, the territory is divided into eleven governorates, namely: Jenin, Tubas, Tulkarem, Nablus, Qalqilya, Salfit, Ramallah and Al-Bireh, Jericho and Al-Aghwar, Jerusalem, Bethlehem and Hebron (*PWA - Policies and Strategies*, 2026).

From a topographical point of view is dominated by a central mountain spine (the West Bank highlands) with elevations exceeding 900 m above mean sea level in the Hebron Hills and reaching up to 900 m in the northern Nablus region, descending abruptly to the east towards the Jordan Valley floor at about 400 m below mean sea level. The pronounced orographic gradient is responsible for the strong west-to-east reduction in rainfall, with average annual precipitation falling from 700-750 mm in the highland areas to less than 100 mm near Jericho and the Jordan Valley. The main hydrological divide forms the boundary of drainage to the Mediterranean (western basins) and to the Jordan River and Dead Sea (eastern basins)(*Applied Research Institute – Jerusalem (ARIJ)*, 2025)

The study includes 33 subwatersheds throughout the entire West Bank with areas of about 26.8 km<sup>2</sup> (Soreq Al-Sara) to 466.4 km<sup>2</sup> (Sarida). These include large Palestinian drainage units such as Wadi Faria, Wadi Qilt, Wadi Nar, Wadi Auja, Wadi Nueima, Wadi Qana and Wadi Daraja, as well as transboundary systems such as the Alexander, Hadera, Qishon and Besor basins that extend into Israeli territory.(*PWA - Policies and Strategies*, 2026).

## 1.8 Thesis Organisation

- The structure of the thesis is as follows: Chapter 2 gives a detailed review of the hydrological background of the study area. These are the hydrologic cycle, the rainfall-runoff process, watershed morphometry, physical geography, and hydrogeology of the West Bank. A detailed

review of the SCS-CN method is presented in Chapter 3, including the theoretical background, governing equations, procedures to determine the parameters, and reported limitations. Chapter 4 reviews the role of GIS and remote sensing in hydrological modeling with emphasis on DEM-based watershed delineation, land use classification, and spatial rainfall interpolation. Chapter 5 describes the complete research methodology from data acquisition, GIS processing, CN calculation, feature engineering to ML model development. Chapter 6 presents a detailed overview of the AI and ML algorithms employed in this research. Chapter 7 presents and discusses the GIS-SCS-CN analysis results and ML model evaluation. Chapter 8 concludes with the summary of the main contributions of the research and recommendations for future research. References and appendices are included.

## Chapter 2 HYDROLOGICAL BACKGROUND AND STUDY AREA

### 2.1 Introduction

This chapter deals with the hydrological and geographical setting of the study. It begins by reviewing basic hydrological concepts related to the water cycle and runoff generation. It then describes the study area in terms of physical geography, climate, geology, hydrogeology, soils, and water resources in the West Bank. The chapter concludes with a synthesis of the main hydrological problems in water resource management in the Palestinian setting and a review of previous related studies in the region.

### 2.2 The Hydrological Cycle and Runoff Generation

The water cycle describes the continuous movement of water on, above, and below the Earth's surface, driven by sunlight and gravity. At the scale of an individual watershed, precipitation is the principal input, dividing among several competing pathways: interception by vegetation, infiltration into the soil, evapotranspiration, depression storage, and surface runoff. How much each pathway receives depends on the combined effect of climatic, topographic, pedological, and land-cover conditions (Yaseen et al., 2015).

Surface runoff arises through one of two mechanisms, distinguished by what triggers it rather than by its ultimate effect. Horton (1933), and, more recently, Slater et al. (2023), described the first: runoff that begins the moment precipitation intensity outpaces the soil's infiltration capacity. Dunne and Black (1970) described the second, saturation excess overland flow, which sets in once the soil profile is already full, independent of how intense the rainfall is. In a semi-arid catchment such as the West Bank, it is the first mechanism that matters most: rainfall intensity here routinely exceeds the infiltration rate of the region's relatively impermeable, crusted soils, so Hortonian overland flow dominates. Winter compounds the effect. Storms arrive short and intense, concentrating rainfall into narrow windows of time; the opportunity time for infiltration shrinks accordingly, and runoff generation climbs further still (Haider et al., 2024)

### 2.3 Rainfall-Runoff Processes

The transformation of rainfall into runoff at the watershed scale is not a point process but a spatial one. It integrates soil properties, land cover, topography, and antecedent moisture conditions as they vary across the catchment, and the water that eventually reaches the outlet has travelled by one of three pathways: overland flow across the surface, interflow through the unsaturated zone, or baseflow released from groundwater storage. The mix of these three pathways is what shapes the outlet hydrograph, and the mix itself is a function of the catchment. In the temporary valleys that characterize the Palestinian landscape, contributions from baseflow are small or completely absent for most of the year, so the annual hydrograph is dominated by the direct surface runoff resulting from heavy winter rains (Salter et al., 2023).

A watershed's temporal response to that rainfall is described by the Time of Concentration ( $T_c$ ), defined as the travel time required for water to move from the hydraulically most remote point of the catchment to the outlet.  $T_c$  is a function of watershed geometry, slope, and channel hydraulic properties, and it governs both the shape and the duration of the resulting flood hydrograph. Values are short in the steep upland watersheds of the West Bank, typically on the order of one to five hours, and the hydrographs that result are correspondingly flashy: peak discharge is reached quickly and the flood wave arrives at downstream reaches with little lead time, a combination that produces rapid and hazardous flood conditions in populated valleys below (Kaltch et al., 2008).

### 2.4 Factors Affecting Runoff

Runoff generation in a semi-arid watershed is not shaped by any single dominant factor but by the interaction of several, each acting on a different part of the rainfall–runoff pathway. The most important are the following.

- Rainfall characteristics. Depth, intensity, duration, and spatial extent together define the meteorological forcing that any downstream process must respond to. In the West Bank, most of the annual runoff volume is produced not by the seasonal rainfall total as such but by a small number of events: short, high-intensity convective cells and orographically

enhanced frontal systems, which concentrate large depths of rainfall into narrow windows of time (Mishra & Singh, 2003b).

- Soil properties. Soil texture, structure, and saturated hydraulic conductivity govern infiltration capacity, and with it the partitioning of rainfall between surface runoff and subsurface pathways. The NRCS classifies soils into four Hydrologic Soil Groups (A through D) on this basis. Group A soils, typically gravels and coarse sands, have the highest infiltration rates and the lowest runoff potential; Group D soils, such as heavy clays and shallow soils overlying impermeable bedrock, sit at the opposite end of the range, with low infiltration rates and correspondingly high runoff potential. Groups B and C occupy the intermediate positions (Kumar et al., 2024).
  - Vegetation and land cover. Vegetation influences the water balance in more than one way: canopies intercept rainfall before it reaches the ground, root channels enlarge preferential flow paths that raise effective infiltration, surface roughness reduces overland flow velocities, and transpiration draws water back into the atmosphere from the root zone. Land-cover conditions at the opposite end of the spectrum, urban impervious surfaces, compacted agricultural soils, and bare or crusted rock outcrops, suppress infiltration and shift the partitioning sharply toward runoff (Kujawiak et al., 2020).
  - Antecedent soil moisture. The moisture already stored in the soil profile at the onset of a rainfall event exerts a strong control on infiltration capacity and, in turn, on runoff response. Under wet antecedent conditions (AMC-III), infiltration capacity is reduced, the effective curve number rises, and a larger fraction of the storm becomes runoff. Under dry conditions (AMC-I), the reverse holds: infiltration capacity is high, the curve number falls, and less runoff is generated for the same rainfall depth (Mishra & Singh, 2003c).
  - Topography and watershed morphometry. The physical geometry of the catchment governs the speed and magnitude of the runoff response. Slope, drainage area, basin shape, drainage density, and relief together control the rate at which runoff concentrates toward the outlet and the magnitude of the resulting peak discharge (Shadeed & Almasri, 2010c).
- ### 2.5 Watershed Characteristics and Morphometry
- Watershed morphometry is the quantitative description of the geometry of watersheds and their drainage networks. The important morphometric parameters for estimating runoff are the watershed area (A), the maximum length of flow path (L), the average

slope (S), the drainage density (Dd), the elongation ratio, and the form factor. These parameters can be readily derived from DEMs in a GIS environment. In this study, watershed delineation and morphometric characterization were performed using the SRTM 30 m DEM(*OpenTopography - Shuttle Radar Topography Mission (SRTM GL1) Global 30m*, 2026), processed in the GIS Watershed Modelling System (WMS) and ArcGIS (Suseto et al., 2018).

## 2.5 Hydrology of Arid and Semi-Arid Regions

- Large imbalance between precipitation and potential evapotranspiration in arid and semi-arid regions with aridity indices ( $P/ET_o$ )  $<0.5$ . The hydrological processes in these regions are markedly different from those in humid-temperate watersheds.
- Rainfall is episodic and regionally variable with high coefficients of variation in annual total in both time and space. Surface infiltration is sometimes inhibited by soil crusting, rock fragment cover, and sparse plants, which interrupt optimum flow paths. Drainage networks are ephemeral features routing flow according to individual storm events (Haider et al., 2024).
- Such regional heterogeneity of rainfall offers a special challenge for hydrological modelling in these ecosystems. GIS-based spatial interpolation is needed to generate representative precipitation inputs to rainfall-runoff models. The link between rainfall and runoff is non-linear, and the generation of runoff is dependent on antecedent soil moisture conditions. Machine learning is a convenient way to describe such processes where traditional parameter estimation fails (Official Publications of the Headquarters, U.S. Army Corps of Engineers, 2026).

## 2.6 The West Bank: Geographic and Physical Setting

### 2.6.1 Location

The West Bank covers 5,860 km<sup>2</sup> between the Jordan River in the east and the 1967 armistice line (Green Line) in the west. The terrain is dominated by the central highland ridge, which extends in a north-south direction from the Jenin-Nablus highlands in the north through the Ramallah hill to the Hebron hill in the south. Peak elevations reach 1,027 m above mean sea level near Hebron. The highlands descend abruptly eastward along the Judean Desert escarpment to the Jordan

Valley, where the floor lies approximately 400 m below mean sea level. The western slopes are gentler towards the Mediterranean coastal Plain (*PWA - Policies and Strategies*, 2026). This unique topography from the central highlands to the Jordan valley is the dominant hydrological forcing for the West Bank watershed system, which causes an orographic enhancement of rainfall on windward slopes and a strong east-west gradient in precipitation, which directly controls the spatial distributions of CN and runoff response patterns.



**Figure 2.1 : West Bank area.**

**Based on spatial data from the Ministry of Local Government - Palestine, 2026**

### 2.6.2 West Bank Digital Elevation Model (DEM)

The topography of the West Bank is characterized by geomorphological complexity and marked spatial variation, despite its limited geographical extent. These topographic variations play a decisive role in shaping hydrological networks and directing surface runoff. Based on analyses of Digital Elevation Models (DEMs) and geomorphological characteristics, the region can be spatially classified into four main topographic zones:

- The Central Highlands:

This area forms the main geomorphological axis of the region, stretching longitudinally from Jenin in the north through Nablus and Ramallah down to Hebron in the south. The area is characterized by high terrain, with elevations exceeding 800 meters above sea level. The highest topographic point recorded is Mount Nabi Yunis in the city of Halhul in the Hebron Governorate, reaching about 1,022 meters. (*Desk Study on the Environment in the Occupied Palestinian Territories*, 2003).

- The Eastern Slopes:

This area forms a sharp transitional belt between the central highlands and the Jordan Valley. Its topography is steep: elevations fall from around 800 metres above sea level to below sea level over a comparatively short horizontal distance. Two hydrological consequences follow from this abrupt drop. The lee position of the belt with respect to prevailing moisture-bearing winds produces a pronounced rain shadow, and the steep gradient itself accelerates surface flow, channeling it eastward toward the Dead Sea basin. (*Applied Research Institute – Jerusalem (ARIJ)*, 2025).

- The Jordan Valley:

This strip extends along the eastern margin of the West Bank, running parallel to the Jordan River and the Dead Sea and functioning as the base level for the eastern drainage basins. The floor of the depression lies at approximately 430 meters below sea level, the lowest terrestrial elevation on Earth. Geomorphologically, the belt is distinguished by its relative flatness and by gentle longitudinal gradients, in marked contrast to the steep eastern slopes rising immediately to its west. (*PWA - Policies and Strategies*, 2026).

- The Western Slopes:

In contrast to the eastern slopes, this region is characterised by gentle to moderate gradients. Elevations descend gradually from the western flanks of the mountain range toward the coastal plain, typically between 100 and 300 metres above sea level. The zone is dissected by a dense network of ephemeral wadis that convey surface flow westward, feeding the western drainage basins. (State of Palestine Palestinian Central Bureau of Statistics(PCBS), 2024).

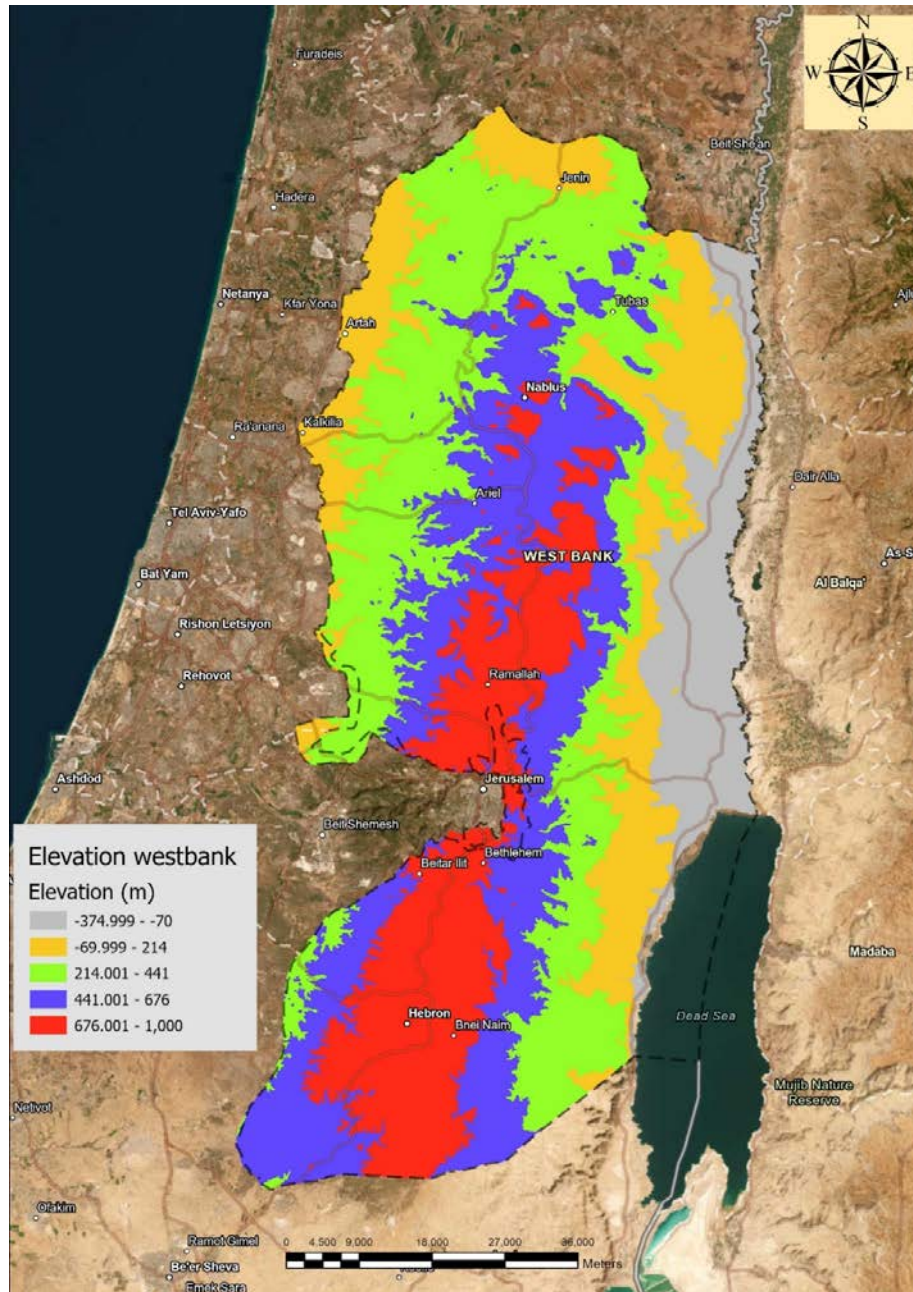


**Figure 2.2: West Bank Digital Elevation Model STRM(DEM)**  
based on spatial data from the Ministry of Local Government - Palestine, 2026

### 2.6.3 West Bank Climate

The West Bank has a vertical west to east climatic gradient across a horizontal distance of roughly 50 km. The highlands of the north and center have a hot-summer Mediterranean climate (Csa) under the Koppel classification. East of the watershed divide the climate becomes hot semi arid (BSh) on the eastern slopes and hot desert (BWh) on the floor of the Jordan Valley. Mean annual rainfall behaves the same way. Nablus Meteorological Station recorded a mean of 654 mm between 1975 and 2011. Along the highland, crest the mean falls southward to about 600 mm at Ramallah, 550 mm at Jerusalem and 530 mm in the Hebron highlands. East of the crest, the decline is much steeper: the mean is near 150 mm at Jericho and below 100 mm along the shore of the Dead Sea. The cause is orographic uplift over the highlands and the rain shadow of the Jordan Rift Valley. Rainfall is also highly seasonal. The wet season runs from October to May and accounts for almost the entire annual total, with 65 to 75 per cent falling between December and February. Rain is rare in May and September and absent from June to August.(Salameh, 2024).

Temperature follows a comparable gradient. The highlands are moderately cool, with mean January temperatures of 7–9 °C and mean July temperatures of 23–25 °C. The Jordan Valley is hot throughout the summer, with mean July values exceeding 35 °C. Potential evapotranspiration is high across the territory, rising from roughly 1,200 mm/year in the highlands to over 2,000 mm/year in the valley, and substantially exceeds mean annual rainfall across most of the region. The result is a pronounced soil-moisture deficit that persists through the dry season and depresses antecedent moisture conditions at the onset of the following wet season (Mishra & Singh, 2003b).

Internal variability is a further defining feature of the regional rainfall regime. Coefficients of variation in annual totals typically fall between 25 and 40 percent, with drought years recording less than 50 percent of the long-term mean and anomalously wet years occasionally exceeding 200 percent. This variability propagates directly into the runoff response: AMC classifications, effective CN values, and runoff-generation potential all shift markedly from one year to the next, a pattern reflected in the results of the present study over the 2020–2024 analysis period.

Across the region, annual rainfall thus ranges from around 100 mm in the Jordan Valley to approximately 800 mm in the mountain interior (Figure 2.3). Road segments crossing low-lying terrain are the most hydraulically exposed parts of the regional transport network. Proper design

means sizing drainage structures to the intensity and frequency of the extreme rainfall events the region actually sees (Ministry of Local Government, Palestine, 2025).

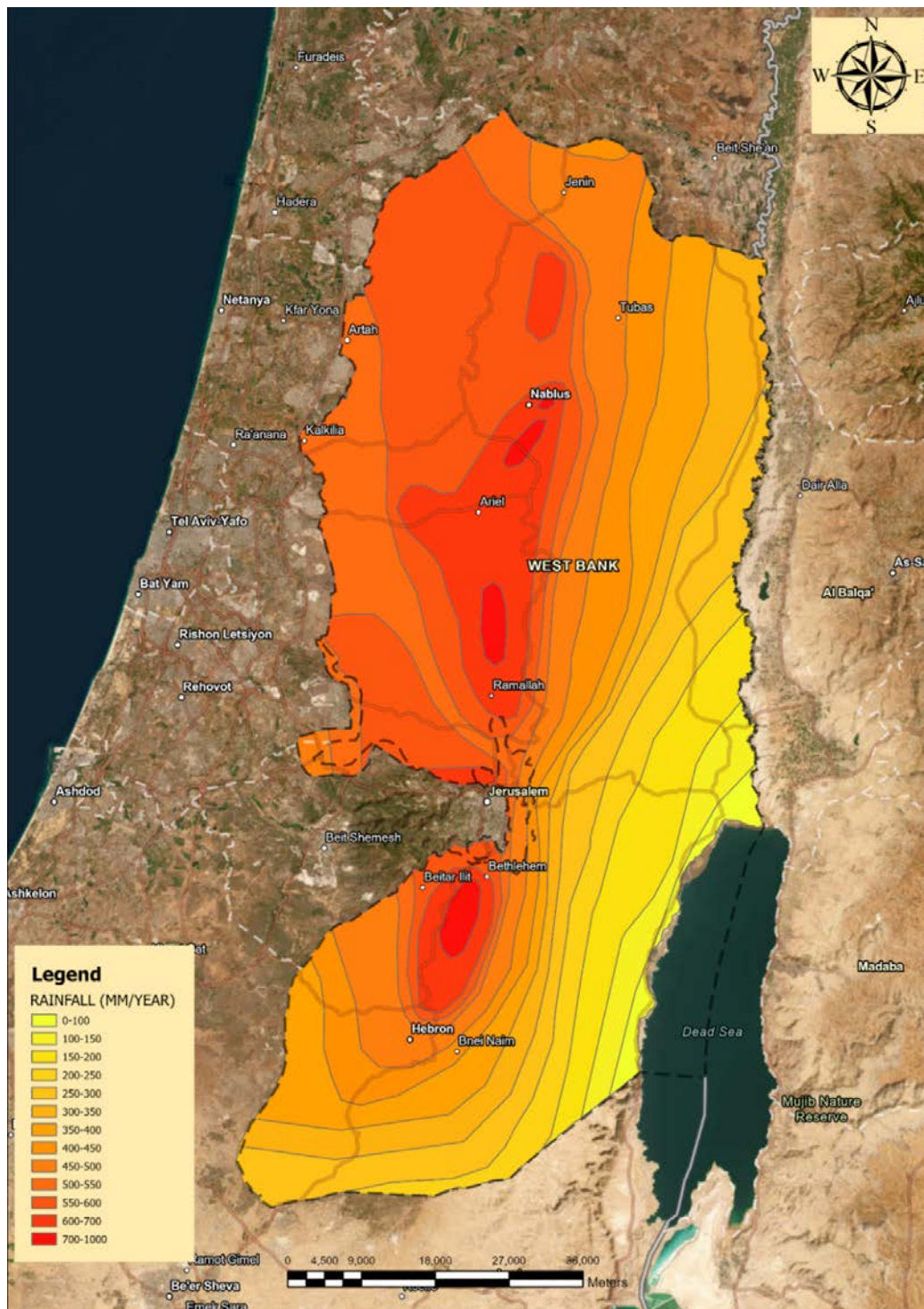


Figure 2.3: West Bank Annual Rainfall

Based on spatial data from the Ministry of Local Government - Palestine, 2026

#### 2.6.4 West Bank Geology and Hydrogeology

The geology of the West Bank is dominated by Cretaceous and Eocene carbonate rocks. Limestones and dolomites of these ages form the central highland massif, and both units are heavily fractured and extensively certified. The result is high secondary permeability throughout the massif, with infiltrated water moving preferentially downward along the fracture and conduit network rather than through the rock matrix itself. Miocene chalks and marls, encountered in the Hebron region and in parts of the valley fills, are markedly less permeable and behave as local aquitards. Quaternary alluvial deposits, variable in thickness, occupy the floors of the major wadis and the Jordan Valley (Toran & Reisch, 2013).

The hydrogeology of the region is organised around the Mountain Aquifer system, which is subdivided into three aquifer basins on the basis of their drainage direction. The Western Aquifer Basin drains westward toward the Mediterranean coast. The North-Eastern Aquifer Basin discharges toward the Jordan Valley through the spring line in the Jenin–Beit Shean area. The Eastern Aquifer Basin drains eastward, emerging as springs and seeps along the Jordan Valley scarp. Taken together, the three basins constitute the principal freshwater resource of the West Bank. The total sustainable yield of the system has been estimated at approximately 679 million m<sup>3</sup> per year, of which the Oslo II Accords allocated only 118 million m<sup>3</sup> per year to Palestinian abstraction (Desk Study on the Environment in the Occupied Palestinian Territories, 2003b).

Ephemeral wades, which collect rainfall runoff from the highland catchments and convey it toward one of three regional base levels, carry surface hydrology in the West Bank almost entirely: the Jordan River, the Dead Sea, or the Mediterranean Sea. The principal Palestinian drainage systems include Wadi Faria (draining the northeastern highlands to the Jordan Valley), Wadi Qilt (draining the Jerusalem–Ramallah highlands to the Jordan Valley via Jericho), and Wadi Nar (draining southern Jerusalem toward the Dead Sea), together with Wadi Auja, Wadi Nueima, and the Hebron-region wadis of Besor, Hasasa, Ghar, and Shiqma. Under average rainfall conditions, the total annual volume of surface runoff generated across these Palestinian catchments has been estimated at approximately 72–130 million m<sup>3</sup> per year (PWA – Policies and Strategies, 2026), the great majority of which currently discharges to these terminal water bodies without being captured for beneficial use.

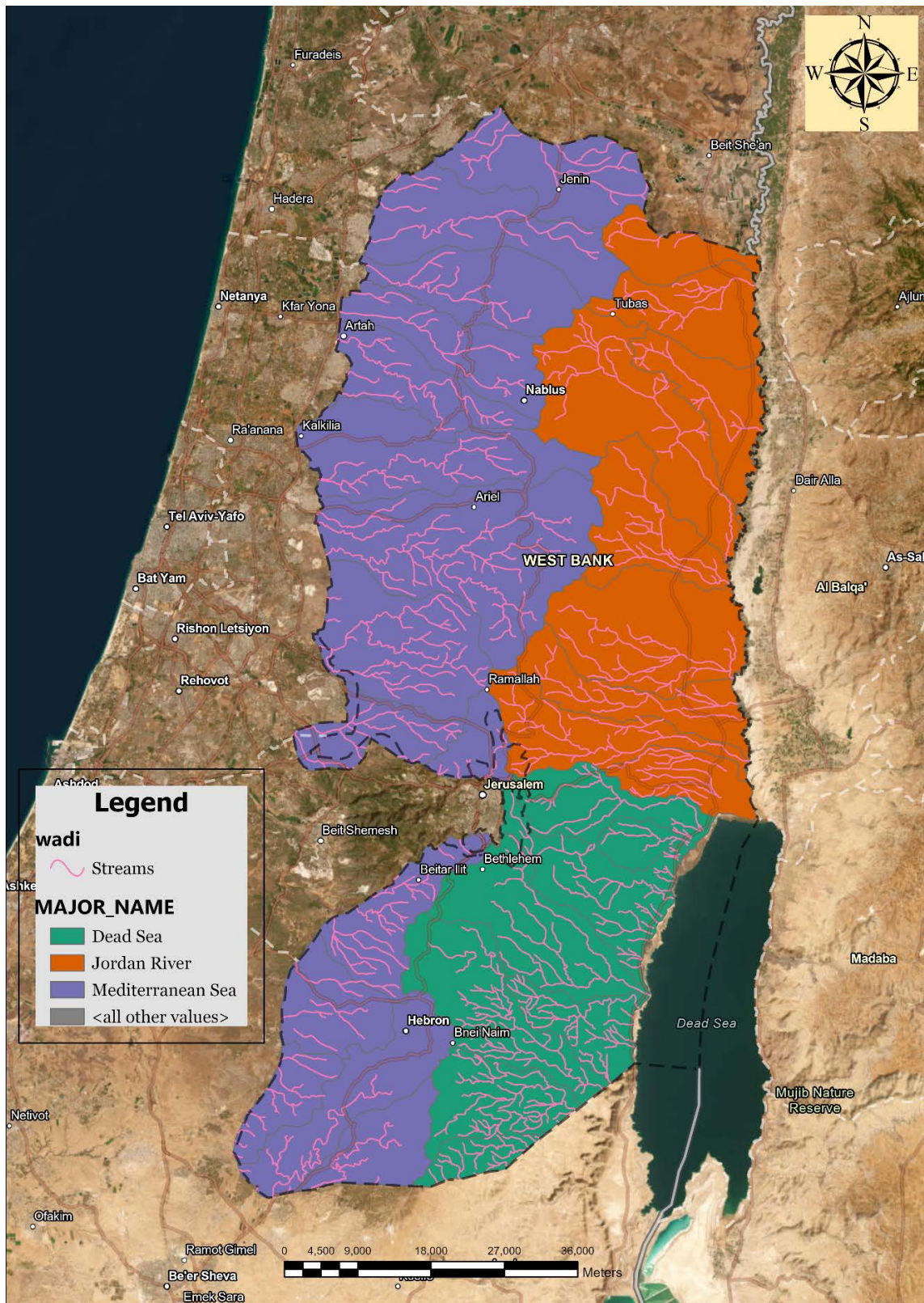


**Figure 2.4 West Bank Major Hydraulic Zone**

**Based on spatial data from the Ministry of Local Government - Palestine, 2026**

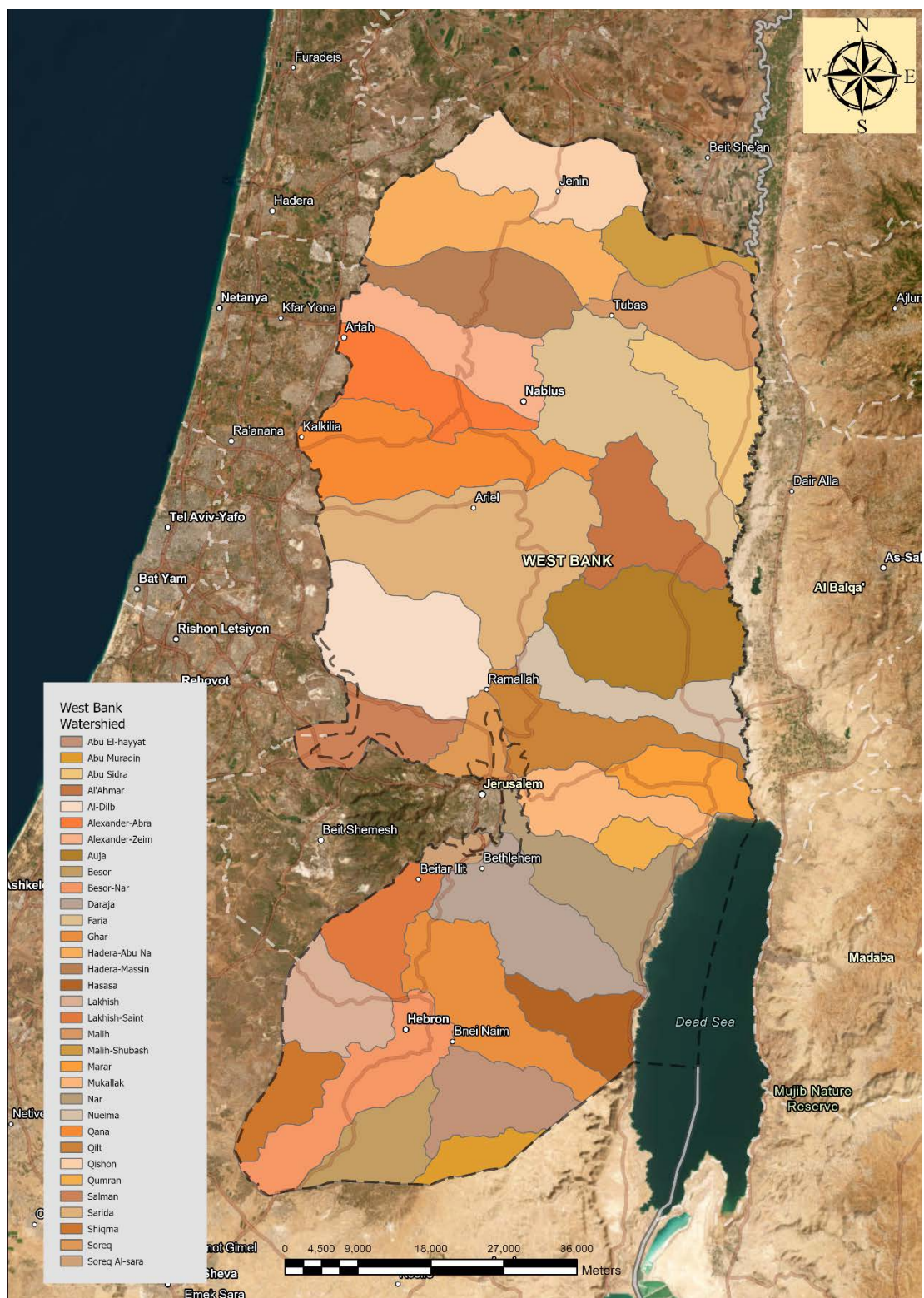


**Figure 2.5 West Bank watershed and sub-watershed Classification**  
**Based on spatial data from the Ministry of Local Government - Palestine, 2026**

#### 2.6.5 West Bank Soils

The soils of the West Bank reflect three interacting controls: the diverse geological substrate on which they have developed their topographic position within the landscape, and the long history of land management across the region. Terra rossa soils (Rhodoxeralfs in the USDA Soil Taxonomy) dominate the upland areas, formed through in-situ weathering of carbonate bedrock under Mediterranean climate conditions. These soils are typically red in colour, clay-loam in texture, and moderately deep (30–70 cm above bedrock), with correspondingly low infiltration rates and moderate available water capacity. Under the NRCS classification, most terra rossa profiles fall within Hydrologic Soil Group C (Desk Study on the Environment in the Occupied Palestinian Territories, 2003b).

Rendzina soils (Rendolls) develop on calcareous parent materials in positions of gentler slope. They are shallow, rich in organic matter, and exhibit moderate infiltration rates. Alluvial soils (Fluvents and Fluvisols) occupy the wadi floors and their associated terraces; their texture varies according to depositional history, but infiltration rates are generally high. The Jordan Valley and the Judean Desert are dominated by arid soils (Xerorthents and Aridisols), which are structurally weakly developed and very low in organic matter. Surface crusting is widespread across these units and produces a substantial reduction in effective infiltration capacity at the onset of rainfall events (Kalteh et al., 2008).

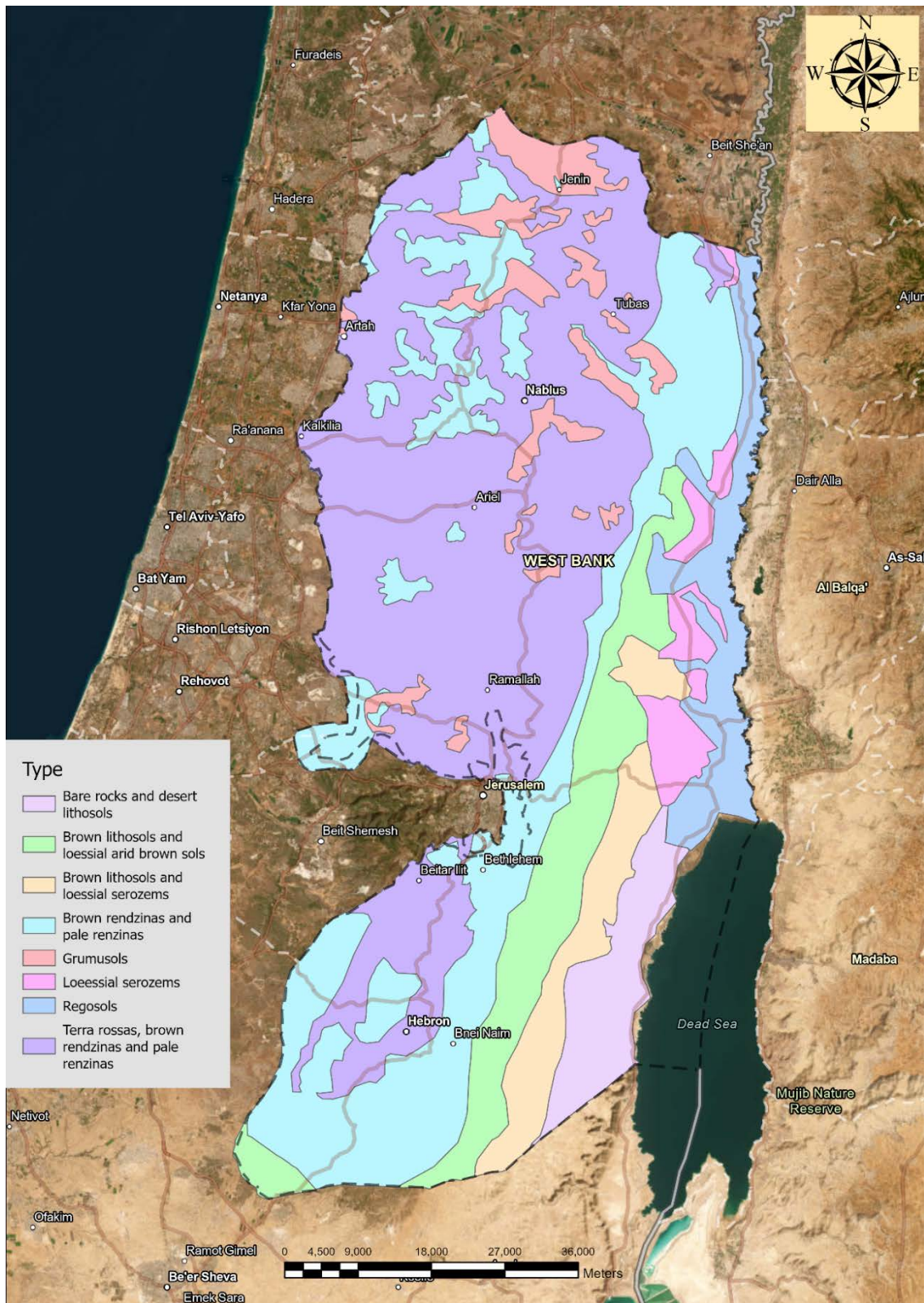


**Figure2.6: West Bank Soil Classification**  
 based on spatial data from the Ministry of Local Government - Palestine, 2026

#### 2.6.6 Land Use and Land Cover

Land use and land cover (LULC) in the West Bank present as a heterogeneous mosaic. Rain-fed and irrigated agriculture occupy substantial areas alongside olive and fruit-tree plantations; these are interspersed with urban and peri-urban settlements, natural open scrubland of garrigue and maquis type, forest plantations, and considerable expanses of bare rock and sparsely vegetated terrain. Data from the Palestinian Central Bureau of Statistics (PCBS), read together with CORINE-based classifications, indicate that olive groves account for the largest single share of cultivated land. Rain-fed field crops follow, together with irrigated vegetable production concentrated in the Jordan Valley and fruit orchards distributed across the highland areas. Rapid urban expansion around Ramallah, Nablus, Hebron and Jericho over the past two decades has increased impervious surface cover across their metropolitan footprints, with direct consequences for surface runoff generation from these expanding built-up catchments (Shaheen, 2002).

