



**Politecnico
di Torino**

Master's degree in

*Environmental and Land Engineering:
Exploiting Sentinel-2 dataset to assess flowing status in
non-perennial rivers. The Sangone river case study.*

Supervisor:

Prof. Paolo Vezza

Co-supervisor:

Dot. Giovanni Negro

Candidate:

Giammarco Manfreda

Academic year 2022 – 2023

Abstract

Non-perennial rivers are the most common type of rivers on Earth today. Due to anthropogenic pressures, such as changes in land uses, water withdrawals and climate change, the shifting from a perennial to a non-perennial condition is becoming faster and the preservation of temporary rivers and streams is in jeopardy. Their ubiquity and crucial role in biodiversity and rivers' ecosystems defense are widely recognized. However, the lack of social perception of their importance, together with the complexity of evaluating their ephemerality, makes it difficult for policymakers to draw up appropriate legislation to establish their degraded condition and promote their preservation. In this thesis, it was used hydrogeologic data, Sentinel-2's images and a Random Forest algorithm to detect and foresee the daily flowing status of three segments of the Sangone river in Piemonte, with the aim of assessing the duration and frequency of each flowing status. By evaluating the reflectance signature of sediments, vegetation, and water in the riverbed, and with the help of ground truth data and high-resolution images, it was found that the false color image with SWIR, NIR, and RED's Sentinel-2 bands allows the optimum discrimination of river water compared with other classes. Furthermore, this composition allows performing the supervised classification of the segments to determine the alternation of three flowing statuses during the years: "Flowing" (F), "Ponding" (P) and "Dry" (D). Completing the dataset with significant meteorological and hydrogeological data, a Random Forest algorithm was implemented to predict the flowing status for days with a cloudy image or no image at all in the period 2015-2021, and the Boruta package to determine the most significant explanatory variables. The outcome of the supervised classification shows an unbalanced dataset, where the image with flowing status was always more than 70% of the total. Thus, the RF model with the best prediction capacity is the oversampling double Boolean model that has firstly distinguished between F/NF (not flowing, where both D and P statuses are included) statuses, and then between D/P. The accuracy obtained for each model is in the range of 0.89-0.99. The outputs highlight an important annual variability of NF status, which goes from 0 to 166 days per year. The cumulative 30-days rainfall [R30] and 90-days rainfall [R90], in some cases also at 10 days [R10], with the average of previous 90-days-maximum air temperature [TMAX90] proved to have good predictive capabilities. Specifically, it was determined the water table level was the most significant explanatory variable to distinguish both F/NF and D/P flowing statuses. The flowrate's measurements have a significant impact in the first F/NF

model, whereas the average of the previous 30-days average relative humidity has a powerful prediction capacity in the D/P model. Furthermore, the models let to determine threshold values for some explanatory variables, which makes it possible for the body in charge of on-time monitoring and prediction of the flowing statuses of Sangone. All the results can be valuable data on fighting the deterioration of the Sangone river and the extinction of non-perennial rivers in general.

Table of contents

Abstract	3
Table of contents	5
List of figures	8
List of tables	18
1. Introduction	20
1.1 Description and characterization of non-perennial rivers	20
1.1.1 Global prevalence of non-perennial rivers and streams	20
1.1.2 The lack in social perception and in legal recognition.....	21
1.1.3 Understanding the Different Definitions of Intermittency in River Systems..	23
1.2 Classification flowing statuses and TREHS regime.....	25
1.3 How to investigate intermittency	27
2. Case study.....	30
2.1 Sangone river.....	30
2.1.1 Sangone description and characterization	30
2.1.2 Hydrologic regime.....	33
2.1.3 Geomorphological characteristics	35
2.1.4 Climate in the area and RCP scenarios for Piemonte region	36
2.1.5 Characterization of surface water body types due to WFD 2000/60/CE	39
2.1.6 Anthropogenic pressures	41
2.1.7 Its intermittency.....	42
2.2 River segmentation	45
2.2.1 Identification of the river section to investigate.....	45
2.2.2 Characterization and segmentation	46
3. Material and methods	49
3.1 Meteo-hydrogeology datasets.....	49
Gauging stations.....	49
Thiessen polygons spatial interpolation	49
Nan values management.....	50
3.2 Remote sensing datasets and field surveys	50

Sentinel-2	51
High resolution images.....	53
Method for field survey acquisition	53
3.3 Temporal scope of the study: exploring the investigated period	55
3.4 Method for the identification of the flowing statuses	57
3.5 Random Forest Classification	58
RF algorithm	58
Variables importance (Boruta package).....	59
Binary models F/NF and D/P	60
Partial Dependence Plots.....	61
3.6 Evaluating thresholds for effective flowing status monitoring	62
3.7 Regime’s type identification with TREHS metrics	63
4. Results and discussion	65
4.1 Spatial interpolation	65
4.2 Spectral signature analysis and false color images	69
4.3 Supervised classification of Sentinel 2 images.....	74
Comparison of false color images with ground truth data	74
Dataset obtained from supervised classification	79
4.4 Random Forest results.....	80
Models F/NF and D/P considering flowrate and water table level series	80
Models F/NF and D/P without flowrate and water table level series.....	83
4.5 Threshold values for effective flowing status monitoring: results and analysis	84
4.6 Duration of flowing status	87
Predictions of flowing statuses.....	87
Classification of regime type with TREHS metrics	89
Comparison of models’ prediction and classification for segments 2	91
5. Conclusion	93
ANNEX I – RF Model Interpretation: Key Charts and Metrics for Understanding Model Output.....	96
ANNEX II – Daily Flowing Status: Predicted vs Observed for 2015-2021 years ..	116

Bibliography.....124

List of figures

Figure 1. The table in figure shows the temporary rivers classification defined in D.M. 131-2008 (implementation of WFD 2000/60/CE).....	22
Figure 2. Schematic presentation of drivers and causes of flow regime alteration in rivers located in Eu-Med regions. The image is taken from the article proposed by Skoulikidis et al. (2017) “Non-perennial Mediterranean rivers in Europe: Status, pressures, and management”.....	24
Figure 3. Evolution of flowing statuses ‘classification for non-perennial rivers. This image is taken from the handbook “Intermittent Rivers and Ephemeral streams: What water managers need to know”.....	26
Figure 4. Distribution of the TREHS regime classes in the Flow-Pools-Dry plot. Qp: Quasi-perennial; AF: Alternate-Fluent; FS: Fluent-Stagnant; St: Stagnant; AS: Alternate-Stagnant; Al: Alternate; Oc: Occasional; Ep: Episodic. The three metrics (triangle altitudes) are from the bottom to the top Mp: flow permanence; from the left side to the right vertex Mf: pool permanence and from the right side to the left vertex Md: dry channel permanence. This image is taken from the handbook “Intermittent Rivers and Ephemeral streams: What water managers need to know”.	27
Figure 5. a) Hydrographic area AI 10 or Sangone basin (in orange) from Piano di Tutela delle Acque, D.C.R. n. 117-10731 del 13-03-2007. b) Piemonte sub-basin division where is possible to see the Sangone-Chisone-Lemina sub-basin (Piano di tutela delle Acque, dicembre 2018 – Tavole di Piano).....	30
Figure 6. Map of groundwater basins from Piano di Tutela delle Acque, D.C.R. n. 117-10731 del 13-03-2007.....	31
Figure 7. Map of Sangone basin that shows the characteristic hourglass shape, from Relazione di tirocinio: Caratterizzazione fisiografica , climatica e del suolo del bacino del Sangone per applicazione di modelli idrologico distribuiti di D., Elena, P. Claps, M. Graziadei 2014.....	32
Figure 8. Piemonte map with the classification of rivers network through the hydrological regime taken from “Processo di implementazione della direttiva 2000/60/CE (WFD) in Piemonte 2009”. In the red circle the Sangone river.	33
Figure 9. Flowrate's trend measured in gauging stations along Sangone river for different time.	34

Figure 10. Localization of gauging stations along Sangone river.	34
Figure 11. High resolution map of upgraded Köpper-Geiger climatic classification of Italy and focus on climate area that insists on Sangone basin.	36
Figure 12. Trends of observed annual mean-temperature during 1901-2021 period in Piemonte taken from Climate knowledge portal.	37
Figure 13. Rainfall’s Anomalies [mm]: difference between the average cumulative rainfall of 2001-2019 and 1971-2000 in Piemonte region, take from “Stato dell’ambiente in Piemonte. Relazione 2020: precipitazioni”.	37
Figure 14. Difference of the average rainy days between 2001-2019 and 1971-2000 in Piemonte region, take from “Stato dell’ambiente in Piemonte. Relazione 2020: precipitazioni”.	37
Figure 15. Graph showing the trend of average rainfall-free days per year for different altitudes, take from “Stato dell’ambiente in Piemonte. Relazione 2020: precipitazioni”.	38
Figure 16. Piemonte division through HECs.....	39
Figure 17. Upper stretch of Sangone characterization due to WFD 2000/60/CE. Its identification code is 0010811pi for the “Piano per la gestione del distretto idrografico del fiume Po and 04SS1N703PI for the regional classification.	39
Figure 18. Second stretch of Sangone characterization due to WFD 2000/60/CE. Its identificational code is 0010812pi for the “Piano per la gestione del distretto idrografico del fiume Po and 04SS2N704PI for the regional classification.	40
Figure 19. Last stretch of Sangone characterization due to WFD 2000/60/CE. Its identificational code is 0010813pi for the “Piano per la gestione del distretto idrografico del fiume Po and 06SS3F705PI for the regional classification.	40
Figure 20. Map with the anthropogenic pressures located along the Sangone river, from the IMPLEMENTAZIONE DELLA DIRETTIVA 2000/60/CE: ANALISI E VALUTAZIONE DEGLI ASPETTI IDROMORFOLOGICI. RELAZIONE SUI CORPI IDRICI ANALIZZATI NEL 2017-2018.....	42
Figure 21. Map that exhibits the “NOT-GOOD” hydrological status of the river for most of the river channel from the IMPLEMENTAZIONE DELLA DIRETTIVA 2000/60/CE: ANALISI E VALUTAZIONE DEGLI ASPETTI IDROMORFOLOGICI. RELAZIONE SUI CORPI IDRICI ANALIZZATI NEL 2017-2018.....	42
Figure 22. a) A man walks on the dry riverbed of Sangone river in Beinasco, Turin, Italy June 19, 2022. REUTERS / Massimo Pinca. b) A woman photographs the dry riverbank	

of the Sangone in Beinasco, Turin, Italy June 19, 2022. REUTERS / Massimo Pinca. c)
A man walks on the dry riverbed of Sangone river, in Beinasco Turin, Italy June 19, 2022.
REUTERS/Massimo Pinca.....44

Figure 23. Map of the river section investigated with the gauging station and water table’s
monitoring wells insisting on it.45

Figure 24. Localization of drains that insists on last part of investigated section.47

Figure 25. Map with the final river segmentation48

Figure 26. a1: example of on field survey with instruments for ponds collection on
September 16,2022. a2: example of on field survey with instruments for ponds collection
on September 11,2022.53

Figure 27. In background the 11/09/2022 satellite’s image in true color composite
visualization (RGB). Zoomed in there are two ponds acquired during the field survey: in
red the “station points” georeferenced manually on GIS software in which the laser
telemeter was located for the acquisition, in blue the points acquired to describe the edge
of the ponds, the pink grid represents the resultant ponds area on Map Stream.54

Figure 28. Different approaches to determine the best ROIs selection for ponds from field
surveys polygons (pink). a1: the ROIs (blue) are determined considering all the pixels.
a2: the ROIs (cyan) are determined considering only significant pixels.....55

Figure 29. Temporal overlapping between meteo-hydrogeology datasets and Sentinel-2
images. With Thiessen are named the variables obtained from Thiessen polygons spatial
interpolation.....56

Figure 30. Scheme of a decision tree.....58

Figure 31. New catchment area compared with the original one.65

Figure 32. New catchment area divided through Thiessen polygons for rainfall measures.
.....65

Figure 33. New catchment area divided through Thiessen polygons for temperature
measures.66

Figure 34. New catchment area divided through Thiessen polygons for relative humidity
measures.67

Figure 35. a) Trend of a spatial interpolated rainfall obtained from Thiessen polygons
method. b) Trend of a spatial temperature (max, mean, min) obtained from Thiessen
polygons method. c) Trend of a spatial interpolated relative humidity (max, mean, min)
obtained from Thiessen polygons method.....68

Figure 36. Graphs with the spectral signatures for the different classes. The signatures are related to the images corresponding to the Google Earth Pro source on late winter. a: spectral signatures for 02/03/2022 image. a: spectral signatures for 07/03/2022 image. 70

Figure 37. Graphs with the spectral signatures for the different classes. The signatures are related to the images corresponding to field surveys source on late summer. c: spectral signatures for 11/09/2022. d: spectral signatures for 16/09/2022 image. a: spectral signatures for 03/10/2022 image.71

Figure 38. The final graph is the resultant from the average of the five spectral signatures of each class. For Sangone flowing statuses class, the classes used were flow from case 1 and 2 and pools from case 3-4-5.....72

Figure 39. a: High resolution image from Google Earth Pro 03/03/2022. b: RGB Sentinel image 02/03/2022. c: FCI Sentinel image 02/03/2022.73

Figure 40. Comparison between FCIs images and field surveys for different date: A is the FCI on 11/09/2022 and a1, a2, a3 are photo taken on field on 11/09/2022; B is the FCI on 16/09/2022 and b1, b2 are photo taken on 16/09/2022; C is the FCI on 03/10/2022 and c1, c2, c3 are photo taken on 03/10/2022. The FCIs images are in scale 1:10000, the same used for the supervised classification. The portion of Sangone river investigated in this figure corresponds to the segment 1.75

Figure 41. Comparison between FCIs images and field surveys for different date: D is the FCI on 11/09/2022 and d1, d2, d3 are photo taken on 11/09/2022; E is the FCI on 16/09/2022 and e1, e2, e3 are photo taken on 16/09/2022; F is the FCI on 03/10/2022 and f1, f2, f3 are photo taken on 03/10/2022. The FCIs images are in scale 1:10000, the same used for the supervised classification. The portion of Sangone river investigated in this figure corresponds to the portion excluded due to the anthropogenic alteration of the flows.77

Figure 42. Comparison between FCIs images and field surveys that underlines the possible misclassification of water presence with shadows in the riverbed. G is the FCI on 03/10/2022 and g1, g2 and g3 are photo taken on 03/10/2022 during the field survey. H is the FCI image on 02/03/2022 and I the Google Earth Image on 02/03/2022.78