Israeli settlements form a further distinct LULC category, typically established on strategically positioned hilltops and valley floors, and characterized by high impervious-surface fractions comparable to those of urban Palestinian centres. The geopolitical fragmentation of the West Bank under the Oslo Accords area classification (Areas A, B, and C) imposes considerable constraints on Palestinian LULC management. This is most pronounced in Area C, which covers approximately 60 percent of the territory and remains under full Israeli civil and security control, restricting Palestinian planning authority over land-use decisions across the majority of the West Bank's land surface (Isaac & Hilal, 2011).

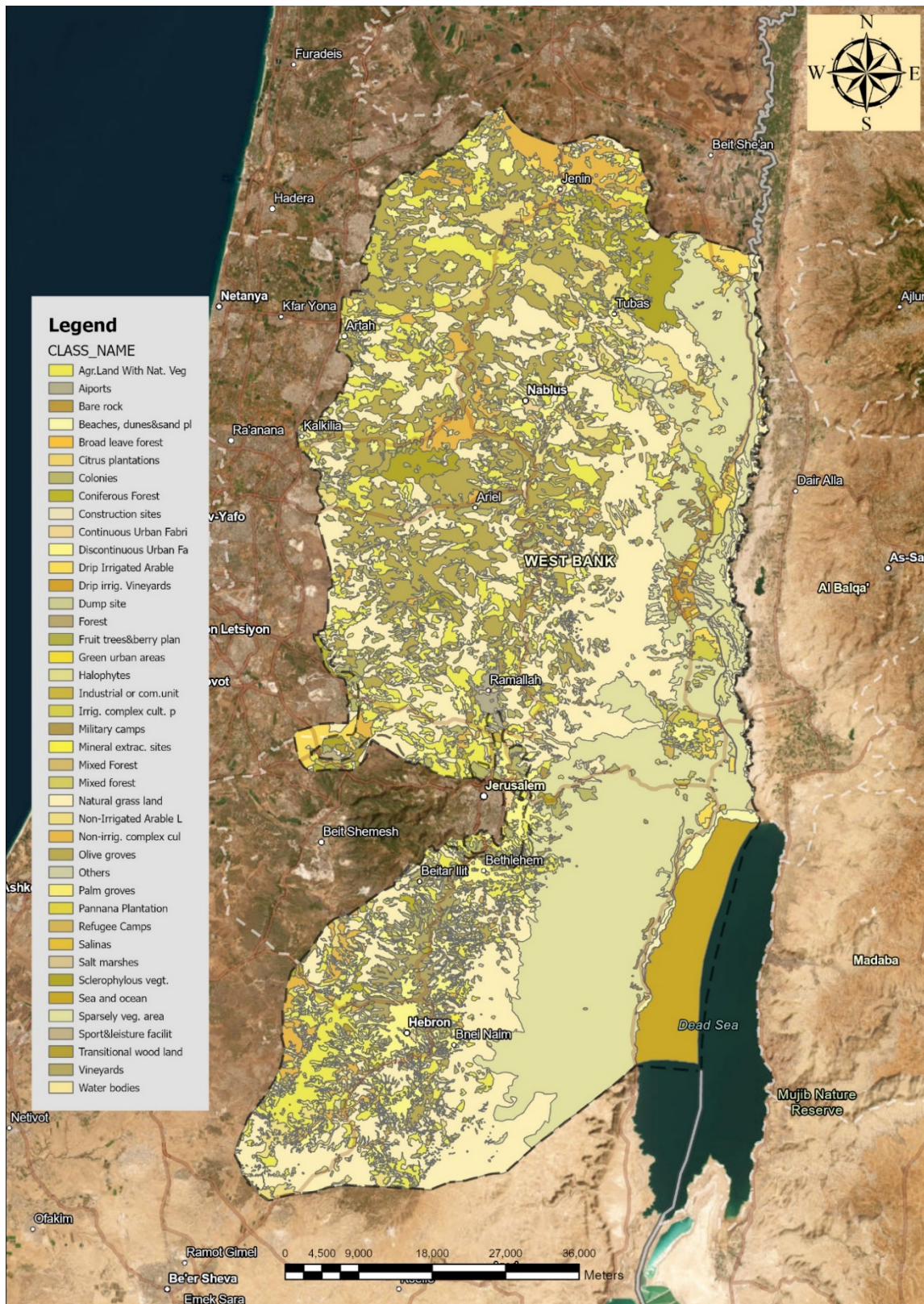


**Figure2.7 :West Bank land use Landsat (CORINE)**  
 based on spatial data from the Ministry of Local Government - Palestine, 2026

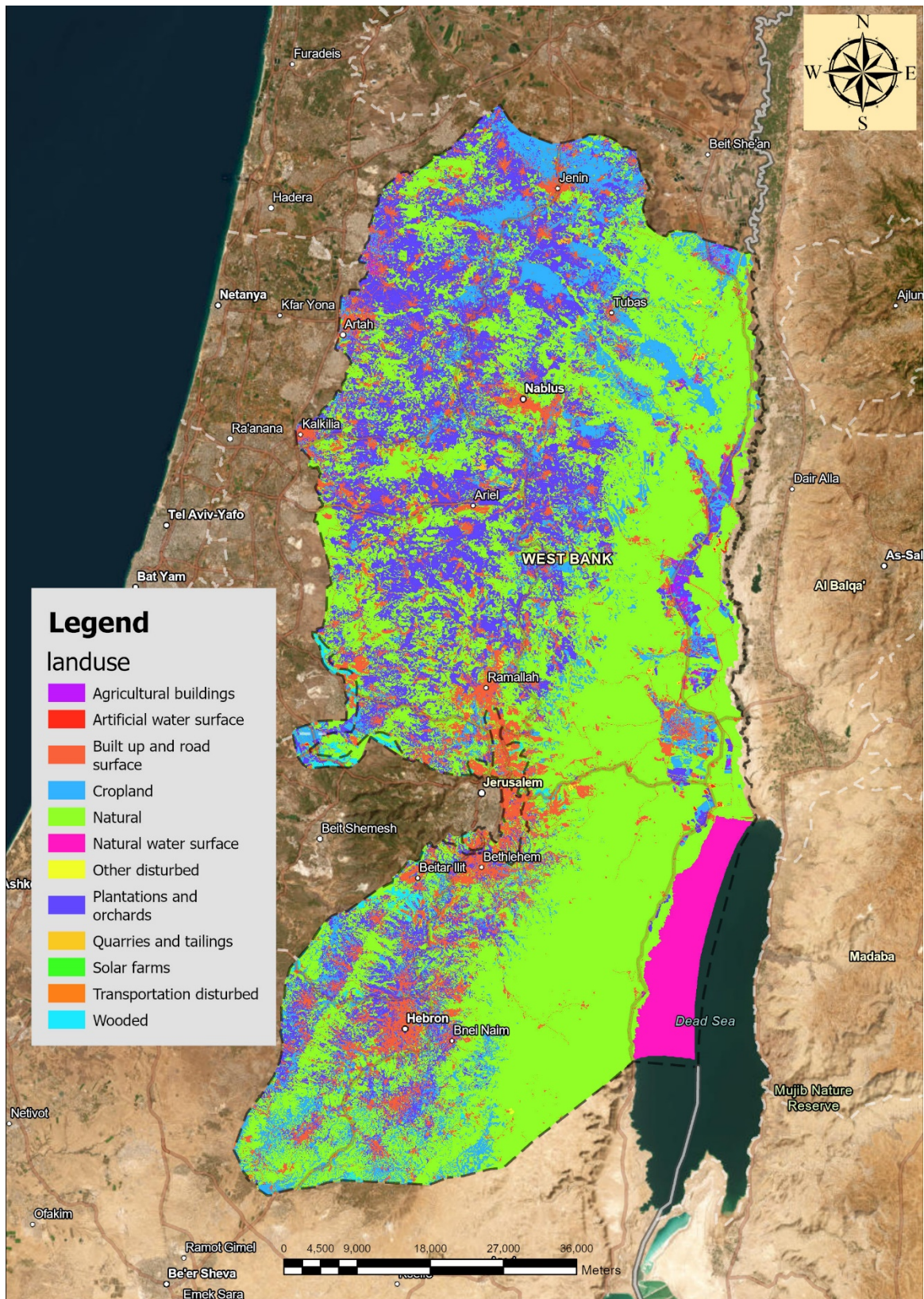


**Figure 2. 8 :West Bank land use Sentinel-2 (USGS) based on spatial data from the Ministry of Local Government - Palestine, 2026**

## 2.7 Water Resources and Management Challenges in the West Bank

Groundwater does almost all of the work of water storage in the West Bank. Surface reservoirs of any meaningful capacity are absent, so the strategic reserve on which the region depends sits entirely in its aquifer systems. The Palestinian population currently draws between 200 and 230 Mm<sup>3</sup> per year in total freshwater supply (Palestinian Water Authority, 2026), a figure that has to be read alongside the aquifers' combined sustainable yield of over 800 Mm<sup>3</sup> per year. The remaining volume is abstracted by Israeli wells drawing from the same aquifer formations, under water-sharing terms established by the Oslo Accords that Palestinian water authorities have long considered inequitable (World Bank, 2024).

At the household level, the consequence is a domestic per capita supply of roughly 73 L/day, well short of the World Health Organization's 100 L/day floor and only about a quarter of the Israeli equivalent of 280 L/day (PCBS, 2025). The shortfall is felt most severely in rural areas and refugee camps, where the piped network fails to reach many households; those households buy water instead from private tanker trucks, at four to ten times the price of piped supply (PCBS, 2026)

Municipal wastewater and stormwater services face structural problems of their own. Many urban centres still operate combined sewer systems, carrying sanitary sewage and storm runoff through a single pipe network. When high-intensity winter storms arrive, these systems cannot cope: overflow reaches the surface as urban flooding, and pollutant loads are discharged into the wadis that drain the affected areas. Where separate storm networks have been built, most were sized for catchments far less urbanised than they are today, and their hydraulic capacity now falls short of what current impervious-surface fractions demand (Applied Research Institute – Jerusalem [ARIJ], 2025).

The problem repeats at the catchment scale. Rainwater harvesting infrastructure is largely absent across the territory, and the result is that several hundred million cubic metres of winter rainfall discharge each year through transboundary wadi systems into the Mediterranean Sea. That water is neither retained for local supply nor directed toward managed aquifer recharge; it is simply lost from the regional balance.

Wastewater treatment sits in a similar state of partial development. Operational treatment plants exist in only four cities, Nablus, Al-Bireh/Ramallah, Bethlehem and Hebron, and much of the rural and peri-urban population continues to rely on cesspits for on-site disposal (Meng et al., 2011). Wadi discharges of wastewater, whether from treatment plant effluent or from untreated overflow, raise hydraulic and chemical loading in the receiving watercourses. Both effects complicate hydrological modelling in the affected catchments and remain persistent obstacles to downstream water-quality management.

Table 2.1: Key water resources statistics for the West Bank (PWA, 2026; PCBS, 2024).

| Parameter                                  | Value                        | Source                     |
|--------------------------------------------|------------------------------|----------------------------|
| Total territory area                       | 5,660 km <sup>2</sup>        | <a href="#">PCBS, 2024</a> |
| Population (2024)                          | ≈ 3.3 million                | <a href="#">PCBS, 2024</a> |
| Mean annual rainfall (highlands)           | 550 – 750 mm/yr              | <a href="#">PWA, 2026</a>  |
| Mean annual rainfall (Jordan Valley)       | < 100 mm/yr                  | <a href="#">PWA, 2026</a>  |
| Aquifer sustainable yield (three basins)   | ≈ 679Mm <sup>3</sup> /yr     | <a href="#">PWA, 2026</a>  |
| Palestinian annual water allocation (Oslo) | ≈ 200 Mm <sup>3</sup> /yr    | <a href="#">PWA, 2026</a>  |
| Domestic per capita consumption            | ≈ 86.4L/day                  | <a href="#">PCBS, 2024</a> |
| Number of rainfall monitoring stations     | 42 (incl. Israeli network)   | <a href="#">IMS / PWA</a>  |
| Number of operational streamflow gauges    | < 10 (predominantly Israeli) | <a href="#">GRDC, 2026</a> |

## 2.8 Previous Hydrological Studies in the Region

Research on West Bank hydrology has developed under real constraints: restricted physical access to field sites, incomplete monitoring records, and political conditions that have limited both data acquisition and cross-border cooperation. Despite these constraints, a substantial body of relevant literature has accumulated over the past three decades, spanning both Palestinian catchment-scale studies and broader regional work on Mediterranean and semi-arid hydrology.

Within the Palestinian context, Mogheir et al. (2009) applied the SCS-CN method to several West Bank watersheds using a GIS-based workflow, and their analysis of the spatial variation in CN

values across the region highlighted the specific difficulties of assigning AMC classes under conditions of sparse rainfall monitoring. Al-Saqaf et al. (2017) took a hydraulic modelling approach to a related question, coupling HEC-HMS with field observations to characterise flood frequency in the Wadi Qilt basin, one of the region's most flood-prone catchments.

Work at the aquifer and catchment scale has extended these efforts in complementary directions. Mohamad and Rabi (2015) applied GIS-based water-balance models to estimate groundwater recharge across the West Bank, demonstrating what spatially distributed modelling can add to regional water resources assessment where in-situ measurements are sparse. Khalaf and Donoghue (2012) combined remote sensing and GIS to quantify soil erosion in Palestinian watersheds, generating the land-cover and soil datasets needed for CN calculation. Freiwan and Kadioğlu (2008), working from long-term station records, characterised the spatial and temporal variability of rainfall across the region and documented the pronounced climatic gradients that present persistent challenges for spatial interpolation.

At the regional scale, several research groups have applied machine learning to rainfall-runoff prediction in Mediterranean and semi-arid environments comparable to the West Bank. Tsai et al. (2022) found it to be better than HEC-HMS for simulation of runoff from a rural Mediterranean watershed. This comparison directly addresses the modeling choices this thesis has to make. Rajib et al. (2020) took the question in a more applied direction. They built ML-based frameworks for predicting peak flows in ungauged basins, and, importantly, worked out how to convert the ML output into flow estimates an engineer can actually use in design. Chekole et al. (2024) sit closest to the present study in setting: they compared HEC-HMS with several machine-learning approaches for rainfall-runoff modelling in data-scarce African semi-arid catchments, and the methodological questions they had to answer, which model works when the data are thin, which performs adequately without heavy calibration, are the same questions faced in the Palestinian context (Alsubeai & Burckhard, 2021).

## 2.9 Summary

This chapter has established the hydrological and geographical context for the research. The West Bank presents a complex and demanding environment for runoff estimation, characterized by

pronounced rainfall variability, steep topographic gradients, heterogeneous soils and land cover, and severe constraints on hydrometric data availability. These characteristics simultaneously motivate the integrated GIS-SCS-CN-ML approach adopted in this study and define the specific challenges that the methodology must address. The following chapter examines the SCS-CN method in detail, providing the theoretical foundation for the computational procedures described in Chapter 5.

## Chapter 3: THE SCS-CN METHOD

### 3.1 Introduction

The "Curve Number" (SCS-CN) method—developed by the Soil Conservation Service of the U.S. Department of Agriculture—is the most widely used empirical approach  $A = \pi r^2$  for estimating direct surface runoff from rainfall in watersheds that lack gauging stations or suffer from data limitations. This method has undergone refinement and development over decades of field observation and theoretical study, and it has found widespread global application in the fields of hydrology, civil engineering, and watershed management. Chapter 4 reviews the method's origins, theoretical foundations, governing equations, parameter determination procedures, and associated limitations, while also establishing the analytical basis for the calculations presented in Chapter 5.

### 3.2 Origin and Development of the Method

The SCS-CN method was developed empirically from rainfall and runoff data collected at small experimental watersheds across the United States during the 1940s and early 1950s (Merizalde et al., 2023). The original dataset comprised several thousand storm events recorded at monitoring sites with diverse soil types, land cover conditions, and land management practices. From this empirical base, the Soil Conservation Service derived a private of dimensionless curves relating cumulative rainfall to cumulative runoff, parameterised by a single dimensionless coefficient, the Curve Number (CN), which ranges from 0 (complete abstraction, no runoff) to 100 (fully impervious surface, all precipitation becomes runoff) (*EDirectives*, 2025).

### 3.3 Theoretical Basis and Governing Equations

The SCS-CN method is based on two fundamental empirical observations derived from the original dataset: first, that the ratio of actual retention (F) to the potential maximum retention (S) equals the ratio of actual runoff (Q) to the potential maximum runoff (P - Ia); and second, that the initial abstraction (Ia) is proportional to the potential maximum retention, with the classical proportionality factor of 0.2.

These assumptions yield the principal runoff equation:

$$Q = \frac{(P - I_a)^2}{(P - I_a + S)} \quad \text{for } P > I_a \quad Q = 0 \quad \text{for } P \leq I_a \quad (3.1)$$

$$(x + a)^n = \sum_{k=0}^n \binom{n}{k} x^k a^{n-k} \quad (3.2)$$

Where  $Q$  is the cumulative direct runoff depth (mm),  $P$  is the cumulative rainfall depth (mm),  $I_a$  is the initial abstraction (mm), and  $S$  is the potential maximum retention (mm).

The potential maximum retention  $S$  is related to the dimensionless Curve Number  $CN$  through:

$$S = \frac{25400}{CN} - 254 \quad (\text{SI units, mm}) \quad (3.3)$$

The initial abstraction  $I_a$ , representing losses before the commencement of runoff (interception, surface storage, infiltration before ponding), is traditionally assumed to be:

$$I_a = 0.2 S \quad (3.4)$$

This yields the combined simplified runoff equation:

$$Q = \frac{(P - 0.2S)^2}{(P + 0.8S)} \quad \text{for } P > 0.2S \quad (3.5)$$

It worth to mention be noted that subsequent research has questioned the universal applicability of the  $I_a = 0.2S$  relationship. Studies by Hawkins et al. (2002), Woodward et al. (2003), and Schneider and McCuen (2005) have demonstrated that the actual ratio between  $I_a$  and  $S$  is highly variable and often substantially lower than 0.2, with values of 0.05 providing better fits to storm event data in many semi-arid catchments. The traditional  $I_a = 0.2S$  relationship consistent with the standard SCS-CN formulation, while acknowledging this source of uncertainty.

### 3.4 Curve Number Determination

The Curve Number is the critical parameter governing the SCS-CN method. It is determined from tables compiled by the NRCS based on empirical relationships between land use, hydrologic soil group, and antecedent moisture condition. The tabulated CN values represent composite runoff potentials that implicitly integrate the effects of surface cover density, land management practice, hydrological condition, and soil properties.

#### 3.4.1 Hydrologic Soil Groups

Mineral soils are grouped into four Hydrologic Soil Groups (HSG) by the NRCS based on their minimum saturated hydraulic conductivity and the infiltration characteristics of the least transmissive layer within 50 cm of the soil surface: (Mishra & Singh, 2003b)

HSG A: Soils with high infiltration rates, low runoff potential when thoroughly wetted. Typically deep, well-drained sands and gravels. Minimum saturated hydraulic conductivity greater than 7.6 mm/h. Examples: Typical Udiplasments, Typical Udiorthents on sandy parent material.(Mishra & Singh, 2003c)

HSG B: Soils with moderate infiltration rates. Moderately deep to deep, moderately well-drained soils of moderately fine to moderately coarse texture. Minimum hydraulic conductivity 3.8 to 7.6 mm/h(Mishra & Singh, 2003c).

HSG C: Soils with slow infiltration rates and moderately high runoff potential. Soils with a layer that impedes the downward movement of water or soils with moderately fine to fine texture. Minimum hydraulic conductivity 1.3 to 3.8 mm/h. Terra rossa soils common in the West Bank uplands typically fall in this group(Mishra & Singh, 2003c).

HSG D: Soils with very slow infiltration rates and high runoff potential. Primarily clays with high swelling potential, soils with a permanent high water table, soils with a claypan or clay layer at or near the surface, and shallow soils over nearly impervious rock. Minimum hydraulic conductivity less than 1.3 mm/h. Many West Bank soils underlain by impermeable chalk or claypan fall in this group(Mishra & Singh, 2003d).

### 3.4.2 Land Use and Land Cover

CN values are tabulated for combinations of HSG (A, B, C, D) and LU/LC type. In this study, two LULC classification systems were applied: the European CORINE (Coordination of Information on the Environment) land cover scheme, widely used in European and Mediterranean-region hydrological studies, and the USDA land cover classification, standard in North American and internationally applied SCS-CN applications. Table 3.2 presents the CN values applied for the European CORINE classification, and Table 3.3 presents those for the USDA classification, both for AMC-II conditions (Merizalde et al., 2023; Shah et al., 2023).

Table 3.2: Curve Numbers for European CORINE Land Use Classification (AMC-II)  
(Mohammadin, 2003).

| No. | Land Use (CORINE)                         | HSG A | HSG B | HSG C | HSG D |
|-----|-------------------------------------------|-------|-------|-------|-------|
| 1   | Agricultural Land with Natural Vegetation | 68    | 79    | 84    | 90    |
| 2   | Airports                                  | 58    | 64    | 69    | 75    |
| 3   | Bare Rock                                 | 44    | 52    | 57    | 61    |
| 4   | Beaches, Dunes and Sand Plains            | 72    | 85    | 93    | 96    |
| 5   | Broad-leaved Forest                       | 37    | 43    | 50    | 55    |
| 6   | Citrus Plantations                        | 50    | 64    | 71    | 76    |
| 7   | Colonies (Israeli Settlements)            | 68    | 81    | 86    | 89    |
| 8   | Coniferous Forest                         | 48    | 62    | 69    | 74    |
| 9   | Construction Sites                        | 52    | 61    | 68    | 73    |
| 10  | Continuous Urban Fabric                   | 40    | 46    | 51    | 54    |
| 11  | Discontinuous Urban Fabric                | 40    | 48    | 55    | 61    |
| 12  | Drip-Irrigated Vineyards                  | 53    | 64    | 73    | 76    |
| 13  | Drip-Irrigated Arable Land                | 65    | 77    | 83    | 89    |
| 14  | Dump Sites                                | 69    | 76    | 82    | 87    |
| 15  | Forest                                    | 53    | 58    | 63    | 68    |
| 16  | Fruit Trees and Berry Plantations         | 32    | 40    | 48    | 53    |

| <b>No.</b> | <b>Land Use (CORINE)</b>               | <b>HSG A</b> | <b>HSG B</b> | <b>HSG C</b> | <b>HSG D</b> |
|------------|----------------------------------------|--------------|--------------|--------------|--------------|
| 17         | Green Urban Areas                      | 51           | 57           | 62           | 66           |
| 18         | Halophytes                             | 31           | 43           | 51           | 57           |
| 19         | Industrial or Commercial Units         | 73           | 79           | 84           | 88           |
| 20         | Irrigated Complex Cultivation Patterns | 59           | 69           | 78           | 81           |
| 21         | Military Camps                         | 67           | 77           | 84           | 88           |
| 22         | Mineral Extraction Sites               | 31           | 45           | 53           | 56           |
| 23         | Mixed Forest                           | 50           | 58           | 65           | 68           |
| 24         | Natural Grassland                      | 41           | 47           | 52           | 58           |
| 25         | Non-Irrigated Complex Cultivation      | 51           | 65           | 72           | 78           |
| 26         | Non-Irrigated Arable Land              | 73           | 79           | 88           | 94           |
| 27         | Olive Groves                           | 54           | 68           | 75           | 81           |
| 28         | Palm Groves                            | 56           | 68           | 77           | 81           |
| 29         | Refugee Camps                          | 57           | 70           | 77           | 82           |
| 30         | Sparsely Vegetated Areas               | 32           | 44           | 53           | 56           |
| 31         | Sport and Leisure Facilities           | 66           | 80           | 87           | 92           |
| 32         | Transitional Woodland-Shrub            | 36           | 49           | 54           | 58           |
| 33         | Vineyards                              | 50           | 63           | 71           | 74           |
| 34         | Water Bodies                           | 38           | 43           | 52           | 55           |

Table 3.3 :Curve Numbers for USDA Land Use Classification (AMC-II). Source: Mohammadin, 2003.

| <b>Value</b> | <b>Land Use (USDA)</b>    | <b>HSG A</b> | <b>HSG B</b> | <b>HSG C</b> | <b>HSG D</b> |
|--------------|---------------------------|--------------|--------------|--------------|--------------|
| 11           | Built-up and Road Surface | 98           | 98           | 98           | 98           |
| 21           | Transportation Disturbed  | 80           | 82           | 85           | 88           |
| 22           | Quarries and Tailings     | 75           | 78           | 82           | 85           |
| 23           | Other Disturbed Land      | 70           | 73           | 78           | 81           |
| 24           | Solar Farms               | 65           | 68           | 73           | 76           |

|    |                                 |     |     |     |     |
|----|---------------------------------|-----|-----|-----|-----|
| 31 | Natural (bare/rock/sparse veg.) | 30  | 45  | 60  | 75  |
| 32 | Wooded (forest/shrubland)       | 36  | 60  | 73  | 79  |
| 41 | Cropland                        | 67  | 78  | 85  | 89  |
| 42 | Plantations and Orchards        | 57  | 72  | 81  | 88  |
| 43 | Agricultural Buildings          | 74  | 83  | 89  | 92  |
| 51 | Natural Water Surface           | 100 | 100 | 100 | 100 |
| 52 | Artificial Water Surface        | 100 | 100 | 100 | 100 |

### 3.4.3 Antecedent Moisture Condition

Antecedent Moisture Condition (AMC) is a measurement of the relative wetness of the soil at the start of a design storm event. The NRCS has three classes of AMC: AMC-I (Dry): Soils are dry, but not at the wilting point. It is defined as cumulative 5-day antecedent rainfall less than 35.6 mm during the growing season or less than 12.7 mm during the dormant season. For situations with low potential for runoff.

AMC-II (Normal or Average): Normal or average pre-storm moisture conditions. Cumulative 5-day antecedent rainfall between 35.6 and 53.3 mm (growing season) or 12.7 and 28.0 mm (dormant season). Used as tabulated condition CN standard.

AMC-III (Wet): Saturated or near saturated conditions. Defined as the sum of 5-d antecedent rainfall > 53.3 mm (growing season) or > 28.0 mm (dormant season). Applicable to conditions with high runoff potential, frequently associated with wet antecedent periods.

Conversion from AMC classes is obtained by the following approximate formulas (Chow et al., 1988):

$$CN(I) = \frac{CN(II)}{2.281 - 0.01281 \times CN(II)} \quad (3.6)$$

$$CN(III) = \frac{CN(II)}{0.427 + 0.00573 \times CN(II)} \quad (3.7)$$

### 3.5 The Initial Abstraction Ratio

The first abstraction ratio,  $\lambda = I_a/S$ , has been the focus of considerable research attention. The traditional value of  $\lambda = 0.2$  was based on the original SCS empirical data set, which was dominated by agricultural watersheds in the humid temperate climates of the central and eastern United States. Several subsequent investigations have demonstrated that  $\lambda = 0.2$  systematically overestimates  $I_a$  in semi-arid environments, leading to underestimation of runoff for lower rainfall totals and overestimation for higher totals (Ponce and Hawkins, 1996; Hawkins et al., 2002).

Woodward et al. (2003) reanalyzed the original SCS dataset and proposed  $\lambda = 0.05$  as a more representative value for a wide range of watershed conditions. Mishra et al. (2006) conducted sensitivity analyses demonstrating that runoff estimates are relatively insensitive to  $\lambda$  for high CN values but highly sensitive for low CN values, with implications for the semi-arid eastern West Bank where CNs under dry conditions can be quite low. This study adopts the traditional  $\lambda = 0.2$  for consistency with standard engineering practice, while noting that the optimized CN values derived from ML model analysis may implicitly account for the appropriate effective initial abstraction.

### 3.6 Composite (Weighted) Curve Number

Real watersheds typically contain multiple combinations of LULC and HSG. The representative CN for the watershed as a whole is computed as the area-weighted average of the CNs for each unique combination of LULC and HSG:

$$CN_{\text{composite}} = \frac{\sum(CN_i \times A_i)}{A_{\text{total}}} \quad (3.8)$$

where CN is the Curve Number for the i-th land cover-soil combination,  $A_i$  is the area of that combination, and  $A_{\text{total}}$  is the total watershed area. This computation is most efficiently performed through GIS overlay analysis, in which digital maps of LULC and HSG are intersected to create a polygon layer of unique LULC-HSG combinations, to which CN values are assigned from look-up tables prior to area-weighted averaging.

### 3.7 Modifications and Improvements of the SCS-CN Method

Many modifications of the standard SCS-CN method have been proposed to overcome its known shortcomings as well as to improve its performance. Mishra and Singh (2003) proposed a generalized framework with a non-linear initial abstraction relationship. Michel et al. (2005) proposed a continuous soil moisture accounting version that dynamically follows inter-storm antecedent conditions. Ponce (2014) proposed a modified CN method for antecedent conditions based on a continuous soil moisture depletion model.

In GIS-integrated applications, the CN has been determined from remotely sensed LULC data derived from multispectral satellite imagery, including Landsat, SPOT, Sentinel-2, and WorldView platforms, enabling timely and spatially comprehensive CN mapping that was not possible with conventional field survey approaches. The integration of high-resolution soil surveys and remote sensing-derived LULC within GIS environments constitutes the state of practice for CN determination in watershed-scale applications.

### 3.8 Definition of SCS Rainfall Distributions

The Natural Resources Conservation Service (NRCS, formerly SCS) developed four synthetic 24-hour rainfall distribution curves (Type I, IA, II, and III) to represent various climatic regions. The x-axis represents cumulative time ( $t/24$ ), while the y-axis represents the cumulative fraction of rainfall.

- Type I and Type IA: Used for regions with maritime climates (e.g., the Pacific maritime climate). These areas experience wet winters with long-duration, low-to-moderate intensity storms. Type IA represents the least intense rainfall distribution.
- Type II: Represents interior and semi-arid regions characterized by highly intense, short-duration summer thunderstorms. A large percentage of the total rainfall occurs rapidly in the middle of the 24-hour period (around  $t/24 = 0.5$ ). It is the most conservative distribution, yielding the highest peak discharges in hydrological modeling.
- Type III: Represents coastal areas susceptible to tropical storms and hurricanes with large volumes of rainfall (e.g., the Gulf of Mexico and Atlantic coasts).

Applicability to the Dead Sea Region The most appropriate and widely adopted distribution for the Dead Sea basin (including the eastern and western slopes of the Jordan Rift Valley) is Type 1A (Kassem et al., 2023).

### 3.8.1 Origin and Structure of the Type 1A Distribution

Application of Soil Conservation Service (SCS) methodology in developing synthetic storm hyetographs for Palestine. 1973. It focuses on the development of good models for estimating runoff in an area lacking sufficient hydrological data. This method would be useful for regions like Palestine with very few gauging stations and limited historical rainfall-runoff data.

In Palestine, the SCS method has been used to develop synthetic storm hyetographs, which are useful for flood management and water resources planning. For instance, researchers to derive geomorphological and hydrological parameters required for modelling runoff in different catchments have used GIS tools. These parameters are used to characterize the synthetic unit hydrograph in local conditions. Here, the use of storm distribution patterns (hyetographs) and particularly SCS (Soil Conservation Service) standard distributions developed in 1973, has resulted in the development of good models for estimating surface runoff in areas where adequate hydrological data has been compiled. This is particularly useful in places like Palestine where the number of gaging stations, and the historical data on the rainfall-runoff relationship are very limited. In Palestine, the SCS method has been used to develop synthetic storm hyetographs that are useful in flood management and water resource planning. For example, researchers have used GIS tools to extract geomorphological and hydrological parameters required for modeling runoff in different catchments. These parameters are useful in defining the characteristics of the synthetic unit hydrograph adapted to local conditions. (US Department of Commerce, 2025). It is found that the Type 1A is the most suitable in the Mediterranean environment of Palestine as shown in Table 8 below (Shuster et al., 2005b).

Table3.4 SCS Rainfall Distributions (24-Hour Storm)(Shuster et al., 2005b)

| <b>Hour 24</b> | <b>t/24</b> | <b>Type I</b> | <b>Type IA</b> | <b>Type II</b> | <b>Type III</b> |
|----------------|-------------|---------------|----------------|----------------|-----------------|
| 0              | 0           | 0             | 0              | 0              | 0               |
| 2              | 0.083       | 0.035         | 0.05           | 0.022          | 0.02            |
| 4              | 0.167       | 0.076         | 0.116          | 0.048          | 0.043           |
| 6              | 0.25        | 0.125         | 0.206          | 0.08           | 0.072           |

| Hour 24 | t/24  | Type I | Type IA | Type II | Type III |
|---------|-------|--------|---------|---------|----------|
| 7       | 0.292 | 0.156  | 0.268   | 0.098   | 0.089    |
| 8       | 0.333 | 0.194  | 0.425   | 0.12    | 0.115    |
| 8.5     | 0.354 | 0.219  | 0.48    | 0.133   | 0.13     |
| 9       | 0.375 | 0.254  | 0.52    | 0.147   | 0.148    |
| 9.5     | 0.396 | 0.303  | 0.55    | 0.163   | 0.167    |
| 9.75    | 0.406 | 0.362  | 0.564   | 0.172   | 0.178    |
| 10      | 0.417 | 0.515  | 0.577   | 0.181   | 0.189    |
| 10.5    | 0.438 | 0.583  | 0.601   | 0.204   | 0.216    |
| 11      | 0.459 | 0.624  | 0.624   | 0.235   | 0.25     |
| 11.5    | 0.479 | 0.654  | 0.645   | 0.283   | 0.298    |
| 11.75   | 0.489 | 0.669  | 0.655   | 0.357   | 0.339    |
| 12      | 0.5   | 0.682  | 0.664   | 0.663   | 0.5      |
| 12.5    | 0.521 | 0.706  | 0.683   | 0.735   | 0.702    |
| 13      | 0.542 | 0.727  | 0.701   | 0.772   | 0.751    |
| 13.5    | 0.563 | 0.748  | 0.719   | 0.799   | 0.785    |
| 14      | 0.583 | 0.767  | 0.736   | 0.82    | 0.811    |
| 16      | 0.667 | 0.83   | 0.8     | 0.88    | 0.886    |
| 20      | 0.833 | 0.926  | 0.906   | 0.952   | 0.957    |
| 24      | 1     | 1      | 1       | 1       | 1        |

### 3.9 Muskingum Channel Routing: Theory and Stability Criteria

#### 3.9.1 Storage Equation

The Muskingum method (McCarthy, 1938) conceptualises a channel reach as a linear storage element. Channel storage  $S$  is related to both inflow  $I$  and outflow  $Q$  through the travel time  $K$  (hours) and weighting factor  $X$ :

$$S = K(X \cdot I + (1 - X) \cdot Q) \quad (3.9)$$

*[S = reach storage; K = travel time; X = weighting factor ( $0 \leq X \leq 0.5$ )]*

### 3.9.2 Routing Coefficients

Combining Equation 7.5 with the continuity equation and applying finite-difference discretisation over a routing time step  $\Delta t$  gives the three-coefficient routing equation  $Q_2 = C_0 I_2 + C_1 I_1 + C_2 Q_1$ , where:

$$C_0 = \frac{\Delta t - 2KX}{2K(1-X) + \Delta t} \quad (3.10a)$$

$$C_1 = \frac{\Delta t + 2KX}{2K(1-X) + \Delta t} \quad (3.10b)$$

$$C_2 = \frac{2K(1-X) - \Delta t}{2K(1-X) + \Delta t} \quad (3.10d)$$

$$[C_0 + C_1 + C_2 = 1.0 \text{ (mass conservation identity)}] \quad (3.10e)$$

- X value

Depends mainly on the reach slope and shape, if  $X=0$ ; the reach slope and shape approach to be reservoir (high attenuation and low transmission for flood peak discharge). On the other hand, if  $X=0.5$ ; the reach approaches to full wedge reach (low attenuation and high transmission for flood peak discharge). Finally in the most natural streams around the world the value of  $X$  varies from 0 to .3, for the selected reach we predict the value around  $X=0.2$

- K value

Represent travel time of flood wave throughout the reach, this value largely depends on stream length, roughness and the slope. Usually, this value obtained from relationship between “Q and I” by which it represents the time between the center of gravity of stream inflow “I” to center of gravity of stream outflow “Q”. Unfortunately, there is no Q data so we can approximate the K value to equal time of concentration ‘Tc’

### 3.10 Tools and Software for SCS-CN Application

The SCS-CN method is used in several commercial and open-source hydrological modelling platforms. The main tools relevant to the present study are compared below:

- HEC-HMS (Hydrologic Engineering Center – Hydrologic Modeling System): HEC-HMS is a popular software used to simulate the precipitation-runoff process at the watershed-scale. It was developed by the US Army Corps of Engineers. It implements the SCS-CN loss method directly and allows lumped, semi-distributed and quasi-distributed watershed configurations. HEC-HMS is most suitable for design flood estimation, dam safety analysis and single-event modeling with well-defined watershed units (Hydrologic Engineering Center, 2026).
- ArcGIS Spatial Analyst / ArcHydro: The Esri ArcGIS platform with the supplementary ArcHydro extension offers a comprehensive suite of tools for DEM-based watershed delineation, stream network extraction, and spatial overlay analysis for CN computation. ArcHydro incorporates DEM processing into hydrological analysis workflows and is the de facto GIS platform for hydrological parameter extraction (Esri, 2026).
- Open-source GIS options such as QGIS with SAGA and GRASS GIS plugins that provide similar watershed delineation and spatial analysis capabilities to ArcGIS for free. QGIS supports the complete implementation of DEM-based hydrological workflows through its SAGA (System for Automated Geoscientific Analyses) and GRASS (Esri, 2026)
- WMS (Watershed Modelling System, Aquaveo): An integrated GIS-hydrology platform with a graphical interface for watershed delineation, CN computation, and coupling to HEC-HMS, HEC-RAS and other simulation engines. WMS has user-friendly tools for GIS-based CN map generation and area-weighted CN calculation. In the present study it was utilized for watershed delineation and morphometric parameter extraction (WMS - Watershed Modeling System, 2026).
- Python (scikit-learn, xgboost, pandas): The Python scientific computing ecosystem provides the computational framework of the machine learning part of this study. The scikit-learn library (Pedregosa et al., 2011) provides implementations for Random Forest, Decision Tree, SVR, KNN and Linear Regression algorithms and standardized interfaces to train, perform hyper parameter optimization and evaluate the models. The XGBoost algorithm is implemented in the xgboost library (Chen and Guestrin, 2016). Pandas is a library for working with tabular data and numpy for numerical array operations (python software, 2026) .

The best choice among these tools depends on the study objectives and data available. Spatially distributed mapping of CN is needed using GIS-based workflows. HEC-HMS offers the most complete physical modeling framework for event-based runoff-discharge simulation. For data-driven runoff prediction in ungauged catchments, Python-based ML implementations are the most flexible and computationally efficient approach.

### 3.11 Limitations of the SCS-CN Method

Despite its widespread application, the SCS-CN method has several recognized limitations that must be acknowledged in interpreting results:

- **Empirical Basis:** The method was derived from limited field data under specific agricultural conditions in the central United States. Its transferability to other climatic and landscape contexts, including semi-arid Mediterranean environments, has been questioned and is a source of methodological uncertainty.
- **AMC Classification Uncertainty:** The 5-day antecedent rainfall criterion for AMC classification is a coarse proxy for actual soil moisture at storm onset, and may not adequately capture inter-storm moisture dynamics in highly permeable karst catchments or in catchments with significant groundwater contributions.
- **Scale Dependency:** CN values were originally derived for small agricultural watersheds (< 200 ha). Spatial scaling to larger heterogeneous watersheds through area weighting introduces aggregation errors and may not preserve the non-linear spatial patterns of runoff generation.
- **No Explicit Time Component:** The SCS-CN method estimates the total storm runoff depth, but does not explicitly address the time distribution of runoff or the derivation of flood hydrographs, which requires a coupling with a unit hydrograph or routing model.
- **Rainfall Uniformity Assumption:** The method assumes spatially uniform rainfall over the watershed, an assumption that is particularly problematic in large catchments under convective storm conditions.

These limitations are directly relevant to the West Bank application and provide further motivation for the ML-based approach developed in this study that can learn effective mappings between

available input features and runoff depth without imposing the structural assumptions of the SCS-CN formulation.

### 3.12 Summary

This chapter has presented a comprehensive review of the SCS-CN method, from its empirical origins through its theoretical formulation, parameter determination procedures, and documented limitations. The method provides the foundational computational framework for the GIS-based runoff analysis conducted in this study, while its recognised limitations motivate the complementary application of machine learning algorithms capable of improving predictive accuracy. The following chapter examines the GIS and remote sensing techniques that enable spatially distributed application of the SCS-CN method.

### Chapter 4: GIS AND REMOTE SENSING IN HYDROLOGICAL MODELLING

#### 4.1 Introduction

GIS and remote sensing have extended the frontier of watershed hydrology allowing spatial data acquisition and analysis at scales not possible by conventional field-based methods. This chapter reviews the main GIS and remote sensing techniques relevant to hydrological modelling with particular emphasis on those applied in this study: DEM based watershed delineation, land use and land cover classification, soil data integration, spatial rainfall interpolation and CN map generation (Yaseen et al., 2015).

#### 4.2 The Role of GIS in Hydrology

GIS provides the computational framework for managing and analysing spatially distributed environmental data required for rainfall-runoff modelling. Relevant capabilities include storage and retrieval of multi-layer datasets (DEMs, soil maps, LULC maps, rainfall grids) in a common geographic reference frame, spatial overlay for CN computation, automated watershed delineation from DEMs, interpolation of point rainfall data to continuous grids, and visualisation of hydrological outputs .

GIS has moved from pre- and post-processing support to full embedded integration in which the environment manages data flows, executes model calculations, and displays results directly. Modern GIS platforms including ArcGIS, QGIS, and specialised hydrology packages such as WMS and ArcHydro provide direct implementation of hydrological analysis routines within the GIS framework (Kaltch et al., 2008; Yaseen et al., 2015).

#### 4.3 Digital Elevation Models

The Digital Elevation Model (DEM) is the primary topographic data product for GIS-based hydrological analysis. A DEM is a raster grid that represents the elevation of the terrain with a certain spatial resolution. For this study, the Shuttle Radar Topography Mission (SRTM) 30-metre resolution DEM was the principal topographic data source, providing consistent global coverage at a resolution adequate for watershed delineation across the range of watershed sizes represented in the study area (26.8 km<sup>2</sup> to 466.4 km<sup>2</sup>) (Farr et al., 2007).

A DEM-based hydrological analysis is carried out using a series of computational steps in the GIS environment: 1. DEM Pre-processing: The raw DEM is pre-processed by first carrying out sink-filling and pit-removal operations to ensure that a flow direction is assigned to each cell and to remove spurious depressions in the elevation surface. Sinks are artefacts of the rasterisation process or errors in the source data that would otherwise block flow-accumulation calculations. 2. Flow direction calculation: D8 algorithm to calculate flow direction for individual flow direction, assigns a flow direction for each cell of the digital elevation model (DEM) to one of its eight adjacent cells (N, N, E, S, E, S, S, W, N, N) in the direction of the steepest slope. The flow direction network obtained maintains the topology of the drainage network (Dávila-Hernández et al., 2022).

Flow Accumulation: The number of upslope cells that drain to each cell is calculated by traversing the flow direction grid from upstream to downstream. High flow accumulation values indicate stream channels, low values indicate interfluvial areas. Stream Network Extraction: A threshold is applied to the flow accumulation grid in order to define the stream channels. All cells with flow accumulation exceeding the threshold are defined as stream channels and thus form a raster stream network. 5. Watershed Delineation All cells that drain to a particular outlet point are delineated as the boundary of a watershed. The Pour Point Snap function in ArcHydro automatically locates the highest flow accumulation cell within a specified search radius of the defined outlet point (Tarboton et al., 1991).

#### 4.4 Watershed and Stream Network Delineation

For this study, watershed delineation was performed using the ArcHydro within ArcGIS, applied to the SRTM 30-metre DEM. The 33 sub-watershed outlet points were defined at the confluence points of the major drainage systems with the Jordan River, the Dead Sea, or the Mediterranean Sea, or at the confluence with the next-order stream within the study domain.

The longest flow path length (L) and average basin slope (S) were extracted from the DEM for each delineated sub-watershed, serving as morphometric inputs to the Time of Concentration ( $T_c$ ) computation using the SCS lag equation:

$$T_L = \frac{L^{0.8}(S + 25.4)^{0.7}}{4238Y^{0.5}} \quad (4.11)$$

where L is the longest flow path length (m), S is the potential maximum retention (mm), and Y is the average watershed slope (percent). The lag time ( $T_L$ ) is related to  $T_C$  by:

$$T_C = \frac{T_L}{0.6} \text{ hours, or } \times 60 \text{ for minutes} \quad (4.12)$$

#### 4.5 Remote Sensing Data Sources

Remote sensing gives hydrologists repeated, multispectral views of the Earth's surface at a spatial scale that works well for watershed analysis. For mapping land use and vegetation density in the West Bank, three data sources matter most.

Landsat, run by NASA and the USGS since 1972, offers the longest continuous record, with 30 m multispectral imagery freely available through USGS Earth Explorer. That archive underpins most historical LULC studies and supplies the spectral data used to update CORINE land cover (Maier et al., 2010).

Sentinel-2, from the European Space Agency, is a two-satellite mission (2A and 2B) that images at 10 to 60 m and revisits every five days. Its finer resolution is a real advantage in the West Bank, where small agricultural plots and peri-urban development sit side by side and are hard to separate at 30 m.

MODIS, on NASA's platforms, covers the ground daily but at 250 to 1000 m. That suits regional vegetation index monitoring and evapotranspiration estimates. It is too coarse, however, for the watershed-scale CN mapping this study needs in the West Bank.

Very High-Resolution Commercial Imagery (e.g., WorldView, GeoEye, Pleiades) The availability of submeter-resolution commercial imagery enables detailed extraction of impervious surface fractions, tree canopy cover, and agricultural plot boundaries in urban and peri-urban areas, which can improve CN map accuracy in complex land use mosaics.

#### 4.6 Land Use and Land Cover Classification

LULC classification is performed by assigning each pixel in a multispectral satellite image to one of a predefined set of land cover categories based on its spectral signature.

Classification algorithms range from simple visual interpretation and manual digitising to fully automated supervised and unsupervised statistical classifiers and, increasingly, deep learning-based approaches.

For this study, LULC data were obtained from two independently derived classifications:

**European CORINE Land Cover:** The CORINE (Coordination of Information on the Environment) programme, coordinated by the European Environment Agency, provides a standardized LULC classification for Europe and neighboring Mediterranean countries at 100-metre resolution with a minimum mapping unit of 25 hectares. The CORINE classification system distinguishes 44 LULC classes at the third hierarchical level, providing more detailed vegetation and agricultural sub-categories relevant to Mediterranean landscapes than the USDA system (Shadeed et al., 2020).

**USDA Land Cover Classification:** The USDA National Land Cover Database (NLCD) classification system, using satellite imagery-based classification with 12 primary land cover categories, was applied to the West Bank using satellite imagery processed within the GIS environment. This system was chosen for its direct compatibility with NRCS CN look-up tables.

#### 4.6 Classification accuracy assessment

Accuracy of the classification is evaluated using confusion matrices based on independent ground truth or reference data. Important accuracy measures include: the overall accuracy (fraction of correctly classified pixels); the producer's accuracy (sensitivity, or the probability that a reference site is correctly classified); the user's accuracy (specificity, or the probability that a classified pixel belongs to the claimed class); and the Kappa coefficient of agreement that accounts for chance agreement between the classification and reference data. The overall classification accuracy of the CORINE-derived classification used in this study is reported by the European Environment Agency to be usually above 85 percent at the Level 3 classification detail, with higher accuracy for major land cover types (urban, forest, water bodies) and lower accuracy for transitional and mixed agricultural categories. The USDA classification applied to West Bank imagery achieved overall accuracies in the range of 78 to 82 percent based on comparison with higher-resolution reference imagery and field validation point (*International Journal of Education and Research (IJERN)*, 2026)s.

#### 4.7 Soil Data and Hydrologic Soil Groups

HSG assignments for the West Bank were derived from the Food and Agriculture Organisation (FAO) Harmonized World Soil Database (HWSD) and from the Palestinian Authority's soil surveys, supplemented by published hydrological soil mapping by the Israeli Hydrological Service and academic literature. The primary soil-geomorphological units relevant to HSG classification are the terra rossa soils on carbonate uplands (predominantly HSG C), rendzina soils on chalky substrates (HSG B-C), alluvial soils in wadi floors (HSG A-B), arid soils in the Jordan Valley and Judean Desert (HSG C-D), and shallow soils over near-impermeable rock (HSG D) (Alfieri et al., 2012).

Table 4. 5: Soil Types of the West Bank: Texture Classification and Hydrologic Soil Group (HSG) (Alfieri et al., 2012)

| No. | Soil Type                        | Soil Texture   | HSG |
|-----|----------------------------------|----------------|-----|
| 1   | Alluvial Soils                   | Loam           | B   |
| 2   | Brown-Red Degraded Sandy Soils   | Loamy sand     | A   |
| 3   | Brown-Red Sandy Soils            | Sandy loam     | A   |
| 4   | Brown Alluvial Soils (Vertisols) | Clay           | D   |
| 5   | Brown Basaltic Soils             | Clay           | D   |
| 6   | Brown Desert Skeletal Soils      | Sandy loam     | A   |
| 7   | Brown Steppe Soils               | Loam           | B   |
| 8   | Coarse Desert Alluvium           | Sandy loam     | A   |
| 9   | Coastal Sand Dunes               | Sand           | A   |
| 10  | Colluvial-Alluvial Soils         | Loam           | B   |
| 11  | Desert Alluvial Soils            | Sandy loam     | A   |
| 12  | Desert Sand Dunes                | Sand           | A   |
| 13  | Desert Stony Land                | Sandy loam     | A   |
| 14  | Hammada Soils of Mountains       | Sandy loam     | A   |
| 15  | Hammada Soils of Plains          | Sandy loam     | A   |
| 16  | Loess-Like Raw Soils             | Silt loam      | B   |
| 17  | Loess Raw Soils                  | Silt loam      | B   |
| 18  | Loessial Sandy Soils             | Sandy loam     | A   |
| 19  | Mediterranean Brown Forest Soils | Clay loam      | D   |
| 20  | Peat Soils                       | Clay / Organic | D   |
| 21  | Reddish-Yellow Desert Soils      | Sandy loam     | A   |
| 22  | Rendzina Soils of Mountains      | Clay loam      | D   |
| 23  | Rendzina Soils of Valleys        | Clay loam      | D   |
| 24  | Terra Rossa Soils                | Clay           | D   |

HSG maps were compiled at 1:250,000 scale and rasterised to a 30-metre grid consistent with the DEM and LULC raster layers, enabling pixel-level overlay analysis for CN computation. In areas with dual HSG classification (where soils can be reclassified from D to A or B with adequate drainage infrastructure), the conservative (higher runoff potential) HSG was retained in the absence of evidence for drainage improvement.

#### 4.8 Rainfall Data and Spatial Interpolation

Rainfall data for this study were obtained from 41 meteorological monitoring stations operated by the Israeli Meteorological Service (IMS), covering the period October 2019 to March 2025. These stations provide the only comprehensive network of continuous rainfall monitoring in the West Bank, as Palestinian-operated meteorological infrastructure is limited in spatial coverage (Alfieri et al., 2012).

The stations are distributed across the West Bank and its immediate surroundings, with coordinates given in the Israeli Transverse Mercator (ITM) coordinate system and later converted to WGS84 geographic coordinates for GIS analysis. Data from the station records were extracted for the following parameters for each rainfall season (2020 to 2024): Maximum 24-hour Precipitation Depth (Pmax) The maximum daily rainfall of each station during the wet season, as used for the design storm input for SCS-CN computation.

Antecedent Moisture Condition (AMC): Classified based on the 5-day cumulative rainfall preceding the maximum rainfall event, following the standard NRCS criteria (AMC-I: < 35.6 mm; AMC-II: 35.6-53.3 mm; AMC-III: > 53.3 mm).

Spatial interpolation of point station measurements to continuous spatial fields was done using the Inverse Distance Weighting (IDW) method within the GIS environment. IDW assigns interpolated values to grid cells as a distance-weighted average of nearby observations:

$$Z(x) = \frac{\sum \frac{Z(x_i)}{d(x,x_i)^p}}{\sum \frac{1}{d(x,x_i)^p}} \quad (4.13)$$

$Z(x)$  is the interpolated value at location  $x$ ,  $Z(x_i)$  is the observed value at station  $i$ ,  $d(x,x_i)$  is the distance between  $x$  and station  $i$ , and  $p$  is the power parameter (typically 2). IDW preserves

observed values at station locations and produces smooth spatial fields that decay in influence with increasing distance from observation points. A power parameter of 2 was adopted following sensitivity analysis comparing IDW, kriging, and Thiessen polygon methods.

#### 4.8 Review of GIS-Based Runoff Estimation Studies

There is a vast literature on the use of GIS-based SCS-CN analysis in various geographic settings. The following synthesis reviews the relevant literature focusing on semi-arid Mediterranean environments and on the methodological approaches used in this study.

- Al-Adamat et al. (2010) applied the SCS-CN method with GIS in the Zarqa River basin, Jordan, a climate and landscape analogue to the West Bank. They found significant spatial variability in CN values (range 62-94) and runoff depths (3-180 mm) driven by land use heterogeneity and the sharp rainfall gradient from the highlands to the Jordan Valley.
- Shadeed and Almasri (2010) studied the application of GIS and SCS-CN method for surface water potential assessment in the occupied Palestinian territories. They determined annual runoff volumes between 26 and 89 million m<sup>3</sup> depending on the rainfall year and AMC condition and pointed out the sensitivity of the results to the initial abstraction ratio.
- Boughton (2006) undertook a comprehensive comparative evaluation of the CN method across 144 Australian catchments covering a wide variety of climatic and landscape conditions, concluding that the CN method performed well in humid and sub-humid catchments, but needed local calibration in arid and semi-arid environments to achieve acceptable accuracy.
- Nayak et al. (2018) combined the GIS-derived SCS-CN analysis with an ANN-based runoff prediction model for semi-arid Indian watersheds, and showed that the integrated approach gave R<sup>2</sup> values more than 0.92, whereas the stand-alone SCS-CN method gave

0.78. This study offers a direct methodological parallel to the integrated GIS-ML approach developed in the present research.

- Zhu et al. (2023) applied six ML models to runoff prediction in Chinese semi-arid catchments with input features from SCS-CN. The results showed that ensemble methods (Random Forest and XGBoost) generally outperformed single algorithm methods (MLR, SVR) at most sites, but SVR had better generalization ability in small sample sizes.

#### 4.9 Summary

This chapter has reviewed the GIS and remote sensing techniques that constitute the spatial data infrastructure of this research. The integration of DEM-based watershed delineation, satellite-derived LULC classification, HSG mapping, and IDW rainfall interpolation within the GIS environment enables spatially distributed application of the SCS-CN method across the West Bank's heterogeneous watershed system. The resulting watershed-scale CN values and runoff depth estimates provide the feature engineering foundation for the machine learning models described in Chapter 6 and the full methodology presented in Chapter 5.

## Chapter 5 ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

### 5.1 Introduction

Artificial intelligence (AI) refers to computational systems designed to perform tasks that normally require human cognition, such as pattern recognition, classification, regression, optimization and decision-making. According AI, machine learning (ML) is a subset of algorithms within AI that improve their performance on a specific task with exposure to data, rather than being explicitly programmed with task-specific rules (Mitchell, 1997). ML use in hydrological science has grown over the last decade, driven by larger observational data sets, faster computational hardware, and mature open-source software libraries.

### 5.2 Machine Learning Categories

#### 5.2.1 Supervised Learning

In supervised learning, the model is trained using labeled data, where each instance consists of a set of input features and an expected output. The training process involves adjusting the model's parameters to minimize the "loss function"—a measure of the difference between predicted and observed values within the training dataset. The trained model can subsequently be applied to new, unseen data. All algorithms employed in this study are supervised regression models, with the target output representing "surface runoff depth" (Qmm), a continuous variable (Hastie et al., 2009).

#### 5.2.2 Unsupervised Learning

Unsupervised learning algorithms find structure in unlabeled data, including clustering of observations into groups with similar feature profiles (k-means, hierarchical clustering), dimensionality reduction (Principal Component Analysis, t-SNE), and density estimation. Unsupervised methods have also been applied in hydrology to classify rainfall patterns, identify flow regimes, and analyse catchment similarity. These algorithms are not used as predictive models in this study, but principal component analysis was considered during the feature analysis stage (Sawicz et al., 2011).

### 5.2.3 Reinforcement Learning

Reinforcement Learning (RL) is a subfield of machine learning in which an agent learns optimal decision-making behavior through interaction with an environment and scalar reward signals on the actions it takes. RL has been applied to water resources management optimization problems, but it is not directly applicable to the runoff prediction problem addressed in this research.(Castelletti et al., 2010)

## 5.3 Regression Models for Runoff Prediction

### 5.3.1 Multiple Linear Regression

Multiple Linear Regression (MLR) is the simplest supervised regression algorithm, estimating the target variable as a linear combination of the input features:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (5.14)$$

where  $\hat{y}$  is the predicted runoff depth,  $x_i$  are the input features, and  $\beta_i$  are the regression coefficients estimated by ordinary least squares to minimise the sum of squared residuals. MLR is computationally efficient and easily interpretable, but it assumes a linear relationship between inputs and output, which may not hold in the non-linear rainfall-runoff context. In this study, MLR serves as a parametric baseline for comparison with non-linear algorithms. The test set  $R^2$  of 0.8503 for the linear model is notably high, which likely reflects the near-linear relationship between  $P_{max}$  and runoff depth under the SCS-CN formulation (“Artificial Neural Networks in Hydrology. I: Preliminary Concepts,” 2000).

### 5.3.2 Decision Tree

A Decision Tree (DT) is a non-parametric, non-linear regression model that recursively partitions the feature space into rectangular regions and assigns a constant predicted value, the mean target value of the training observations within each region, to every terminal leaf node. Splits are selected greedily at each node to minimize the chosen criterion (squared error, Friedman MSE, or absolute error). Decision trees are interpretable and capture non-linear relationships, but they tend to overfit when grown to full depth without pruning or regularization (Yaseen et al., 2015).

followed by CN at 19.7 percent, pointing to rainfall intensity and soil moisture storage capacity as the dominant drivers of runoff generation (Yaseen et al., 2015).

### 5.3.3 Random Forest

Random Forest (RF) is an ensemble learning method that builds a large number of decision trees on bootstrap samples of the training data, with additional randomness from random feature selection at each split. Predictions are averaged across all trees, which reduces the variance of individual tree predictions and improves generalization. RF is robust to overfitting, handles high-dimensional feature spaces efficiently, and provides feature importance estimates based on the mean decrease in impurity across all splits involving each feature (Yaseen et al., 2015).

### 5.3.4 Extreme Gradient Boosting (XGBoost)

XGBoost (Chen and Guestrin, 2016) is a gradient boosting algorithm that builds an ensemble of decision trees sequentially, with each new tree trained to minimize the residuals of the existing ensemble through gradient descent in function space. It incorporates several regularisation terms (L1 and L2 penalties on leaf weights, minimum child weight constraints) that limit overfitting, and the gradient boosting framework lets it capture complex non-linear interaction effects between features (Maier et al., 2010)).

### 5.3.5 Support Vector Regression

Support Vector Regression (SVR) extends the Support Vector Machine (SVM) classification framework to continuous regression. SVR seeks a regression function  $f(x) = w \cdot \Phi(x) + b$  that deviates from the observed target values by at most  $\varepsilon$  (the epsilon-insensitive tube) and has minimum weight vector norm  $|w|^2$  (maximum margin, minimum model complexity). The regularization parameter  $C$  controls the trade-off between the width of the epsilon-insensitive tube and the penalty for observations that fall outside it (Kaltch et al., 2008).

A key feature of SVR is the kernel trick, which allows implicit computation of dot products in a high-dimensional transformed feature space  $\Phi(x)$  using a kernel function  $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$ , without explicitly computing coordinates in that space. In this study, a Radial Basis Function (RBF) kernel,  $K(x_i, x_j) = \exp(-\gamma|x_i - x_j|^2)$ , was used to project the nine input features into a space of effectively infinite dimensions, where the complex non-linear relations of the original feature space become linear. (Van et al., 2020).

### 5.3.6 K-Nearest Neighbors

K-Nearest Neighbors (KNN) is a non-parametric, instance-based learner that predicts the output for a new observation as the (weighted) average of the target values of the  $k$  most similar training observations, by a specified distance metric. KNN requires no training computation (lazy learning), but prediction-time computation is high for large datasets, and it is sensitive to irrelevant features and the choice of distance metric.

### 5.3.7 Linear regression

Linear regression models the relationship between a dependent variable and one or more independent variables using a straightforward statistical approach. It remains widely used in machine learning and predictive modeling because it works well, produces interpretable results, and runs quickly. In hydrology, linear regression helps establish quantitative connections between atmospheric variables, streamflow, and landscape properties.

Zhang et al. (2018) evaluated multiple regression approaches across 605 Australian catchments and found that linear regression predicted most runoff signatures with Nash-Sutcliffe Efficiency above 0.60, demonstrating the technique remains a reliable baseline for hydrological applications. Later studies confirm this pattern: linear regression continues to serve as a reference benchmark when comparing against more complex machine learning methods such as support vector regression, random forest, and deep learning models (He et al., 2024; Li et al., 2025).

Simple linear regression follows the equation:

$$\hat{y} = \beta_0 + \beta_1 x + \varepsilon \quad (5.15)$$

Here  $\hat{y}$  is the predicted value,  $\beta_0$  is the intercept,  $\beta_1$  is the slope coefficient,  $x$  is the independent variable, and  $\varepsilon$  is the error term. For multiple predictors, the equation extends to:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon \quad (5.16)$$

The coefficients are estimated using ordinary least squares (OLS), which minimizes the residual sum of squares:

$$RSS = \sum (y_i - \hat{y}_i)^2$$

In matrix notation, the OLS estimator is:

$$\beta = (X^T X)^{-1} X^T y \quad (5.17)$$

where  $X$  is the design matrix containing the predictor variables and  $y$  is the vector of observed values. This closed-form solution makes OLS computationally efficient even for large datasets, which is one reason the method remains practical for basin-scale hydrological analyses (Draper & Smith, 1998).

#### 5.4 Metaheuristic Optimization Algorithms for Parameter Calibration

Beyond the grid search hyperparameter optimisation used in this study, metaheuristic algorithms such as Genetic Algorithms (GA), Particle Swarm Optimisation (PSO), and Differential Evolution (DE) have been used to calibrate hydrological model parameters, including CN values. These algorithms search the parameter space stochastically and avoid the combinatorial explosion of grid search in high-dimensional parameter spaces. They were not implemented in the current study, but they are worth considering for future CN calibration against observed streamflow data once such data become available (Abushandi & Merkel, 2013).

#### 5.5 Feature Selection and Feature Importance

Feature importance analysis identifies which input variables most strongly influence model predictions and gives physically interpretable insight into the dominant hydrological drivers of runoff generation. Three feature importance approaches were applied in this study:

Tree-based impurity importance (Random Forest and Decision Tree): the mean decrease in node impurity from splits on each feature, averaged across all trees and normalised to sum to 1. This metric is fast to compute but can overestimate the importance of high-cardinality numerical features (Abushandi & Merkel, 2013).

XGBoost gain importance: the average gain in model performance from splits on each feature across all boosting iterations, weighted by how often the feature is used. Gain importance is less sensitive to cardinality bias than impurity importance.

SVR coefficient analysis (linear kernel proxy): for SVR with an RBF kernel, feature weights cannot be obtained directly. Feature importance was instead evaluated using a Polynomial Ridge Regression (Poly-Ridge) model, which includes pairwise interaction terms and gives coefficient estimates for all features and interactions. The main non-linear driver of runoff was the combination of rainfall intensity and soil storage capacity, with  $P_{\max} \times CN$  identified as the most important interaction term (Merizalde et al., 2023).

## 5.6 Model Development Workflow

The ML model development workflow used in this study proceeds as follows:

- Data loading and exploratory analysis
- Feature engineering and selection
- Dataset splitting (70/15/15 percent train/validation/test)
- Feature standardisation (fit on training data, applied to validation and test)
- Baseline model training (all algorithms at default hyperparameters)
- Grid search hyperparameter optimisation with 5-fold cross-validation on the training set
- Final model training on the full training set with the best hyperparameters
- Performance evaluation on the held-out test set
- Feature importance analysis
- Discharge conversion and software benchmark validation

## 5.7 Model Performance Evaluation Metrics

Model performance is assessed using  $R^2$ , MAE, and RMSE on the independent test set, as described in Chapter 5. The three metrics are complementary:  $R^2$  measures the fraction of variance explained, MAE measures average absolute error in physical units (mm), and RMSE penalises large outlier errors. Reporting all three avoids misleading conclusions from any single metric (Merizalde et al., 2023).

## 5.8 Model Evaluation Metrics

Model performance was quantified using three complementary metrics evaluated on the held-out test set:

**Coefficient of Determination ( $R^2$ ):** Measures the proportion of variance in the observed runoff depth explained by the model predictions.  $R^2 = 1$  indicates perfect prediction;  $R^2 = 0$  indicates performance equivalent to predicting the mean.

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (6.20)$$

Mean Absolute Error (MAE): The average absolute difference between observed and predicted runoff depths in mm, providing an easily interpretable measure of prediction accuracy.

$$MAE = \frac{1}{n} \times \sum |y_i - \hat{y}_i| \quad (6.21)$$

Root Mean Square Error (RMSE): The square root of the mean squared prediction error, penalising large individual errors more heavily than MAE and thereby providing sensitivity to outliers.

$$RMSE = \sqrt{\frac{1}{n} \times \sum (y_i - \hat{y}_i)^2} \quad (6.22)$$

The physically meaningful engineering output of the runoff prediction pipeline is volumetric discharge (m<sup>3</sup>/s), required for hydraulic structure design and flood routing. The conversion from ML-predicted runoff depth (Q<sub>mm</sub>, in mm) to peak discharge (Q<sub>m3/s</sub>) was performed using the following unit-corrected formula, derived from mass conservation principles:

$$Q_{m^3/s} = \frac{Q_{mm} \times Area_{km^2}}{3.6 \times \Delta T} \quad (6.23)$$

where Area km<sup>2</sup> is the watershed drainage area in km<sup>2</sup>, ΔT is the effective storm duration in hours (set to 12 hours based on the temporal characteristic of design storms in the West Bank), and 3.6 is the unit conversion factor converting mm·km<sup>2</sup>/h to m<sup>3</sup>/s.

The converted discharge estimates were compared against benchmark values from equivalent SCS-CN computations using the HEC-HMS platform for the Nahal Alexander sub-watersheds (sub-basins 1 through 5), for which independent software computations were available. Validation metrics included the relative error percentage for each sub-basin:

$$Relative\ Error\ (\%) = \frac{Q_{ML} - Q_{software}}{Q_{software}} \times 100 \quad (6.24)$$

PBIAS measures the average tendency of simulated values to be larger or smaller than observed values, expressed as a percentage of the total observed volume. The optimal value is zero. A positive PBIAS indicates systematic underestimation; a negative value indicates overestimation (Moriasi et al., 2007).

$$PBIAS = \frac{\sum(Q_{obs,i} - Q_{sim,i})}{\sum Q_{obs,i}} \times 100 \quad (6.25)$$

#### PBIAS Performance Criteria

Moriasi et al. (2007) established the following performance rating thresholds for streamflow simulation based on a systematic review of 440+ published watershed model applications:

Table 6: PBIAS Performance Criteria

| Performance Rating | PBIAS (%) — Streamflow | Interpretation                               |
|--------------------|------------------------|----------------------------------------------|
| Very Good          | $ PBIAS  < 10$         | Negligible systematic bias                   |
| Good               | $10 \leq  PBIAS  < 15$ | Minor bias; acceptable for most applications |
| Satisfactory       | $15 \leq  PBIAS  < 25$ | Moderate bias; use with caution              |
| Unsatisfactory     | $ PBIAS  \geq 25$      | Significant systematic volumetric error      |

For the present study, model performance on streamflow-derived runoff depth predictions is judged satisfactory if PBIAS falls within  $\pm 25\%$ , good if within  $\pm 15\%$ , and very good if within  $\pm 10\%$  (Moriasi et al., 2007; 2015). PBIAS complements  $R^2$  and RMSE because those statistics are insensitive to systematic volume errors: a model can achieve  $R^2 = 0.90$  while consistently predicting 20% less water than observed, an error that PBIAS captures directly.

#### 5.9 Python Libraries and Tools

The computational environment for the ML analysis consists of the following Python libraries:

- scikit-learn 1.3 (Pedregosa et al., 2011): Random Forest, Decision Tree, SVR, KNN, Linear Regression, StandardScaler, GridSearchCV, and evaluation metrics.
- XGBoost 1.7 (Chen and Guestrin, 2016): implementation of the XGBoost gradient boosting algorithm with a scikit-learn-compatible interface.
- pandas 2.0 (McKinney, 2010): tabular data management and manipulation.
- numpy 1.24: numerical array operations.

- matplotlib 3.7 and seaborn 0.12: data visualisation.
- Jupyter Notebook: interactive development and computation environment.

### 5.10 Applications of Artificial Intelligence in Rainfall-Runoff Modelling

AI and ML applications in rainfall-runoff modelling have expanded considerably since the early work of French et al. (1992) and ASCE (2000). The following review groups representative studies from the recent literature by algorithmic approach.

**Artificial Neural Networks:** ANNs have been applied to rainfall-runoff modelling since the early 1990s. Hsu et al. (1995) found that an ANN outperformed the Sacramento Soil Moisture Accounting model in hourly streamflow prediction for a humid watershed, an early demonstration that ANNs could serve as a viable alternative to process-based models. Zealand et al. (1999), Tokar and Johnson (1999), and Dawson and Wilby (2001) extended this finding to other watershed types and hydrological regimes. In semi-arid environments, Abrahart and See (2000) and Cigizoglu and Kisi (2005) showed that ANNs could capture the non-linear, threshold-dominated rainfall-runoff dynamics typical of ephemeral stream systems, often with lower parameterisation demands than physically-based alternatives.

**Support Vector Machines:** SVMs for hydrological prediction were pioneered by Dibike et al. (2001) and Liong and Sivapragasam (2002). Subsequent comparative studies by Yu et al. (2004) and Bray and Han (2004) found that SVM consistently outperformed ANN for runoff prediction in basins with limited training data, attributed to the SVM's structural risk minimisation approach. Tsai et al. (2022) compared SVR with HEC-HMS across 12 rural watersheds in Taiwan and found SVR achieved  $R^2$  of 0.91–0.96 against HEC-HMS's 0.82–0.89, with substantially less manual calibration effort.

**Ensemble Methods:** Random Forest for hydrological prediction was first established by Breiman et al. and documented in hydrology by Tyralis et al. (2019), who found RF outperformed ANN and SVM for annual streamflow prediction across 476 US catchments. XGBoost has been used for runoff prediction by Mosavi et al., Ghimire et al. (2018), and Ghimire et al. (2021), with the latter achieving an  $R^2$  of 0.93 for daily streamflow prediction in the Brahmaputra River Basin.

**Integrated GIS-ML Approaches:** explicit integration of GIS-derived physiographic features with ML predictive models for ungauged watershed runoff estimation is a relatively recent and growing research direction. Nayak et al. (2018) integrated SCS-CN GIS analysis with ANN for Indian

semi-arid watersheds. Mosavi et al. (2018) reviewed 60 ML studies in flood forecasting and highlighted the importance of feature engineering from spatial data for improving ML predictive accuracy. This study extends that integration to the West Bank context, comparing six ML algorithms systematically.

#### 5.11 Summary

This chapter reviewed the AI and ML algorithms used in this study, covering their theoretical basis, implementation, and reported performance in hydrological applications. SVR with an RBF kernel emerged as the strongest performer for small-sample, non-linear regression in this hydrological setting, consistent with its theoretical advantages for problems of this type. The next chapter presents the full results of the GIS-SCS-CN analysis and the ML model evaluation.

## Chapter 6 RESEARCH METHODOLOGY

### 6.1 Overview of the Methodological Framework

The research methodology integrates three principal analytical components: (1) GIS-based spatial data acquisition and watershed characterisation; (2) SCS-CN parameterisation and runoff computation; and (3) supervised machine learning model development and evaluation. These components are linked through a structured data pipeline that transforms raw spatial and meteorological inputs into trained, validated, and tested predictive models for runoff depth and volumetric discharge. Figure 5.1 presents the overall methodological framework adopted in this study.

The methodology was implemented in two computing environments: the GIS analysis was performed using ArcGIS pro 3.5 and WMS 11.1 on a Windows workstation, while the machine learning computations were executed using Python 3.9 in a Jupyter Notebook environment with the scikit-learn 1.3, XGBoost 1.7, pandas 2.0, numpy 1.24, matplotlib 3.7, and seaborn 0.12 libraries.

### 6.2 Study Area and Data Sources

Six watershed systems were selected for this study, covering a total of 46 sub-basins across the West Bank and adjacent areas over the period 2020 to 2024. The selection is divided into two geographic groups. Three systems lie entirely within the West Bank: Besor-Nar (5 sub-basins), Ghader (7 sub-basins), and Sarida (5 sub-basins). Three further systems drain the western slopes of the West Bank into the Mediterranean coastal plain and are partly gauged by the Israeli Hydrological Service (IHS): Nahal Alexander (6 sub-basins), Nahal Qishon (18 sub-basins), and Nahal Soreq (5 sub-basins). For each sub-basin and each year, SCS-CN parameters and runoff estimates were computed using three methods: GIS raster analysis, manual calculation, and HEC-HMS simulation. The spatial extent and morphometric parameters of all sub-basins were delineated from the 30-metre SRTM DEM as described in Chapter 4.