Figure 43. Using of RIVALTA_EXSTAB_WTL and TRANA_FLOWRATE to define threshold values for underline not flowing and flowing area. In green the uncertainty zone where there is the overlapping of statuses: a1 flowrate thresholds for segment 1, a2 water table level thresholds for segment 1, b1 flowrate thresholds for segment 2, b2 water table

level thresholds for segment 2, c1 flowrate thresholds for segment 3 and c2 water table level thresholds for segment 3.....	84
Figure 44. Scatterplots of RIVALTA_EXSTAB_WTL and TRANA_FLOWRATE to visualize the not-flowing area and the uncertainty zones for each segment: d1 segment 1, d2 segment2 and d3 segment 3. In cyan the not flowing area and in green the uncertainty zones.	85
Figure 45. Duration of flowing statuses per year for each segment. a: segment 1, b: segment 2, c: segment 3.....	87
Figure 46. Comparison between resultant flowing statuses durations of segment 2 using flowrate and water table level as predictor variables or not. a: model with flowrate and water table level as predictors, b: model without flowrate and water table level as predictors.	91
Figure 47. a: Results for Spearman analysis to defines the evaluate the correlation between variables. b: definition of variable importance through Boruta algorithm.	96
Figure 48. Partial dependence plots for TRANA_FLOWRATE variable: c1, flowing status; c2, not-flowing status.	96
Figure 49. Partial dependence plots for RIVALTA EXSTAB WTL variable: d1, flowing status; d2, not-flowing status.	97
Figure 50. Partial dependence plots for T MAX MEAN CUM 90 variable: e1, flowing status; e2, not-flowing status.	97
Figure 51. Partial dependence plots for RAINFALL CUM 10 variable: f1, flowing status; f2, not-flowing status.....	97
Figure 52. Partial dependence plots for RAINFALL CUM 30 variable: g1, flowing status; g2, not-flowing status.	98
Figure 53. Confusion matrix of F/NF model for segment 1 through which is possible to visualize the accuracy, the sensitivity and the specificity of the classification. 1 is the not flowing status and 3 is the flowing one.	98
Figure 54. a: Results for Spearman analysis to defines the evaluate the correlation between variables. b: definition of variable importance through Boruta algorithm.	99
Figure 55. Partial dependence plots for TRANA_FLOWRATE variable: c1, ponding status; c2, dry status.....	99
Figure 56. Partial dependence plots for RIVALTA_EXSTAB_WTL variable: d1, ponding status; d2, dry status.....	99

Figure 57. Partial dependence plots for T MAX MEAN CUM 90 variable: e1, ponding status; e2, dry status.....	100
Figure 58. Partial dependence plots for RAINFALL CUM 10 variable: f1, ponding status; f2, dry status.	100
Figure 59. Partial dependence plots for H AVR MEAN CUM 30 variable: g1, ponding status; g2, dry status.....	100
Figure 60. Confusion matrix of F/NF model for segment 1 through which is possible to visualize the accuracy, the sensitivity and the specificity of the classification. 1 is the dry status and 2 is the ponding one.	101
Figure 61. a: Results for Spearman analysis to defines the evaluate the correlation between variables. b: definition of variable importance through Boruta algorithm.	102
Figure 62. Partial dependence plots for TRANA_FLOWRATE variable: c1, flowing status; c2, not-flowing status.	102
Figure 63. Partial dependence plots for RIVALTA_EXSTAB_WTL variable: d1, flowing status; d2, not-flowing status.	103
Figure 64. Partial dependence plots for T MAX MEAN CUM 90 variable: e1, flowing status; e2, not-flowing status.	103
Figure 65. Partial dependence plots for RAINFALL CUM 30 variable: f1, flowing status; f2, not-flowing status.	103
Figure 66. Partial dependence plots for RAINFALL CUM 90 variable: g1, flowing status; g2, not-flowing status.	104
Figure 67. Confusion matrix of F/NF model for segment 2 through which is possible to visualize the accuracy, the sensitivity and the specificity of the classification. 1 is the not flowing status and 3 is the flowing one.	104
Figure 68. a: Results for Spearman analysis to defines the evaluate the correlation between variables. b: definition of variable importance through Boruta algorithm.	105
Figure 69. Partial dependence plots for H_AVR_MEAN_CUM_30 variable: c1, ponding status; c2, dry status.....	105
Figure 70. Partial dependence plots for RIVALTA_EXSTAB_WTL variable: d1, ponding status; d2, dry status.....	105
Figure 71. Partial dependence plots for T MAX MEAN CUM 90 variable: e1, ponding status; e2, dry status.....	106
Figure 72. Partial dependence plots for RAINFALL CUM 30 variable: f1, ponding status; f2, dry status.	106

Figure 73. Confusion matrix of D/P model for segment 2 through which is possible to visualize the accuracy, the sensitivity and the specificity of the classification. 1 is the dry status and 2 is the ponding one.	106
Figure 74. a: Results for Spearman analysis to defines the evaluate the correlation between variables. b: definition of variable importance through Boruta algorithm.	107
Figure 75. Partial dependence plots for TRANA_FLOWRATE variable: c1, flowing status; c2, not-flowing status.	107
Figure 76. Partial dependence plots for RIVALTA_EXSTAB_WTL variable: d1, flowing status; d2, not-flowing status.	108
Figure 77. Partial dependence plots for T_MAX_MEAN_CUM_90 variable: e1, flowing status; e2, not-flowing status.	108
Figure 78. Partial dependence plots for RAINFALL_CUM_30 variable: f1, flowing status; f2, not-flowing status.	108
Figure 79. Partial dependence plots for RAINFALL_CUM_90 variable: g1, flowing status; g2, not-flowing status.	109
Figure 80. Confusion matrix of F/NF model for segment 3 through which is possible to visualize the accuracy, the sensitivity and the specificity of the classification. 1 is the not flowing status and 3 is the flowing one.	109
Figure 81. a: Results for Spearman analysis to defines the evaluate the correlation between variables. b: definition of variable importance through Boruta algorithm.	110
Figure 82. Partial dependence plots for H_AVR_MEAN_CUM_30 variable: c1, ponding status; c2, dry status.	110
Figure 83. Partial dependence plots for RIVALTA_EXSTAB_WTL variable: d1, ponding status; d2, dry status.	110
Figure 84. Partial dependence plots for T_MAX_MEAN_CUM_90 variable: e1, ponding status; e2, dry status.	111
Figure 85. Partial dependence plots for RAINFALL_CUM_30 variable: f1, ponding status; f2, dry status.	111
Figure 86. Confusion matrix of D/P model for segment 3 through which is possible to visualize the accuracy, the sensitivity and the specificity of the classification. 1 is the dry status and 2 is the ponding one.	111
Figure 87. a: Results for Spearman analysis to defines the evaluate the correlation between variables. b: definition of variable importance through Boruta algorithm.	112

Figure 88. Partial dependence plots for T MAX MEAN CUM 90 variable: c1, flowing status; c2, not-flowing status.	112
Figure 89. Partial dependence plots for RAINFALL CUM 30 variable: d1, flowing status; d2, not-flowing status.	112
Figure 90. Partial dependence plots for RAINFALL CUM 90 variable: e1, flowing status; e2, not-flowing status.	113
Figure 91. Confusion matrix of the second F/NF model for segment 3 through which is possible to visualize the accuracy, the sensitivity and the specificity of the classification. 1 is the not flowing status and 3 is the flowing one.	113
Figure 92. a: Results for Spearman analysis to defines the evaluate the correlation between variables. b: definition of variable importance through Boruta algorithm.....	114
Figure 93. Partial dependence plots for H_AVR_MEAN_CUM_30 variable: c1, ponding status; c2, dry status.....	114
Figure 94. Partial dependence plots for RAINFALL CUM 30 variable: d1, ponding status; d2, dry status.....	114
Figure 95. Partial dependence plots for RAINFALL CUM 90 variable: e1, ponding status; e2, dry status.	115
Figure 96. Confusion matrix of the second D/P model for segment 2 through which is possible to visualize the accuracy, the sensitivity and the specificity of the classification. 1 is the dry status and 2 is the ponding one.	115
Figure 97. Evolution of observed and predicted flowing status for segment 1 during 2015.	116
Figure 98. Evolution of observed and predicted flowing status for segment 1 during 2016.	116
Figure 99. Evolution of observed and predicted flowing status for segment 1 during 2017.	116
Figure 100. Evolution of observed and predicted flowing status for segment 1 during 2018.	117
Figure 101. Evolution of observed and predicted flowing status for segment 1 during 2019.	117
Figure 102. Evolution of observed and predicted flowing status for segment 1 during 2020.	117
Figure 103. Evolution of observed and predicted flowing status for segment 1 during 2021.	117

Figure 104. Evolution of observed and predicted flowing status for segment 2 during 2015.....	118
Figure 105. Evolution of observed and predicted flowing status for segment 2 during 2016.....	118
Figure 106. Evolution of observed and predicted flowing status for segment 2 during 2017.....	118
Figure 107. Evolution of observed and predicted flowing status for segment 2 during 2018.....	119
Figure 108. Evolution of observed and predicted flowing status for segment 2 during 2019.....	119
Figure 109. Evolution of observed and predicted flowing status for segment 2 during 2020.....	119
Figure 110. Evolution of observed and predicted flowing status for segment 2 during 2021.....	119
Figure 111. Evolution of observed and predicted flowing status for segment 3 during 2015.....	120
Figure 112. Evolution of observed and predicted flowing status for segment 3 during 2016.....	120
Figure 113. Evolution of observed and predicted flowing status for segment 3 during 2017.....	120
Figure 114. Evolution of observed and predicted flowing status for segment 3 during 2018.....	121
Figure 115. Evolution of observed and predicted flowing status for segment 3 during 2019.....	121
Figure 116. Evolution of observed and predicted flowing status for segment 3 during 2020.....	121
Figure 117. Evolution of observed and predicted flowing status for segment 3 during 2021.....	121
Figure 118. Evolution of observed and predicted flowing status for segment 2 during 2015, second model.....	122
Figure 119. Evolution of observed and predicted flowing status for segment 2 during 2016, second model.....	122
Figure 120. Evolution of observed and predicted flowing status for segment 2 during 2017, second model.....	122

Figure 121. Evolution of observed and predicted flowing status for segment 2 during 2018, second model.	123
Figure 122. Evolution of observed and predicted flowing status for segment 2 during 2019, second model.	123
Figure 123. Evolution of observed and predicted flowing status for segment 2 during 2020, second model.	123
Figure 124. Evolution of observed and predicted flowing status for segment 2 during 2021, second model.	123

List of tables

Table 1. Gauging stations along Sangone river.....	33
Table 2. Monthly flowrate values for 2002-2021 years in Trana’s gauging stations.....	35
Table 3. Monthly flowrate values for 2021 year in Trana’s gauging stations.....	35
Table 4. Characteristics of Sentinel-2 MSI.	52
Table 5. Threshold values for Mf, Mp and Md for each different regime.	63
Table 6. Area and relative weight respect the overall catchment area for each rainfall gauging stations.....	66
Table 7. Area and relative weight respect the overall catchment area for each temperature gauging stations.....	66
Table 8. Area and relative weight respect the overall catchment area for each relative humidity gauging stations.....	67
Table 9. Resume of ground truth source, date and related Sentinel-2 image.....	69
Table 10. The table reports the dimension of the pools acquired during the field surveys related to figure 40.....	76
Table 11. The table reports the dimension of the pools acquired during the field surveys related to figure 41.....	78
Table 12. Results of supervised classification.....	79
Table 13. Trend of revisit time during the years.	79
Table 14. Characterization of binary models, considering flowrate and water table level measures, for each segments through selected most significant variables (in order of importance from top-down direction) and statistical parameters (accuracy, sensitivity, specificity and TSS).....	81
Table 15. Characterization of binary models, without flowrate and water table level measures, for segment 2 through selected most significant variables (in order of importance from top-down direction) and statistical parameters (accuracy, sensitivity, specificity and TSS).....	83
Table 16. The table reports the significant values of flowrate and water table level underlined in figure 43.	85
Table 17. The table reports the threshold values that define the Not-flowing area and the uncertainty zones.	86
Table 18. Regime classification of segment 1 through TREHS classification.....	89
Table 19. Regime classification of segment 2 through TREHS classification.....	90

Table 20. Regime classification of segment 3 through TREHS classification.....90

Table 21. Comparison TREHS classification between models on segment 2 with and without flowrate and water table level as predictive variables. The last column is the precedent classification shown in table 13 from the model with the original datasets. In yellow are underlined year with different classification.92

1. Introduction

1.1 Description and characterization of non-perennial rivers

1.1.1 Global prevalence of non-perennial rivers and streams

A non-perennial river or stream is a freshwater body that does not flow continuously, but rather only during a certain period of the year or in response to events such as rainfall or snowmelt. It is important to clarify that non-perennial rivers are all rivers in which some data testify to the absence of water along the riverbed. This definition includes a wide range of different intermittencies: from the river with episodic water presence to quasi-perennial rivers.¹ These rivers were usually located in regions with a semi-arid or arid climate, where the dry season is longer each year than in other regions.^{2,3} However, nowadays, it is incorrect to figure out whether this phenomenon is localized only in particular climate regions.

All the newest research that has tried to quantify the global presence of non-perennial rivers around the world gives back an estimation for which the condition of intermittency, as defined before, is the most common. Some estimates report a percentage of non-perennial rivers around 41% for rivers with MAF (mean annual flow) > 1 . If low-order rivers with MAF between 1 are included, the percentage goes between 51% (conservative approach) to 60%.⁴ These results return a rethinking of the general concept of rivers for which the most common condition is to be temporary and not perennial, as common perception might lead us to think.

Moreover, the separation between perennial and non-perennial is not fixed in time. Anthropogenic pressures, such as climate change,^{5,6} changes in land use and water withdrawals define a significant shift from perennial to non-perennial conditions. The restoration of precedent conditions is a remote possibility.⁷ Also, in the last 50 years many of the biggest and most famous rivers flow, such as the Nile, Yellow, Indus, and Colorado, that used to run continuously, have started to have stretches where water ceases to flow.⁸ Last summer, it became commonly known in Italy that this phenomenon also affected the peninsula's most prominent and longest river, the Po.

Furthermore, increased evaporation and decreased precipitation are direct effects of the rising global temperatures and the result of an intermittent river is that water will flow less frequently or ceases to exist entirely. Additionally, many frequent droughts and heatwaves can make an intermittent river's water flow more unpredictable and unreliable.

Also, the timing and intensity of precipitation are affected by climate change, which alters the timing of snowmelt and runoff that feed the river and can mess with an intermittent river's natural flow patterns.⁹ Communities that depend on the river for resources such as water, food, and other resources as well as the ecosystem could be significantly touched by these changes.

The risk of wildfires, which can obliterate the vegetation that helps the river's watershed retain water, can also rise as a direct consequence of climate change.¹⁰ As a result, the river may experience increased sedimentation and erosion as well as a decrease in water flow.

Concerning safeguarding intermittent rivers, communities and ecosystems that depend on them, it is essential to take steps to lessen the effects of climate change and to adapt to the changes that are already taking place.

1.1.2 The lack in social perception and in legal recognition

Despite the burgeoning literature on the theme in the last years, the lack of consciousness on the importance of the non-perennial rivers for the river network,^{11,12} their role in aquifer regulation, their contribution to local and regional biodiversity and biogeochemical integrity,¹ results in a subordination of the attention of national legislation and policy to the protection of these systems, which still lags rather than given to perennial rivers.^{1,4}

The literature has shown how the social perception of a river's importance is closely related to the permanence of flow.¹¹ Sociocultural difficulties still exist in including a dry riverbed in the definition of a river. In addition, devaluation occurs in communities due to reduced recreational and economic value when water stops to flow. Moreover, fishers

are important stakeholders that address the politics of restoration and preservation more on perennial reaches where biota species are more present.¹² Furthermore, especially in an arid and semi-arid region, there is a widespread presence of ephemeral streams that are immediately reconnected to the dangers of flooding during the rainfall season.¹ Only in recent years has the growing attention on non-perennial rivers led researchers to investigate new indicators to evaluate their ecosystem service toward the environment and the society.¹³

All these cultural biases on non-perennial rivers determine legislation that still lags in updating the management and protection of non-perennial rivers.^{4,14} These rivers fall under the same regulations as perennial rivers (e.g., the Water Framework Directive in the EU, the Water Act in the U.S.)¹² or may not be considered a river at all and their management is specifically evaluated.¹⁵

In Europe, WFD 200/60/EC determines an important update in river conservation by defining reference conditions (RC) for each type of water body through which to compare and evaluate the actual biological, hydro-morphological and ecological status of European rivers.¹⁶ For non-perennial rivers, there is no robust method for assessing RCs and thus, all these factors affect the regulatory gap for non-perennial rivers. As a result, non-perennial rivers are not mentioned in the WFD, but each member state could include them in national jurisdiction.

In Italy, law 131-2008 (implementation of WFD) defines a classification of non-perennial rivers (figure 1). The bill establishes that all rivers with a catchment area less than 10 km² must be included in the “River Basin Management Plan” (RBMP) except the episodic streams that carry water less than once every 5 years.^{15,17} The episodic streams are, in the end, excluded from WFD and devaluated at water bodies without an ecologic relevance for the societies.

Stream type	Description
Temporary	Watercourse that can dry out completely and/or at some stretches
• Intermittent	Water is present more than 8 months a year. It may dry out in some river stretches and/or several times a year
• Ephemeral	Water is present <8 months a year. Disconnected pools may remain
• Episodic	Water only present after heavy rains, once every 5 years

Figure 1. The table in figure shows the temporary rivers classification defined in D.M. 131-2008 (implementation of WFD 2000/60/CE).

1.1.3 Understanding the Different Definitions of Intermittency in River Systems

Scientific research will be able to successfully link different fields of research if the communication of research is done consistently and using clear definitions for key terms.

Thus, it was essential to define a universal and commonly accepted classification of the different types of intermittencies that a non-perennial river can present. Over the years, many epithets were used to refer to some kinds of the non-perennial river like “arid, discontinuous, dry, ephemeral, episodic, intermittent, interrupted, irregular, non-perennial, non-permanent, seasonal, and temporary”.

*Busch et al.*² proposed a review of most used epithets for different kind of non-perennial river to define universal and general definitions:

- *Non-perennial*: any lotic, freshwater system that periodically ceases to flow and/or is dry at some point in time and/or space.
- *Intermittent*: a non-perennial river or stream with a considerable connection to the groundwater table, having variable cycles of wetting and flow cessation, and with flow that is sustained longer than a single storm event. These waterways are hydrologically gaining most of the time when considering long term flow patterns.
- *Ephemeral*: a type of non-perennial river or stream without a considerable groundwater connection that flows for a short period of time, typically only after precipitation events. These waterways are hydrologically losing most of the time when considering long term flow patterns.

Moreover, to safeguard the non-perennial river network, it is crucial to have consciousness of the causes that determine the intermittency, not only the type of its intermittency.

Despite the attention on the intermittent river still arising in multiple research fields, the focus remains mainly on rivers and streams where the alternation between flowing and not flowing conditions is a characteristic of the natural hydrological regime.

To have complete knowledge of the non-perennial rivers network, it is essential taking into account the classification between natural flow intermittence (NFI) and anthropogenic flow intermittence (AFI).¹⁰ In figure 2 a schematic of anthropogenic drivers that can affect flow regime.

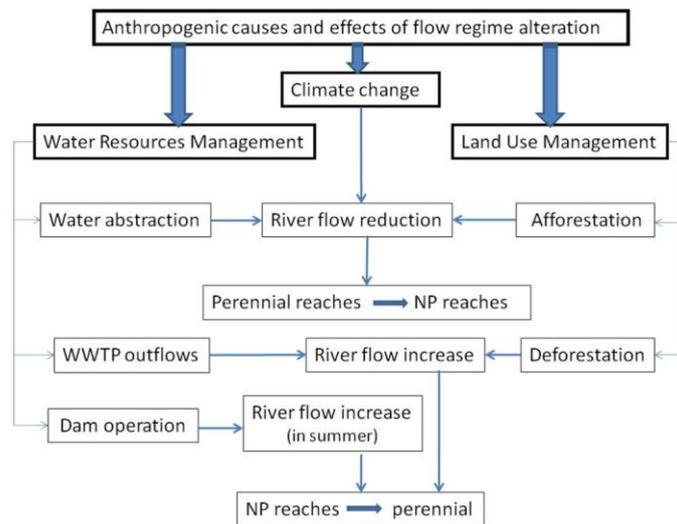


Figure 2. Schematic presentation of drivers and causes of flow regime alteration in rivers located in Eu-Med regions. The image is taken from the article proposed by Skoulikidis et al. (2017) "Non-perennial Mediterranean rivers in Europe: Status, pressures, and management".

It is not always easy to assess the difference between anthropic and natural drivers which determines the intermittency of a river, especially when the results are a decrease in flow discharge.¹⁸

On the other hand, the unnatural flow increases, when non-perennial rivers become perennials, which is a condition typically determined by anthropogenic drivers, like Wastewater Treatment Plants (WWTPs) or urban, civil and industrial discharges.¹⁹ In that case, even if the general perception could figure out a positive effect, especially during years with the frequency arise of drought periods, the alteration of the natural flow regime and its characteristics (timing, frequency, seasonality) could enhance ecological changes such as the increase of non-native species.²⁰

1.2 Classification flowing statuses and TREHS regime

The capacity to define a general and valid classification of different flowing statuses of non-perennial rivers was a fundamental step to be able to study their evolution, determine their ecology status and draw specific guidelines for their management.

Their main characteristic is their high hydrologic variability both on a spatial and temporal scale. The flowing status can vary during the year and along the river channel from the source to its end. Moreover, the temporariness processes determine a spatial variability that considers both longitudinal and lateral directions.

The difficulty of describing the flow regime of a non-perennial river, the lack of longitudinal connectivity and the greater importance that the perennial river still holds for the water manager determine the situation whereby non-perennial rivers, especially non-perennial stretches, are not surveyed.²¹

Nevertheless, the gauging stations for the flow's measurements are point measurements from which is difficult to extract valid information for upper or downstream reaches. In addition, their installation and maintenance are characterized by high costs and can be limited in detecting extremely low flows or detecting side pools.²²

Considering these limits, field surveys as the best method to evaluate the real-time condition of a non-perennial river.¹ Moreover, the need for a large amount of fields survey promotes citizen science projects and crowdsourcing, like the project "CrowdWater",²³ as possibilities to enlarge the possibilities of field data. On the other hand, involving the population is important to increase awareness of climate change and the limits of water availability.^{24,25} Thus, the classification based more on a qualitative than a quantitative approach of standardization of different statuses was more suitable for this method.

Over the years, different approaches (figure 3) were proposed to standardize the possible different conditions in which the riverbed of non-perennial rivers could appear.^{1,23,26,27}



Figure 3. Evolution of flowing statuses 'classification for non-perennial rivers. This image is taken from the handbook "Intermittent Rivers and Ephemeral streams: What water managers need to know".

The six aquatic states can describe all the different possible conditions that can occur during a field survey of a non-perennial river. However, the main disadvantage was their great accuracy in classification, information that was impossible to obtain at this level of detail in years past.

Thus, Gallart et al., (2017)²⁷ proposed a more simplified classification base on three different aquatic phases (dry, pools, and flow) or flowing statuses (dry, ponding, flowing).

Both aquatic phases and flowing statuses emphasized the importance of the intermediate condition that occurs when the water present in the river channel is only in non-interconnected pools. This status could be a simple transient phase between dry and

flowing but, especially when there is an important exchange with the groundwater, could be stable.

In the general definition of different non-perennial rivers, the ponding status was not considered. Thus, *Gallart et al.*, (2017),²⁷ through the development of a free software tool TREHS (Temporary Rivers Ecological and Hydrological Status), determines nine different regimes considering both the three statuses metrics and the permanence of each status during the year. This final classification allows taking into account both spatial and temporal variability of non-perennial rivers. In figure 4 are reported the TREHS chart with threshold values of flowing, ponding and dry permanence (Mf, Mp, Md) that determine the passage from a regime to another.

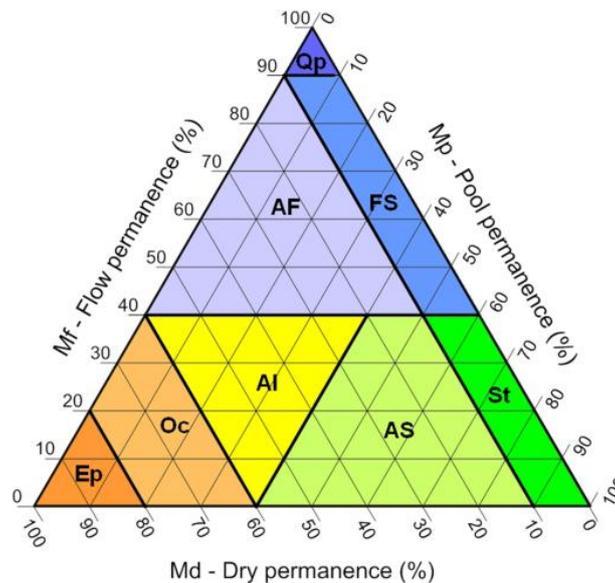


Figure 4. Distribution of the TREHS regime classes in the Flow-Pools-Dry plot. Qp: Quasi-perennial; AF: Alternate-Fluent; FS: Fluent-Stagnant; St: Stagnant; AS: Alternate-Stagnant; Al: Alternate; Oc: Occasional; Ep: Episodic. The three metrics (triangle altitudes) are from the bottom to the top Mf: flow permanence; from the left side to the right vertex Mp: pool permanence and from the right side to the left vertex Md: dry channel permanence. This image is taken from the handbook “Intermittent Rivers and Ephemeral streams: What water managers need to know”.

1.3 How to investigate intermittency

The high spatial and temporal variability defines pivotal obstacles to studying non-perennial river intermittency. To have clear patterns of river intermittency is fundamental to know the possible flowing statuses that rivers can present in different periods, the frequency, the duration, and the seasonality of these changes.

Furthermore, it is possible to understand the evolution of intermittency over time and the alteration of the natural flow regime only by comparing real-time data with a long-term dataset or natural that can be used as RC of the river.¹

Over the years, several methods and approaches have developed to study the evolution and characterize the regime and the intermittency of non-perennial rivers.

Field surveys

Field surveys are one of the best solutions for acquiring real-time data on the metrics of the flow regime and can assess with high accuracy which flowing status is present on time.¹ The main obstacles are the limited possibilities to replicate the campaigns. Citizen science and crowdsourcing could be viable alternatives when the non-perennial stretches are easy to reach.

Gauging stations

Gauging stations are the best solution to obtain long-term data and evaluate the possible evolution of flow discharge over the years. The main disadvantages are the rare presence of gauging stations in non-perennial stretches and their problem of measuring small flows or ponding status.²²

Logger sensors

Another method to obtain real-time and medium-term data are field loggers that can measure water temperature, electrical conductivity, or both of them.²⁸ These instruments may detect the movement of wetting and drying fronts²⁹ but could have difficulties distinguishing between flowing and standing water. In addition, the drawback is the possibility of instruments being swept away or buried during floods or their integrity being compromised by vandalism.¹

Environmental Dna

Environmental Dna (or eDna) found an application evaluating non-perennial rivers highlighting the exchange between surface water and groundwater.³⁰

Hydrologic modeling

Actual hydrologic models are still biased in predicting variability of flow discharge in non-perennial rivers. They overestimate zero-flow events and still lack predicting the spatial variability of flowing statuses.

Remote sensing

Remote sensing has defined significant opportunities in monitoring the conditions of the non-perennial river. The airborne surveys allow executing rapid and extended surveys on intermittent reaches, even if the riverbed is complicated to reach personally.³⁰ The satellite images can return periodical, sometimes with revisit time shorter than a week, multispectral image of the entire river network, through which is possible to monitor constantly the evolution of aquatic phases.^{31,32} The main drawback of the satellite image is the spatial resolution, at a maximum of 10m, which hinders the application for narrow rivers and streams.³³

The literature has defined the combination of these approaches as the best option and the choice usually depends on the method's limits and characteristics of non-perennial rivers. The variety of possible diverse conditions (type, regime, stable flowing status, dimension, anthropogenic causes) that can differentiate non-perennial rivers requires a priori analysis through which the expert could define the best possible fitting method.