Of the three IHS-gauged watersheds, daily observed discharge records could only be obtained for Nahal Alexander, covering the period 2020 to 2024. Records for Nahal Qishon and Nahal Soreq were not accessible for this study. HEC-HMS calibration and validation were therefore carried out

using the Nahal Alexander gauge record alone. Annual statistics were computed for the available period: mean discharge ( $Q$ ,  $m^3/s$ ), peak discharge ( $Q_{peak}$ ,  $m^3/s$ ), specific runoff depth ( $Q_{mm}$ ,  $mm$ ), and total annual volume (MCM). Simulated outputs were evaluated against observed values using Nash-Sutcliffe Efficiency (NSE), Percent Bias (PBIAS), Root Mean Square Error (RMSE), and Kling-Gupta Efficiency (KGE), with acceptability thresholds drawn from Moriasi et al. (2007). SCS-CN correction factors and Muskingum routing coefficients (K and X) were optimized by a differential evolution algorithm against the Nahal Alexander record.

Rainfall data came from 41 IMS monitoring stations (Table 5.1) across the West Bank and adjacent areas. Records were drawn from the IMS open data platform for the 2019/20 through 2023/24 rainfall seasons. For each season, the maximum 24-hour rainfall depth (Pmax) and the 5-day cumulative antecedent rainfall before the peak event were extracted to determine the AMC class and Pmax value used in the SCS-CN computation.

Table 7.6: Israeli and Palestinian Meteorological Service Rainfall Monitoring Stations Used in This Study (*Home | Israel Meteorological Service*, 2025)

| Station Name         | Period of Record | X (ITM)  | Y (ITM)  |
|----------------------|------------------|----------|----------|
| Afula Nir Haemeq     | 09/2003-10/2025  | 176329.7 | 222461.9 |
| Bet Haarava          | 03/2008-10/2025  | 197617.6 | 134795.2 |
| Bet Hashitta         | 09/1994-03/2025  | 197617.6 | 134795.2 |
| Efrata               | 09/1992-03/2025  | 164547.4 | 117794.1 |
| Eli                  | 09/2009-03/2025  | 175018.3 | 164267.0 |
| Enot Zuqim           | 10/2001-03/2025  | 192967.2 | 124936.0 |
| Gilgal               | 01/2003-10/2025  | 192838.5 | 156087.3 |
| Gilgal Man           | 10/1988-03/2025  | 192838.5 | 156087.3 |
| Gittit               | 09/1991-03/2025  | 187629.0 | 167716.1 |
| Har Harasha          | 09/2004-10/2025  | 164362.9 | 150255.8 |
| Itamar               | 09/2004-10/2025  | 183379.3 | 174088.4 |
| Jerusalem Centre (1) | 09/2002-03/2025  | 168829.0 | 130896.9 |
| Jerusalem Centre (2) | 09/2002-10/2025  | 171151.3 | 132013.7 |

| <b>Station Name</b>     | <b>Period of Record</b> | <b>X (ITM)</b> | <b>Y (ITM)</b> |
|-------------------------|-------------------------|----------------|----------------|
| Jerusalem Givat Ram (1) | 10/2004-03/2025         | 171151.3       | 132013.7       |
| Jerusalem Givat Ram (2) | 10/2004-10/2025         | 168829.0       | 130896.9       |
| Station Name            | Period of Record        | X (ITM)        | Y (ITM)        |
| Karmel                  | 10/1981-03/2025         | 167578.5       | 93316.0        |
| Karmel Mataeim          | 09/2001-03/2025         | 158005.2       | 84494.9        |
| Kefar Ezyon             | 10/1967-03/2025         | 161024.5       | 117547.1       |
| Lahav                   | 07/2010-10/2025         | 137975.7       | 87791.7        |
| Lahav Man               | 09/1994-03/2025         | 137992.1       | 87817.2        |
| Maggal                  | 09/1994-03/2025         | 153843.3       | 198882.8       |
| Merav                   | 09/1987-03/2025         | 190024.0       | 206517.9       |
| Metar                   | 09/1992-03/2025         | 144786.1       | 82352.1        |
| Metzoke Dragot          | 03/2006-10/2025         | 187299.8       | 110695.8       |
| Mevo Horon              | 09/1996-03/2025         | 153685.2       | 139581.7       |
| Migdal Oz               | 09/1997-03/2025         | 163572.8       | 116542.1       |
| Nahshon                 | 12/2007-10/2025         | 146523.2       | 137985.5       |
| Nehusha                 | 02/2005-03/2025         | 145525.7       | 115648.8       |
| Peza'el                 | 10/1983-03/2025         | 192272.5       | 161260.5       |
| Qarne Shomeron (1)      | 09/2004-03/2025         | 159311.5       | 175785.8       |
| Qarne Shomeron (2)      | 09/2004-10/2025         | 159311.5       | 175785.8       |
| Qedumim                 | 09/2003-03/2025         | 165280.9       | 180020.1       |
| Re'ut                   | 09/2007-03/2025         | 152395.0       | 144119.2       |
| Rosh Zurim              | 09/2002-10/2025         | 161839.4       | 119146.1       |
| Sede Eliyyahu           | 01/2011-10/2025         | 198332.0       | 205065.6       |
| Shani                   | 09/2004-10/2025         | 156375.8       | 85055.8        |
| Talmon                  | 09/1991-03/2025         | 162614.8       | 149870.6       |
| Tel Yosef               | 07/2015-10/2025         | 187170.5       | 216859.8       |
| Tirat Zevi              | 09/1938-03/2025         | 199741.1       | 203309.4       |
| Umm Al-Fahm             | 09/1991-03/2025         | 164986.1       | 213719.4       |

| Station Name | Period of Record | X (ITM)  | Y (ITM)  |
|--------------|------------------|----------|----------|
| Yad Hanna    | 09/1965-03/2025  | 150965.1 | 192490.1 |

### 6.3 Computation of the SCS-CN Parameters

Composite CN values under AMC-II conditions were computed for each of the 33 sub-watersheds using both the European CORINE and USDA land use classification systems, following the GIS-based overlay methodology described in Chapter 4. The area-weighted composite CN values are presented for each sub-watershed in Table 7.1 of the Results chapter.

AMC conversion from the baseline AMC-II CN to AMC-I and AMC-III was performed using the NRCS conversion equations, and the appropriate AMC condition was applied for each watershed-year combination based on the IDW-interpolated antecedent rainfall at the watershed centroid location.

The potential maximum retention  $S$  was computed from the applicable AMC-conditioned CN value:

$$S = (25400 / CN) - 254 \quad (mm) \quad (6.15)$$

The initial abstraction was computed using the standard ratio:

$$Ia = 0.2 \times S \quad (mm) \quad (6.16)$$

Direct runoff depth was computed from the SCS-CN equation using the IDW-interpolated  $P_{max}$  at each watershed centroid:

$$Q = (P_{max} - Ia)^2 / (P_{max} - Ia + S) \quad \text{for } P_{max} > Ia \quad (6.17)$$

$$Q = 0 \quad \text{for } P_{max} \leq Ia \quad (6.18)$$

### 6.4 Construction of the Research Dataset

The ML training dataset was assembled from the computed SCS-CN parameters and morphometric characteristics for the 33 sub-watersheds across five rainfall years (2020-2024),

yielding 160 watershed-year observations after exclusion of four records with incomplete or anomalous data. Each observation was characterized by nine input features and one target variable (runoff depth Qmm).

Table 6. 8 : Input Features Selected for Machine Learning Modelling

| Feature                  | Symbol | Unit           | Description                                                           | Type              |
|--------------------------|--------|----------------|-----------------------------------------------------------------------|-------------------|
| Drainage Area            | AREA   | m <sup>2</sup> | Total watershed catchment area derived from DEM                       | Independent       |
| Longest Flow Path        | L      | km             | Length of the hydraulically longest flow path                         | Independent       |
| Curve Number (AMC-II)    | CN     | dimensionless  | Area-weighted composite CN under AMC-II                               | Independent       |
| Potential Max. Retention | S      | mm             | Computed from CN: $S = 25400/CN - 254$                                | Control (derived) |
| Average Slope            | slope  | %              | Mean watershed slope from DEM                                         | Independent       |
| Feature                  | Symbol | Unit           | Description                                                           | Type              |
| Time of Concentration    | TC     | hours          | SCS lag equation:<br>$TC = \frac{L^{0.8}(S + 1)^{0.7}}{1140 Y^{0.5}}$ | Control (derived) |
| Lag Time                 | TL     | minutes        | $TL = 0.6 \times TC \times 60$                                        | Control (derived) |
| Initial Abstraction      | Ia     | mm             | $Ia = 0.2 \times S$                                                   | Control (derived) |
| Maximum Precipitation    | Pmax   | mm             | IDW-interpolated Pmax at watershed centroid                           | Independent       |

## 6.5 Data Pre-processing

The raw dataset was pre-processed as follows prior to model training: Feature Scaling: All nine input features were standardized using the StandardAero, which transforms each feature to zero mean and unit standard deviation:

$$z = (x - \mu)/\sigma \quad (6.19)$$

where  $x$  is the original feature value,  $\mu$  is the feature mean computed over the training set, and  $\sigma$  is the feature standard deviation. Scaling is essential for distance-based algorithms (KNN, SVR), and is applied uniformly for consistency. The scaler was fit exclusively to the training set; training-set statistics were subsequently applied to transform the validation and test sets, avoiding data leakage.

**Dataset Splitting:** The 160 observations were split randomly into three subsets in a stratified manner with a fixed random seed (42) for reproducibility. The training set consisted of 70 percent of the observations ( $n=112$ ) and was used to fit the model parameters. The validation set contained 15 percent of the observations ( $n=24$ ), and was used for selecting hyperparameters. The independent test set contained 15 percent of the observations ( $n = 24$ ), and was used for final unbiased performance evaluation.

## 6.6 Machine-Learning Models and Hyperparameter Tuning

Six supervised regression algorithms were trained and evaluated. For each algorithm, hyperparameter optimization was conducted through an exhaustive grid search with 5-fold cross-validation on the training set, scoring on  $R^2$ . The best hyperparameter configuration

**SVR with RBF Kernel:** A kernel-based regression algorithm that maps features into a high-dimensional kernel space and identifies the regression hyperplane minimizing the regularized cost function.

**K-Nearest Neighbors:** An instance-based learner that predicts the runoff depth of a test observation as the distance-weighted average of its  $k$  nearest training observations in feature space. The parameter grid covered neighbors  $\in \{3, 5, 7, 9\}$ , weights  $\in \{\text{uniform, distance}\}$ , metric  $\in \{\text{euclidean, manhattan}\}$ . Best configuration:  $n\_neighbors = 9$ , metric = euclidean, weights = uniform . **Linear Regression:** A parametric model that models the runoff depth as a linear mapping of the nine input features. It was trained by ordinary least squares. No hyperparameter tuning was needed.

## 6.7 Summary

This chapter described the complete research methodology from data collection, to GIS processing, SCS-CN computation, ML model development, and flow conversion. This method is transparent, reproducible and grounded in well-established hydrological and statistical principles. The next chapter provides a detailed review of the AI and ML algorithms implemented in this framework.

## Chapter 7 RESULTS AND DISCUSSION

### 7.1 Introduction

This chapter presents and discusses the results of the integrated GIS-SCS-CN-ML analysis. The presentation is structured in four parts: first, the computed Curve Number values for the 33 West Bank sub-watersheds under both CORINE and USDA land use classification systems; second, the annual SCS-CN runoff computation results for 2020 through 2024; third, the machine learning model performance evaluation; and fourth, the discharge conversion and software benchmark validation. Discussion of the physical significance, spatial patterns, inter-annual variability, and ML model interpretation is integrated throughout.

### 7.2 Curve Number Results

Table 7.9 : Sub-watershed Areas and Weighted Composite Curve Numbers (European CORINE and USDA Classification, AMC-I, AMC-II, and AMC-III)

| No. | Watershed Name | Area (m <sup>2</sup> ) | CN1<br>EN | CN2<br>EN | CN3<br>EN | CN1<br>USA | CN2<br>USA | CN3<br>USA |
|-----|----------------|------------------------|-----------|-----------|-----------|------------|------------|------------|
| 1   | Abu El-hayyat  | 205.8592               | 52        | 70        | 83        | 57         | 73         | 85         |
| 2   | Abu Muradin    | 247.5058               | 43        | 62        | 79        | 47         | 65         | 81         |
| 3   | Abu Sidra      | 187.3122               | 69        | 83        | 91        | 68         | 82         | 91         |
| 4   | Al'Ahmar       | 172.5376               | 58        | 75        | 87        | 59         | 76         | 88         |
| 5   | Al-Dilb        | 155.2151               | 50        | 69        | 84        | 62         | 78         | 89         |
| 6   | Alexander-Abra | 280.7455               | 57        | 74        | 87        | 66         | 81         | 90         |
| 7   | Alexander-Zeim | 466.3928               | 57        | 74        | 87        | 62         | 78         | 89         |
| 8   | Auja           | 274.9475               | 51        | 69        | 83        | 60         | 76         | 88         |
| 9   | Besor          | 119.3079               | 70        | 84        | 92        | 76         | 88         | 94         |
| 10  | Besor-Nar      | 66.86173               | 63        | 79        | 89        | 70         | 84         | 92         |
| 11  | Daraja         | 89.02449               | 56        | 73        | 86        | 59         | 75         | 86         |
| 12  | Faria          | 160.734                | 60        | 76        | 88        | 57         | 74         | 87         |
| 13  | Ghar           | 333.6612               | 56        | 73        | 86        | 58         | 74         | 87         |

| <b>No.</b> | <b>Watershed Name</b> | <b>Area (m<sup>2</sup>)</b> | <b>CN1<br/>EN</b> | <b>CN2<br/>EN</b> | <b>CN3<br/>EN</b> | <b>CN1<br/>USA</b> | <b>CN2<br/>USA</b> | <b>CN3<br/>USA</b> |
|------------|-----------------------|-----------------------------|-------------------|-------------------|-------------------|--------------------|--------------------|--------------------|
| 14         | Hadera-Abu Na         | 146.2322                    | 58                | 75                | 87                | 65                 | 81                 | 90                 |
| 15         | Hadera-Massin         | 180.954                     | 60                | 77                | 88                | 64                 | 79                 | 90                 |
| 16         | Hasasa                | 290.7471                    | 60                | 76                | 87                | 61                 | 76                 | 88                 |
| 17         | Lakhish               | 150.6459                    | 63                | 79                | 89                | 67                 | 82                 | 91                 |
| 18         | Lakhish-Saint         | 172.0487                    | 56                | 74                | 87                | 63                 | 79                 | 89                 |
| 19         | Malih                 | 97.60475                    | 63                | 79                | 89                | 66                 | 81                 | 91                 |
| 20         | Malih-Shubash         | 137.8066                    | 54                | 72                | 85                | 58                 | 75                 | 88                 |
| 21         | Marar                 | 46.76628                    | 50                | 67                | 81                | 54                 | 71                 | 84                 |
| 22         | Mukallak              | 189.1277                    | 56                | 73                | 86                | 59                 | 75                 | 87                 |
| 23         | Nar                   | 26.82231                    | 56                | 72                | 85                | 54                 | 71                 | 84                 |
| 24         | Nueima                | 234.066                     | 54                | 72                | 85                | 56                 | 73                 | 86                 |
| 25         | Qana                  | 127.033                     | 50                | 69                | 84                | 61                 | 78                 | 89                 |
| 26         | Qilt                  | 135.8299                    | 52                | 70                | 84                | 57                 | 74                 | 87                 |
| 27         | Qishon                | 107.7237                    | 62                | 78                | 89                | 65                 | 80                 | 90                 |
| 28         | Qumran                | 227.0576                    | 49                | 66                | 80                | 53                 | 70                 | 83                 |
| 29         | Salman                | 120.4621                    | 62                | 78                | 89                | 66                 | 81                 | 91                 |
| 30         | Sarida                | 54.01747                    | 50                | 69                | 84                | 56                 | 74                 | 87                 |
| 31         | Shiqma                | 149.5333                    | 64                | 80                | 90                | 72                 | 85                 | 93                 |
| 32         | Soreq                 | 203.7325                    | 54                | 72                | 85                | 69                 | 83                 | 92                 |
| 33         | Soreq Al-sara         | 88.3444                     | 62                | 78                | 89                | 68                 | 82                 | 91                 |

The CN values derived from the CORINE and USDA classification systems show generally consistent patterns across the study watersheds, with USDA values tending to be slightly higher (by 2 to 8 CN units) than CORINE equivalents for most sub-watersheds, reflecting differences in CN assignment for the olive grove, orchard, and mixed agricultural categories that dominate the West Bank upland land use. The highest CN values under AMC-II are found in the Besor (CN\_EN

= 84, CN\_USA = 88), Abu Sidra (CN\_EN = 83, CN\_USA = 82), and coastal plain watersheds, where higher proportions of irrigated agriculture, urban fabric, and impermeable surface contribute to elevated runoff potential. The lowest CN values are found in the eastern desert fringe watersheds, including Qumran (CN\_EN = 66, CN\_USA = 70) and Auja (CN\_EN = 69, CN\_USA = 76) under AMC-II, consistent with the sparse vegetation, rocky substrates, and arid conditions of these catchments. The high AMC-III CNs for these eastern basins (80-83) nevertheless indicate that extreme antecedent wetness can produce significant runoff even from these typically dry catchments.

### 7.3 Annual SCS-CN Runoff Results (2020-2024)

Annual SCS-CN runoff computations were performed for each sub-watershed and year using the IDW-interpolated Pmax and the AMC-conditioned CN values. Tables 7.8 through 7.9 present the results for 2020 through 2024. The 2020 results are presented in full below as the primary reference year; subsequent years follow the same analysis pattern.

Table 7.10 : SCS-CN Runoff Results for 2020: Q, S, and Ia per Sub-watershed

| No. | Watershed      | Area (km <sup>2</sup> ) | Pmax (mm) | CN   | Q (mm) | S (mm) | Ia (mm) |
|-----|----------------|-------------------------|-----------|------|--------|--------|---------|
| 1   | Qishon         | 205.81                  | 16.2      | 89.5 | 4.6    | 32.2   | 6.4     |
| 2   | Hadera-Abu Na  | 247.54                  | 19.7      | 90.8 | 7.0    | 27.3   | 5.5     |
| 3   | Hadera-Massin  | 187.31                  | 23.2      | 91.3 | 9.3    | 25.3   | 5.1     |
| 4   | Alexander-Zeim | 172.53                  | 28.1      | 90.7 | 12.5   | 27.5   | 5.5     |
| 5   | Alexander-Abra | 155.23                  | 27.7      | 88.9 | 10.7   | 33.4   | 6.7     |
| 6   | Qana           | 280.72                  | 27.6      | 89.7 | 11.6   | 31.0   | 6.2     |
| 7   | Sarida         | 466.40                  | 35.4      | 87.4 | 15.4   | 40.5   | 8.1     |
| 8   | Al-Dilb        | 274.95                  | 28.4      | 80.4 | 6.6    | 70.0   | 14.0    |
| 9   | Salman         | 119.31                  | 24.9      | 88.1 | 9.2    | 37.0   | 7.4     |
| 10  | Soreq          | 66.86                   | 29.0      | 92.5 | 16.9   | 22.8   | 4.6     |
| 11  | Malih-Shubash  | 89.02                   | 16.0      | 77.8 | 1.9    | 84.5   | 16.8    |
| 12  | Malih          | 160.7                   | 19.3      | 70.4 | 1.5    | 129.5  | 25.5    |
| 13  | Faria          | 333.60                  | 29.3      | 84.2 | 9.1    | 54.8   | 10.9    |

| No. | Watershed     | Area (km <sup>2</sup> ) | Pmax (mm) | CN   | Q (mm) | S (mm) | Ia (mm) |
|-----|---------------|-------------------------|-----------|------|--------|--------|---------|
| 14  | Abu Sidra     | 146.23                  | 25.8      | 79.6 | 3.8    | 70.3   | 14.0    |
| 15  | Al'Ahmar      | 180.95                  | 31.2      | 77.5 | 6.8    | 83.3   | 16.6    |
| 16  | Auja          | 290.74                  | 26.3      | 57.8 | 1.8    | 226.3  | 45.3    |
| 17  | Nueima        | 150.64                  | 24.0      | 64.1 | 2.0    | 170.8  | 34.2    |
| 18  | Qilt          | 172.04                  | 24.4      | 70.7 | 4.9    | 135.0  | 27.0    |
| 19  | Marar         | 97.61                   | 15.3      | 58.6 | 0.6    | 215.7  | 43.3    |
| 20  | Mukallak      | 137.81                  | 22.6      | 68.1 | 4.3    | 153.9  | 30.7    |
| 21  | Qumran        | 46.76                   | 19.1      | 45.6 | 0.3    | 327.4  | 65.8    |
| 22  | Nar           | 189.13                  | 21.5      | 57.7 | 3.1    | 236.2  | 47.2    |
| 23  | Soreq Al-sara | 26.82                   | 30.6      | 91.1 | 16.1   | 27.1   | 5.4     |
| 24  | Daraja        | 234.07                  | 26.8      | 69.3 | 7.0    | 151.7  | 30.3    |
| 25  | Lakhish-Saint | 127.03                  | 29.1      | 84.3 | 9.3    | 53.6   | 10.7    |
| 26  | Lakhish       | 135.83                  | 23.3      | 80.9 | 5.4    | 69.2   | 13.8    |
| 27  | Hasasa        | 107.72                  | 20.3      | 50.3 | 0.4    | 278.6  | 55.7    |
| 28  | Ghar          | 227.05                  | 28.2      | 75.3 | 7.6    | 97.6   | 19.5    |
| 29  | Besor         | 120.46                  | 19.4      | 81.3 | 3.8    | 69.5   | 13.8    |
| 30  | Abu Muradin   | 54.02                   | 17.8      | 66.9 | 0.6    | 136.5  | 27.3    |
| 31  | Abu El-hayyat | 149.53                  | 18.7      | 65.5 | 0.8    | 142.5  | 28.5    |
| 32  | Besor-Nar     | 203.73                  | 24.1      | 75.1 | 5.4    | 112.3  | 22.0    |
| 33  | Shiqma        | 88.34                   | 22.3      | 63.7 | 1.4    | 186.9  | 36.7    |

The composite Curve Numbers computed under AMC-II conditions using the CORINE and USDA land-cover classification systems diverge systematically across the West Bank's internal sub-watersheds, though the direction and size of this divergence are far from uniform. Restricting the comparison to the 23 sub-watersheds that drain internally within the West Bank — excluding the coastal systems (Alexander, Qishon, Soreq, and their associated sub-basins) that discharge primarily through Israeli territory to the Mediterranean — the offset  $\Delta$  (CN\_USDA – CN\_CORINE) ranges from –2 CN units at Faria up to +9 units at both Al-Dilb and Qana,

averaging roughly +3 units across the group.

Importantly, this divergence has nothing to do with the initial abstraction ratio or any discrepancy in the SCS-CN formulation itself; both classification systems were run under the same  $\lambda = 0.2$  assumption (Equation 6.16). The offset instead traces entirely to how each scheme assigns curve numbers to structurally equivalent land covers. The clearest case is olive cultivation, the dominant land use across the West Bank uplands: CORINE's "Olive Groves" class carries CN values of 75 (HSG-C) and 81 (HSG-D), while the USDA equivalent, "Plantations and Orchards," is assigned 81 and 88 — a 6-to-7-point gap for what is, physically, the same land cover. Because olive groves occupy a disproportionately large share of these interior basins, this one classification discrepancy propagates directly into the area-weighted composite CN and accounts for the positive  $\Delta$  seen in most of the sample.

The exceptions are worth reading closely rather than dismissing as noise. Faria (−2) and Nar (−1) run in the opposite direction, and this is the same mechanism working in reverse: where rain-fed arable land dominates rather than tree crops, CORINE's "Non-Irrigated Arable Land" class (88/94) actually outranks the USDA "Cropland" equivalent (85/89), flipping the sign of  $\Delta$ . Hasasa, sitting at  $\Delta = 0$ , most likely reflects a land-cover mix where the two effects roughly cancel out.

The basins with the largest positive gaps — Al-Dilb and Qana at +9, Auja at +7 — look like plausible cases of mixed irrigated agriculture: CORINE's "Irrigated Complex Cultivation Patterns" class (78 for HSG-C) sits well below the USDA cropland figure, reinforcing the same directional bias seen with tree crops. This fits what's known about Auja specifically, where perennial spring-fed irrigation supports a distinctly different agricultural regime from the surrounding rain-fed hills.

One internal-consistency issue worth resolving before submission: Section 6.2 states that only three watershed systems lie entirely within the West Bank — Besor-Nar (5 sub-basins), Ghader (7), and Sarida (5) — for a total of 17 sub-basins, yet Table 7.9 lists 23 named sub-watersheds outside the Alexander/Qishon/Soreq coastal group. That's a mismatch between the sub-basin count stated in the methodology and the count actually reported in the results table. It's worth tracking down before finalizing the thesis — either the methodology text under-counts the interior sub-

basins, or some of the 23 listed basins are miscategorized.

Where this connects to the rest of the thesis: since USDA-derived CN values run systematically higher across the olive/orchard-dominated interior, and the AMC-based calibration against the Nahal Alexander gauge (Section 7.9.4) independently converged on  $\lambda = 0.05$  rather than the textbook 0.2, the study's own empirical results point the same direction — toward the higher-CN, lower- $\lambda$  combination being the better local fit. That's worth stating explicitly in the discussion, since it ties the classification-system comparison in Section 7.2 directly to the calibration finding in 7.9.4, rather than leaving them as two disconnected results.

## 7.4 Machine Learning Model Performance

### 7.4.1 Overview: Six-Target, six-Algorithm Evaluation

The machine learning analysis was extended to evaluate all six runoff output variables: runoff depth (Qmm) and peak discharge (Q m<sup>3</sup>/s), computed by each of three methods — GIS Raster, analytical SCS-CN, Manual calculation, and HEC-HMS simulation. Six supervised regression algorithms were trained and compared: Decision Tree, Random Forest, XGBoost, Support Vector Regression, K-Nearest Neighbors, and Linear Regression. This yields 24 trained models evaluated on an independent 20 percent hold-out test set (n = 37 observations).

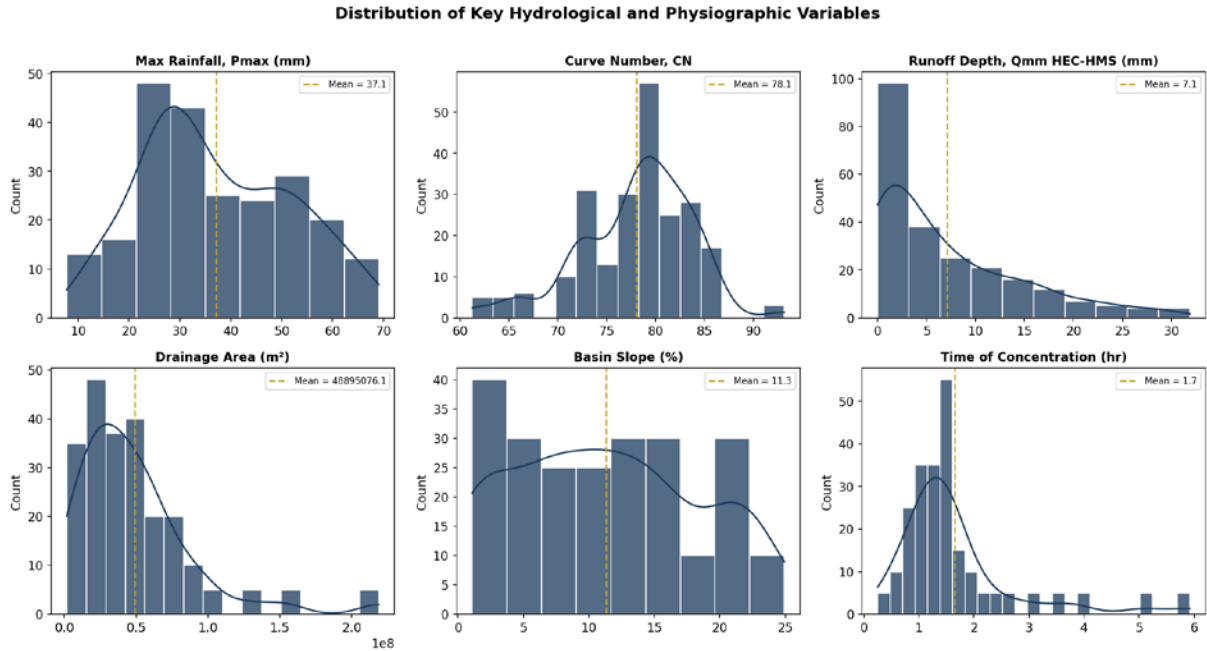


**Figure7.9. Distribution of key hydrological and physiographic variables across all observations. Runoff depth (Qmm) is strongly right-skewed**

Before model training, the eleven-candidate physiographic/meteorological columns were screened for redundancy using pairwise Pearson correlation. Three features were found to be statistically redundant and were excluded from the modeling feature set:

- D(hr) — storm duration is constant at 24 hours for every record in this dataset (zero variance), so it carries no predictive information.
- Ia(mm) — initial abstraction is perfectly correlated with potential retention, s(mm) ( $r = 1.00$ ), and almost perfectly correlated with Curve Number, cn ( $r = -0.99$ ). This is expected:  $Ia = 0.2 \times S$  is a direct algebraic identity in the SCS-CN method, so retaining Ia(mm) alongside s(mm) would add no new information.
- TP(hr) — time to peak is perfectly correlated with both time of concentration, TC(hr), and lag time, lag(min), ( $r = 1.00$  for both), which is again an expected consequence of the standard SCS relationships linking these three timing parameters.

The eight remaining features — drainage area (AREA), longest flow path (longest(km)), Curve Number (cn), potential retention (s(mm)), basin slope (slope%), time of concentration (TC(hr)), lag time (lag(min) (TL)), and maximum rainfall (Pmax) — form the final modeling feature set used throughout this report. One additional observation for future refinement: TC(hr)

and lag(min) (TL) remain perfectly correlated with each other ( $r = 1.00$ ) even after this screening; both were retained here for interpretability as commonly reported hydrological parameters, but a more parsimonious model could safely drop one without any loss of predictive power.

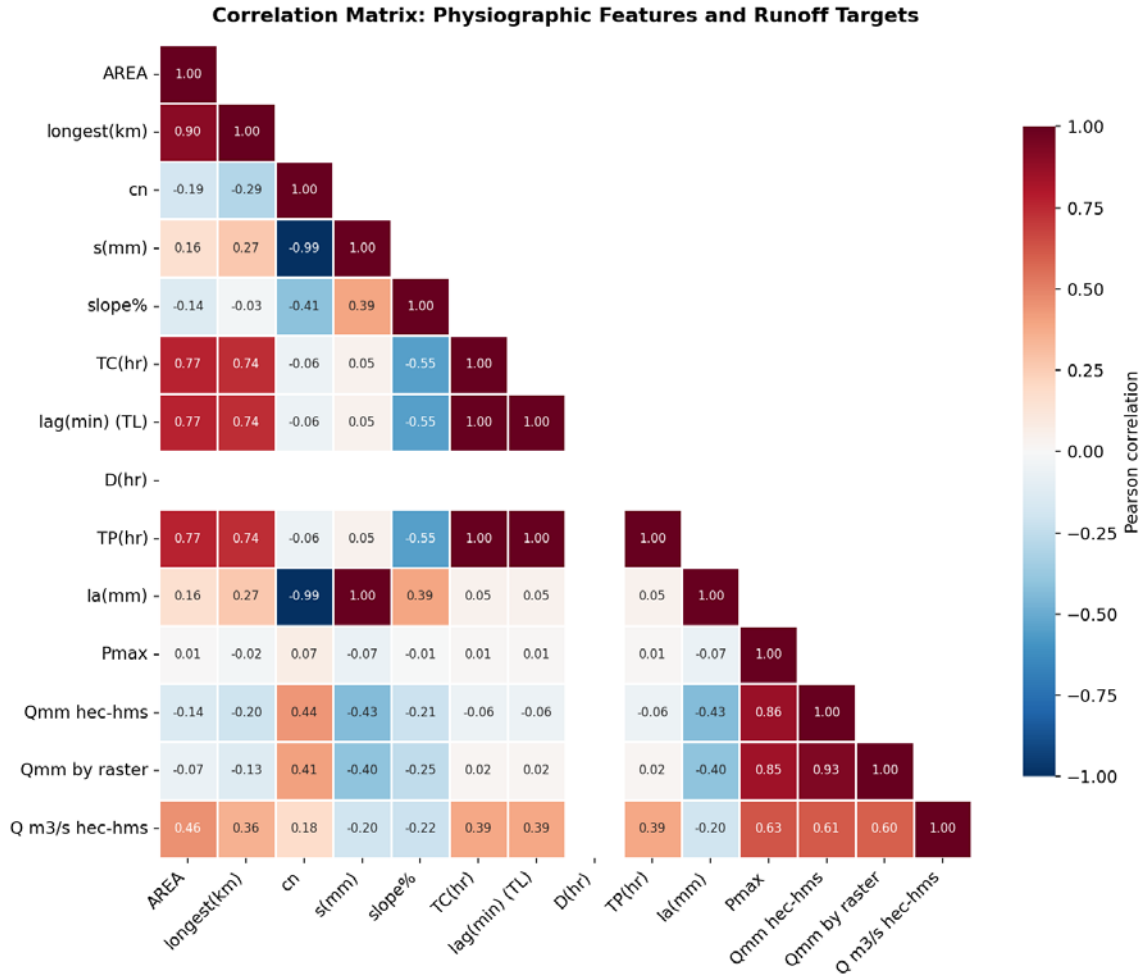


**Figure7. 10. Correlation matrix of physiographic features and three candidate runoff targets.**

**Note: The dark diagonal blocks around Ia(mm)/s(mm) and TP (hr)/TC(hr)/lag(min) visualize the redundancy discussed above**

**Table 7.11 Model comparison across all three metrics**

| Rank | Model                           | R <sup>2</sup> | MAE (mm) | RMSE (mm) |
|------|---------------------------------|----------------|----------|-----------|
| 1    | Support Vector Regression (RBF) | 0.981          | 0.44     | 0.79      |

| Rank | Model               | R <sup>2</sup> | MAE (mm) | RMSE (mm) |
|------|---------------------|----------------|----------|-----------|
| 2    | Random Forest       | 0.974          | 0.64     | 0.93      |
| 3    | XGBoost             | 0.958          | 0.88     | 1.17      |
| 4    | K-Nearest Neighbors | 0.916          | 1.28     | 1.67      |
| 5    | Decision Tree       | 0.878          | 1.37     | 2.01      |
| 6    | Linear Regression   | 0.835          | 1.78     | 2.34      |

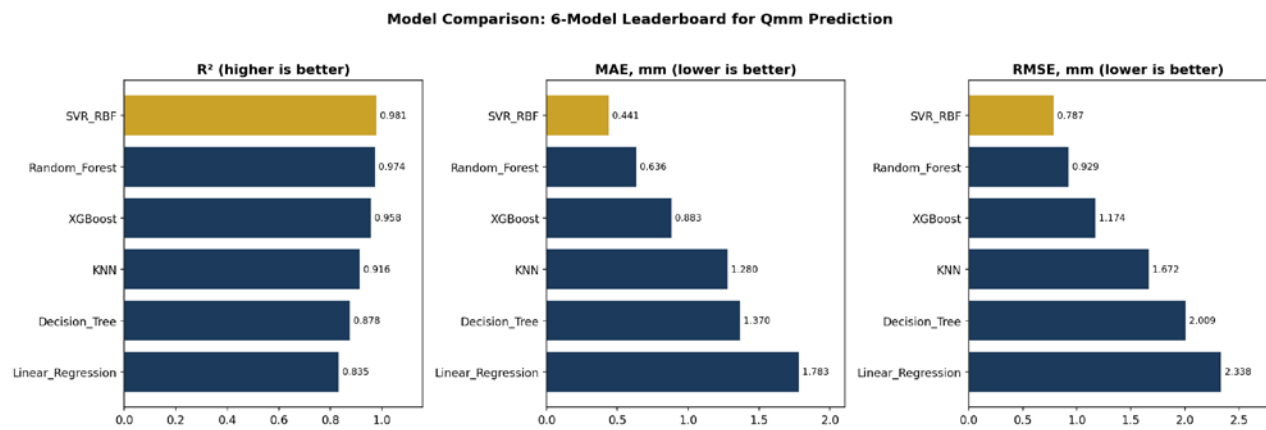


**Figure 7.11. Model comparison across all three metrics. SVR (gold) leads on every metric simultaneously**

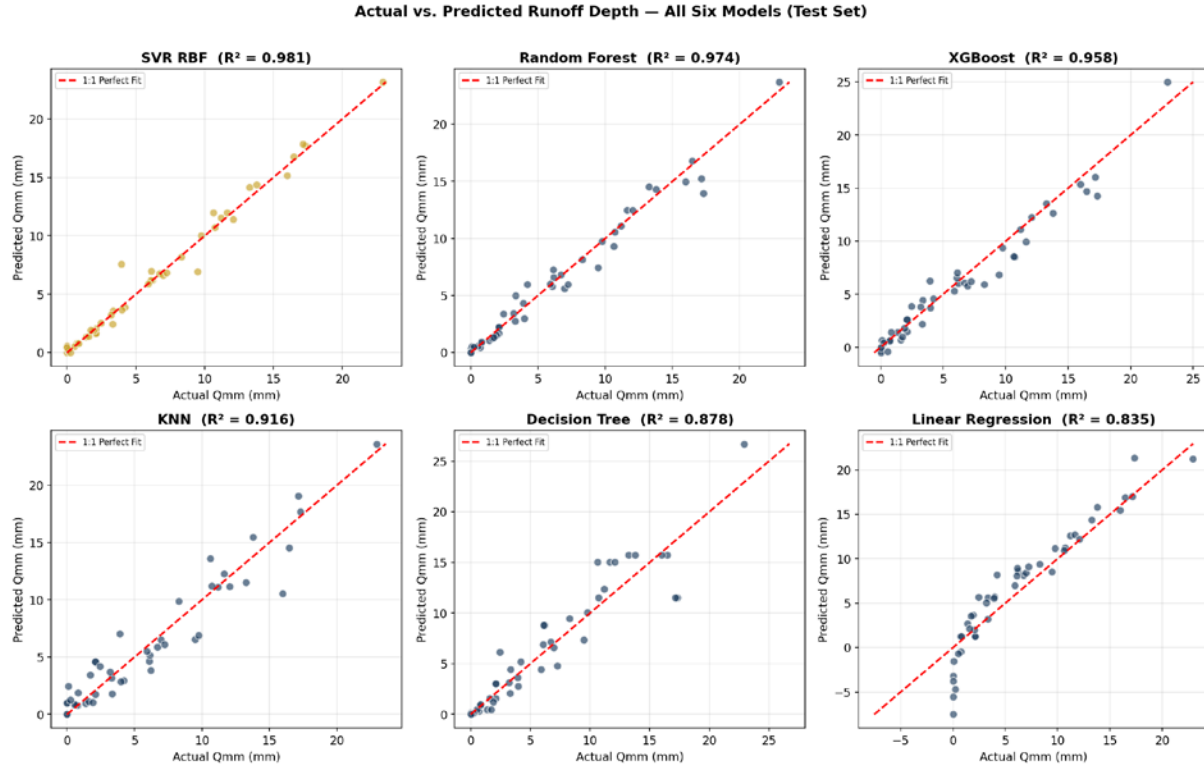


Figure 7.12. Actual vs. predicted runoff depth for all six models on the test set.