For this research, following the combined approach applied by al,*Cavallo et al*,³¹ on several rivers in Campania, it was used a combination method between field survey, remote sensing analysis (high-resolution image and Sentinel-2 archive), and a machine learning algorithm (Random forest) to predict the flowing statuses of Sangone river on a daily scale.

2. Case study

2.1 Sangone river

2.1.1 Sangone description and characterization

The Sangone river is a large river located in Piemonte, it starts from the Cottian Alps (the source is Fontana Mura), 2000 m above the sea level and flows through the Val Sangone, between the Val di Susa (north) and the Val Chisone (south) until reaching the PO river a Moncalieri, 220 m above the sea level. The river's watercourse is long around 48 km, the average slope is 3.4% and the catchment area is 258 km².^{34,35} The Sangone river was classified as part of the AI 10 "Hydrographic Area"³⁶ and then regrouped in the sub-basin Sangone-Chisola-Lemina.^{37,38} In the map in figure 5a is possible to see the localization of the Sangone basin in the regional river network and figure 5b the sub-basin division of Piemonte.

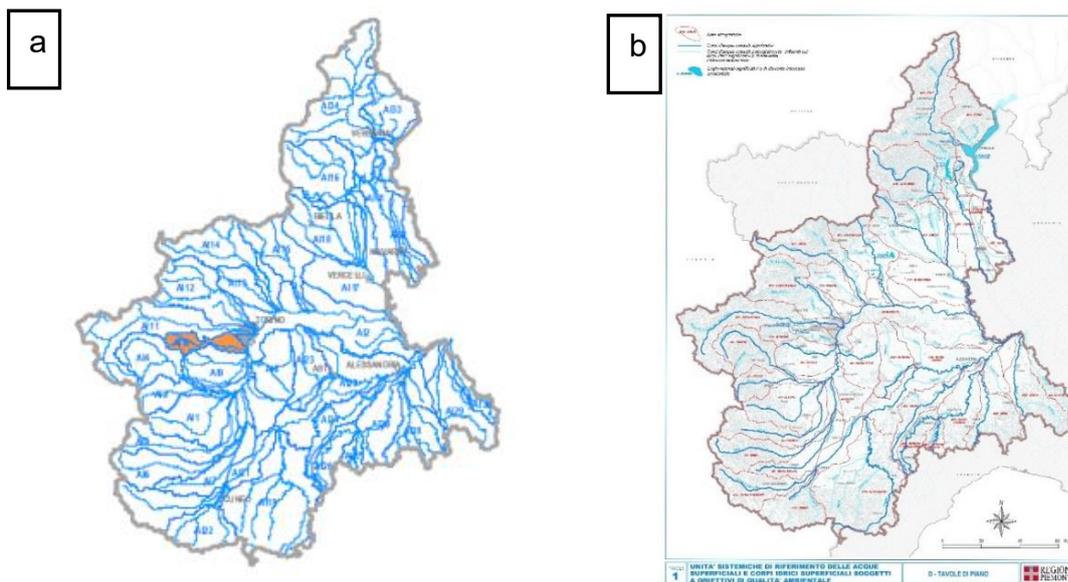


Figure 5. a) Hydrographic area AI 10 or Sangone basin (in orange) from Piano di Tutela delle Acque, D.C.R. n. 117-10731 del 13-03-2007. b) Piemonte sub-basin division where is possible to see the Sangone-Chisola-Lemina sub-basin (Piano di tutela delle Acque, dicembre 2018 – Tavole di Piano).

According to the lithologic units' classification of the region, a predominant siliceous geological composition characterizes the Sangone catchment area.³⁹ Furthermore, the river exhibits a relevant exchange with the groundwater basin present in the lower area of the catchment area as can be seen in figure 6. The aquifer is, with the melting of snow

deposits and meteorological phenomena, a fundamental recharge for surface water bodies.³⁶

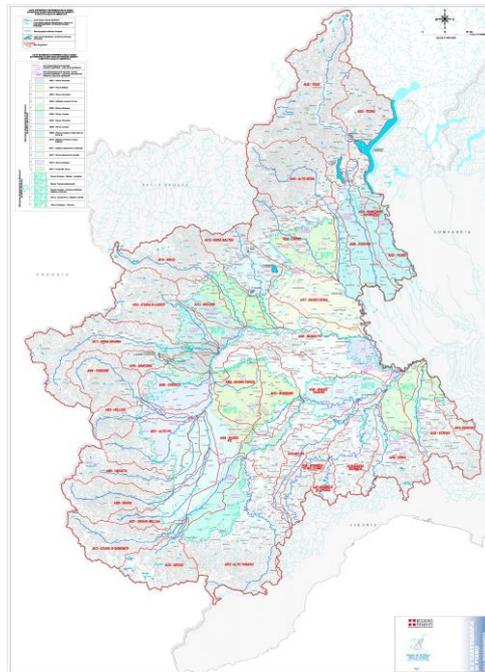


Figure 6. Map of groundwater basins from Piano di Tutela delle Acque, D.C.R. n. 117-10731 del 13-03-2007.

The catchment area is characterized by a densely-distributed drainage network (above 3.2 km/km⁻²). Especially, the river presents a widespread presence of tributaries, mostly in the upper valley.

The principal ones are:

- from hydrographic left:
 - Rio Ricciavrè, Forno di Coazze (905 m s.l.m.);
 - Torrente Sangonetto, at Sangonetto (699 m s.l.m);
 - Rio Ollasio at Trana (400 m s.l.m);
- from hydrographic right:
 - Rio Tronera at Pontepietra (598 m s.l.m.);
 - Rio Romarolo: at Giaveno (490 m s.l.m.).

There are other minor streams tributaries like rii Costabruna, Arpone, Casasse, Tovalera, Maleselle Balma.

The river basin presents an hourglass shape in which the bottleneck is at Trana how it is underline by the figure 7.³⁵



Figure 7. Map of Sangone basin that shows the characteristic hourglass shape, from Relazione di tirocinio: Caratterizzazione fisiografica , climatica e del suolo del bacino del Sangone per applicazione di modelli idrologico distribuiti di D., Elena, P. Claps, M. Graziadei 2014.

The river changes its characteristics before and after this bottleneck, reflecting the differences in the geological, hydrological, and human influences along its course.

Before, the Sangone River typically flows through a more mountainous and rural landscape, with a steeper gradient and higher water flow rate. The river may be narrower and shallower in this section, with more rapid changes in water level and flow rate. Moreover, all the main tributaries flow into the Sangone ahead of Trana. Additionally, the river in this section may be surrounded by a more natural landscape, with less human influence on the riparian zone.

After Trana, the Sangone River typically flows through a more urbanized and industrialized landscape, a typical flatland with a gentler gradient and lower water flow rate. The river may be wider and deeper in this section, with more stable water levels and flow rates. Additionally, the river in this section may be surrounded by a more developed landscape, with greater human influence on the riparian zone, such as concrete channels, water withdrawals and wastewater discharge.

2.1.2 Hydrologic regime

The map of the hydrologic regimes of the Piemonte region highlights how the Sangone changes its hydrologic regime along its course.³⁹

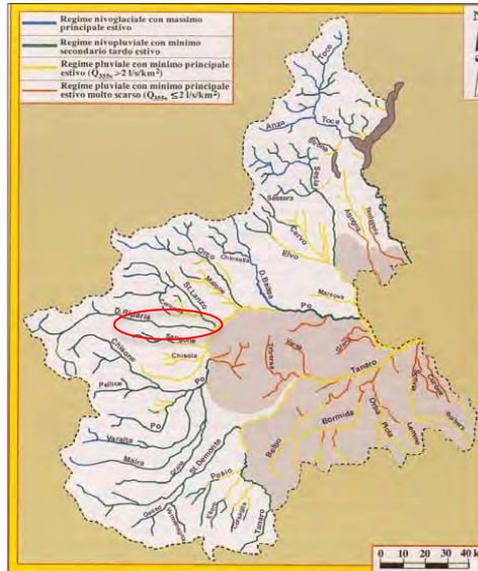


Figure 8. Piemonte map with the classification of rivers network through the hydrological regime taken from “Processo di implementazione della direttiva 2000/60/CE (WFD) in Piemonte 2009”. In the red circle the Sangone river.

The map in figure 8 shows how the first part of the river is classified as nivo-pluvial with the secondary minimum in the last part of summer and the second one in the pluvial regime with a principal minimum in summer.

Along the river, there are two active gauging stations in Trana and in Torino from the ARPA Piemonte network that gives us information about the Sangone flowrate. Moreover, there are also old measurements from two deactivated stations in Rivalta and Moncalieri.⁴⁰ Table 1 reports all the gauging stations for flowrate.

Table 1. Gauging stations along Sangone river.

Gauging station	Municipality	Org.	ARPA Cod	Registration starting year	Act/Dis
TRANA	Trana	ARPA	278	2002	Act
TORINO	Torino	ARPA	S7421	2015	Act
MONCALIERI	Moncalieri	ARPA	365	2005	Dis
RIVALTA	Rivalta di Torino	CMTO	-	2014	Dis

In figure 9 is possible to see the temporal series of flowrate and in figure 10 the location of Sangone gauging stations.

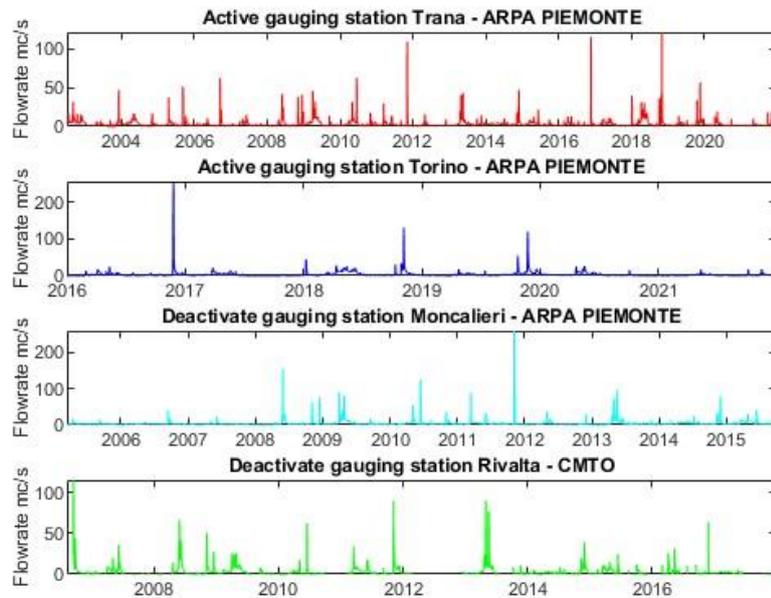


Figure 9. Flowrate's trend measured in gauging stations along Sangone river for different time.

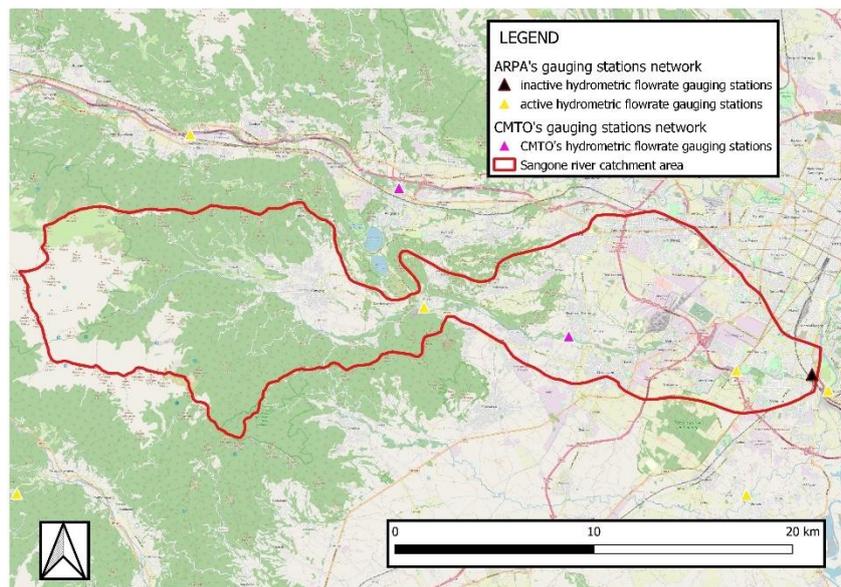


Figure 10. Localization of gauging stations along Sangone river.

The monthly average flowrate during years, presented in table 2, lets for investigating the seasonal variation of the flowrate during the year.⁴⁰ In table 3, there are the flowrate during the 2021, the last year with validated data.

Table 2. Monthly flowrate values for 2002-2021 years in Trana's gauging stations.

Parameter	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dic
QMIN [m ³ /s]	0.42	0.51	0.4	0.33	1.37	0.7	0.01	0.01	0.01	0.01	0.36	0.38
QAVR [m ³ /s]	1.85	1.69	2.57	5.41	6.84	4.54	1.99	1.33	2.35	3.18	5.57	2.78
QMAX [m ³ /s]	39.6	7.41	29.9	45.8	43.5	63	17.2	14.3	62.5	39.2	122	47.3

Table 3. Monthly flowrate values for 2021 year in Trana's gauging stations.

Parameter	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dic
QMIN [m ³ /s]	0.74	1.05	0.87	1.05	1.62	1.5	1.28	0.97	0.91	1.93	2.03	1.16
QAVR [m ³ /s]	1.07	1.21	0.99	1.45	3.06	1.9	1.54	1.13	1.04	3.34	4.82	1.49
QMAX [m ³ /s]	1.81	1.96	1.21	2.44	10.1	3.6	2.05	1.59	1.59	18.3	17.8	2.1

The seasonality of the river is highlighted from the values in table 2 and table 3, in particular from the average values of flowrate. As an alpine river, is fed by snow-melting phenomena in spring and the rainfall precipitation in Autumn, instead, winter and summer as the dryer periods in the year.

2.1.3 Geomorphological characteristics

The geomorphological studies on this basin highlight how the upper Sangone Valley is characterized by a diffuse covering of detritus with an origin related to moraine deposits and typical calcschist, mica-schist, and serpentine bedrock. Moreover, the basin has a significant slope gradient and a poor presence of vegetation. This condition determines an important instability that can bring about the occurrence of rockfall phenomena. During extreme rainfall, the high availability of debris in the basin could induce with ease relevant mass transport phenomena in the riverbed.⁴¹

On the other hand, the lower part of the Sangone catchment area is characterized by a lower slope gradient in the area and the river channel. The higher urbanization of the area and the man's actions determines a widespread presence of cyclopic rocks with an important decrease capacity of the river to induce relevant mass transport phenomena.^{41,42}

2.1.4 Climate in the area and RCP scenarios for Piemonte region

The climate in Piemonte, is temperate and semi-continental, with cold, snowy winters and warm, sunny summers. The region presents relevant precipitation throughout the year, with the highest amounts occurring in autumn and winter. The proximity of the region to the Alps also contributes to its diverse and changeable weather patterns.

According with the upgraded Köpper-Geiger climate classification,^{43,44} it is possible to see how also the climate shows a switch at the Trana bottleneck. Figure 11 shows how the upper part of the Sangone basin has generally colder temperatures and the portion of the land is included in four different climate areas (ET, Dfc, Dfb, Cfb, and Cfa) in which the most predominant is the cold, no dry season, warm summer (Dfb) whereas the basin after Trana is all embedded in the temperate, no dry, hot summer climate (Cfa), typical for all Po Valley.

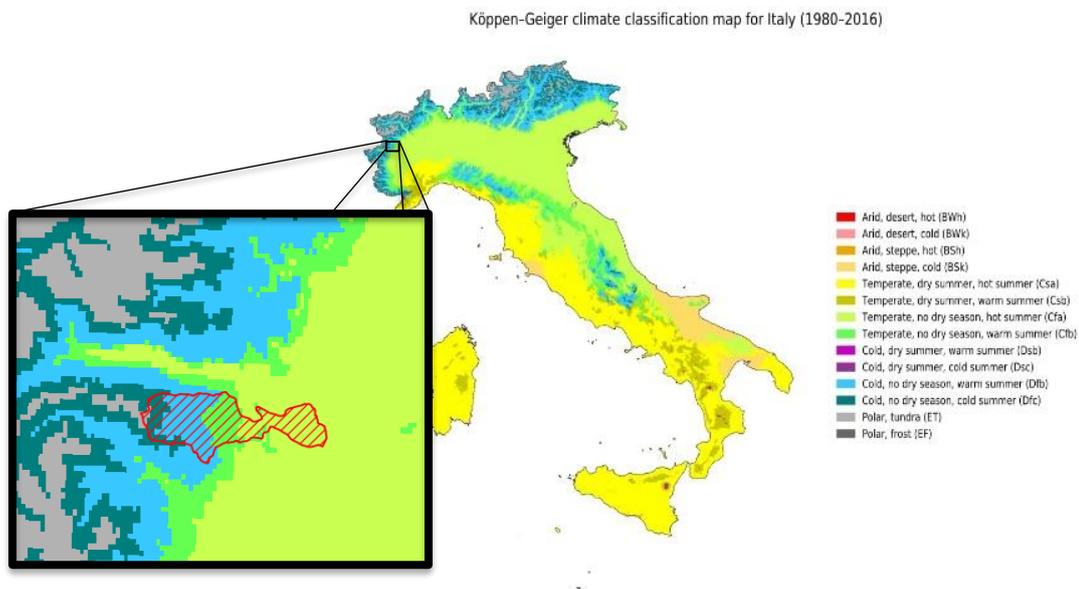


Figure 11. High resolution map of upgraded Köpper-Geiger climatic classification of Italy and focus on climate area that insists on Sangone basin.

Due to climate changes, the climate is shifting to higher temperatures, longer periods of drought, and general dryer conditions. Figure 12 highlights how the trend to higher average temperature is substantial in Piemonte and how the last years still arise in speed.⁴⁵

Observed Annual Mean-Temperature, 1901-2021
Piemonte, Italy

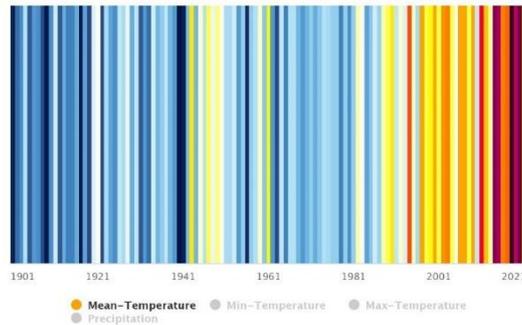


Figure 12. Trends of observed annual mean-temperature during 1901-2021 period in Piemonte taken from Climate knowledge portal.

Indeed, the rise of temperature is greater than the global temperature increase, especially in the Alpes which are typically *hot spots* for climate change. Between 1958 and 2019, the increase of maximum temperature is around 2.3°C and in the last period (1981-2019) around 0.58°C/10 years.⁴⁶

Furthermore, the comparison between the last 18 years and the reference's period 1971-2000, showed in figure 13 and 14, outlines a remarkable decrease in the rainy days (with rainfall > 1mm) with larger dryer periods and an overall decrease in rainfall.⁴⁷ Instead, figure 15 shows the trends for rainfall-free days.

Anomalie di precipitazione(mm): differenza della cumulata media tra 2001-2019 e 1971-2000

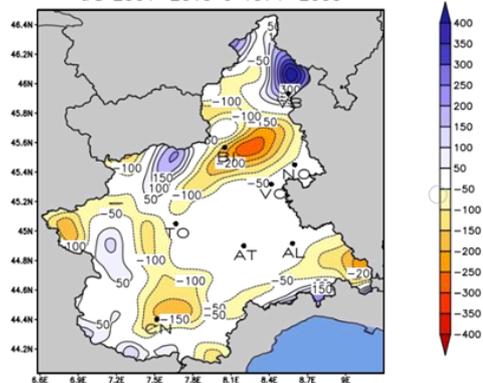


Figure 13. Rainfall's Anomalies [mm]: difference between the average cumulative rainfall of 2001-2019 and 1971-2000 in Piemonte region, take from "Stato dell'ambiente in Piemonte. Relazione 2020: precipitazioni".

Differenza del numero medio di giorni piovosi tra 2001-2019 e 1971-2000

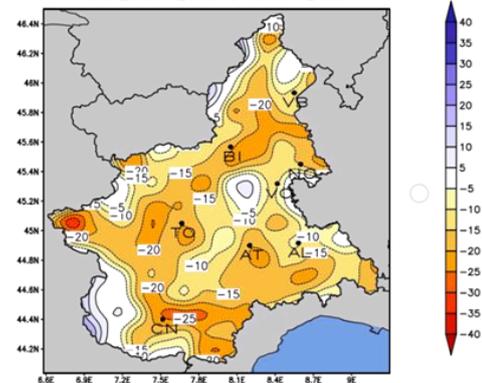


Figure 14. Difference of the average rainy days between 2001-2019 and 1971-2000 in Piemonte region, take from "Stato dell'ambiente in Piemonte. Relazione 2020: precipitazioni".

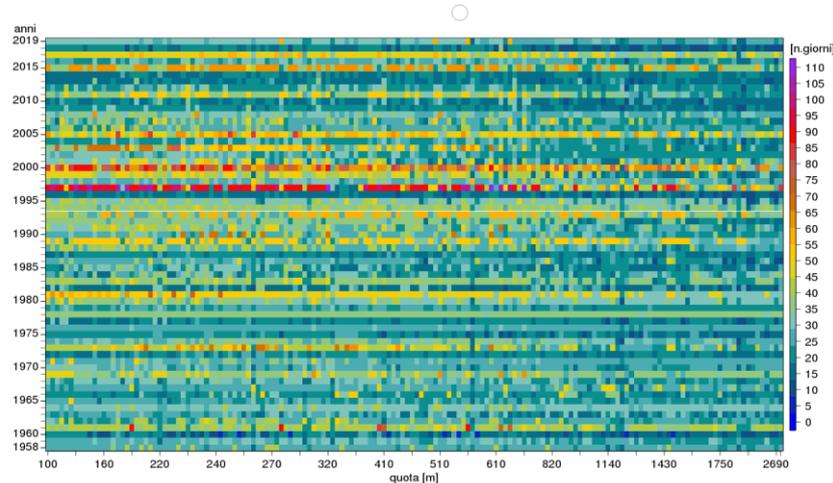


Figure 15. Graph showing the trend of average rainfall-free days per year for different altitudes, take from “Stato dell’ambiente in Piemonte. Relazione 2020: precipitazioni”.

The possible climate evolution scenarios for the end of the century for the Piemonte region, based on IPCC’s model, foresee a dangerous degradation of the situation both in RCP 4.5 scenario – mitigation scenario -, based on the respect to the Paris Agreement, and RCP 8.5 scenario – trend scenario -, with higher greenhouse gases emissions.^{48,49}

The temperature will continue to arise for 0.2°C/10 years for the mitigation scenario and 0.5°C/10 years for the second one. The cumulative annual precipitation does not show a significant decrease but a critical variation in distribution during the year. Often rainfall will happen in the winter and, on the other hand, there will be dryer springs and summer. The rainy days will continue to diminish to reach an average of 5-8 days without rainfalls at the end of the century in a better scenario to 15 days in the worse one.⁴⁸

The potential of the evapotranspiration process will continue to increase until reaching an augmentation between 8% (for RCP 4.5) and 15% (RCP 8.5).⁴⁸

The actual and possible trends for climate point out the critical pressure that will attempt a river’s survival, especially with a non-perennial river such as Sangone.

2.1.5 Characterization of surface water body types due to WFD 2000/60/CE

Due to the WFD 2000/60/CE,¹⁶ it has been done the typology classification of regional river networks (particularly for the rivers with catchment area above 10 km²) based on the following abiotic descriptors:³⁹

- geographic localization,
- morphometric descriptors,
- climatic descriptors,
- geologic descriptors.

According to these indicators, the European territory was divided into homogeneous hydro-ecoregions (HER) where more detail descriptors (such as intermittency, persistency, river's origin, and distance from the origin) in the second step of river classification were

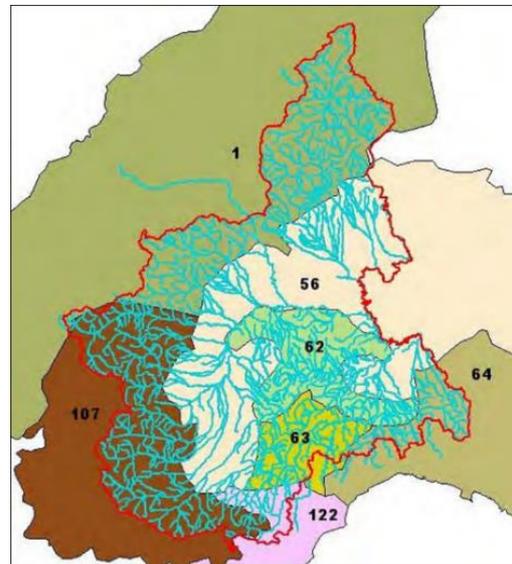


Figure 16. Piemonte division through HECs.

applied. In Piemonte, as can be seen in figure 16, there are six different HECs.^{38,39} Due to this method, the Sangone has been divided in three different stretches.^{37,39}

The upper stretch of Sangone (figure 17), the closest to the river source, is in a typical mountain area with a torrential regime. This stretch is long around 5.4 km, and its slope is higher than 6%. Its course is natural, with no significant anthropic pressures.



Figure 17. Upper stretch of Sangone characterization due to WFD 2000/60/CE. Its identification code is 0010811pi for the "Piano per la gestione del distretto idrografico del fiume Po and 04SS1N703PI for the regional classification.

The second part of the Sangone (figure 18) is the is the longest one with 27.74 km of the river channel. This stretch passes from a mountain through hill lands until reaches flatlands after the bottleneck in Trana.

Its course is natural but there are significant pressures as hydroelectric facilities and withdrawals for civil, irrigation, and industrial uses. There is an alteration of the hydrologic regime and an alteration of the biological quality.



Figure 18. Second stretch of Sangone characterization due to WFD 2000/60/CE. Its identificational code is 0010812pi for the “Piano per la gestione del distretto idrografico del fiume Po and 04SS2N704PI for the regional classification.

The final stretch (figure 19) flows only in flatlands (slope lower than 1%) through an area with intensive urbanization. It extends for 14.24 km. The area presents important anthropic pressure like livestock farms and wastewater discharge that compromise the biological status and chemical and physical characteristics of the water.



Figure 19. Last stretch of Sangone characterization due to WFD 2000/60/CE. Its identificational code is 0010813pi for the “Piano per la gestione del distretto idrografico del fiume Po and 06SS3F705PI for the regional classification.

2.1.6 Anthropogenic pressures

Although the last evaluation made by ARPA has shown the reaching of a “GOOD” chemical status and at least a “SUFFICIENT” status from an ecologic point of view in all the stretches,³⁸ the relations of 2018 on hydro-morphologic features of regional rivers, as an implementation of the Water Frame Directive 2000/60/CE drafted by ARPA Piemonte,⁵⁰ still underlines a critical impact of the withdrawals on the river that determined a “NOT GOOD” status from a hydrologic point of view.

The Sangone was always a pivotal source for communities that grew around its banks. The river has served as a water resource for domestic, civil, and industrial uses, irrigation, and urban development along its course. Moreover, it has a meaningful role for recreational and touristic purposes.

The area is characterized by a strong presence of the industrial sector, especially the engineering sector. The level of quantitative impairment of the surface water resources determined a serious pressure on the resource. The withdrawals, combined with the type of hydrological regime of the basin, determined a substantial issue of dry spells, especially in the summer.³⁷

Several agricultural activities are present along the river; winter cereals, corn, and forage grasses are the most widespread crops. Livestock activity is also of modest interest, involving mainly herds of cattle.³⁶

Although this freshwater body has always had a pivotal role for its communities, the actual consequence of anthropogenic pressure determines significant negative impacts on the river.

The ARPA's relation⁵⁰ highlights how several water shunts for irrigation and energy production insist on the river channel. The map in figure 20 shows as the main ones are before Rivalta. The principal withdrawals are in the municipalities of Coazze, Giaveno, and Trana.

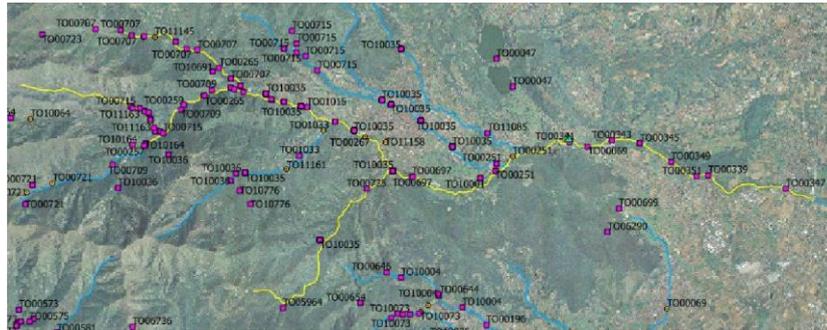


Figure 20. Map with the anthropogenic pressures located along the Sangone river, from the IMPLEMENTAZIONE DELLA DIRETTIVA 2000/60/CE: ANALISI E VALUTAZIONE DEGLI ASPETTI IDROMORFOLOGICI. RELAZIONE SUI CORPI IDRICI ANALIZZATI NEL 2017-2018.