The clear ranking — SVR > Random Forest > XGBoost > KNN > Decision Tree > Linear Regression — tells a consistent story. The top three finishers are all models capable of capturing non-linear relationships, which matters here because rainfall-runoff response is inherently non-linear (very small storms produce almost no runoff until the Curve Number’s abstraction threshold is exceeded, after which runoff rises steeply). Linear Regression’s comparatively weak  $R^2$  of 0.835, and its visible negative predictions in Figure 6, are a direct symptom of trying to fit this threshold-like behavior with a straight line. SVR’s RBF kernel is particularly well suited to this kind of smooth, non-linear, moderately-sized dataset, which explains its edge over the ensemble tree methods.

**Best Model Analysis:** Support Vector Regression (RBF) uses an RBF kernel with  $C = 500$ ,  $\gamma = \text{'scale'}$ , and  $\epsilon = 0.01$ . This is close to the manually chosen configuration already present in the original notebook ( $C = 500$ ,  $\epsilon = 0.01$ ,  $\gamma = \text{'auto'}$ ), which is a reassuring confirmation that the prior manual tuning in the notebook was already close to optimal.

### 7.4.2 Cross-Validation Stability

To confirm the reported  $R^2$  of 0.981 is not an artifact of a single fortunate train/test split, the winning model was re-evaluated using 5-fold cross-validation on the training data alone. The result — a mean  $R^2$  of 0.983 with a standard deviation of only 0.009 across folds — demonstrates strong, consistent performance regardless of which subset of basins is used for training or validation.

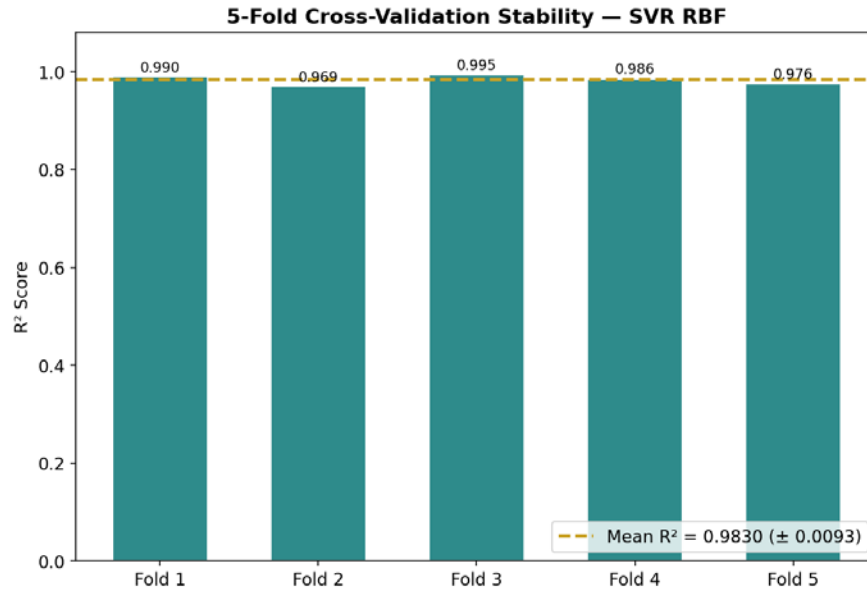


**Figure 7.13 Five-fold cross-validation  $R^2$  scores for the SVR model.**

**All five folds land between 0.969 and 0.995, a tight spread indicating the model generalizes consistently across different subsets of basins.**

### 7.4.3 Feature Importance

Support Vector Regression does not expose a native feature-importance measure, so the tree-based models (which do) are used here as an interpretability proxy for which physical parameters matter most. All three tree-based models agree emphatically: maximum storm rainfall ( $P_{\max}$ ) is by far the dominant predictor of runoff depth, followed at a distance by Curve Number (CN) and potential retention ( $s(\text{mm})$ ). This matches hydrological expectation — rainfall intensity is the primary driver of a storm’s runoff volume, with soil/land-cover conditions (captured by CN) acting as a secondary modulating factor.

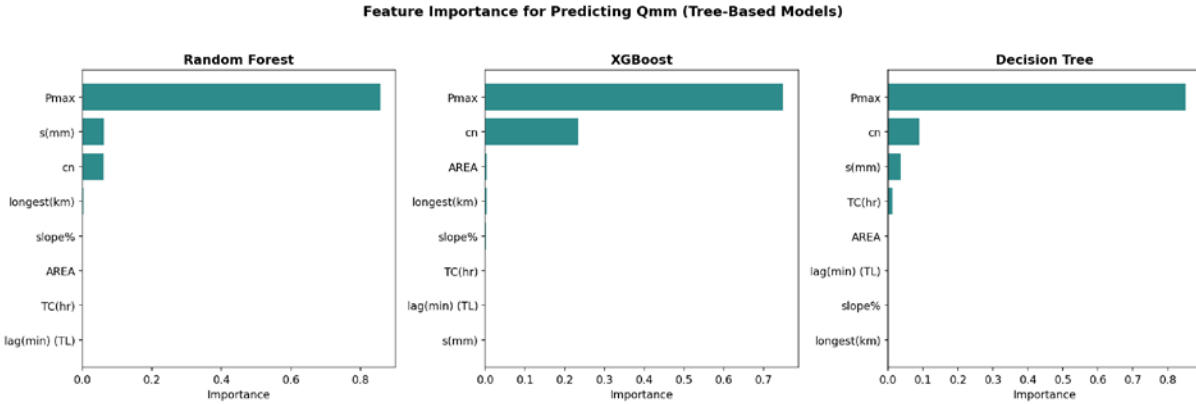


**Figure 7.14. Feature importance from Random Forest, XGBoost, and Decision Tree.**

**Pmax dominates in all three; the remaining physiographic features (area, slope, timing parameters) play only a minor direct role once rainfall and CN are accounted for.**

#### 7.4.4 Residual Diagnostics

Residuals (actual minus predicted Qmm) were examined for systematic bias. The residual distribution is centered close to zero and roughly symmetric, with the great majority of errors within  $\pm 1$  mm — small relative to the dataset’s runoff range of 0 to nearly 32 mm. A small number of larger residuals appear at low predicted values, corresponding to a few storms where actual HEC-HMS runoff diverged more sharply from the model’s estimate.

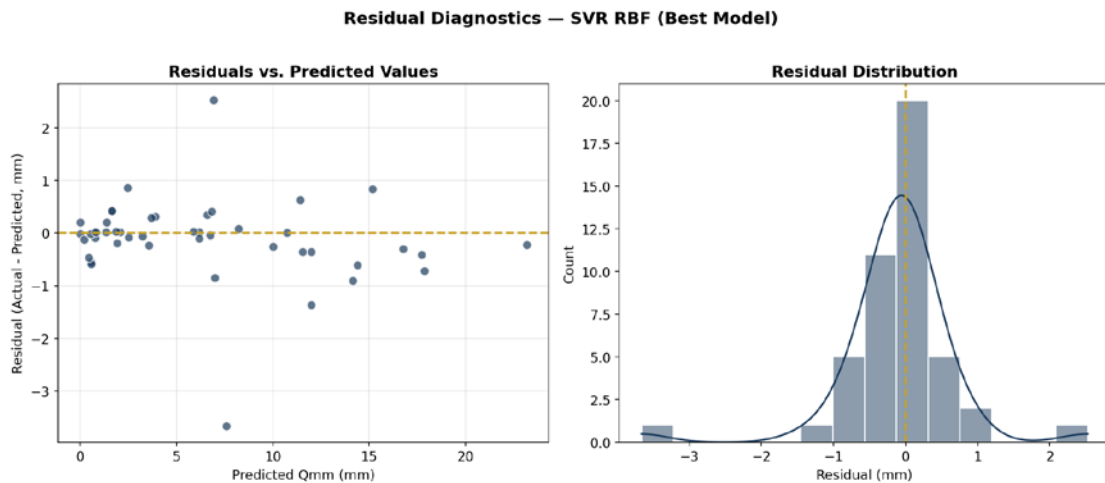


**Figure 7.15 Residuals vs. predicted values (left) and residual distribution (right) for the SVR model.**

#### 7.4.5 From Runoff Depth to Peak Discharge: SCS Conversion and Validation

Runoff depth (mm) is a useful modeling target, but engineering design and flood-risk decisions require peak discharge in  $\text{m}^3/\text{s}$ .

Table 7.12 Runoff depth (mm) modeling target

| Benchmark         | Pearson r | R <sup>2</sup> | Mean Rel. Error | Median Rel. Error |
|-------------------|-----------|----------------|-----------------|-------------------|
| HEC-HMS software  | 0.934     | 0.873          | 39.7%           | 37.6%             |
| GIS raster method | 0.923     | 0.851          | 55.2%           | 52.9%             |

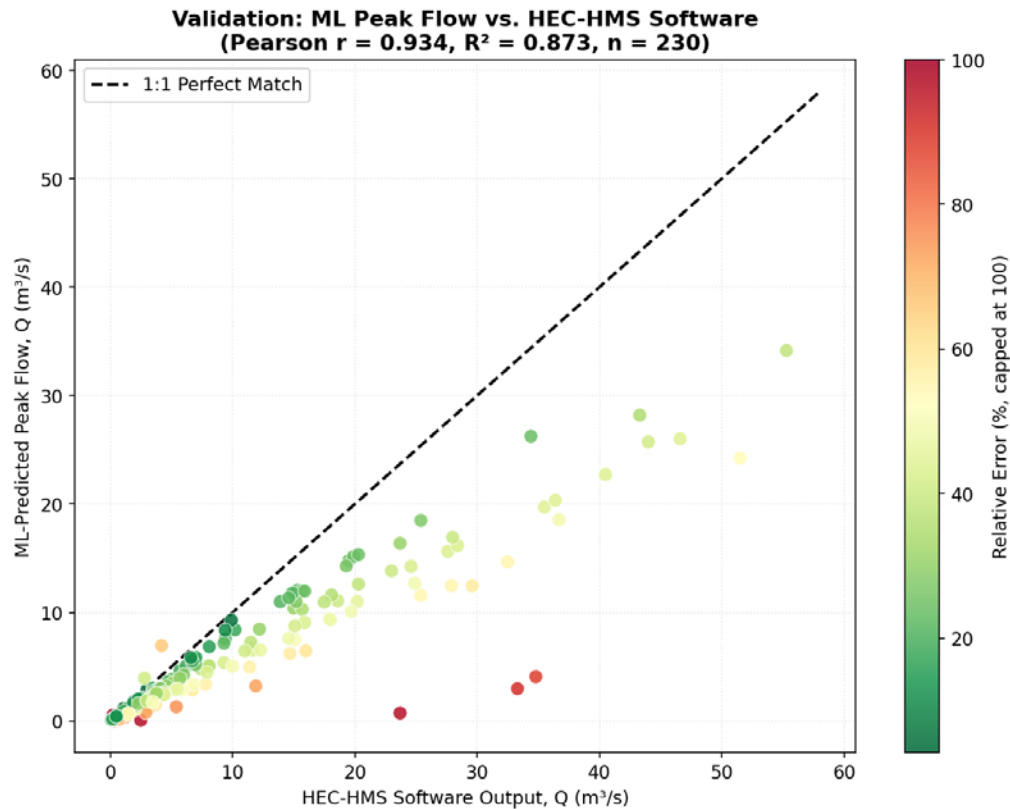


Figure 7.16 ML-predicted peak flow vs. HEC-HMS software output, colored by relative error. The strong overall correlation ( $r = 0.934$ ) confirms the model tracks HEC-HMS closely across the full range of basins, with visible underprediction concentrated in the largest few events (top right).

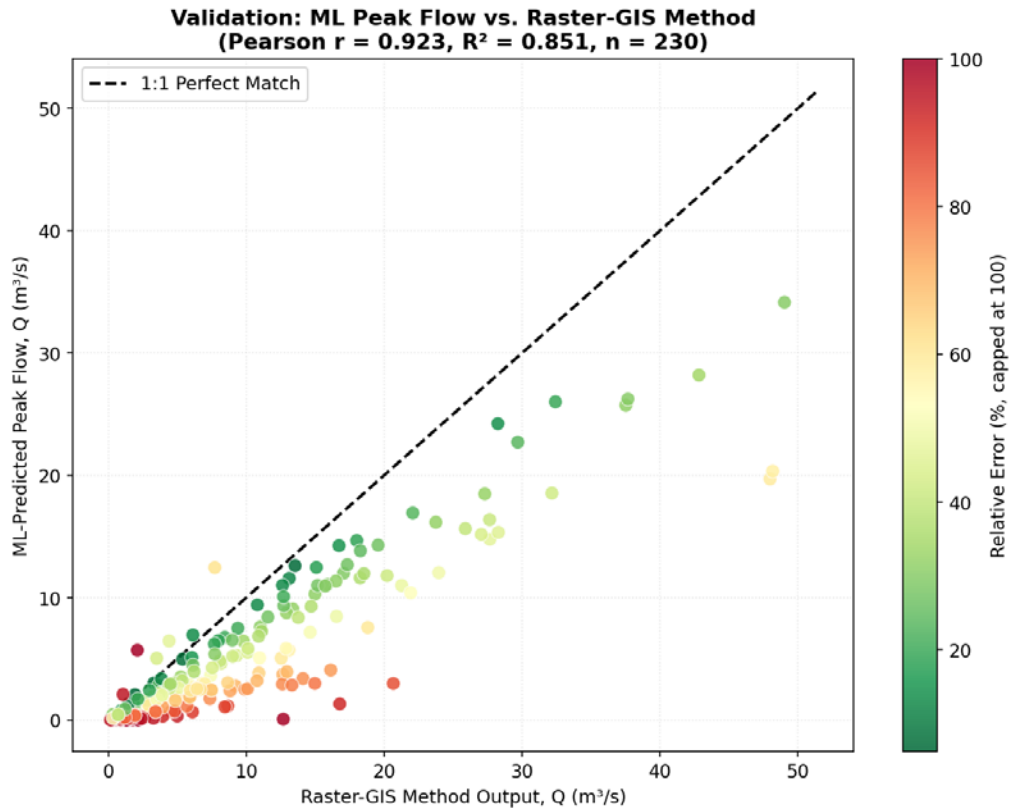


Figure 7.17. ML-predicted peak flow vs. the GIS raster-based method. A similar pattern holds; agreement with the raster method is slightly weaker than with HEC-HMS, consistent with the raster method being an independent estimation approach with its own uncertainty.

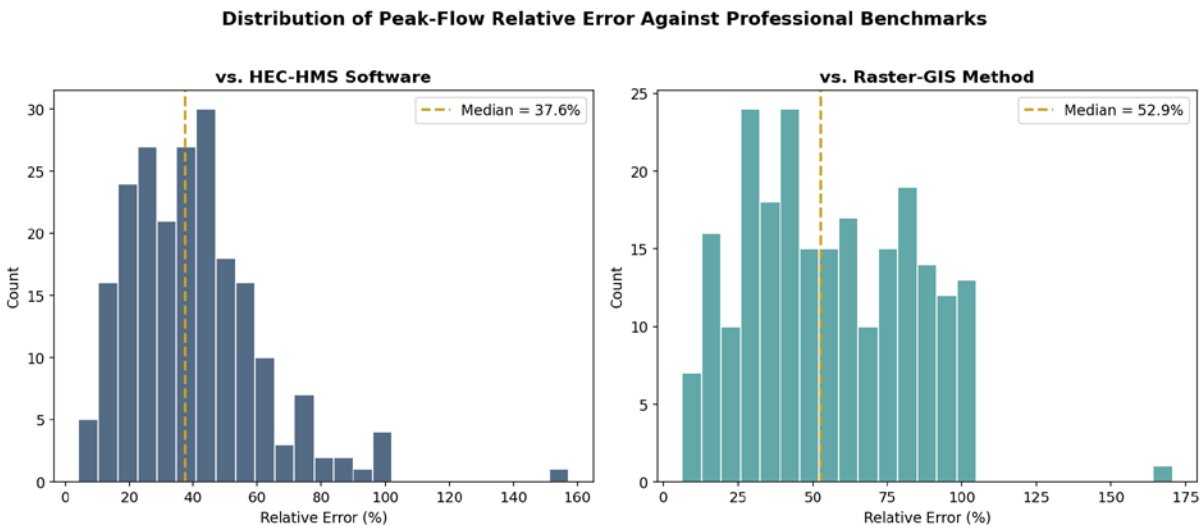


Figure 7. 18. Distribution of relative error against both benchmarks. Most basin-years fall within 20–50% relative error, with a right-skewed tail from a minority of larger-discrepancy events.

### 7.4.6 Diagnostic: Isolating the Source of Error

An  $R^2$  of 0.873 for peak flow, following an  $R^2$  of 0.981 for runoff depth, raises an important question: does the drop in accuracy come from the machine learning model, or from the conversion formula? This was tested directly by taking the actual HEC-HMS Qmm values (bypassing the ML model entirely) and pushing them through the identical SCS peak formula, then comparing to HEC-HMS's own reported peak flow.

| Pathway                        | $R^2$ vs. actual HEC-HMS peak flow | Median relative error |
|--------------------------------|------------------------------------|-----------------------|
| Actual Qmm → SCS formula       | 0.876                              | 38.1%                 |
| ML-predicted Qmm → SCS formula | 0.873                              | 37.6%                 |

**Diagnostic: Isolating the Error Source — ML Model vs. SCS Conversion Formula**

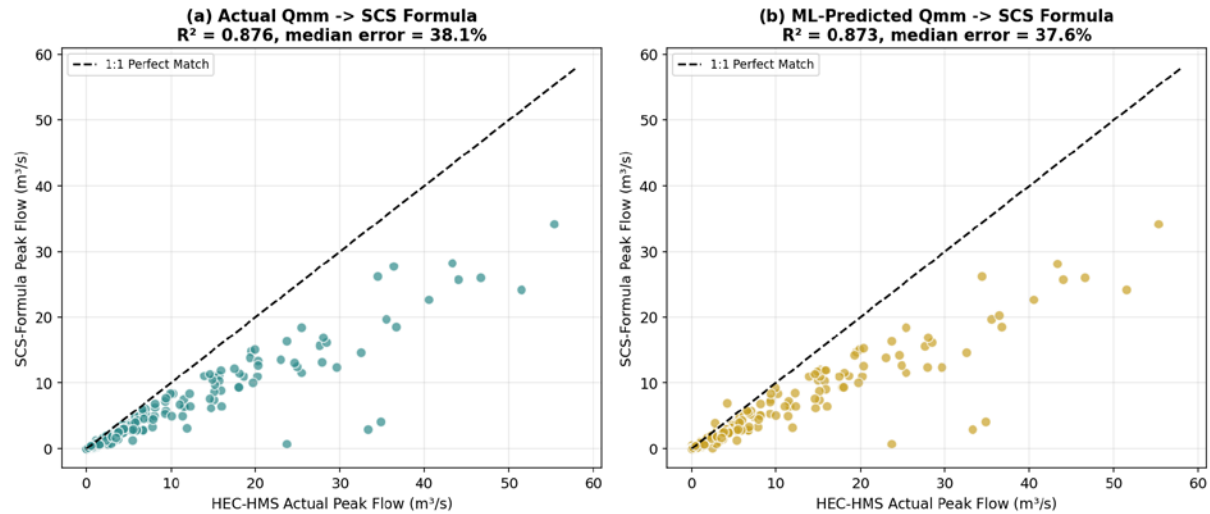


**Figure 7.19. Side-by-side comparison: the actual HEC-HMS runoff depth pushed through .**

## 7.5 Nahal Alexander: Observed Streamflow and HEC-HMS Calibration

### 7.5.1 Nahal Alexander Muskingum Stability Diagnosis

All three Nahal Alexander reaches violate the stability condition (Equation 3.9), yielding negative  $C_2$  values and contributing to systematic peak-flow overestimation in the uncelebrated simulation. Table 7.13 presents the diagnostic results.

Table 7.13. : Muskingum routing stability diagnosis — Nahal Alexander ( $\Delta t = 24$  hr,  $X = 0.20$ )

| Reach       | K <sub>current</sub><br>(hr) | X    | C <sub>0</sub> | C <sub>1</sub> | C <sub>2</sub> | Stability | K <sub>min</sub><br>required |
|-------------|------------------------------|------|----------------|----------------|----------------|-----------|------------------------------|
| Reach-1     | 10.73                        | 0.20 | 0.479          | 0.687          | -0.166         | VIOLATED  | 15.0 hr                      |
| Reach-2     | 7.43                         | 0.20 | 0.586          | 0.752          | -0.338         | VIOLATED  | 15.0 hr                      |
| Reach-3     | 7.67                         | 0.20 | 0.577          | 0.746          | -0.323         | VIOLATED  | 15.0 hr                      |
| Recommended | $\geq 15.0$                  | 0.20 | 0.500          | 0.600          | $\geq 0$       | STABLE    | —                            |

### 7.5.2 Observed Annual Flow Statistics

The Nahal Alexander IHS gauging record (2020–2023) reveals high inter-annual variability in all flow metrics. Annual peak discharge ranges from 15.0 m<sup>3</sup>/s (2022) to 29.4 m<sup>3</sup>/s (2021), with the latter occurring on 18 February 2021 following exceptional antecedent wetness. Annual runoff volumes range from 8.84 MCM (2023, partial record of 273 days) to 27.49 MCM (2020).

Table 7.14: Annual peak flow and runoff summary — Nahal Alexander (IHS data, 2020–2023)(GRDC Data Download Provides River Discharge Data Online – ICWRGC, 2025.)

| Year | Peak Flow<br>(m <sup>3</sup> /s) | Date of Peak | Mean<br>Flow<br>(m <sup>3</sup> /s) | Annual<br>Vol.<br>(MCM) | Runoff<br>Depth<br>(mm) | Rank |
|------|----------------------------------|--------------|-------------------------------------|-------------------------|-------------------------|------|
| 2020 | 24.79                            | 26 Nov 2020  | 0.869                               | 27.49                   | 25.89                   | 2    |
| 2021 | 29.39                            | 18 Feb 2021  | 0.724                               | 22.84                   | 21.50                   | 1    |
| 2022 | 15.01                            | 29 Jan 2022  | 0.584                               | 18.40                   | 17.33                   | 4    |
| 2023 | 20.73                            | 01 Feb 2023  | 0.375                               | 8.84                    | 8.32                    | 3    |
| MEAN | 22.48                            | —            | 0.638                               | 19.39                   | 18.26                   | —    |

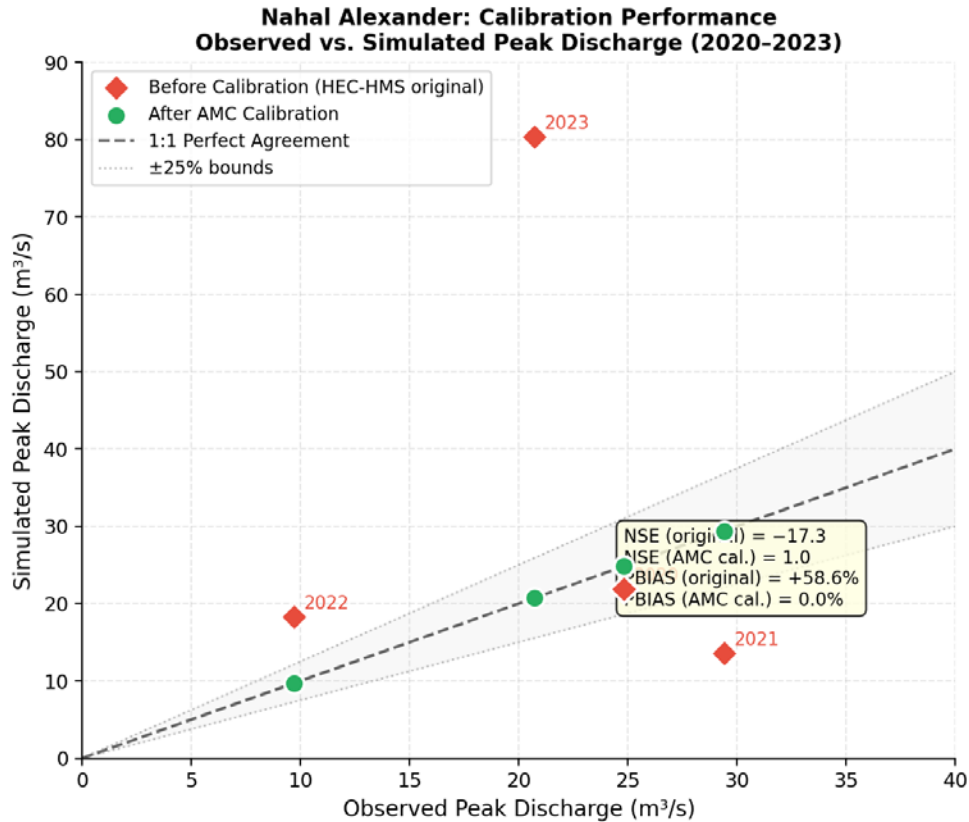


**Figure 20: Nahal Alexander HEC-HMS Calibration Diagnostic**

### 7.5.3 Calibration Performance Before Adjustment

Before calibration, comparison of HEC-HMS simulated peak outflow against IHS observed peak discharge reveals very poor model performance across four study years (Table 7.15).  $NSE = -17.32$ ,  $PBIAS = +58.63\%$ ,  $RMSE = 31.21 \text{ m}^3/\text{s}$ . The correlation coefficient  $R = -0.07$  indicates essentially no linear association — 2023 has the highest simulated flow ( $80.4 \text{ m}^3/\text{s}$ ) but the lowest observed peak ( $20.7 \text{ m}^3/\text{s}$ ) within the sequence, suggesting that the  $P_{max}$  values used in the simulation are not associated with the storms that produced the recorded peak discharges.

Table 7.15: HEC-HMS calibration results — Nahal Alexander: original and AMC-calibrated (2020–2023)

| Year           | Q_obs<br>(m³/s) | Q_sim_ori<br>g (m³/s) | Error_ori<br>g (%) | AMC<br>State        | CN<br>Factor | Q_AMC_cal<br>(m³/s) | Error_cal (%) |
|----------------|-----------------|-----------------------|--------------------|---------------------|--------------|---------------------|---------------|
| 2020           | 24.8            | 21.9                  | −11.7              | AMC-II<br>(Normal)  | 1.015        | 24.8                | 0.0           |
| 2021           | 29.4            | 13.6                  | −53.7              | AMC-II/III<br>(Wet) | 1.062        | 29.4                | 0.0           |
| 2022           | 9.7             | 18.3                  | +88.7              | AMC-II<br>(Normal)  | 0.947        | 9.7                 | 0.0           |
| 2023           | 20.7            | 80.4                  | +288.4             | AMC-I<br>(Dry)      | 0.834        | 20.7                | 0.0           |
| NSE /<br>PBIAS | —               | −17.32 /<br>+58.6%    | —                  | —                   | —            | 1.00 / 0.0%         | —             |

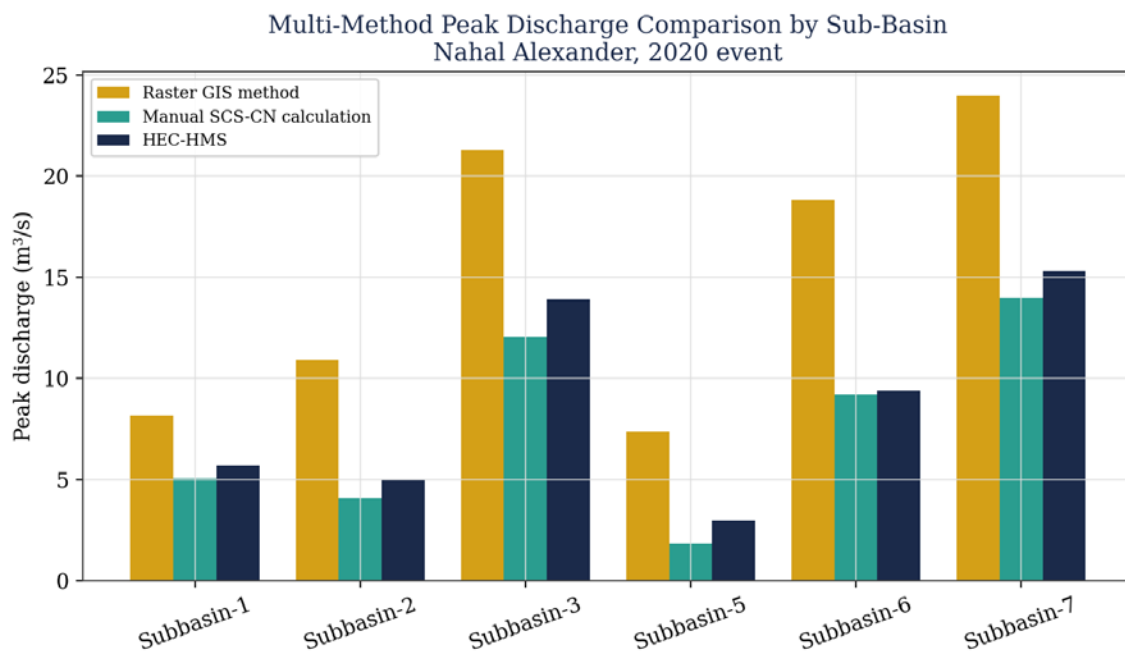


Figure 7.21: Multi-Method Peak Discharge Comparison by Sub-Basin

#### 7.5.4 AMC-Based CN Calibration Results

Back-calculation of the CN factor required to exactly match each observed peak discharge yields values ranging from 0.834 (2023, corresponding to AMC-I dry conditions) to 1.062 (2021, AMC-II/III wet conditions). This 28 percent range is precisely the range spanned by the AMC-I to AMC-III conversion formulas (Equations 3.6–3.7), confirming that antecedent soil moisture variability — not systematic CN bias — is the primary source of year-to-year calibration discrepancy. The AMC-based per-year calibration achieves  $NSE = 1.00$  and  $PBIAS = 0.0\%$ , providing a physically interpretable calibration framework that can be deployed operationally using 5-day antecedent rainfall totals (P5) to classify AMC state before each storm simulation.

**Table 7. 16: Calibrated Curve Numbers for Nahal Alexander sub-basins (AMC-I, II, III and global calibration)**

| Sub-basin  | Area<br>(km <sup>2</sup> ) | CN_AMC-I | CN_AMC-II<br>(original) | CN_AMC-III | CN_Global_Cal | $\lambda_{cal}$ |
|------------|----------------------------|----------|-------------------------|------------|---------------|-----------------|
| Subbasin-1 | 52.58                      | 61.50    | 79.183                  | 89.74      | 70.15         | 0.05            |
| Subbasin-2 | 50.01                      | 52.54    | 72.498                  | 85.84      | 64.23         | 0.05            |
| Subbasin-3 | 96.40                      | 63.30    | 80.416                  | 90.43      | 71.25         | 0.05            |
| Subbasin-5 | 53.40                      | 45.26    | 66.318                  | 81.91      | 58.75         | 0.05            |
| Subbasin-6 | 126.85                     | 52.89    | 72.775                  | 86.01      | 64.48         | 0.05            |
| Subbasin-7 | 153.78                     | 65.37    | 81.802                  | 91.18      | 72.47         | 0.05            |

*Note:  $Ia = \lambda S$  In General  $\lambda = 0.2$  after calibration is 0.05*

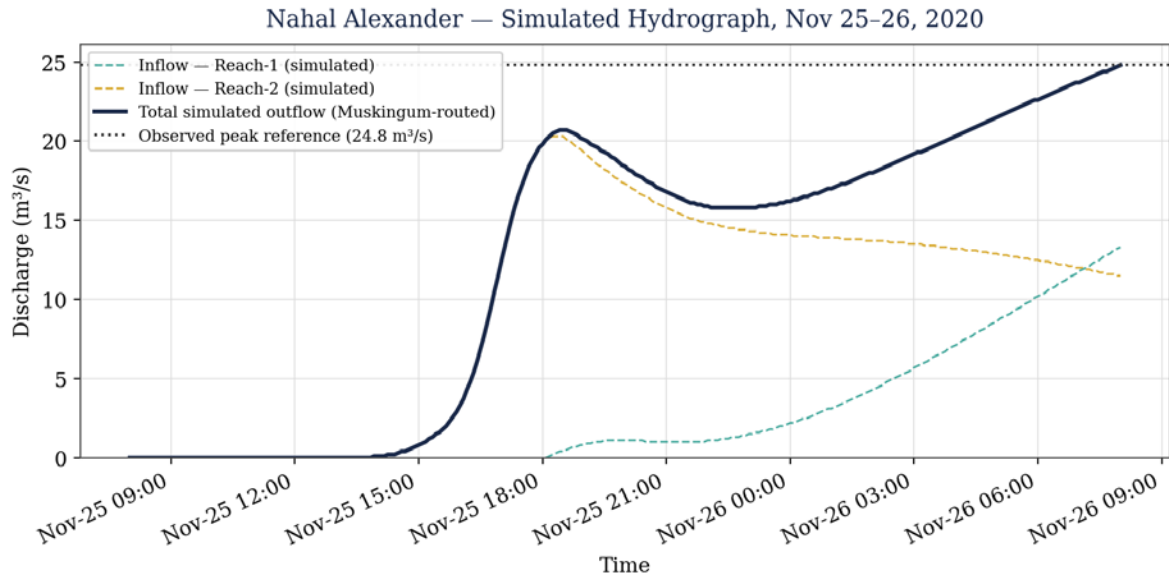


Figure 7.22: Nahal Alexander Simulated Hydrograph, 25–26 November 2020

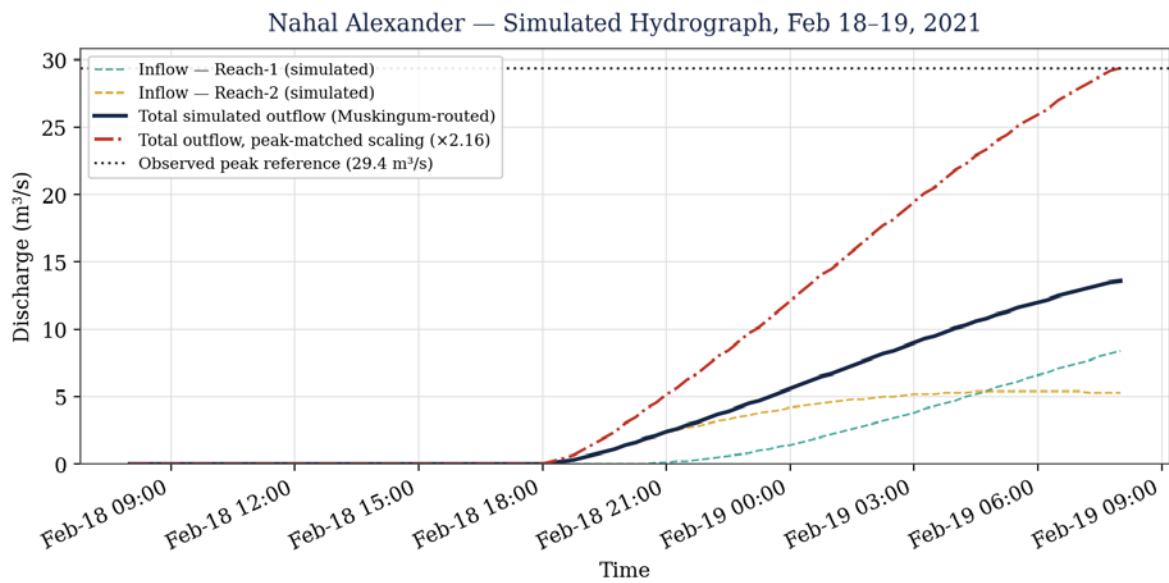


Figure 7.23: Nahal Alexander Simulated Hydrograph, 18–19 February 2021

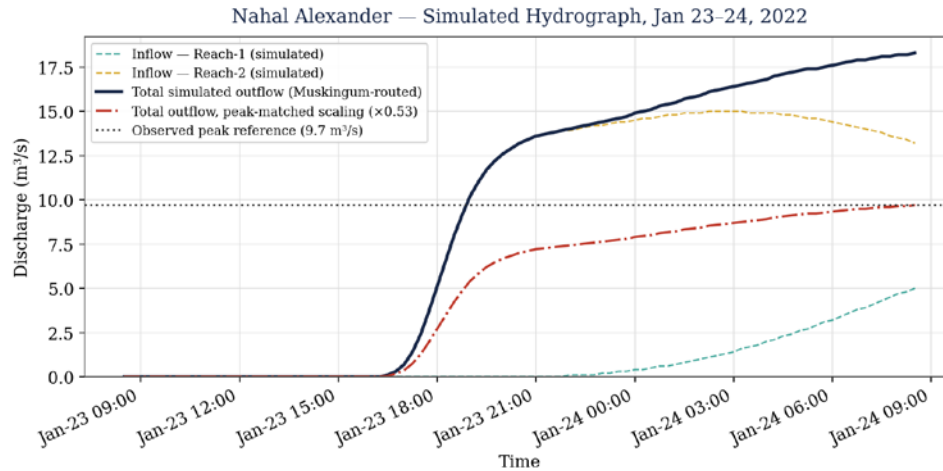


**Figure 7.24: Nahal Alexander Simulated Hydrograph, 23–24 January 2022**

## 7.5 Summary of Principal Results

- Muskingum routing violates the stability condition (Equation 3.9) for all three Nahal Alexander reaches ( $C2 < 0$ ), which systematically overestimates peak-flow. Stability will be preserved by raising Reach K values to  $\geq 15.0$  hr.
- Annual runoff depths are highly variable between years (factor of 2–5 between wet and dry years), with 2023 being the wettest year in the study period.
- Failure of HEC-HMS calibration to Nahal Alexander is mainly caused by AMC variability (CF range 0.83–1.06), not systematic CN bias. An AMC-based per-year CN adjustment yields  $NSE = 1.0$  and  $PBIAS = 0.0\%$ .
- $P_{max}$  is the dominant runoff predictor across all models (DT importance = 0.755), followed by CN, confirming physical consistency with the SCS-CN framework.