The hydroelectric withdrawals return water to the river, thus it is possible to evaluate the measurements of Trana gauging stations as values of the natural flow. Meanwhile, the withdrawal for irrigation uses determines a loss in water that never returns to the river.⁵⁰

In the end, Sangone was still declared as at risk of failing to meet quality objectives under Water Frame Directive 2000/60/EC. In particular, the most compromised section starts from the municipalities of Rivalta and Orbassano.³⁷ From figure 21 is possible to see in red how almost all the Sangone river presents a “NOT GOOD”

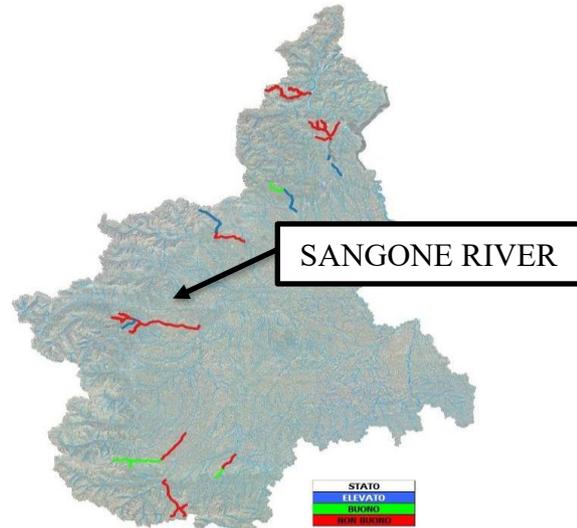


Figure 21. Map that exhibits the “NOT-GOOD” hydrological status of the river for most of the river channel from the IMPLEMENTAZIONE DELLA DIRETTIVA 2000/60/CE: ANALISI E VALUTAZIONE DEGLI ASPETTI IDROMORFOLOGICI. RELAZIONE SUI CORPI IDRICI ANALIZZATI NEL 2017-2018.

2.1.7 Its intermittency

The Sangone river, as usual for alpine rivers is typically fed by snowmelt and precipitation, with high flow rates in spring, and lower flow rates in the winter months.

The river is also influenced by groundwater charge and the presence of wetlands, which can contribute to the overall flow of the river.

However, particularly in the basin region where there is a Cfa climate region, the summer season, between the melting season – typically between April and May – and the rainfall in autumn, could determine a second period in the year with a critical flow rate. The area characterized by hot summer, with negligible rainfall, induced significant processes of evapotranspiration and dryer land, which arise their capacity of absorption of surface water.

These natural characteristics of these waterbodies determine naturally that stretches of the Sangone could present a cease of flowing water during the year. According to the non-perennial river classification⁴, Sangone can be considered an intermittent river with an active flow for the main part of the year and period in which the evapotranspiration processes exceeds feeding processes with shifting to a not-flowing status.

Moreover, on Sangone it is possible to highlight how anthropogenic pressures determine additional uncertainty on the hydrologic continuity of flow during the year. It is a perfect example of intermittency in which both natural and anthropogenic pressure could establish an increase in frequency and duration of the intermittency.

The heatwaves and droughts periods that characterized last spring and summer brought an unreal landscape on the Sangone's riverbed during June and July, carrying him and the Po River into the national and international press spotlight.⁵¹



Figure 22. a) A man walks on the dry riverbed of Sangone river in Beinasco, Turin, Italy June 19, 2022. REUTERS / Massimo Pinca. b) A woman photographs the dry riverbank of the Sangone in Beinasco, Turin, Italy June 19, 2022. REUTERS / Massimo Pinca. c) A man walks on the dry riverbed of Sangone river, in Beinasco Turin, Italy June 19, 2022. REUTERS/Massimo Pinca.

The photos, shown in figure 22, and, in general, the experience of the last years⁵²⁻⁵⁴ are in contrast with the statement declared in the last report of 2018³⁸ for which all the rivers in the Piemonte region are classified as *perennial*.

This inconsistency shows the urgency and importance to find and adopt a general and clear method for the quantification of intermittency and evaluation of its evolution both in time and space.

Due to climate change and possible trends predicted by RCP scenarios for the region, the resulting rise in temperatures, and the overall trend of precipitation reduction,⁴⁸ the survival of the river is jeopardized. The frequency, the duration of intermittency, and the length of intermittent stretches could increase tragically with significant consequences for the ecosystems and the economy of the communities that depend on it.

Even today is not possible to quantify objectively how the situation of its temporariness is serious because there is no prior information through which to compare the actual situation and possible evolution due to the climate future trends.

This research aims to give the relevant authority the capacity to monitor the status of intermittency and act to mitigate it.

2.2 River segmentation

2.2.1 Identification of the river section to investigate

To reach the aim of the research, the river section to investigate must have some specific characteristics:

- frequent and significant phenomena of ceasing of flowing during the year,
- limited anthropogenic pressures that can disturb our analysis,
- easy to reach and investigate for ground truth data.

According to these characteristics, at first, for instance, exploiting the river through preliminary field surveys and satellite images, the stretch underlined in figure 23 has been chosen as the case study.



Figure 23. Map of the river section investigated with the gauging station and water table's monitoring wells insisting on it.

This portion of the Sangone river is straddling the second (code 04SS2N704PI) and third (code 06SS3F705PI) stretches of types of characterization. The segment considered develops in flatlands and it was between Rivalta di Torino and Orbassano, the portion of the river that the ARPA's report on hydrological status identified as the most affected by intermittency during the year.⁵⁰

In addition, there are a large number of stations for the depth to groundwater in this area: Parco Via Piossasco, Stabilimento ex Chimica Industriale from the CMTO's monitoring network and Orbassano from ARPA's network. Sangano and Bruino do not have values for the water table level. Moreover, the section is located immediately after the TRANA_SANGONE hydrometric stations and there is also the Sangone a Rivalta hydrometric station from CMTO's network. Both water table level and flowrate can be valuable data for predicting river intermittency.

The presence of a quarry determines the end point of the section due to the possible presence of some discharges that can affect the dependence of water presence from natural events.

Moreover, the new "Parco del Sangone", that surrounds the banks of this area of the Sangone river, is easy to reach and there are multiple points where the riverbed can be accessible.

2.2.2 Characterization and segmentation

The first part of the considered section, belonging to the 04SS2N704PI stretch, has a slope of around 0.85% and is classified as sinuous and unconfined. The riverbed is large around 44.15 m, its morphology is flat and there is a predominance of sediments with a size like a boulder.⁴²

Either the second part, embedded in the 06SS3F705PI stretch, is unconfined and sinuous with a slope of around 0.52%. The average riverbed's width is around 42 m and, as before, its morphology is flat but in this case, the sediments present a lower diameter with pebble size.⁵⁵

The field surveys allowed us to verify that several drains insist on this last part of the investigated section, as shown in figure 24.

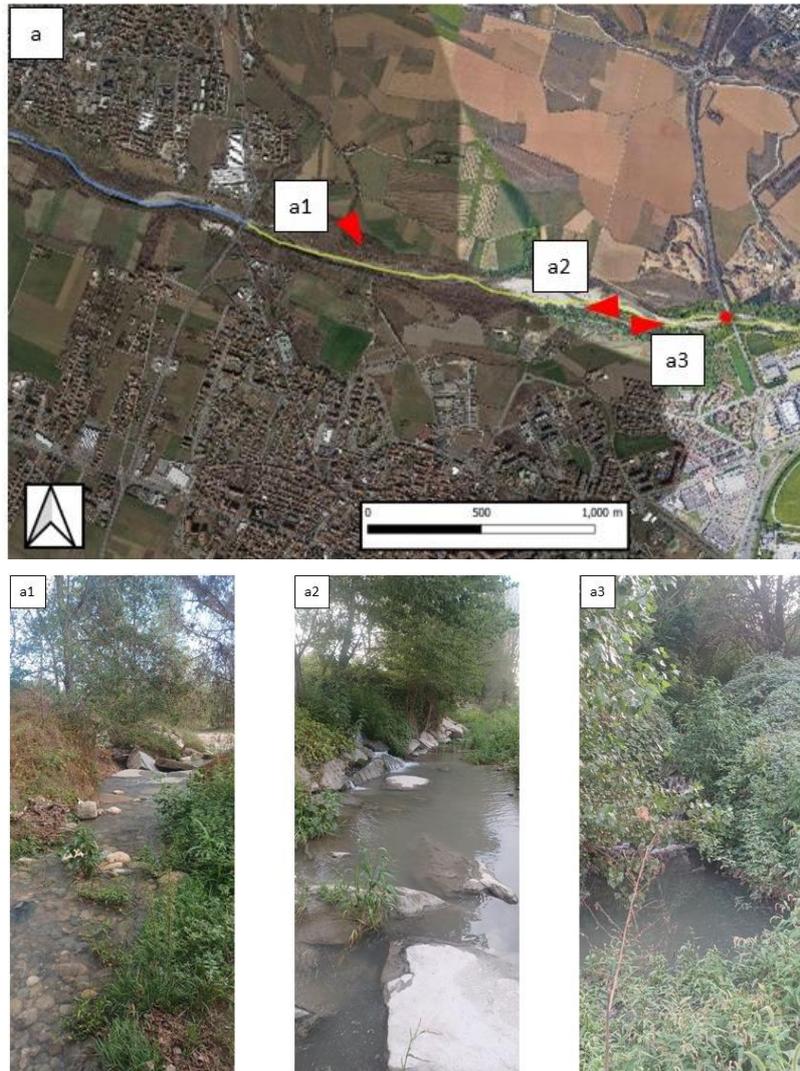


Figure 24. Localization of drains that insists on last part of investigated section.

In this case the modification of the natural flowrate does not consist in enhancing the frequency and duration of cease of flowing water but, on the opposite, leading to decreasing natural frequency and duration of its ephemerality.

Indeed, this latter part of the section is where most of the pools were found during field surveys.

Due to the characteristics described, the remaining section taken into account was divided into three different segments (figure 25). In this case, the division was based both on the different trends to change the flowing status that has been observed during the field surveys and on the acknowledgment of the satellite's image for this area.



Figure 25. Map with the final river segmentation

The three segments are long respectively around 2085 m, 2240 m and 1220 m.

3. Material and methods

3.1 Meteo-hydrogeology datasets

Gauging stations

The meteo-hydrogeology data were obtained from regional^{40,56,57} and local^{58,59} datasets. The variables used are daily rainfall, air temperature (maximum, mean, minimum), relative humidity (maximum, mean, minimum), flowrate measurements and water table level, all on a daily time scale. The rainfall and water table level datasets allow to assess possible recharge processes that may influence the flowing status of the river. Relative humidity and air temperature can describe the evapotranspiration process on the riverbed. The flowrate is a point measure that can describe the correlation between the presence of water in the investigated river section with the amount of flow in an upper location.

For the flowrate measurements, the series that have been used are from the stations: TRANA SANGONE (TRANA_FLOWRATE) and Sangone a Rivalta (RIVALTA_FLOWRATE), while for water table level the observation wells considered are Rivalta Parco Piosasco (RIVALTA_PP_WTL), Rivalta ex stabilimento Chimica Industriale (RIVALTA_EXSTAB_WTL) and the one in Orbassano (ORBASSANO_WTL) (figure 23).

Thiessen polygons spatial interpolation

For the meteorological series, the Thiessen polygon technique was used for spatial interpolation of the drainage area. To obtain better and more accurate results, spatial interpolation was applied to a new area, built up on QGIS from the regional DTM, using GRASS tools and defining the outlet section at the end of the investigated area. Its extension is equal to 192.47 km². In this way, the portion of the basin draining water for the downstream river channel was not considered.

The interpolated meteorological series are then used to evaluate the corresponding 3,5,7,10,30, and 90-day cumulative series, which, taking into account not only the daily value, encapsulates information that can help in describing processes with timescales greater than the day.

Nan values management

The main problem of these all-time series considered is the presence of Nan values for some days that can hinder the capacity to predict on a daily scale the flowing status. Thus, for Nan values, temperature, relative humidity, water table level, and flowrate, which have more linear trends on a daily time scale, a linear interpolation between the first no-Nan values before and after the day without a valid measurement.

For rainfall, which has a higher variability on a daily scale, a different approach was preferred in which the final value was calculated through spatial interpolation only between the gauging stations that present no-Nan values.

3.2 Remote sensing datasets and field surveys

The large volume of remote sensing datasets defines new perspectives for fluvial geomorphologists to observe and study the dynamics of the river.^{60,61} The unprecedented spatiotemporal scales of satellite images allow observing and studying phenomena undergoing rapid changes as the study of non-perennial rivers and the swift evolution from one flowing status to another.

Nonetheless, the limits in the spatial and temporal resolution still exist and, thus, the choice of the most suitable sensor to adopt is still important considering the spatial characteristics and the speed of the phenomena that would be investigated.

For this research, the spatial resolution must be adequate to distinguish the presence of water in both flowing and ponding status, and, following Jiang et al⁶², the width of the riverbed must be at least three times the image resolution. In this case, with an average segment width of around 40m, the spatial resolution can not be higher than 10 m.

Due to the high variability on a daily scale of flowing status changes, the time resolution can not be higher than a few days.

Sentinel-2

These criteria determine the Sentinel-2 satellites and their multispectral sensors as the most appropriate. Yet, the use of their images' archive for these kinds of studies finds a burgeoning scientific literature.^{31,32,63,64}

Sentinel-2 is a satellite mission developed by the European Space Agency (ESA) as part of the Copernicus Program, which aims to provide continuous and high-resolution imagery of Earth's surface for a wide range of applications. The Sentinel-2 satellite carries a multispectral imaging instrument with 13 spectral bands, with a spatial resolution of 10 to 60 meters and a wide swath width of 290 km. The last characteristic guarantees the presence of all the Sangone segments in the same image.

The Sentinel mission started on June 23, 2015, with the launch of the first satellite Sentinel-2A on a sun-synchronous orbit. Less than two years after on March 17, 2017, a second satellite Sentinel-2B, with the same multispectral imaging instrument, same wide swath width, and same revisit time, was launched. Both satellites have a revisit time of 5 days and, considering an offset of 2 days between Sentinel-2A and Sentinel-2B, the overall mean revisits time from March 2017 of 2.5 days. Naturally, it should be noted that this data is purely theoretical because the cloud coverage can decrease the image exploitable for observation.

The 13 bands cover a range from visible, near-infrared (NIR), and short-wave infrared (SWIR). Four of them are collected with a spatial resolution at 10 m (B2, B3, B4, B8), six at 20 m (B5, B6, B7, B8A, B11, B12), and the last three at 60 m (B1, B9, B10). The following table 4 summarizes these characteristics.

Moreover, Sentinel-2 data is freely available and can be accessed through various online platforms, making it an accessible dataset for researchers around the world.

Table 4. Characteristics of Sentinel-2 MSI.

Bands	Wavelength range [nm]	Spatial resolution [m]	Spectral region
B1	423-463	60	Coastal aerosol
B2	458-523	10	Blue
B3	543-578	10	Green
B4	650-680	10	Red
B5	698-713	20	Vegetation Red Edge
B6	733-748	20	Vegetation Red Edge
B7	773-793	20	Vegetation Red Edge
B8	785-899	10	NIR
B8A	855-875	20	Narrow NIR
B9	925-965	60	Water - Vapour
B10	1350-1410	60	SWIR - Cirrus
B11	1565-1655	20	SWIR
B12	2100-2280	20	SWIR

The satellite launch defines the beginning of the studied period as 2015. Thus, all the images from 2015 to the end of 2021 are freely downloaded from Copernicus Open Access Hub³⁶: the first image until November 1, 2016, has been processed through an atmospheric correction tool Sen2cor, from the Sentinel Application Platform, because they present top-of-the-atmosphere (TOA) reflectance instead of bottom-of-the-atmosphere (BOA). Always from SNAP a bilinear interpolation was applied to resample all bands to a 10 m spatial resolution.

High resolution images

High resolutions satellite images from Google Earth Pro and field surveys are used for ground truth data to acquire verified regions of interest (ROIs) for the different classes: vegetation, sediments and water in the river channel. The Google Earth Pro image considered is the last one available for the investigated area: on March 3, 2022. The field surveys started on September 2022 to end on October 3, 2022, due to the start of the stable flowing status period.

Method for field survey acquisition

The field surveys are executed on a sunny day in which one of the two Sentinel satellites would have acquired a free-cloud image of the segments investigated after significant meteorological events.

The aim was to acquire the georeferenced polygons that represent the ponds in the river channel. For this purpose, the field instrumentation was:

- a laser telemeter,
- a handheld computer with GIS software integrated with Map Stream plug-in,
- a tripod,
- a compass.

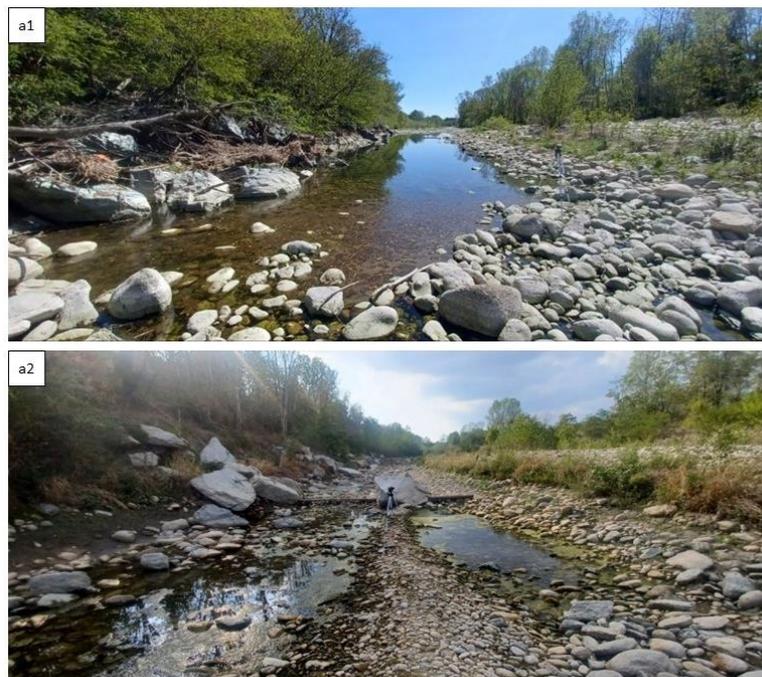


Figure 26. a1: example of on field survey with instruments for ponds collection on September 16,2022. a2: example of on field survey with instruments for ponds collection on September 11,2022.

Following the ISPRA's lineguide⁶⁵ and Map Stream Manual⁶⁶, three fields survey are performed on September 11, September 16, and October 3, 2022, with the acquisition of 17 ponds along the river channel. Examples of the instruments on field are shown in figure 26. Figure 27 shows example of field surveys' acquisition.



Figure 27. In background the 11/09/2022 satellite's image in true color composite visualization (RGB). Zoomed in there are two ponds acquired during the field survey: in red the "station points" georeferenced manually on GIS software in which the laser telemeter was located for the acquisition, in blue the points acquired to describe the edge of the ponds, the pink grid represents the resultant ponds area on Map Stream.

Figure 27 underlines how the polygons obtained from field surveys to represent water ponds in the riverbed can cover some pixels integrally and others only partially. For this reason, to understand if and how the selection of the pixel influences the final spectral signature for water in the river channel, the choice of pixels that must be included in the ROIs was performed with two different approaches for each pond:

- considering all the pixels that present a minimum portion covered by the polygon (so called "pools" class),
- considering only the pixels mostly covered by the polygon and which grants to resume the form of the pond, (so called "completed pools" class).

In figure 28 are reported the two approaches for pixel selection.



Figure 28. Different approaches to determine the best ROIs selection for ponds from field surveys polygons (pink). a1: the ROIs (blue) are determined considering all the pixels. a2: the ROIs (cyan) are determined considering only significant pixels.

3.3 Temporal scope of the study: exploring the investigated period

The figure 29 shows the temporal availability for all the different variables, including the Sentinel-2 acquiring period.

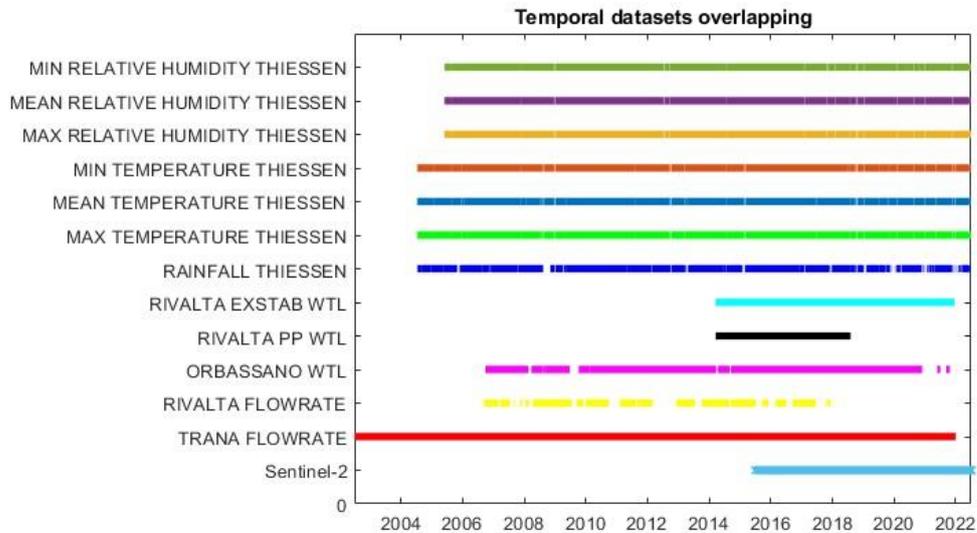


Figure 29. Temporal overlapping between meteo-hydrogeology datasets and Sentinel-2 images. With Thiessen are named the variables obtained from Thiessen polygons spatial interpolation.

The validated data for TRANA FLOWRATE and RIVALTA ESTAB WTL variables stop at the end of 2021. No data were available for further periods. For meteorological datasets (rainfall, temperatures and relative humidity) the end of available datasets is the 30th of June of 2022. Thus, the 31st of December 2021 was defined as the end of the studying period to prevent six months of Nan values for crucial variables as flowrate and water table level. Moreover, the choice to end the investigated period at the end of 2021 allow to avoid a more unbalanced datasets without considering part of a year in which is not included the typical period of ceasing of flowing water phenomena.

Considering the beginning of Sentinel-2 images acquisition, the investigated period was defined between 2015-2021.

In addition, the graph in figure 29 shows the temporal inconsistency of RIVALTA FLOWRATE due to the repeated and significant period with no data, the too-short period of measurements of RIVALTA PP WTL and, thus, these variables were not used for further models.

3.4 Method for the identification of the flowing statuses

The different flowing statuses' identification was performed through a supervised classification of the Sangone segments. To ensure higher accuracy for the method, three spectral bands were searched to guarantee a great separability between the three classes.³²

The field surveys and the study of Google Earth Pro images allow for defining accurate ROIs for vegetation, sediments, and water in the riverbed. Furthermore, another class of ROI, taken from known general water bodies (such as Avigliana lakes or Po River), was added to the spectral signature analysis to underline the difference between the spectral signature of water from the Sangone river channel, rather than the traditional spectral signature of water.

It used Google Earth Engine to compare the spectral signature of different ROIs' classes for the same image.^{67,68} The outcomes grant to determine the best three bands. Also, the analysis was executed for different seasons taking into account the seasonality variation of the spectral signature of classes: the corresponding field surveys' images, between September and October, in late summer and the beginning of the autumn, and the Google Earth Pro Image at the end of winter.

In the end, the False Color Image (FCI), obtained from this triplet of bands, is the best visualization to distinguish the classes and decreases the number of possible misclassifications. To standardize the image classification process, the same interpreter classified the whole image archive and the scale used was always 1:10000.

The definition for each flowing status used for the supervised classification is:

- Flowing, continuity of water's flows along the whole segment,
- Ponding, at least a trait in which the continuity of flow is interrupted, or it can not be observed from the FCIs,
- Dry, absence of water along the whole segment or it can not be observed from the FCIs.

3.5 Random Forest Classification

The dataset obtained from the supervised calibration was used to calibrate and evaluate the accuracy of a daily prediction of flowing status in segments, performed through a Random Forest algorithm, implemented on R.⁶⁹

RF algorithm

The Random Forest (RF) is a machine learning algorithm that used a nonparametric method to resolve both classification problems and regression problems. It is based on Classification And Regression Trees algorithm (CART) but allows it to overcome its drawbacks.

CART algorithm used only a unique decision tree built up splitting the dataset considering the explanatory variables that allow it to reach the best separation into different classes (or minimizing the variance for regression problems). The scheme of a decision tree is shown in figure 30. This method encores runs into overfitting problems.

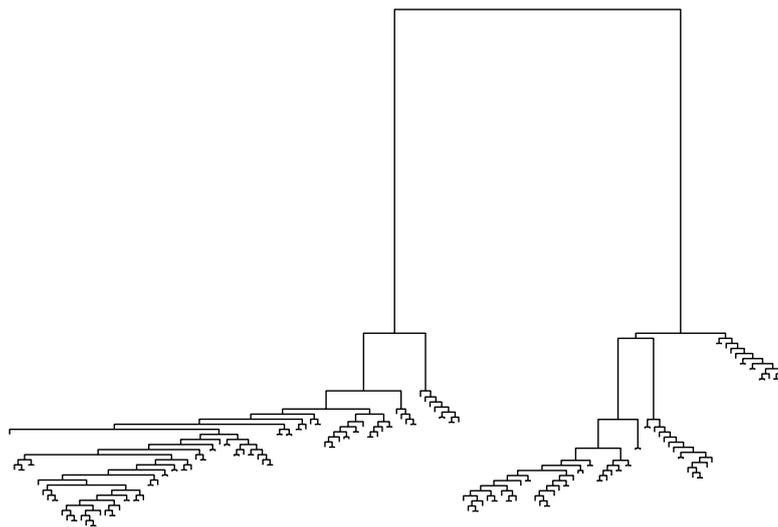


Figure 30. Scheme of a decision tree

Therefore, RF used a multitude (a “forest”) of decision trees to make powerful predictions, less prone to overfitting problems. This is possible thanks to randomness in the creation of each tree and the final combination of the entire forest. Each tree is built upon different subset of the original dataset that is randomly obtained through the “Bootstrap aggregating” (or *Bagging*) method. Moreover, at each node of a tree the “Random Inputs” method defines a random selection of a new subset of explanatory

variables and, between the variables in this subset, looks for the one which guarantee the best split¹. In the end, the resultant trees' predictions are aggregated to obtain the final predictions.

Furthermore, the RF algorithm can handle high-dimensional datasets and noisy data.

Variables importance (Boruta package)

Another important characteristic of RF is the capacity to evaluate the importance of a variable. The algorithm is capable to define a hierarchy between explanatory variables based on their capacity in defining more homogenous nodes when used for splitting. For this research, the Boruta package was used for defining the relevant variables. This method was used because is directly built around the RF. Defining the most significant variables is crucial since many machine-learning algorithms can lose accuracy with a high number of variables and because the selection process of the most significant variables allows a better interpretation of the results.

To optimize the predictive performances, minimizing the out-of-bag error (E_{OOB})², the following four hyperparameters was set⁷⁰:

- the number of decision trees (*n_{tree}*), increasing the number more accurate will be the results, for this research was equal to 1000 (two times the default value) to be sure that the EOOB was stabilized,
- the number of variables randomly sampled in each node (*m_{try}*), is the most influenced parameter and for the classification problem is defined as equal to the square root of the total number of significant variables considered in each model, with 2 for a model with less than 4 variables,
- the observations drawn to training each tree (*sampsiz*e) to maximize model performance, a lower number determines less correlation between the trees (positive effect on prediction accuracy) but also a loss in accuracy for the result of the single tree, indeed is defined as a problem dependent parameter,

¹ The best split for classification problems is the one that decreases the presence of misclassification in the following nodes. It can be measured through the Gini purity function.

² The out-of-bag error (E_{OOB}) is a cross-validation of RF classification using the values observation that was remained out of the bootstrap sampled used as subset (usually 37% of total).