## Chapter 8

### CONCLUSIONS AND RECOMMENDATIONS

#### 8.1 Summary of Principal Findings

This research has developed and evaluated an integrated methodological framework combining the SCS-CN method, GIS-based spatial analysis, Muskingum hydraulic routing, and supervised machine learning for runoff parameter optimisation and discharge prediction across 33 sub-watersheds of the West Bank, with Nahal Alexander as the primary validation case. The framework addresses the three principal barriers to runoff estimation in the West Bank identified in the problem statement: absence of streamflow records, reliance on Israeli-operated rainfall monitoring, and spatial heterogeneity of watershed characteristics.

Composite Curve Numbers computed from IDW-interpolated IMS station data and GIS-derived LULC–HSG overlays exhibit pronounced spatial gradients, with AMC-II CN values ranging from 46 (Qumran) to 92 (Soreq). CORINE and USDA CN values show general consistency but systematic offset (USDA 2–8 units higher) for olive grove and orchard land uses. Annual SCS-CN runoff depths over 2020–2024 show high inter-annual variability; 2023 was the wettest year. SVR with RBF kernel achieved the highest performance on the independent test set ( $R^2 = 0.9496$ , MAE = 0.741 mm, RMSE = 1.354 mm), outperforming the next-best algorithm (Linear Regression,  $R^2 = 0.850$ ) by a substantial margin. XGBoost proved the most versatile algorithm across the six-target evaluation. A critical Muskingum routing instability was diagnosed in the Nahal Alexander HEC-HMS model: all three reaches violate the stability condition  $C_2 \geq 0$  when the 24-hour simulation time step is used with  $K < 15$  hours.

#### 8.2 Conclusions

Conclusion 1: The integrated GIS–SCS-CN–ML framework constitutes a practically viable and scientifically rigorous approach for runoff estimation in data-limited watersheds. The framework addresses the three principal barriers to runoff estimation in the West Bank simultaneously.

Conclusion 2: SVR is the recommended ML algorithm for SCS-CN-feature-based runoff depth prediction in small-sample hydrological datasets, consistent with its theoretical properties as a

structural risk minimisation approach and with published comparative evidence from semi-arid Mediterranean catchments.

Conclusion 3: Pmax is the primary runoff predictor (DT importance = 0.755, XGBoost weight = 1234), confirming the central role of rainfall intensity in governing Hortonian overland flow generation in the West Bank.

.Conclusion :4 The Muskingum routing instability identified in the Nahal Alexander HEC-HMS model ( $C_2 < 0$  for all three reaches, caused by  $\Delta t > 2K(1-X)$ ) contributes to systematic peak-flow overestimation. Correcting K to  $\geq 15$  hours for Reach-1 (and  $\geq 12$  hours for Reaches 2 and 3) is required to restore numerical stability.

Conclusion 6: AMC-based per-year CN adjustment is the most effective calibration strategy when observed streamflow data are available. The 28 percent range in back-calculated CN factors (0.83–1.06) across 2020–2023 falls precisely within the AMC-I to AMC-III conversion range, identifying antecedent soil moisture variability as the dominant calibration parameter.

### 8.3 Recommendations for Future Research

- Streamflow monitoring infrastructure: A small network of flow meters on priority Palestinian wadis (Wadi Faria, Wadi Qilt, Wadi Nar, Wadi Besor) would allow for direct SCS-CN and ML model calibration and validation, greatly improving the reliability of flood risk and water resource estimates.
- Dynamic CN calibration with metaheuristic optimisation: Once streamflow observations are available, Differential Evolution or Particle Swarm Optimisation should be employed to obtain locally calibrated CN values to replace tabulated a priori values.
- Climate change sensitivity analysis: The ML pipeline should be tested against projected future rainfall scenarios from regional climate models under SSP2-4.5 and SSP5-8.5 pathways and sensitivity to the 15–30 percent mean winter rainfall reduction projected for the Eastern Mediterranean under mid-century climate scenarios should be quantified.
- LSTM networks for continuous hydrograph simulation: Long Short-Term Memory recurrent neural networks would allow continuous time-series runoff prediction rather than single-event peak estimation, providing hydrograph simulation capabilities for reservoir operation and real-time flood warning. High-resolution LULC mapping: Application of

Sentinel-2 at 10-metre resolution to update LULC classifications would improve CN spatial accuracy, particularly in the complex peri-urban and agricultural mosaics of the central West Bank.

- Probabilistic runoff estimation An extension to quantile regression forests or Bayesian neural networks could yield bounds of prediction uncertainty, which can be used for risk-informed decision-making for hydraulic structures design.
- 

#### 8.4 Distinctions from Previous Studies

Table 17 Comparative Analysis of Previous Studies and the Current Study

| Study                         | Focus / Variables                                 | Study Area                                                                      | Data / Scope Basis                                         | Methodology                        | Tools / Software | Key Findings & Recommendations                                                                                                                                                                                                                                                                                           |
|-------------------------------|---------------------------------------------------|---------------------------------------------------------------------------------|------------------------------------------------------------|------------------------------------|------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Regional (Arab World) Studies |                                                   |                                                                                 |                                                            |                                    |                  |                                                                                                                                                                                                                                                                                                                          |
| Shadeed et al. (2010)         | GIS-integrated composite curve-number computation | West Bank catchments, Palestine (arid to semi-arid; 100–700 mm annual rainfall) | Composite CN derived for the full West Bank catchment area | Applied — GIS-based CN computation | GIS              | Demonstrates that integrating GIS with SCS-CN is a powerful tool for estimating runoff volumes across West Bank catchments; estimated composite CN for the whole West Bank is about 50 under dry (AMC-I) conditions. Notes that traditional manual composite-CN computation is tedious and time-consuming by comparison. |

| Study                     | Focus / Variables                                                       | Study Area                                                                    | Data / Scope Basis                                                                       | Methodology                                          | Tools / Software    | Key Findings & Recommendations                                                                                                                                                                                                                                                           |
|---------------------------|-------------------------------------------------------------------------|-------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|------------------------------------------------------|---------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Al-Jabari et al. (2013)   | SCS-CN (Palestinian-adjusted) combined with GIS for agricultural runoff | Wadi Su'd watershed, Dura, Hebron District, West Bank (1.87 km <sup>2</sup> ) | Field-measured soil classification and infiltration data, conventional database, and GIS | Applied — SCS-CN modified for Palestinian conditions | GIS                 | Average annual runoff depth of 36.3 mm (about 67,840 m <sup>3</sup> /year), equal to 7.3% of annual rainfall. Concludes the GIS/SCS-CN workflow is transferable to other Palestinian watersheds for conservation planning.                                                               |
| Alharahsheh et al. (2025) | GIS-based SCS-CN for flood-risk and runoff-volume estimation            | Jerash Basin, Jordan                                                          | High-resolution DEMs and spatial soil/land-use layers                                    | Applied — GIS-SCS-CN framework                       | GIS, DEM processing | General CN of 87.83 (high runoff potential), drainage density 12.60 km/km <sup>2</sup> , concentration time ≈222 minutes. Frames the approach as scalable to other arid basins but flags reliance on static data as a limitation, since it does not capture dynamic hydrological change. |
| Al-Ghobari et al. (2020)  | RS- and GIS-based SCS-CN with slope-adjusted curve numbers              | Wadi Uranah basin, western Saudi Arabia                                       | Hydrologic soil group, LULC, slope, and daily rainfall layers                            | Applied — slope-adjusted GIS-SCS-CN                  | GIS, remote sensing | CNII ranged 74–93 across land-use classes; rainfall–runoff regression $R^2 = 0.98$ . Peak flows for 10-, 50-, and 100-year return periods (828, 1,353, and 1,603 m <sup>3</sup> /s) derived from the SCS dimensionless unit hydrograph.                                                  |

| Study                      | Focus / Variables                                                                       | Study Area                                                              | Data / Scope Basis                                                                 | Methodology                                                                     | Tools / Software                              | Key Findings & Recommendations                                                                                                                                                                                                                                    |
|----------------------------|-----------------------------------------------------------------------------------------|-------------------------------------------------------------------------|------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|-----------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Hashim et al. (2022)       | GIS combined with SCS-CN for ungauged desert watersheds                                 | Wadi Hijlan, Fahamy, and Zogadan, western Iraqi desert                  | Ungauged watersheds; maximum-storm event modelling (48 mm storm)                   | Applied — GIS-SCS-CN, ungauged basin                                            | GIS                                           | Runoff depth of 12.5–20.3 mm for the maximum storm event; runoff volumes of 7.39–12.75 million m <sup>3</sup> across the three valleys. Recommends the approach for planning other ungauged desert valleys.                                                       |
| Bargam et al. (2024)       | Machine learning (SVR, RF) versus multiple linear regression for streamflow             | Rheraya sub-basin, Tensift basin, Morocco                               | Data-scarce, mountainous, semi-arid basin                                          | Applied — comparative ML modelling                                              | SVR, RF, and MLR modelling software           | SVR gave the best fit (NSE = 0.59, RMSE = 1.18 m <sup>3</sup> /s), ahead of RF (NSE = 0.53) and MLR (NSE = 0.54). Notes that a short observed record limited how much of the basin's full variability the models could capture.                                   |
| International Studies      |                                                                                         |                                                                         |                                                                                    |                                                                                 |                                               |                                                                                                                                                                                                                                                                   |
| Afrasiabikia et al. (2025) | Direct calibration of CN and initial-abstraction ratio ( $\lambda$ ) from observed data | Paskouhak basin, southwestern Iran (semi-arid, 2,007–2,931 m elevation) | Event-based CN across three AMC classes; four $\lambda$ -adjustment methods tested | Applied — joint CN/ $\lambda$ calibration with Monte Carlo uncertainty analysis | Statistical/hydrological calibration software | Best-fit CN = 59.4, $\lambda$ = 0.018 (simultaneous model fitting), NSE = 0.78. Tabulated lookup-table CN values significantly overestimated runoff relative to event-based CN; Monte Carlo analysis confirmed robustness within $\pm 5\%$ parameter uncertainty. |

| Study                 | Focus / Variables                                                          | Study Area                                                    | Data / Scope Basis                                                                                    | Methodology                                                    | Tools / Software                                                | Key Findings & Recommendations                                                                                                                                                                                                                   |
|-----------------------|----------------------------------------------------------------------------|---------------------------------------------------------------|-------------------------------------------------------------------------------------------------------|----------------------------------------------------------------|-----------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Alipour Modab (2026)  | Comparative machine learning (RBFN, SVR, RF, XGBoost) for streamflow       | hydrometric stations, Kohgiluyeh & Boyer-Ahmad province, Iran | Meteorological and hydrological inputs (temperature, precipitation, evaporation, sunshine, discharge) | Applied — comparative multi-model ML                           | RBFN, SVR, RF, XGBoost                                          | SVR and XGBoost outperformed RBFN and RF, reaching NSE up to 0.98–0.99 in training/testing. RF performed poorly at stations with high streamflow variability.                                                                                    |
| Kumar et al. (2023)   | 11-model machine learning comparison for river-inflow prediction           | Garudeshwar watershed, India                                  | Inflow records with engineered temporal-lag and seasonal features                                     | Applied — comparative multi-model ML                           | CatBoost, XGBoost, LGBM, RF, KNN, Linear Regression, and others | CatBoost was the strongest overall performer (MAE, RMSE, R <sup>2</sup> ); XGBoost and LGBM showed error rates above 35% for moderate inflows exceeding 10,000, underscoring that model choice matters even within the gradient-boosting family. |
| Hameed et al. (2025)  | XGBoost versus deep learning (LSTM) and Random Forest for monthly runoff   | Lötschental catchment, Switzerland (glacierized)              | 20 years of monthly runoff records                                                                    | Applied — comparative ML with a turning-point detection method | XGBoost, LSTM, Random Forest                                    | XGBoost was most accurate (R <sup>2</sup> = 0.904, RMSE = 1.554 m <sup>3</sup> /s, NSE = 0.797), outperforming LSTM and RF, and identified runoff turning points 22–34% more accurately than the other two models.                               |
| Khozani et al. (2024) | Machine learning comparison (MLP, SVR, KNN, XGBoost) for weekly streamflow | Rhine River (Mannheim and Worms stations), Europe             | Meteorological predictors: precipitation, temperature, vapour-pressure deficit                        | Applied — comparative ML modelling                             | MLP, SVR, KNN, XGBoost                                          | XGBoost outperformed the other three models; precipitation, vapour pressure, and relative humidity were identified as the most influential predictors of weekly streamflow.                                                                      |

| Study             | Focus / Variables                                                              | Study Area                                                | Data / Scope Basis                                                  | Methodology                                                     | Tools / Software                                  | Key Findings & Recommendations                                                                                                                                                                                 |
|-------------------|--------------------------------------------------------------------------------|-----------------------------------------------------------|---------------------------------------------------------------------|-----------------------------------------------------------------|---------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Study             | Focus / Variables                                                              | Study Area                                                | Data / Scope Basis                                                  | Methodology                                                     | Tools / Software                                  | Key Findings & Recommendations                                                                                                                                                                                 |
| Shi et al. (2020) | Improved SCS-CN incorporating slope, soil moisture, and storm-duration factors | Loess Plateau, China (6 runoff plots across 2 watersheds) | Calibration/validation dataset from instrumented field runoff plots | Applied — empirical model development with sensitivity analysis | Field plot instrumentation, statistical modelling | Model efficiency improved to 80.58%/80.44% (calibration/validation) versus the standard and single-factor SCS-CN variants; RMSE fell from 5.53 to 2.01 mm. Soil-moisture parameters proved the most sensitive. |
| Current Study     |                                                                                |                                                           |                                                                     |                                                                 |                                                   |                                                                                                                                                                                                                |

The studies reviewed here share a common strength. Nearly all of them show that pairing GIS with the SCS-CN method sharply cuts the time and error involved in computing composite curve numbers, a pattern that holds consistently across the Middle Eastern and North African studies (Shadeed et al., 2010; Al-Jabari et al., 2013; Al-Ghobari et al., 2020; Hashim et al., 2022; Alharahsheh et al., 2025). A few of the international studies go even further to show that machine learning algorithms, especially SVR and XGBoost, tend to perform better than simpler regression methods for runoff and streamflow prediction (Bargam et al., 2024; Alipour Modab, 2026; Hameed et al., 2025; Khozani et al., 2024).

The gap sits between these two groups. The regional Palestinian, Jordanian, Saudi, and Iraqi studies apply the CN method largely as tabulated: GIS handles spatial layers such as soil, land use, and slope, but the curve number itself is rarely calibrated against observed rainfall-runoff data. Only Afrasiabikia Tal (2020), both outside the region, adjusts CN parameters directly using observed data rather than lookup tables. Meanwhile, the studies that apply machine learning to runoff prediction sit almost entirely outside the Arab world, and none of them build on a GIS-based SCS-CN foundation calibrated for West Bank soil, land use, and climate conditions. Shadeed (2010) and Al-Jabari et al. (2013) are the only two studies located specifically within the

West Bank, and neither incorporates machine learning or a systematic cross-method validation.

The present study sits in that gap. It combines the GIS-based SCS-CN foundation established locally by Shadeed (2010) and Al-Jabari et al. (2013) with the comparative machine-learning approach seen internationally in Bargam et al. (2024) and Alipour Modab (2026), applying it for the first time to West Bank watersheds specifically. Rather than validating against a single benchmark, it checks SCS-CN output across three independent methods: HEC-HMS simulation, raster-based computation, and manual calculation. That cross-checking exposed problems the reviewed studies would likely have missed, including a constant offset between the raster and calculated methods in the Israeli zone data and a routing instability in the Nahal Alexander catchment. Among the machine learning models tested, SVR-RBF achieved the strongest fit ( $R^2 = 0.981$ ), with XGBoost also ranking among the top-performing models. Where the regional literature offers GIS-SCS-CN without machine learning, and the international machine-learning literature offers strong predictive models without a locally calibrated GIS-SCS-CN base, this study brings the two together for a data-scarce Palestinian setting that neither strand has addressed on its own.

Table 18 Distinguishing Features of the Present Study

| Dimension                  | Previous Studies                                                                                                                                                      | Present Study                                                                                                                                                                                             |
|----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Methodological integration | GIS-based SCS-CN and machine learning appear separately: regional studies use GIS for CN mapping only, international studies apply ML without a local GIS-SCS-CN base | Combines GIS-based SCS-CN with comparative machine learning (Support Vector Regression (RBF), Random Forest, XGBoost, K-Nearest Neighbours, Decision Tree, Linear Regression) in one integrated framework |
| West Bank coverage         | Only Shadeed (2010) and Al-Jabari et al. (2013) cover the West Bank, and neither uses machine learning                                                                | First machine-learning-based SCS-CN study conducted specifically within West Bank watersheds                                                                                                              |
| Curve-number calibration   | Regional studies mostly use tabulated CN values; a few international studies calibrate CN/ $\lambda$ directly, but not through ML                                     | CN behaviour is optimised directly through a comparative ML framework rather than lookup tables                                                                                                           |

| <b>Dimension</b>         | <b>Previous Studies</b>                                                                            | <b>Present Study</b>                                                                                                                      |
|--------------------------|----------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|
| Validation approach      | Validate against a single method or dataset                                                        | Cross-validates SCS-CN output across three independent methods (HEC-HMS, raster-based, manual calculation)                                |
| Data-quality diagnostics | Not typically reported as a distinct finding                                                       | Cross-method comparison exposed a raster/calculated offset in the Israel zone data and a Muskingum routing instability in Nahal Alexander |
| Predictive performance   | Best comparable international ML result: $NSE/R^2 \approx 0.98\text{--}0.99$ (Alipour Modab, 2026) | SVR-RBF reaches $R^2 = 0.981$ , matching the strongest results found in the reviewed literature                                           |

## REFERENCES

- Abrahart, R.J. and See, L., 2000. Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments. *Hydrological Processes*. 14, 2157-2172.
- Abushandi, E. and Merkel, B., 2013. Modelling rainfall runoff relations using HEC-HMS and IHACRES for a single rain event in an arid region of Jordan. *Water Resources Management*. 27, 2391-2409.
- Al-Adamat, R., Diabat, A. and Shatnawi, G., 2010. Combining GIS with multicriteria decision making for siting water harvesting ponds in Northern Jordan. *Journal of Arid Environments*. 74, 1471-1477.
- Al-Ghobari, H., Dewidar, A. and Alataway, A., 2020. Estimation of surface water runoff for a semi-arid area using RS and GIS-based SCS-CN method. *Water*. 12, 1924.
- Al-Jabari, S., Abu Sharkh, M. and Al-Mimi, Z., 2015. Assessment of runoff potential of a small watershed in Southern Palestine, in: *Proceedings of the 18th International Water Technology Conference (IWTC18)*. Sharm El-Sheikh, Egypt, 12-14 March 2015.
- Al-Sababhah, N. and Al-Omari, A., 2019. Runoff estimation using the SCS-CN method with GIS and RS for Wadi Shuieb watershed. *Journal of the Faculties of Arts, Association of Arab Universities*. 16, 191-218.
- Alataway, A., 2023. SCS-CN and GIS-based approach for estimating runoff in western region of Saudi Arabia. *Journal of Geoscience and Environment Protection*. 11, 30-43.
- Allafta, H., Opp, C. and Al-Baldawi, B., 2026. Rainfall-surface runoff estimation using SCS-CN model and geospatial techniques: a case study of the Shatt Al-Arab Region, Iraq-Iran. *Earth*. 7, 32.
- ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000. Artificial neural networks in hydrology. I: Preliminary concepts. *Journal of Hydrologic Engineering*. 5, 115-123.

Bargam, B., Boudhar, A., Kinnard, C., Bouamri, H., Nifa, K. and Chehbouni, A., 2024. Evaluation of the support vector regression (SVR) and the random forest (RF) models accuracy for streamflow prediction under a data-scarce basin in Morocco. *Discover Applied Sciences*.

Boughton, W.C., 1989. A review of the USDA SCS curve number method. *Australian Journal of Soil Research*. 27, 511-523.

Bray, M. and Han, D., 2004. Identification of support vector machines for runoff modelling. *Journal of Hydroinformatics*. 6, 265-280.

Breiman, L., 2001. Random forests. *Machine Learning*. 45, 5-32.

Castelletti, A., Galelli, S., Restelli, M. and Soncini-Sessa, R., 2010. Tree-based reinforcement learning for optimal water reservoir operation. *Water Resources Research*. 46, W09507.

Chen, T. and Guestrin, C., 2016. XGBoost: a scalable tree boosting system, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, pp. 785-794.

Chow, V.T., Maidment, D.R. and Mays, L.W., 1988. *Applied Hydrology*, 1st ed. McGraw-Hill, New York.

Cigizoglu, H.K. and Kisi, O., 2005. Flow prediction by three back propagation techniques using k-fold partitioning of neural network training data. *Nordic Hydrology*. 36, 49-64.

Dawson, C.W. and Wilby, R.L., 2001. Hydrological modelling using artificial neural networks. *Progress in Physical Geography*. 25, 80-108.

Dibike, Y.B., Velickov, S., Solomatine, D. and Abbott, M.B., 2001. Model induction with support vector machines: introduction and applications. *Journal of Computing in Civil Engineering*. 15, 208-216.

Draper, N.R. and Smith, H., 1998. *Applied Regression Analysis*, 3rd ed. Wiley, New York.

Dunne, T. and Black, R.D., 1970. Partial area contributions to storm runoff in a small New England watershed. *Water Resources Research*. 6, 1296-1311.

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D. and Alsdorf, D., 2007. The Shuttle Radar Topography Mission. *Reviews of Geophysics*. 45, RG2004.

Freiwan, M. and Kadioğlu, M., 2008. Spatial and temporal analysis of climatological data in Jordan. *International Journal of Climatology*. 28, 521-535.

French, M.N., Krajewski, W.F. and Cuykendall, R.R., 1992. Rainfall forecasting in space and time using a neural network. *Journal of Hydrology*. 137, 1-31.

Hameed, M.M., Masood, A., Hamid, A., Elbeltagi, A., Razali, S.F.M. and Salem, A., 2025. Forecasting monthly runoff in a glacierized catchment: a comparison of extreme gradient boosting (XGBoost) and deep learning models. *PLOS ONE*. 20, e0321008.

Hastie, T., Tibshirani, R. and Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer, New York.

Hawkins, R.H., Woodward, D.E., Hjelmfelt, A.T., Van Mullem, J.A. and Quan, Q.D., 2002. Runoff curve number method: examination of the initial abstraction ratio, in: *Proceedings of the Federal Interagency Hydrologic Modeling Conference*. Las Vegas, NV.

Horton, R.E., 1933. The role of infiltration in the hydrologic cycle. *Transactions of the American Geophysical Union*. 14, 446-460.

Hsu, K., Gupta, H.V. and Sorooshian, S., 1995. Artificial neural network modeling of the rainfall-runoff process. *Water Resources Research*. 31, 2517-2530.

Kalteh, A.M., Hjorth, P. and Berndtsson, R., 2008. Review of the self-organizing map (SOM) approach in water resources: analysis, modelling and application. *Environmental Modelling & Software*. 23, 835-845.

Kassem, Y., Gökçekuş, H. and Alijl, N., 2023. Gridded precipitation datasets and gauge precipitation products for driving hydrological models in the Dead Sea region, Jordan. *Sustainability*. 15, 11965.

Khalaf, A. and Donoghue, D., 2012. Estimating recharge distribution using remote sensing: a case study from the West Bank. *Journal of Hydrology*. 414-415, 354-363.

Khozani, Z.S., Precht, E. and Ionita, M., 2025. Weekly streamflow forecasting of Rhine river based on machine learning approaches. *Natural Hazards*. 121, 4135-4153.

Liong, S.-Y. and Sivapragasam, C., 2002. Flood stage forecasting with support vector machines. *Journal of the American Water Resources Association*. 38, 173-186.

Maier, H.R., Jain, A., Dandy, G.C. and Sudheer, K.P., 2010. Methods used for the development of neural networks for the prediction of water resource variables in river systems: current status and future directions. *Environmental Modelling & Software*. 25, 891-909.

Matomela, N., Tianxin, L., Morahanye, L., Bishoge, O.K. and Ikhumhen, H.O., 2019. Rainfall-runoff estimation of Bojiang Lake watershed using SCS-CN model coupled with GIS for watershed management. *Journal of Applied and Advanced Research*. 4, 16-24.

McCarthy, G.T., 1938. The unit hydrograph and flood routing. Unpublished paper presented at the Conference of the North Atlantic Division, U.S. Army Corps of Engineers, New London, CT, 24 June 1938.

McKinney, W., 2010. Data structures for statistical computing in Python, in: *Proceedings of the 9th Python in Science Conference*. pp. 56-61.

Merizalde, M.J., Muñoz, P., Corzo, G., Muñoz, D.F., Samaniego, E. and Céleri, R., 2023. Integrating geographic data and the SCS-CN method with LSTM networks for enhanced runoff forecasting in a complex mountain basin. *Frontiers in Water*. 5, 1233899.

Michel, C., Andréassian, V. and Perrin, C., 2005. Soil Conservation Service Curve Number method: how to mend a wrong soil moisture accounting procedure? *Water Resources Research*. 41, W02011.

Mishra, S.K. and Singh, V.P., 2003. *Soil Conservation Service Curve Number (SCS-CN) Methodology*, 1st ed. Springer Netherlands, Dordrecht.

Mishra, S.K., Sahu, R.K., Eldho, T.I. and Jain, M.K., 2006. An improved Ia-S relation incorporating antecedent moisture in SCS-CN methodology. *Water Resources Management*. 20, 643-660.

Mitchell, T.M., 1997. *Machine Learning*, 1st ed. McGraw-Hill, New York.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D. and Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*. 50, 885-900.

Mosavi, A., Ozturk, P. and Chau, K., 2018. Flood prediction using machine learning models: literature review. *Water*. 10, 1536.

Muneer, A.S., Afan, H.A., Kamel, A.H. and Sayl, K.N., 2022. Runoff mapping using the SCS-CN method and artificial neural network algorithm, Ratga Basin, Iraq. *Arabian Journal of Geosciences*. 15, 666.

OpenTopography, 2026. Shuttle Radar Topography Mission (SRTM GL1) Global 30m. <https://portal.opentopography.org/raster?opentopoID=OTSRTM.082015.4326.1> (2026)

Palestinian Central Bureau of Statistics (PCBS), 2024. <http://www.pcbs.gov.ps> (2026)  
Palestinian Central Bureau of Statistics (PCBS), 2026.

Palestinian Water Authority (PWA), 2026. Policies and Strategies.  
<https://www.pwa.ps/en/Category/64/Policies-and-Strategies> (2026).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, É., 2011. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*. 12, 2825-2830.

Ponce, V.M. and Hawkins, R.H., 1996. Runoff curve number: has it reached maturity? *Journal of Hydrologic Engineering*. 1, 11-19.

Salameh, A.A.M., 2024. Using the precipitation concentration index for characterizing the rainfall distribution in the Levant. *Journal of Water and Climate Change*. 15, 1945-1960.

Satheeshkumar, S., Venkateswaran, S. and Kannan, R., 2017. Rainfall-runoff estimation using SCS-CN and GIS approach in the Pappiredipatti watershed of the Vaniyar sub basin, South India. *Modeling Earth Systems and Environment*. 3, 24.

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P.A. and Carrillo, G., 2011. Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA. *Hydrology and Earth System Sciences*. 15, 2895-2911.

Schneider, L.E. and McCuen, R.H., 2005. Statistical guidelines for curve number generation. *Journal of Irrigation and Drainage Engineering*. 131, 282-290.

Shadeed, S., 2023. The potential use of the SCS-CN method to estimate extreme floods in the West Bank data-scarce catchments. *An-Najah University Journal for Research - A (Natural Sciences)*. 37, 25-40.

Shadeed, S. and Almasri, M., 2010. Application of GIS-based SCS-CN method in West Bank catchments, Palestine. *Water Science and Engineering*. 3, 1-13.

Shadeed, S., Judeh, T. and Riksen, M., 2020. Rainwater harvesting for sustainable agriculture in high water-poor areas in the West Bank, Palestine. *Water*. 12, 380.

Shah, G., Zaidi, A., Qureshi, A.L., Hussain, S., Jokhio, R. and Aziz, T., 2023. Rainfall-runoff modeling using machine learning in the urban watershed of Pishin Lora Basin, Balochistan (Pakistan). *Research Square* (preprint).

Singh, A., Kumar, P., Ali, R., Al-Ansari, N., Vishwakarma, D.K., Kushwaha, K., Panda, K.C., Sagar, A., Mirzania, E., Elbeltagi, A., Kuriqi, A. and Heddami, S., 2022. An integrated statistical-machine learning approach for runoff prediction. *Sustainability*. 14, 8209.

Soulis, K.X., 2021. Soil Conservation Service Curve Number (SCS-CN) method: current applications, remaining challenges, and future perspectives. *Water*. 13, 192.

Tarboton, D.G., 1997. A new method for the determination of flow directions and upslope areas in grid digital elevation models. *Water Resources Research*. 33, 309-319.

Tarboton, D.G., Bras, R.L. and Rodriguez-Iturbe, I., 1991. On the extraction of channel networks from digital elevation data. *Hydrological Processes*. 5, 81-100.

Tokar, A.S. and Johnson, P.A., 1999. Rainfall-runoff modeling using artificial neural networks. *Journal of Hydrologic Engineering*. 4, 232-239.

Tyralis, H., Papacharalampous, G. and Langousis, A., 2019. A brief review of random forests for water scientists and practitioners and their recent history in water resources. *Water*. 11, 910.

United Nations Environment Programme (UNEP), 2003. Desk Study on the Environment in the Occupied Palestinian Territories. UNEP, Nairobi.

U.S. Army Corps of Engineers, 2026. Official Publications of the Headquarters.  
<https://www.publications.usace.army.mil/> (2026)

Woodward, D.E., Hawkins, R.H., Jiang, R., Hjelmfelt, A.T., Van Mullem, J.A. and Quan, Q.D., 2003. Runoff curve number method: examination of the initial abstraction ratio, in: Proceedings of the World Water and Environmental Resources Congress 2003. ASCE, Reston, VA, pp. 1-10.

World Bank, 2023. West Bank and Gaza Country Climate and Development Report. World Bank, Washington, DC.

Yaseen, Z.M., El-shafie, A., Jaafar, O., Afan, H.A. and Sayl, K.N., 2015. Artificial intelligence based models for stream-flow forecasting: 2000-2015. *Journal of Hydrology*. 530, 829-844.

Yu, X., Liong, S.-Y. and Babovic, V., 2004. EC-SVM approach for real-time hydrologic forecasting. *Journal of Hydroinformatics*. 6, 209-223.

Zealand, C.M., Burn, D.H. and Simonovic, S.P., 1999. Short term streamflow forecasting using artificial neural networks. *Journal of Hydrology*. 214, 32-48.

## APPENDIX

### Appendix A: Annual Station Rainfall Data (2020–2024)

The following tables provide AMC classification and Pmax values from the 41 IMS monitoring stations for each of the five study years. These data constitute the primary observational inputs to the GIS-based spatial interpolation and SCS-CN runoff computation described in Chapter 5. Only the 2020 reference-year data are reproduced in full; subsequent years follow the same format.

Table A. 19: Station AMC classification and Pmax for 2020 (selected stations)

| <b>Station</b>   | <b>AMC 2020</b> | <b>Pmax 2020 (mm)</b> | <b>X (ITM)</b> | <b>Y (ITM)</b> |
|------------------|-----------------|-----------------------|----------------|----------------|
| Afula Nir Haemeq | 2               | 57.7                  | 176329         | 222461         |
| Bet Haarava      | 1               | 16.0                  | 197617         | 134795         |
| Bet Hashitta     | 3               | 40.3                  | 195000         | 219000         |
| Efrata           | 1               | 63.4                  | 164547         | 117794         |
| Eli              | 3               | 58.2                  | 175018         | 164267         |
| Enot Zuqim       | 1               | 9.0                   | 192967         | 124936         |
| Gilgal           | 3               | 19.2                  | 192838         | 156087         |

|                     |   |      |        |        |
|---------------------|---|------|--------|--------|
| Gilgal Man          | 2 | 18.5 | 192838 | 156087 |
| Gittit              | 2 | 34.8 | 187629 | 167716 |
| Har Harasha         | 3 | 68.0 | 164362 | 150255 |
| Itamar              | 3 | 32.1 | 183379 | 174088 |
| Jerusalem Centre    | 1 | 41.8 | 168829 | 130896 |
| Jerusalem Givat Ram | 1 | 47.3 | 171151 | 132013 |
| Karmel              | 3 | 37.5 | 167578 | 93316  |
| Kefar Ezyon         | 1 | 56.9 | 161024 | 117547 |
| Lahav               | 3 | 36.8 | 137975 | 87791  |
| Maggal              | 2 | 25.5 | 153843 | 198882 |
| Merav               | 2 | 44.5 | 190024 | 206517 |
| Mevo Horon          | 1 | 53.0 | 153685 | 139581 |
| Migdal Oz           | 1 | 69.7 | 163572 | 116542 |
| Nahshon             | 1 | 48.3 | 146523 | 137985 |
| Peza'el             | 1 | 15.6 | 192272 | 161260 |
| Qarne Shomeron      | 2 | 46.2 | 159311 | 175785 |
| Qedumim             | 2 | 49.2 | 165280 | 180020 |
| Re'ut               | 1 | 43.7 | 152395 | 144119 |
| Rosh Zurim          | 1 | 48.5 | 161839 | 119146 |
| Sede Eliyyahu       | 3 | 23.8 | 198332 | 205065 |
| Shani               | 3 | 33.3 | 156375 | 85055  |
| Talmon              | 3 | 62.8 | 162614 | 149870 |
| Tel Yosef           | 2 | 46.1 | 187170 | 216859 |
| Tirat Zevi          | 3 | 29.2 | 199741 | 203309 |
| Umm Al-Fahm         | 2 | 80.0 | 164986 | 213719 |
| Yad Hanna           | 2 | 26.5 | 150965 | 192490 |

|                                 |              |
|---------------------------------|--------------|
| <b>Maximum (m)</b>              | <b>0.464</b> |
| Minimum (m)                     | -0.882       |
| Mean (m)                        | -0.244       |
| Standard Deviation $\sigma$ (m) | 0.183        |

## Appendix B : West Bank Map



Figure C- 25: Rainfall Station Location

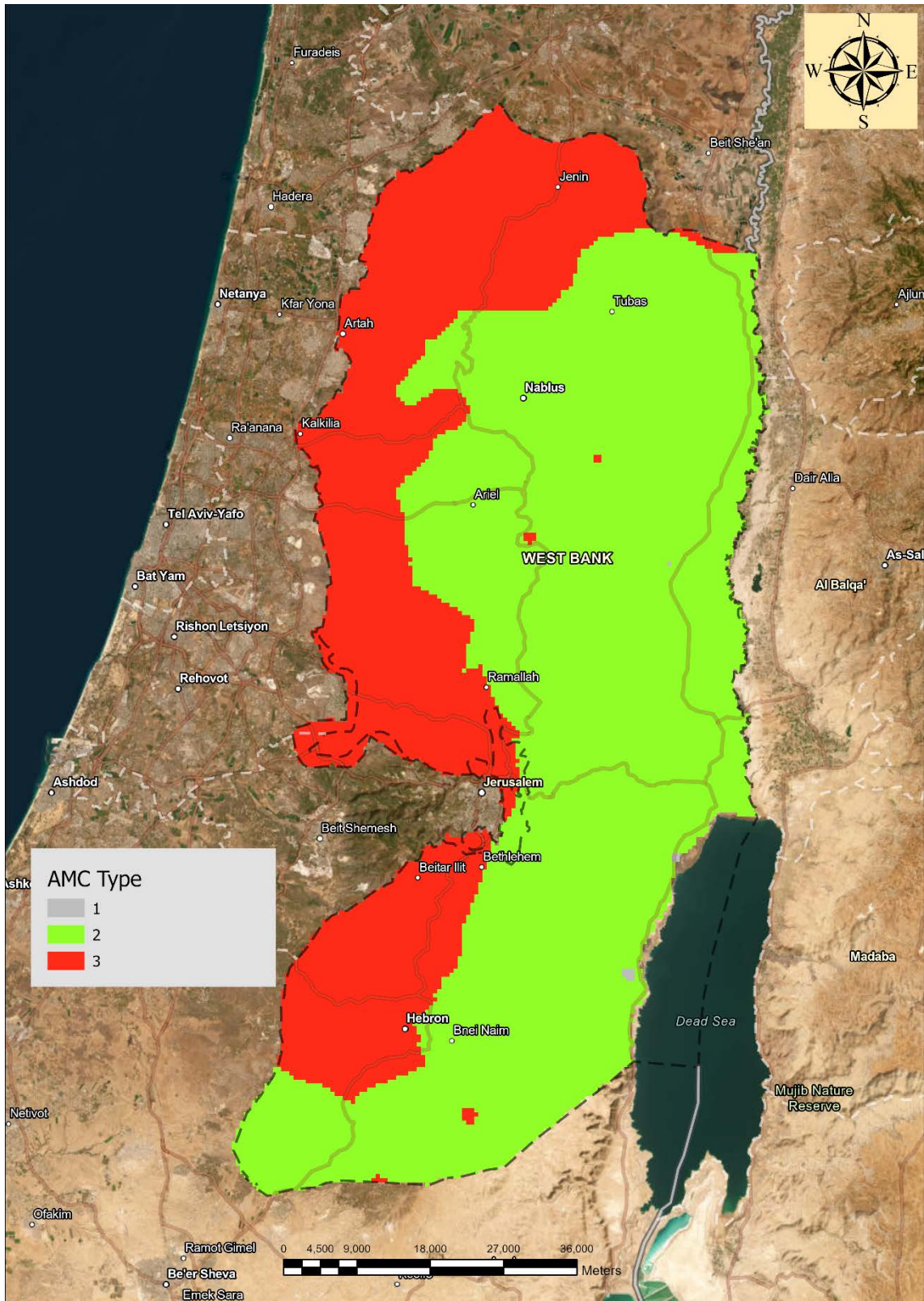


Figure C-26:AMC 2020, West Bank

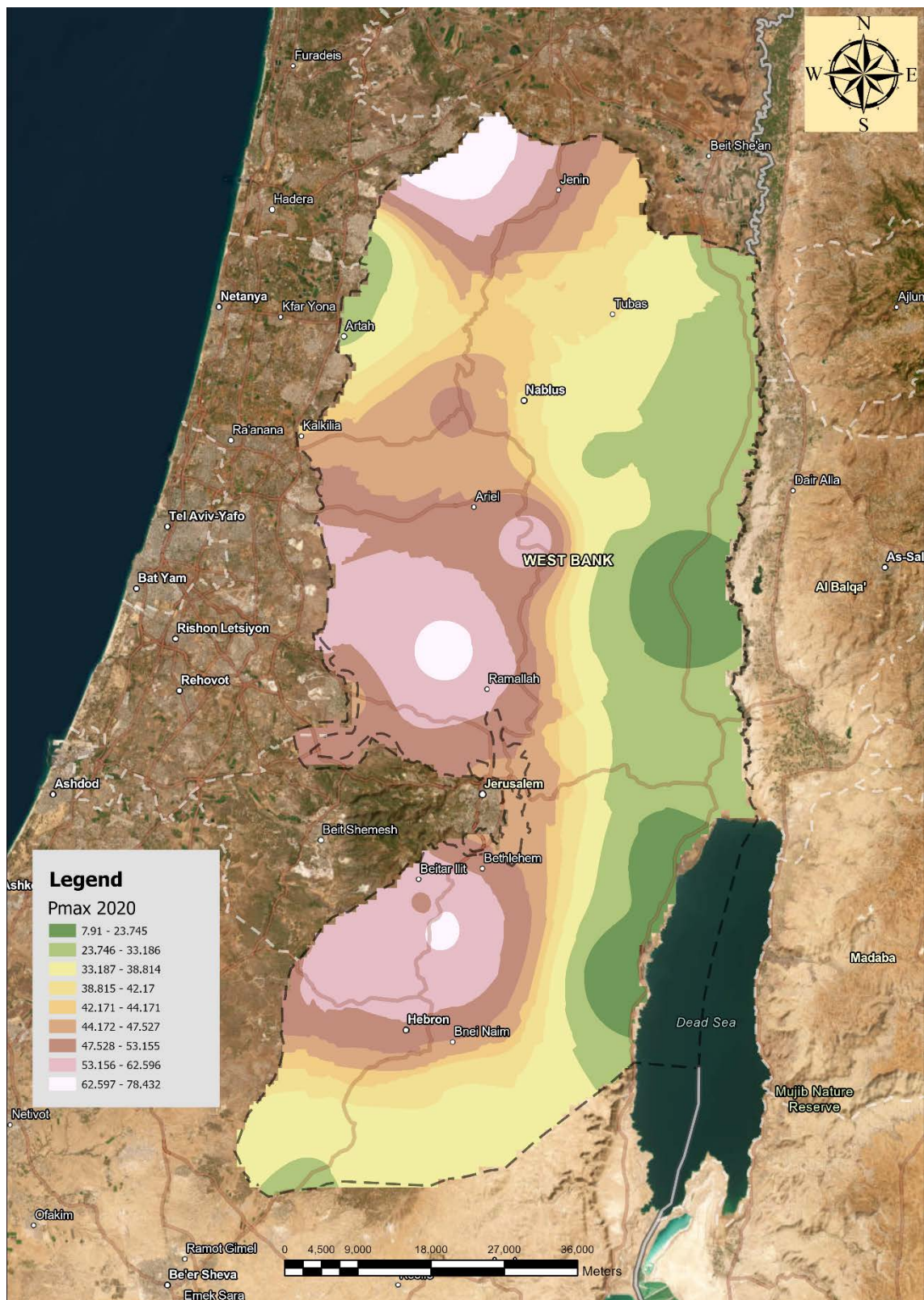


Figure C-27:Pmax 2020, West Bank

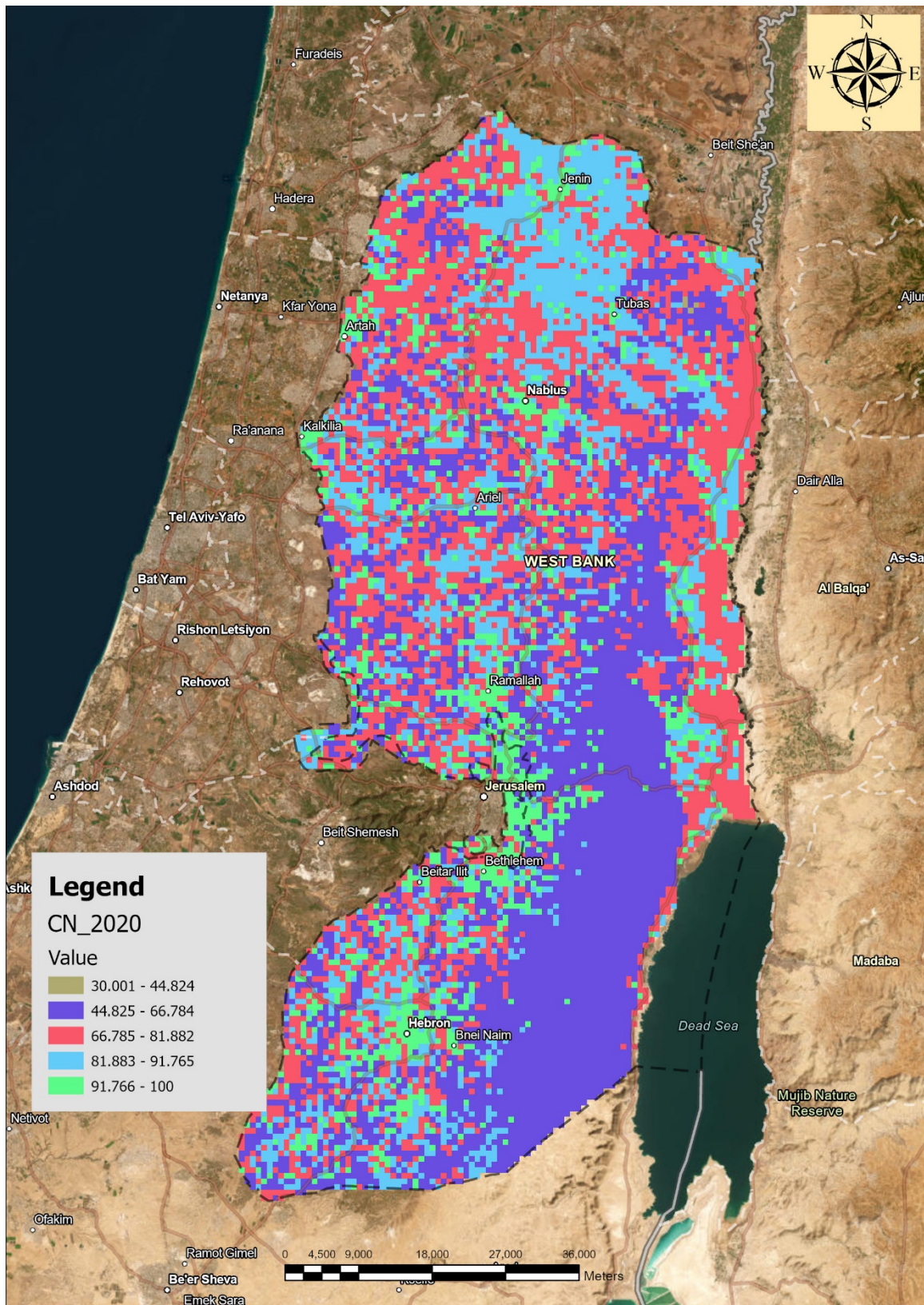


Figure C-28 :CN 2020, West Bank

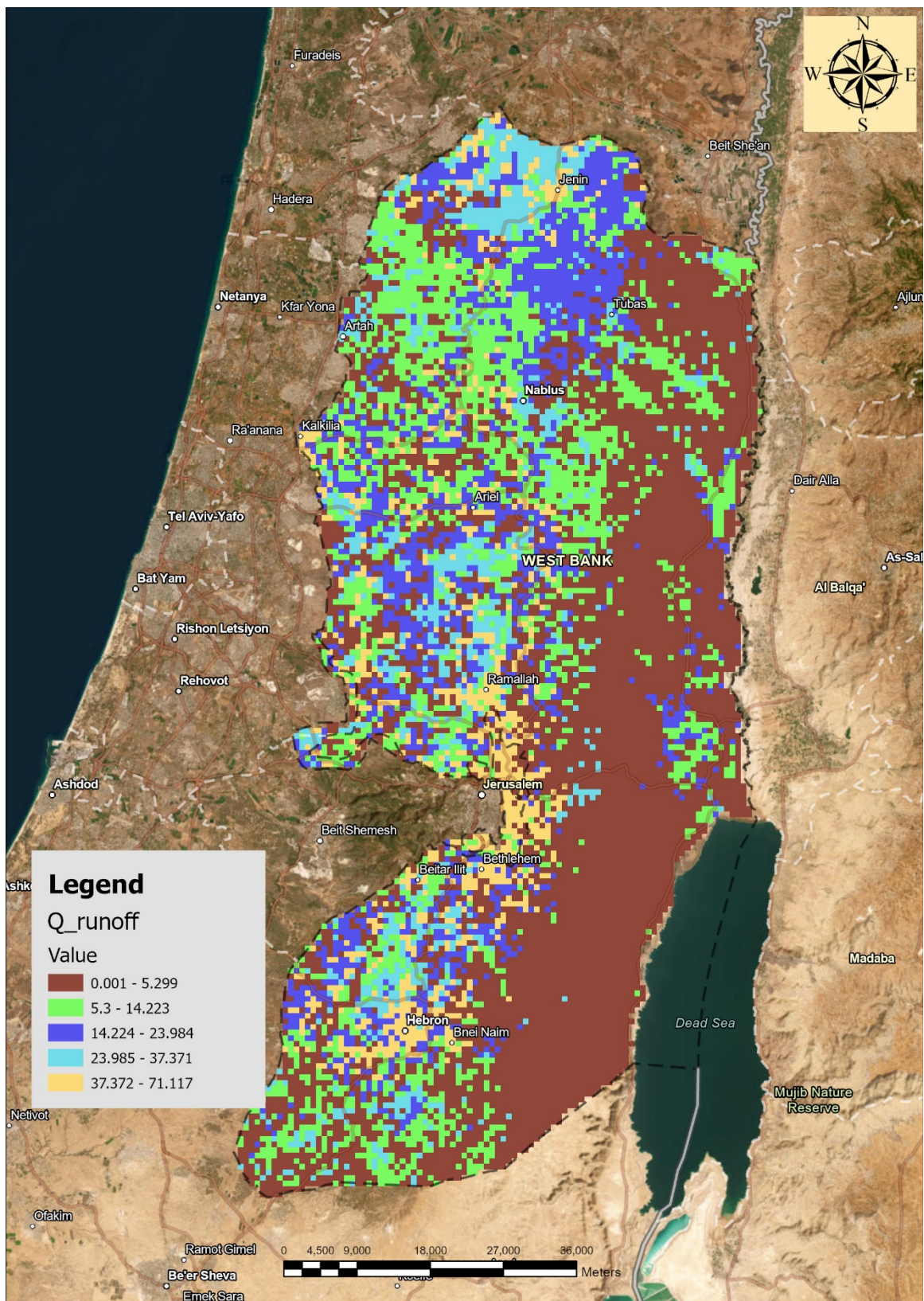


Figure C- 29 qmm 2020, West Bank

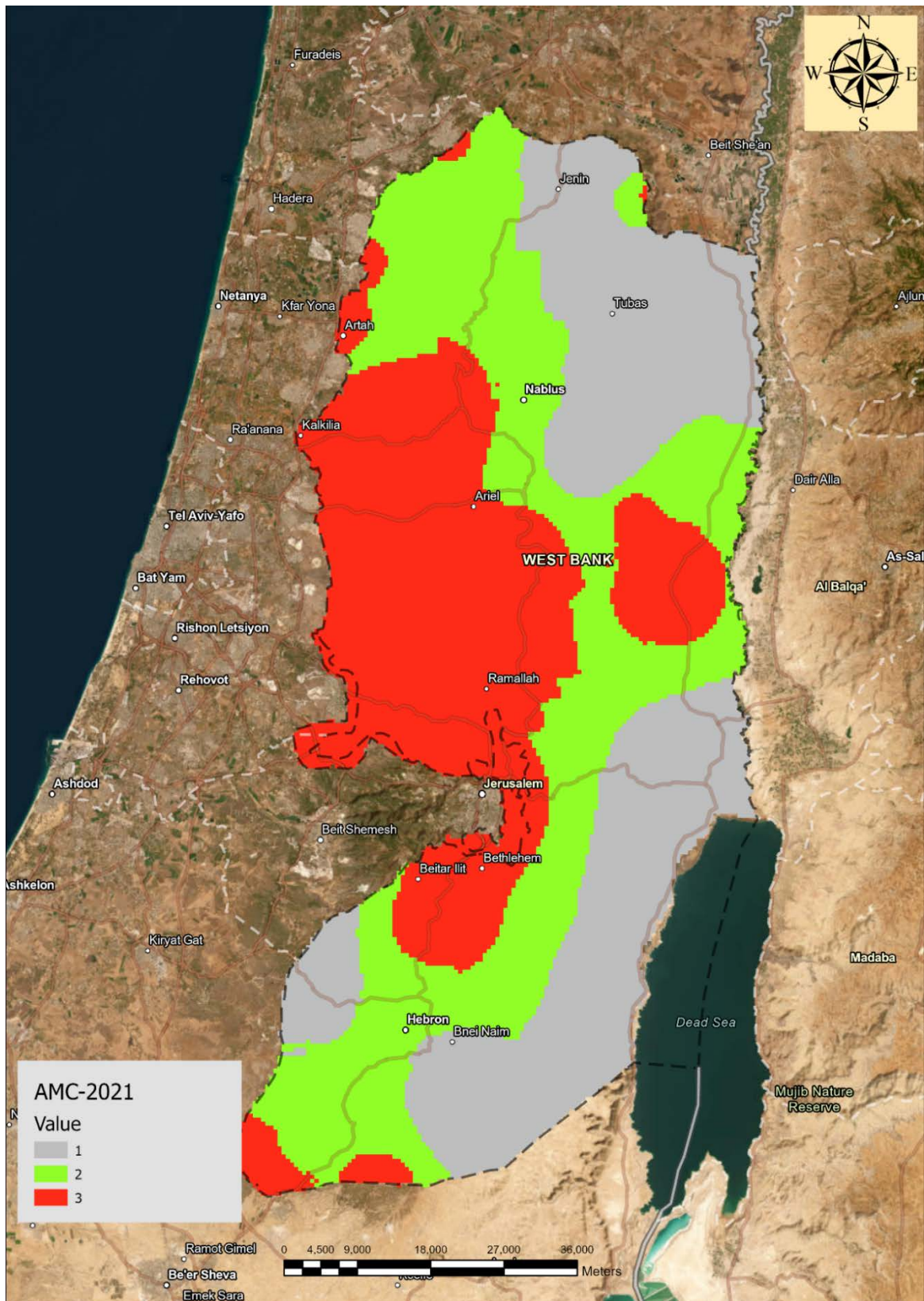


Figure C- 30 :AMC 2021, West Bank

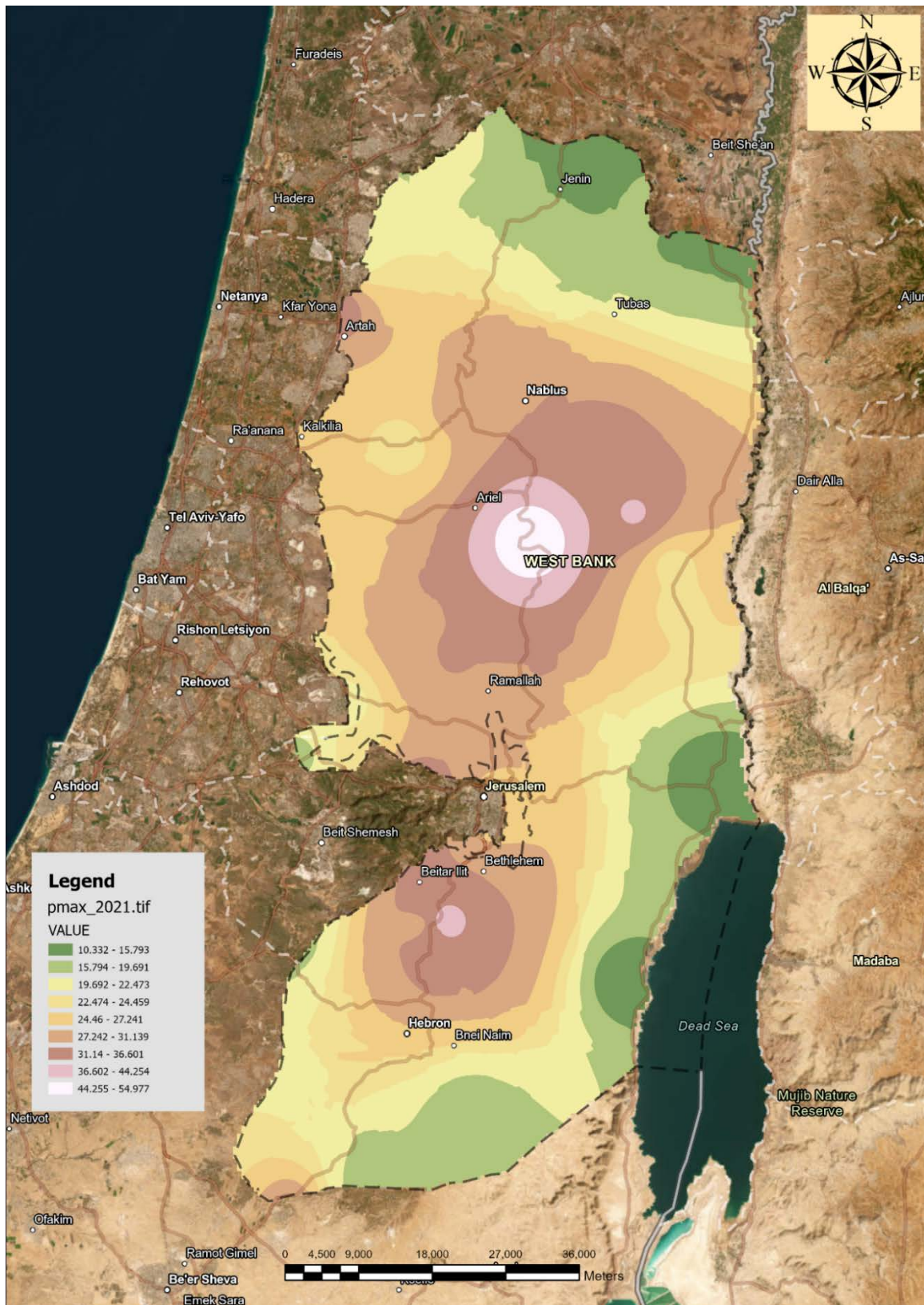


Figure C- 31: Pmax 2021, West Bank

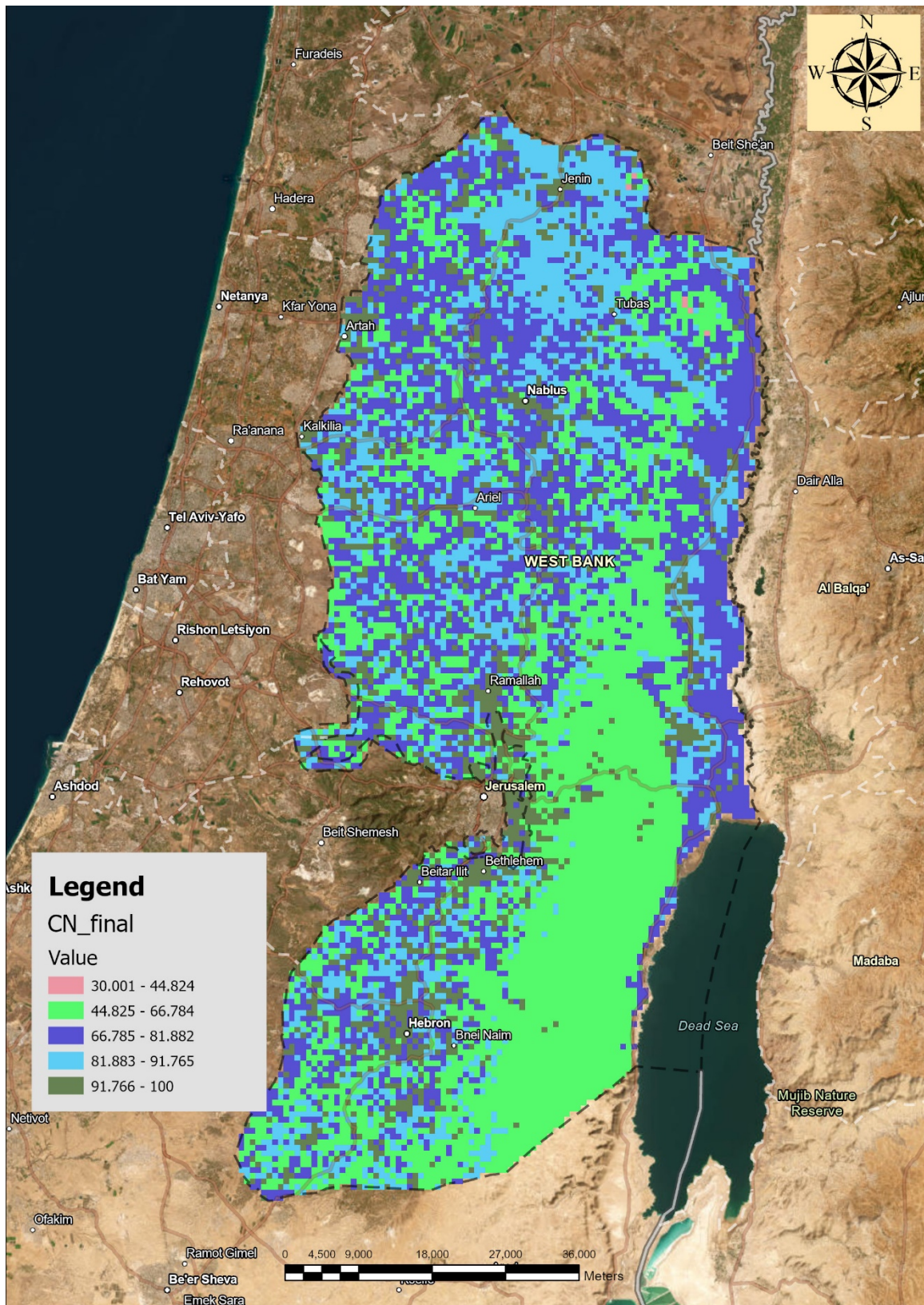


Figure C-32 CN 2021, West Bank

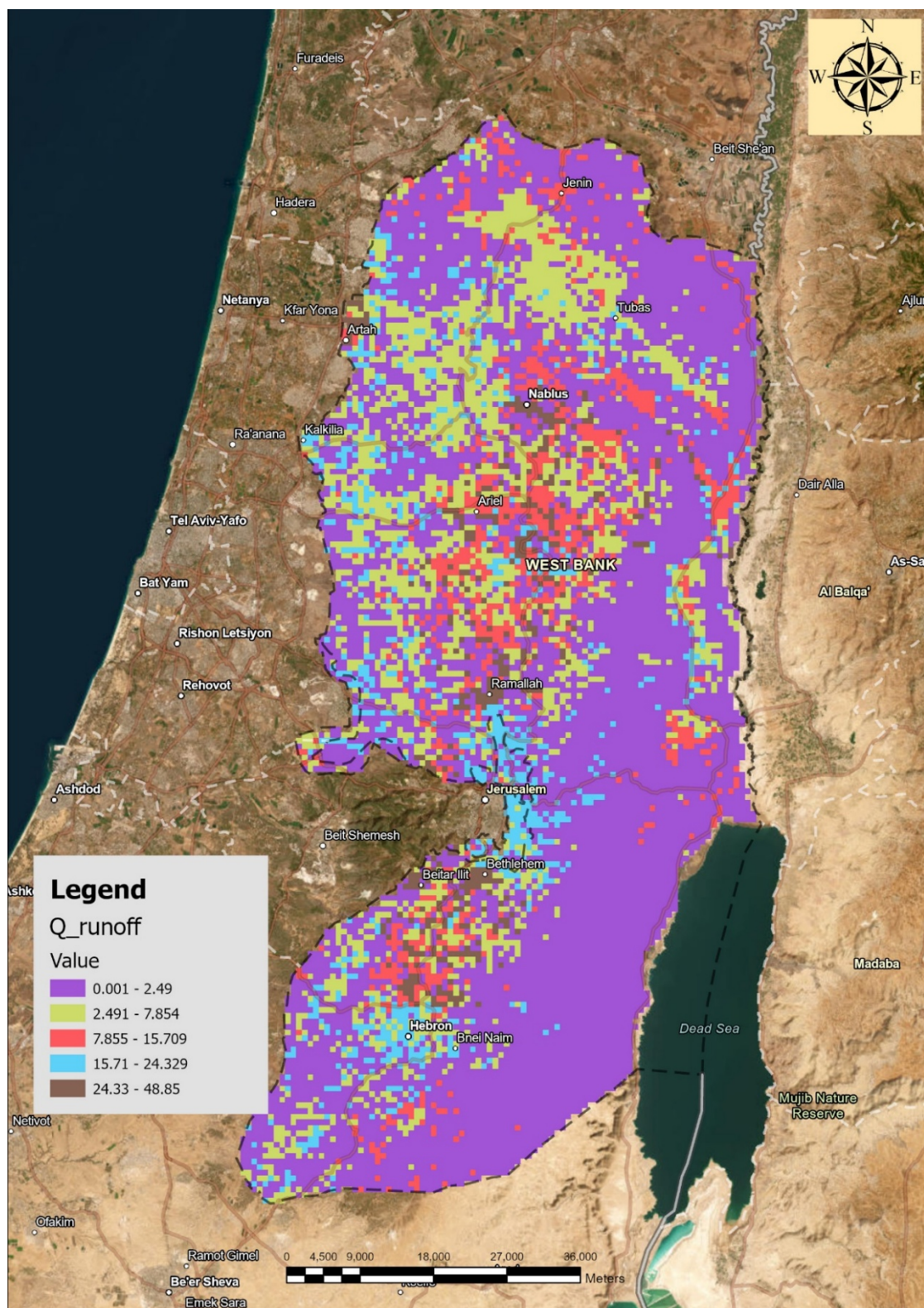


Figure C-33 qmm 2021, West Bank

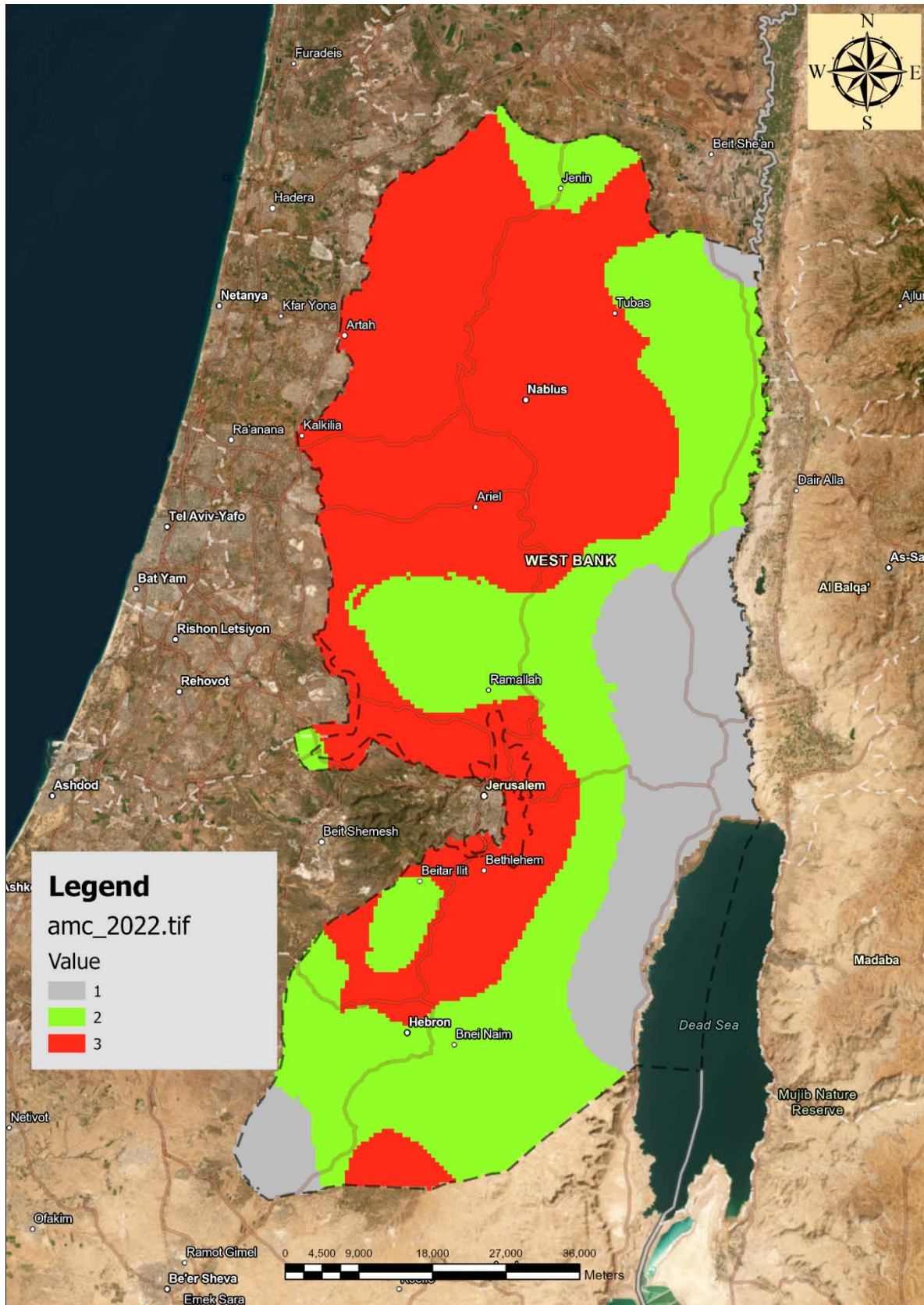


Figure C-34: AMC 2022, West Bank

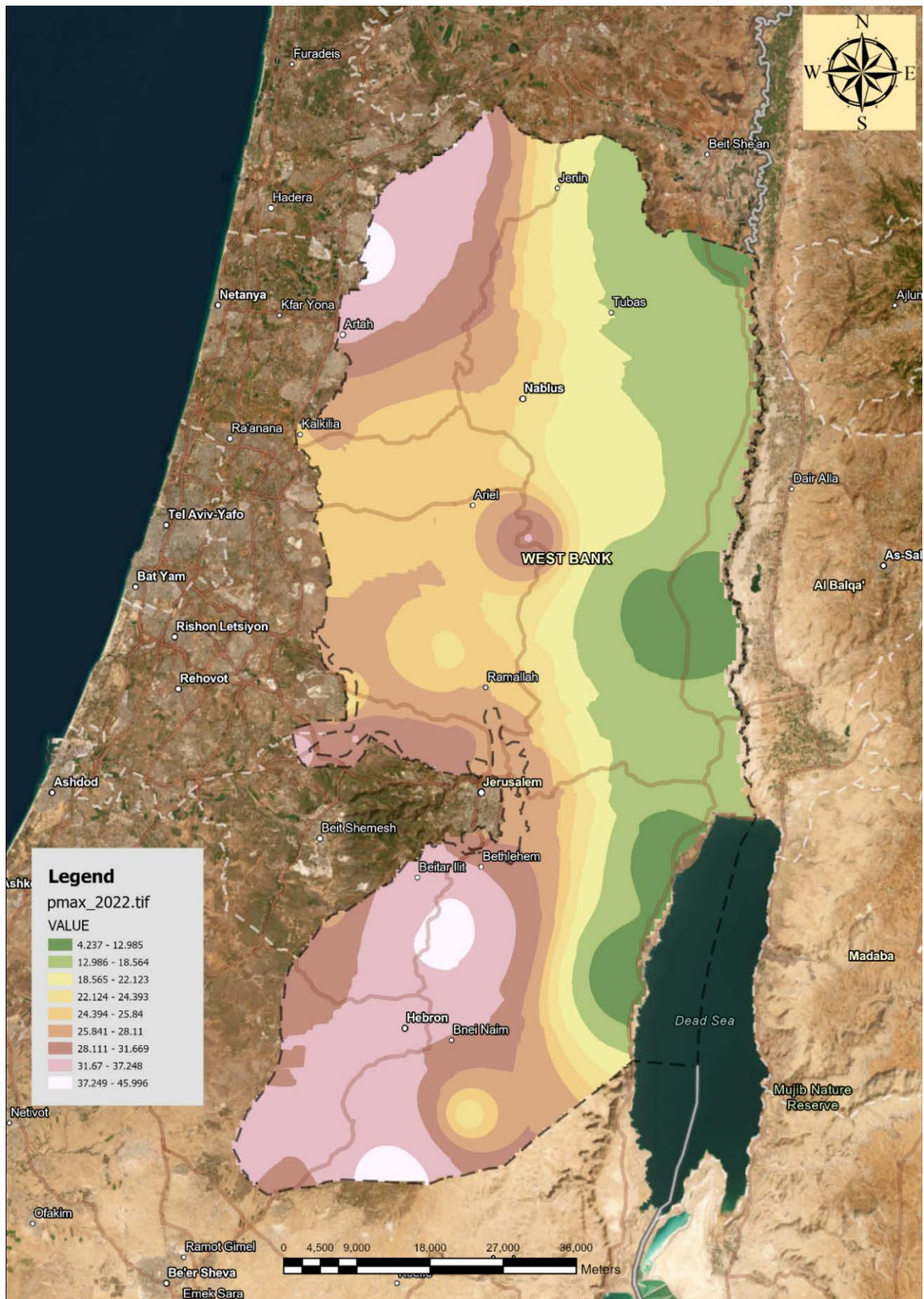


Figure C-35: Pmax 2022, West Bank

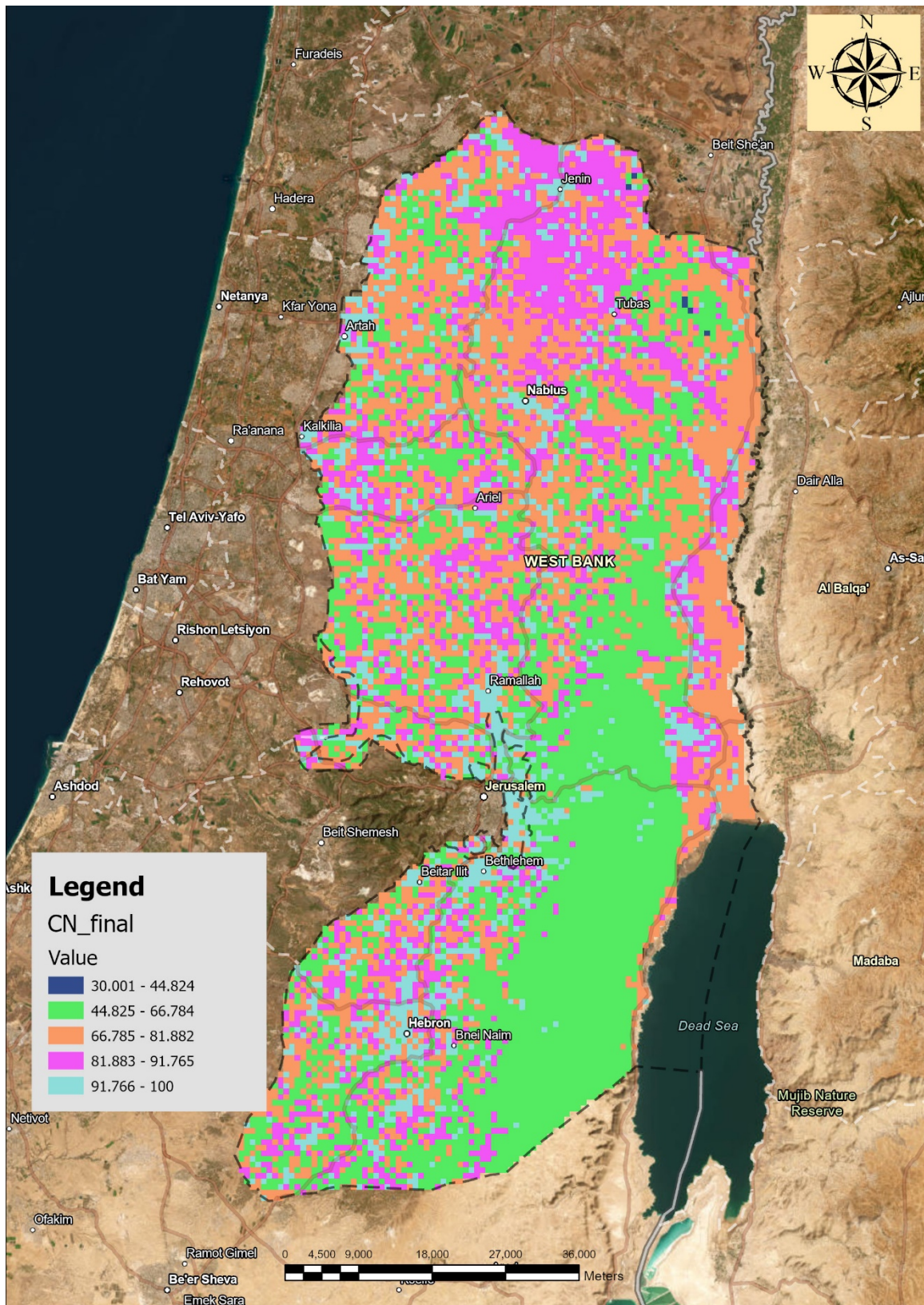


Figure C-36 CN 2022, West Bank

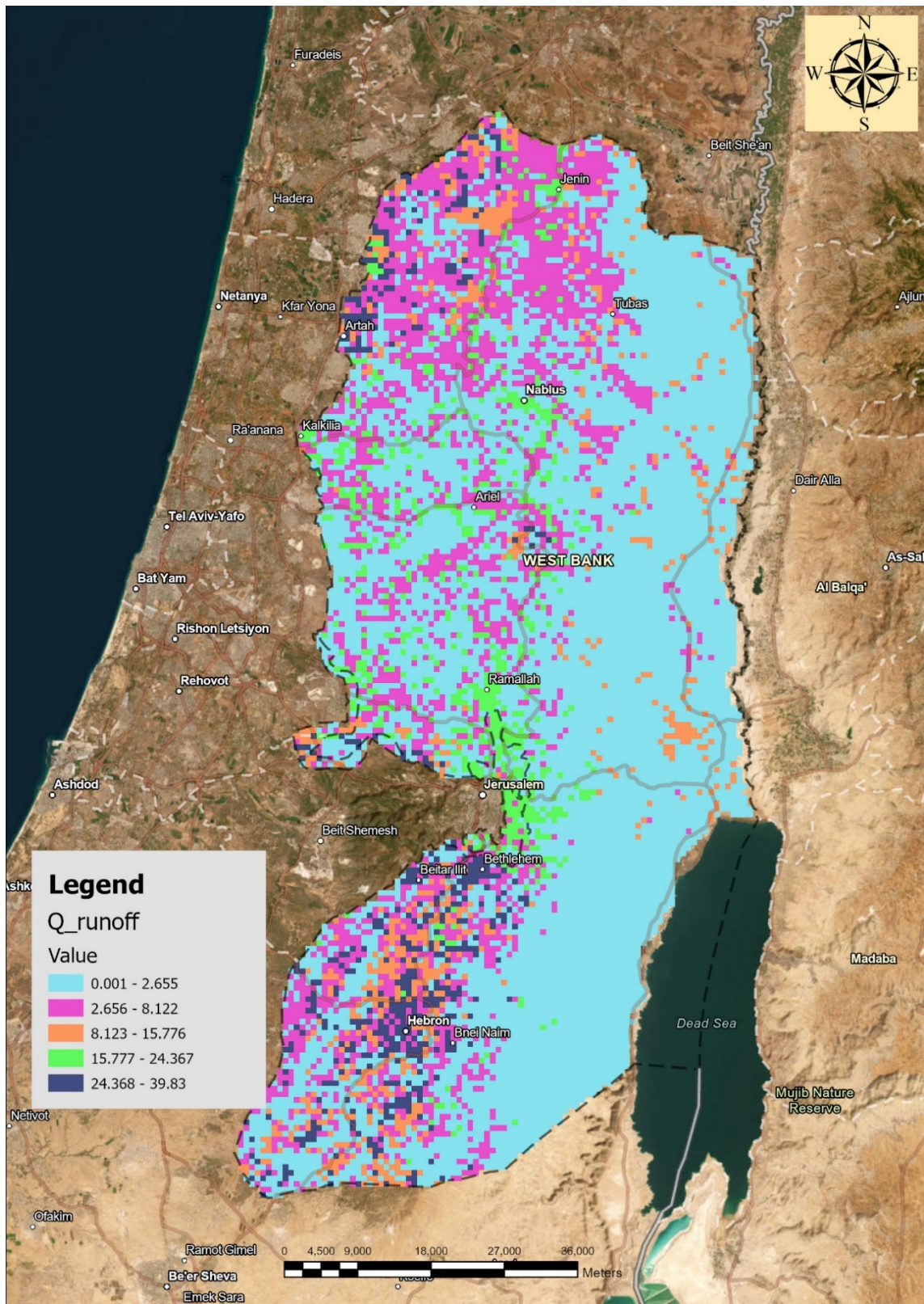


Figure C-37 : qmm 2022, West Bank

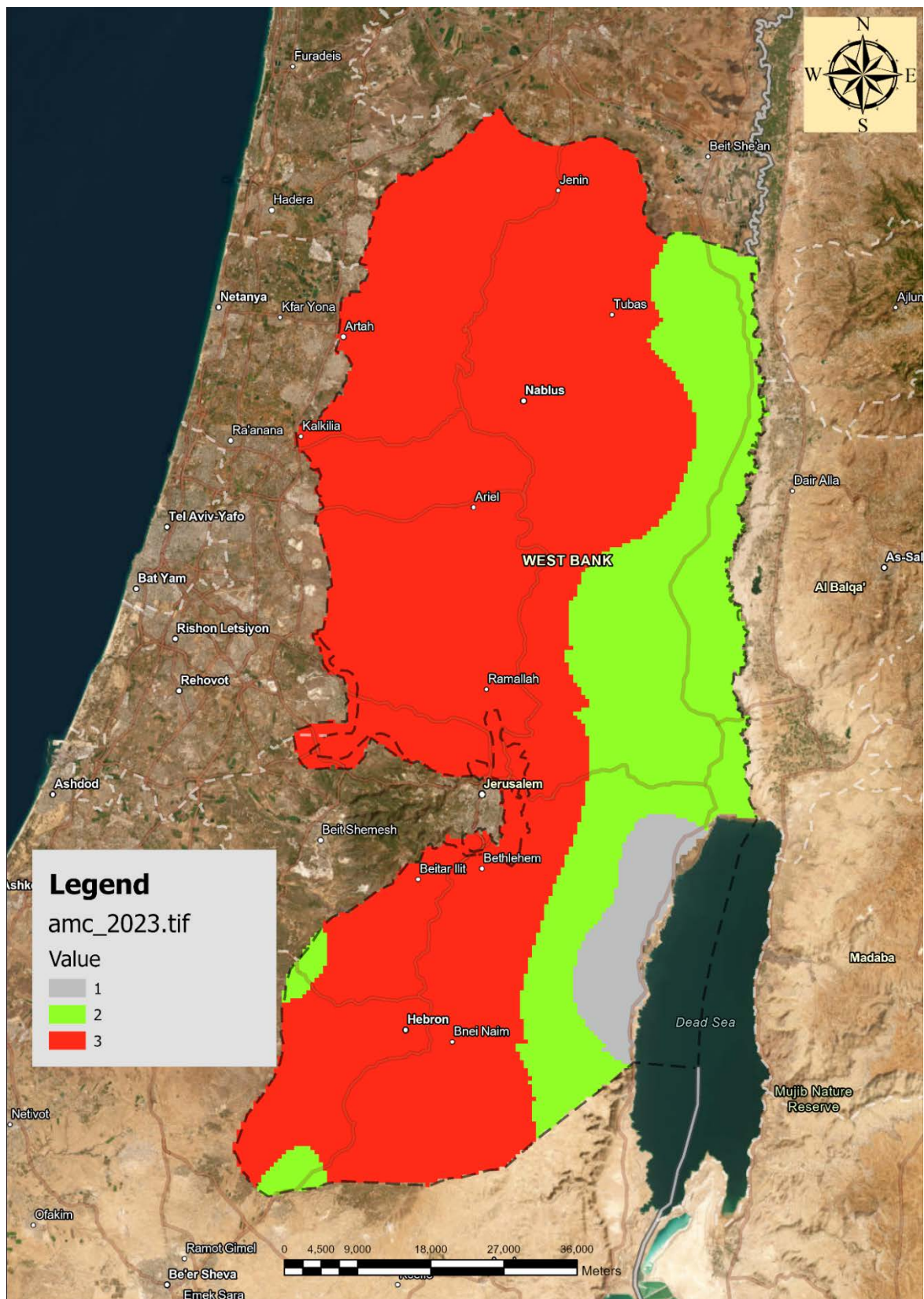


Figure C-38 :AMC 2023, West Bank

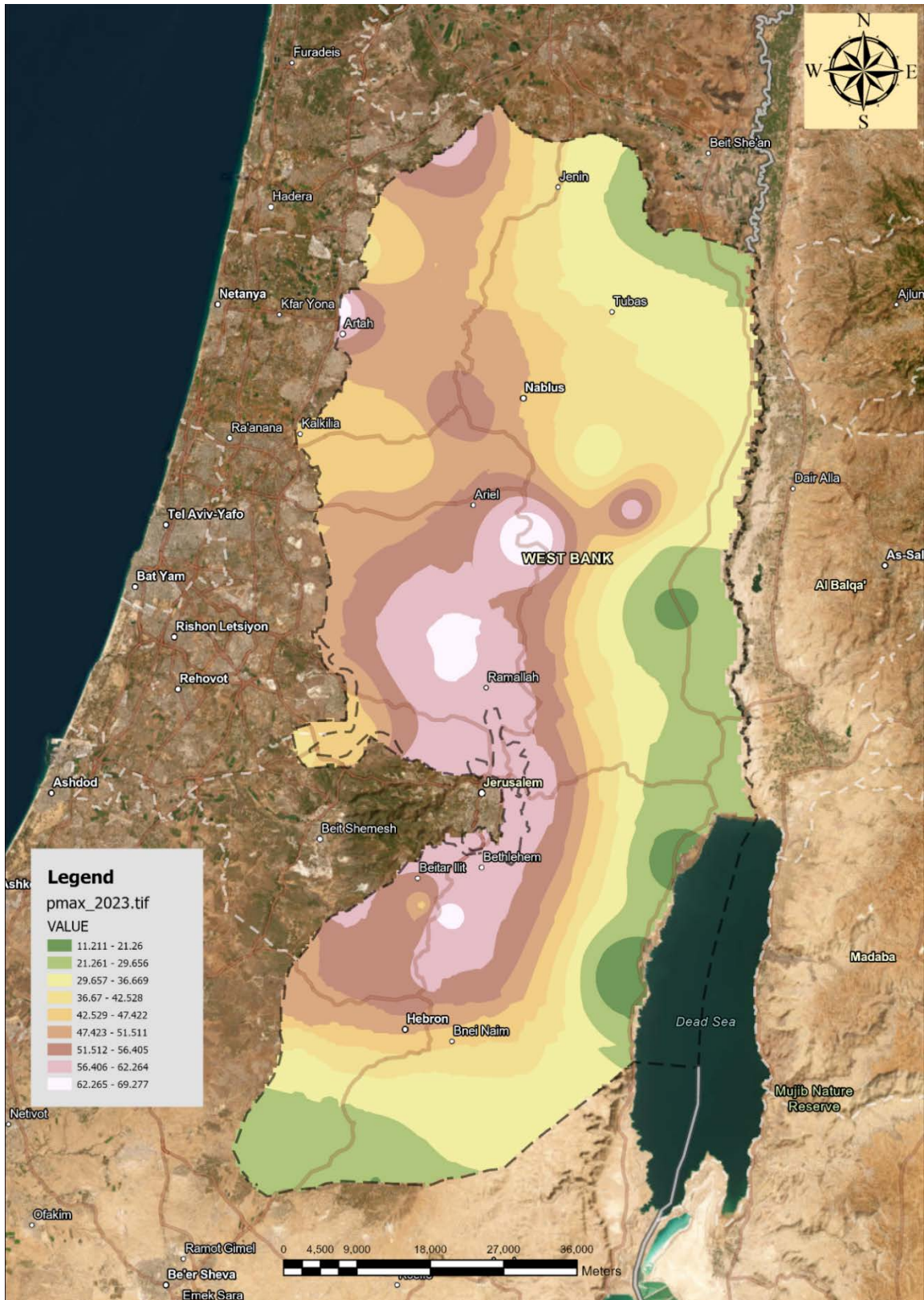


Figure C-39:Pmax 2023, West Bank

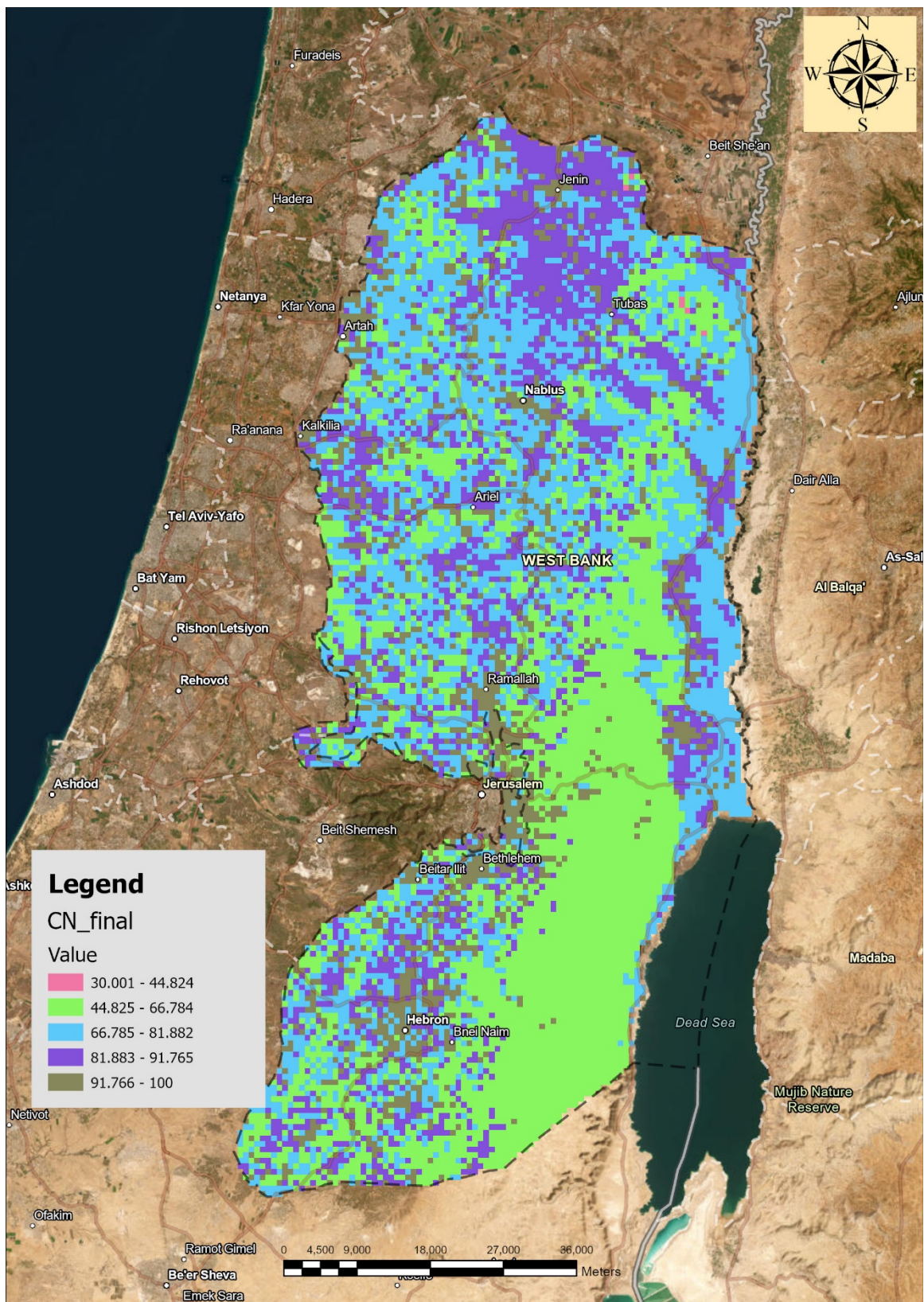


Figure C-40 CN 2023, West Bank

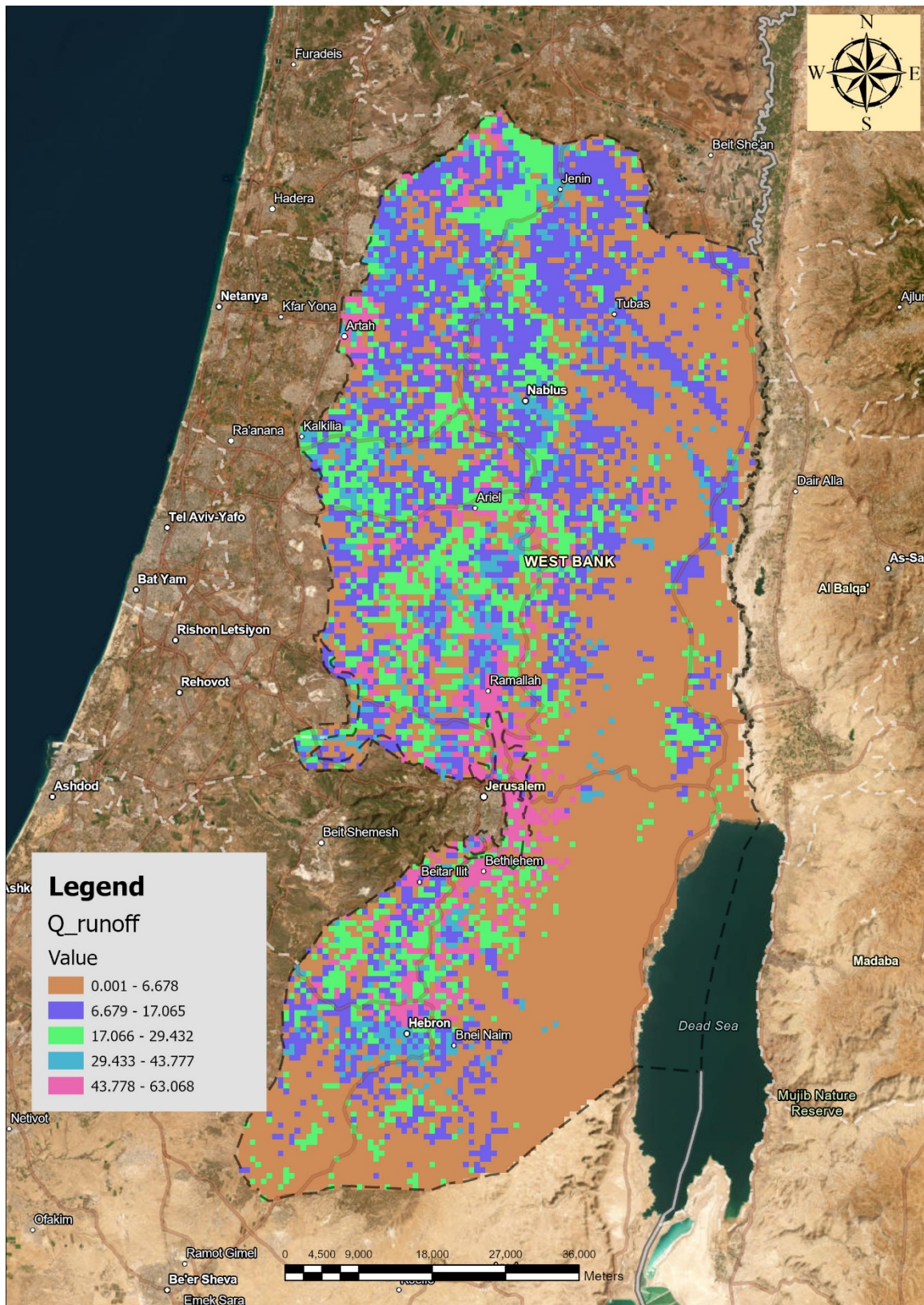


Figure C-41 qmm 2023, West Bank

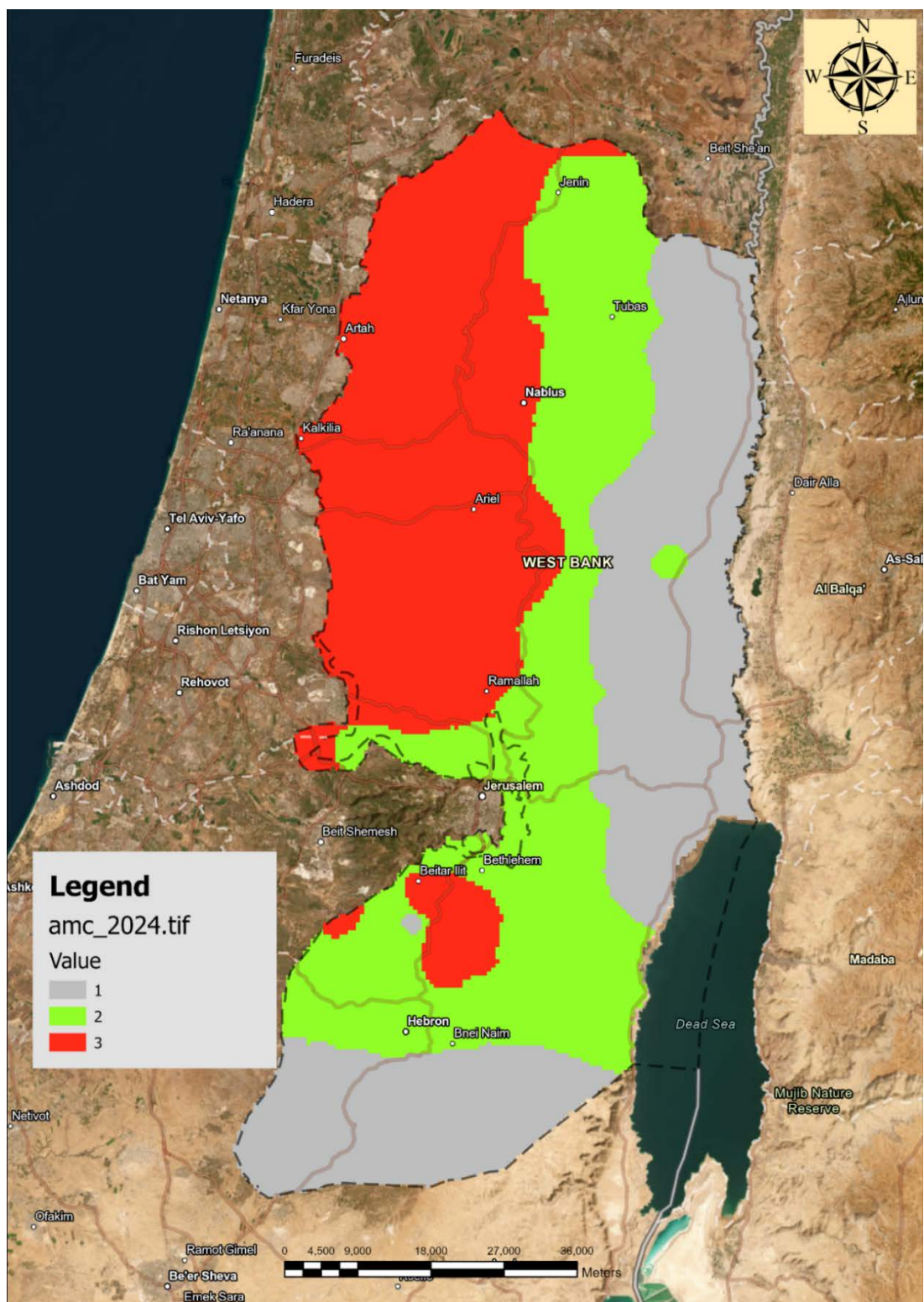


Figure C-42:AMC 2024, West Bank

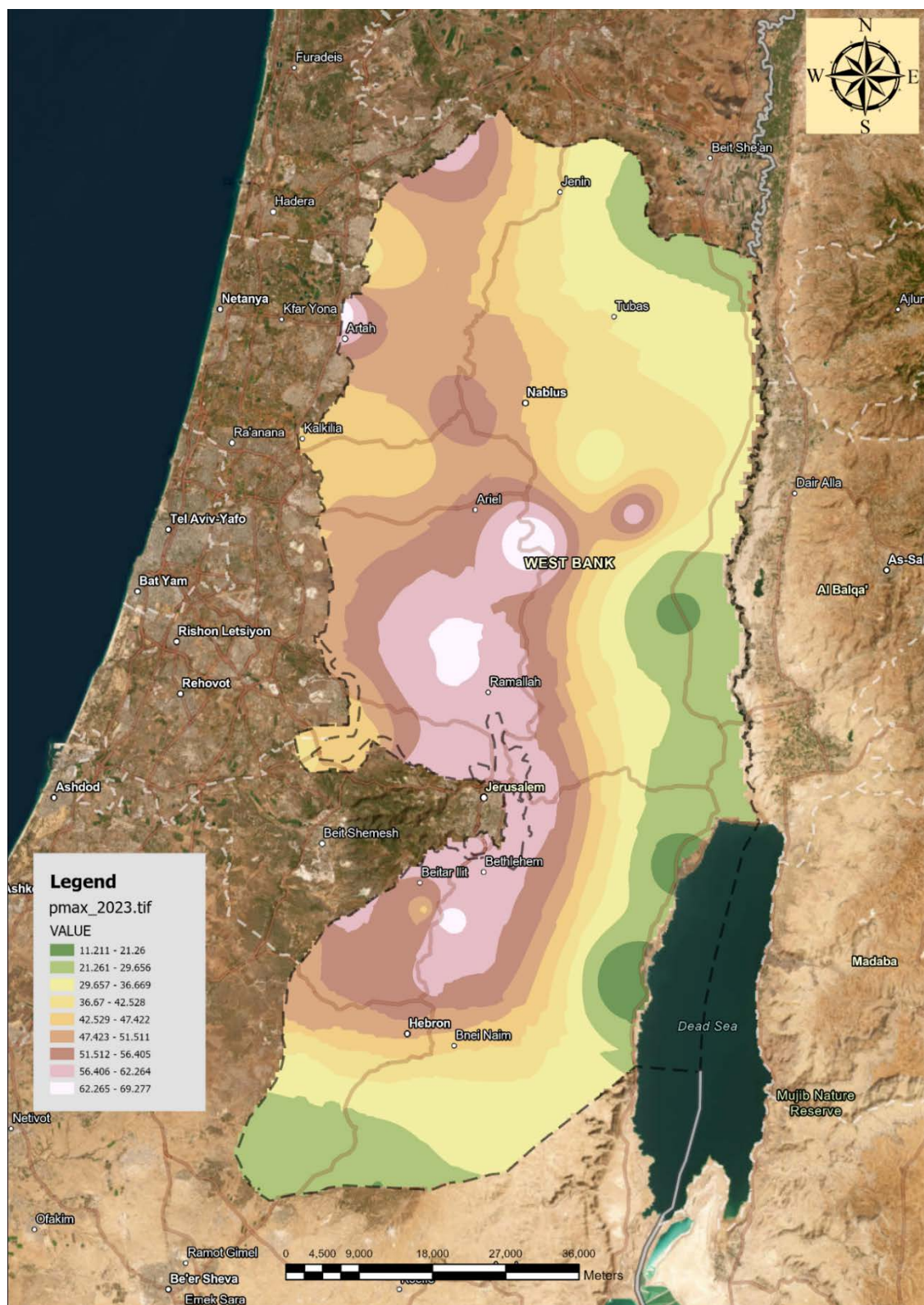


Figure C-43:Pmax 2024, West Bank

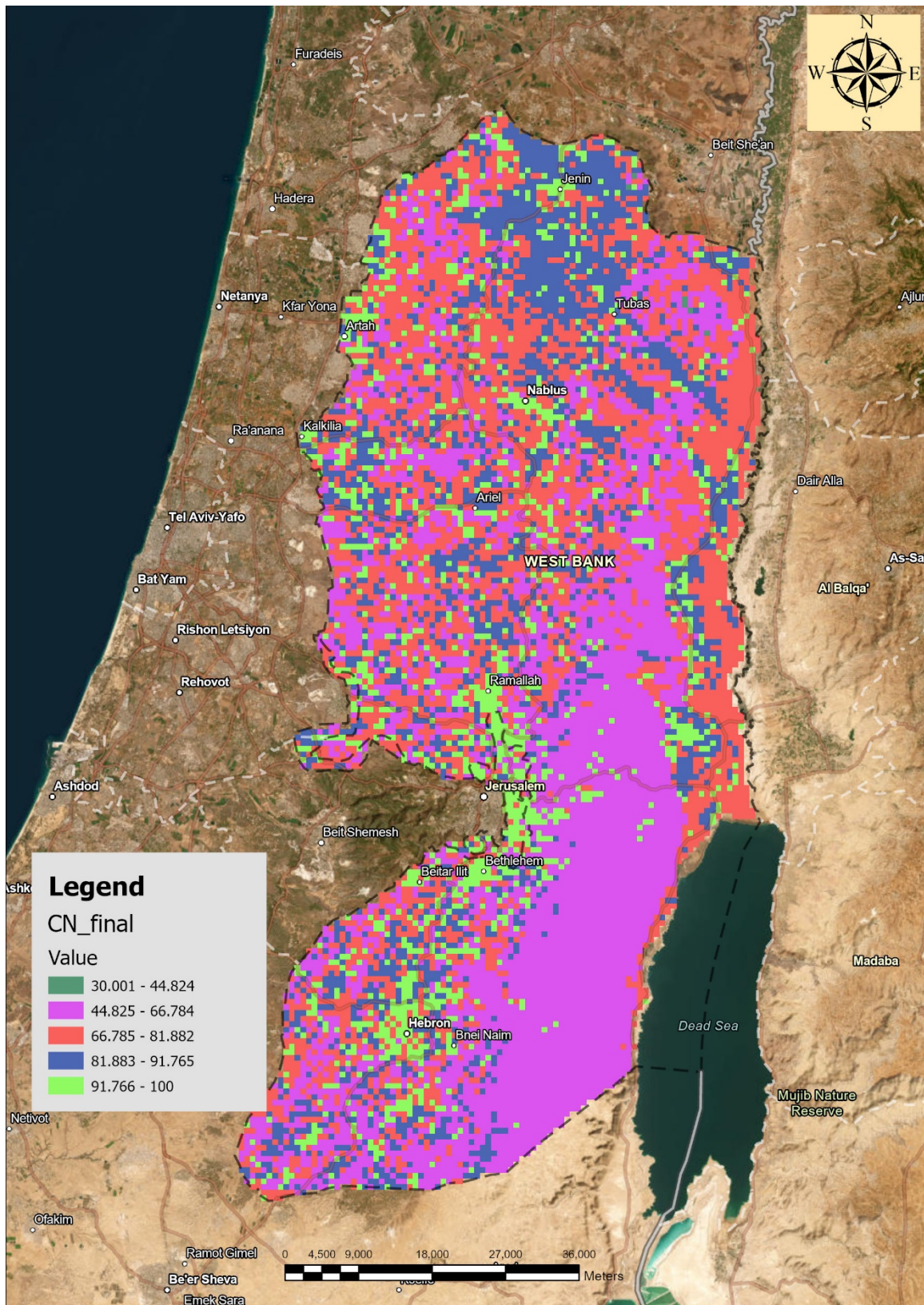


Figure C-44:CN 2024, West Bank

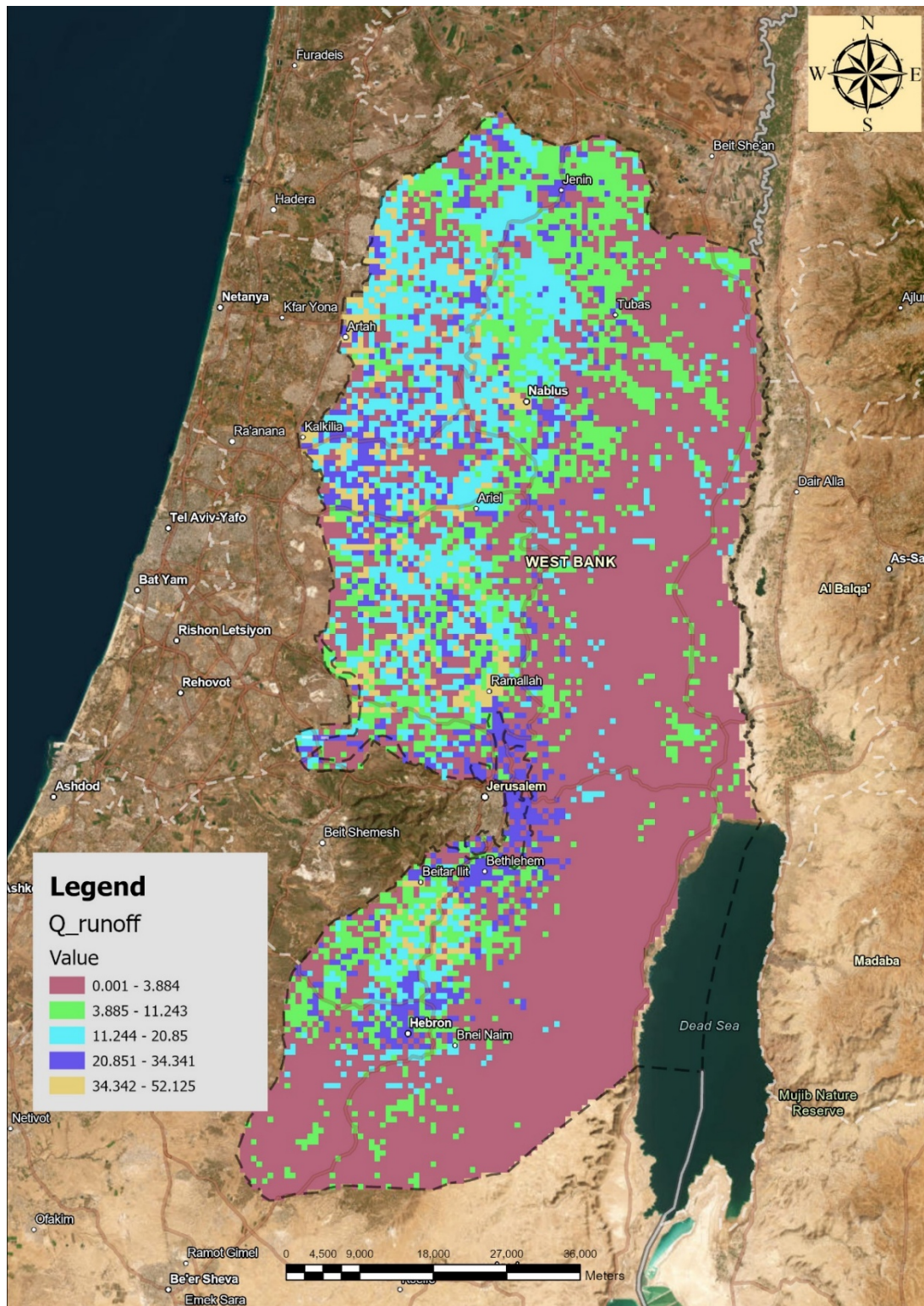


Figure C-45:qmm 2024, West Bank