- the minimum number of observations in a terminal node (*nodesize*) and for classification problems is equal to 1.

The possibility to have a high correlation between selected variables is avoided by defining the condition of Spearman ρ 's coefficient³ lower than 0.7.

All these characteristics described for RF determine a wide range of problems in which this method has found application. Especially, this method has already taken place in ecology research^{71,72} and, especially, it was already used for non-perennial rivers classification problems.⁴

Binary models F/NF and D/P

Moreover, resulting from the supervised classification of a strongly unbalanced dataset of an image, the RF was improved through a random oversampling of poor classes to evaluate its prediction on a more balanced dataset.

Therefore, the RF algorithm was first tested evaluating its prediction capacity using the subset corresponding to the day with a classified satellite's image. The predictive variables used were ORBASSANO_WTL, RIVALTA_EXSTAB_WTL, TRANA_FLOWRATE, the rainfall, temperatures (max, mean, min) and relative humidity (max, mean, min) spatially interpolated with the Thiessen polygons and their respective cumulative (for rainfall) and average (for temperature and relative humidity) series at 3-5-7-10-30-90 days.

Starting from the whole explanatory dataset, through an iterative process the most relevant and not highly correlated variables were selected.

In the end, the obtained model was used to predict the daily flowing status of the segments.

³ Spearman ρ 's coefficient, also called the rank-correlation coefficient, is a particular case of Pearson correlation coefficient. The main difference is that the Spearman coefficient is a nonparametric measure size-independent that evaluates the monotonic relationship between two variables. It can be equal to -1 or 1 for strongly negative or positive monotonic relation and 0 for no correlation.⁷⁴

After this cyclical process was performed on a D/P/F model, better results resulted from a double Boolean model that firstly classifies each day between Flowing (F) and Non-flowing (NF) statuses, where non-flowing represents both D and P conditions. The final states of NF were in turn classified between D and P states.

Due to the lack of gauging stations that typically affects non-perennial rivers, a second double model was executed for segment 2, the one with less unbalanced datasets, considering only meteorological datasets (rainfall, temperature max, mean, min, relative humidity max, mean, min and their respective cumulative series). The exclusion of point measurements, such as flow rate and water table depth, provides insight into the possible applications of this method in the absence of these types of predictive variables. Therefore, considering the actual situation, if this model gives back a robust result, this determines the general application without considering the presence (or distance from) of observation wells and discharge gauging stations.

The prediction capabilities of all these models were evaluated through statistical parameters such as accuracy, sensitivity, specificity and TSS (true skill statistics), obtained by the formula: $TSS = \text{Sensitivity} + \text{Specificity} - 1$.^{72,73}

Due to the double Boolean models (F/NF and D/P) used, the D/P depends on the F/P results. Thus, the statistical parameters that RF will evaluate are relative values. To obtain the absolute statistical values for D/P is necessary to calculate the product of each statistical parameter with the corresponding F/NF model.

Partial Dependence Plots

The RF models can determine how the classification is affected by each of the predictive variables separately. The so-called Partial Dependence Plots, introduced by Jerome Friedman in 2001, shows the dependence of a class varies with a predictive variable.

The Partial plots are powerful tools to visualize and understand if and how there is a strong relationship between the evolution of a predictive variable and the classification probability of a flowing status.

They are generated by maintaining the other predictive variables constant while the investigated variable modifies its values. Then, the plots are built up with the predicted results showing how they changed as the predictor variable varied.

The trends plotted in these graphs can help the interpretation of the most significant predictors for the models. Furthermore, the visualization of strong dependency between a flowing status and a predictive variable could assist to individuate the existence of threshold values for some variables that present a significant and rapid change in the affection of the classification.

3.6 Evaluating thresholds for effective flowing status monitoring

The RF models' results, especially from the interpretation of Partial Dependence Plots, underline the significant relationship between some flowing statuses and some predictive variables.

To assess possible turning points, the daily measurements of the chosen variable(s) and the corresponding classified flowing status for that day were plotted in graphs and scatter plots to visualize the researched values.

For the study of threshold values, the investigation has preferred to concentrate the analysis on the distinction between F/NF considering critical values of point measurements, like flowrate and groundwater, both due to their marginal effects trends that clearly show a step from a certain value, that could be interpreted as the turning points, and because these measurements can be used immediately and must not be manipulated applying spatial interpolations or evaluating their cumulative series over time.

Moreover, even if all flowing statuses are important from an ecological point of view because in determining a specific habitat for biota species, real-time predictions could be

very useful for the real-time management of withdrawals for which the difference between flowing and not-flowing statuses is crucial.

The difficulties in evaluating the exact extensions, their deep and the presence of all pools due to the limits in spatial resolution of Sentinel-2 images makes problematic the evaluation of water resource present in a ponding reach of a river and, thus, for water management, the distinction between F/NF is the most important

3.7 Regime's type identification with TREHS metrics

By predicting on a daily scale the flowing statuses for the period 2015-2021, it is possible to evaluate the permanence of each condition for each year and for the whole investigated period (M_f , M_p , M_d). Considering the boundaries and the metrics defined for TREHS classification,^{26,27} it is possible to evaluate the hydrologic regime for each segment. This classification is simple and limited but can give immediate and clear knowledge about the characteristics of intermittency of a river and the permanence of each flowing status. The threshold values are reported in table 5.

Table 5. Threshold values for M_f , M_p and M_d for each different regime.

Regime	M_f	M_p	M_d
Perennial (Pe)	$0.99 < M_f \leq 1.00$	$0.00 \leq M_p < 0.01$	$0.00 \leq M_d < 0.01$
Quasi-perennial (Qp)	$0.90 < M_f \leq 1.00$	$0.00 \leq M_p < 0.10$	$0.00 \leq M_p < 0.10$
Fluent-Stagnant (FS)	$0.40 < M_f \leq 0.90$	$0.00 \leq M_p < 0.60$	$0.00 \leq M_d < 0.10$
Alternate-Fluent (AF)	$0.40 < M_f \leq 0.90$	$0.00 \leq M_p < 0.50$	$0.10 \leq M_d < 0.60$
Stagnant (St)	$0.00 < M_f \leq 0.40$	$0.50 \leq M_p < 1.00$	$0.00 \leq M_d < 0.10$
Alternate-Stagnant (AS)	$0.00 < M_f \leq 0.40$	$0.40 \leq M_p < 0.90$	$0.10 \leq M_d < 0.60$
Alternate (Al)	$0.00 < M_f \leq 0.40$	$0.00 \leq M_p < 0.40$	$0.20 \leq M_d < 0.60$

Occasional (Oc)	$0.00 < Mf \leq 0.40$	$0.00 \leq Mp < 0.40$	$0.60 \leq Md < 0.80$
Episodic (Ep)	$0.00 < Mf \leq 0.20$	$0.00 \leq Mp < 0.20$	$0.80 \leq Md < 1.00$

It is important to emphasize that this classification is not dependent on geomorphology, considering only the presence of water in the riverbed. Moreover, the definition of thresholds could be more accurate in the future if the critical values, shown in Table 5, are updated with studies that consider regional and local information on morphology, sediment, flood events, and biological knowledge of the local river ecosystem.

4. Results and discussion

4.1 Spatial interpolation

In figure 31 there is the comparison between the whole Sangone catchment area and new one obtained specific for the investigated area.

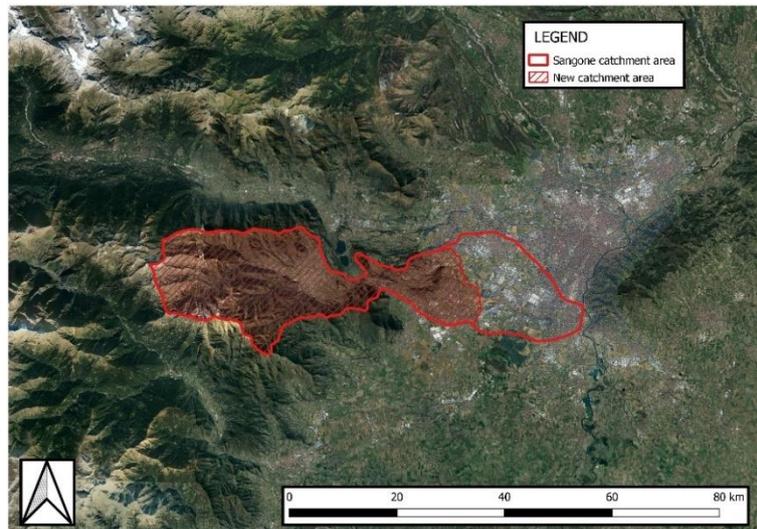


Figure 31. New catchment area compared with the original one.

Then, the Thiessen polygons division of new catchments was performed on rainfall, temperature, and relative humidity gauging stations are realized and it was evaluated the corresponding weight of each station.

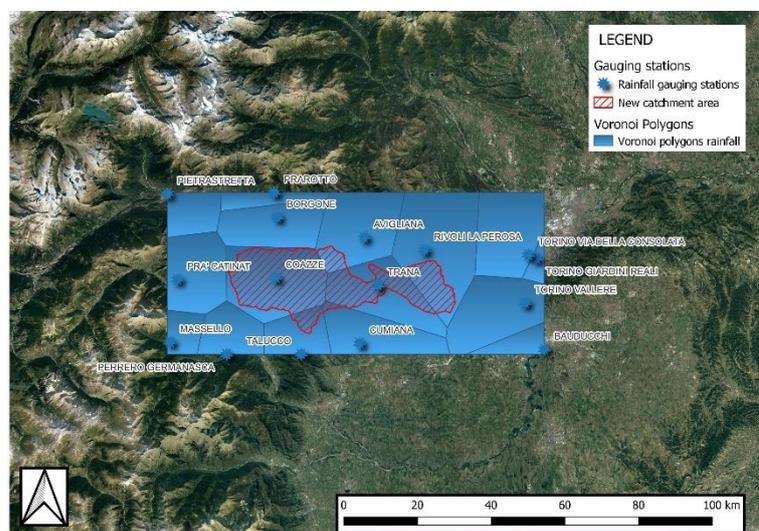


Figure 32. New catchment area divided through Thiessen polygons for rainfall measures.

In figure 32 there is the Thiessen polygons for rainfall stations and in table 6 their extension and weight in percentage.

Table 6. Area and relative weight respect the overall catchment area for each rainfall gauging stations.

Gauging station	Area [km ²]	Weight [%]
AVIGLIANA	9.44	4.90
BORGONE	0.01	0.01
COAZZE	97.12	50.46
CUMIANA	2.08	1.08
RIVOLI LA PEROSA	27.53	14.30
TALUCCO	8.79	4.57

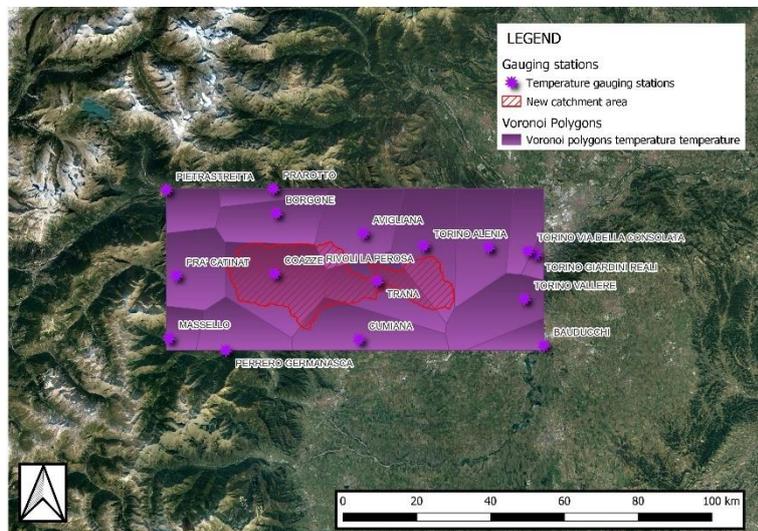


Figure 33. New catchment area divided through Thiessen polygons for temperature measures.

In figure 33 there is the Thiessen polygons for temperature stations and in table 7 their extension and weight in percentage.

Table 7. Area and relative weight respect the overall catchment area for each temperature gauging stations.

Gauging station	Area [km ²]	Weight [%]
AVIGLIANA	9.44	4.9

BORGONE	0.01	0.01
COAZZE	102.64	53.33
CUMIANA	5.36	2.78
RIVOLI LA PEROSA	27.53	14.3
TRANA	47.5	24.68

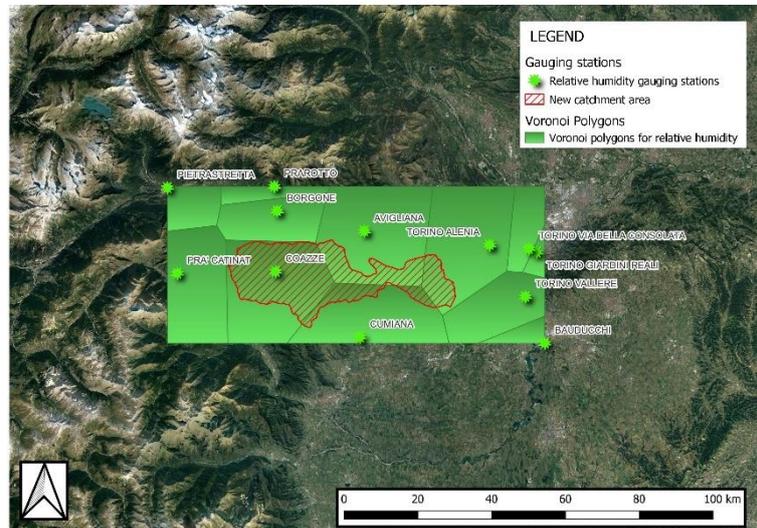


Figure 34. New catchment area divided through Thiessen polygons for relative humidity measures.

In figure 34 there is the Thiessen polygons for relative humidity stations and in table 8 their extension and weight in percentage.

Table 8. Area and relative weight respect the overall catchment area for each relative humidity gauging stations.

Gauging station	Area [km ²]	Weight [%]
AVIGLIANA	41.15	21.38
BORGONE	0.01	0.01
COAZZE	104.74	54.41
CUMIANA	25.34	13.16
TORINO ALENIA	21.25	11.04

Considering the weight of each gauging stations, it was applied the spatial interpolation for rainfall, temperature and relative humidity. Their trends are shown in figure 35.

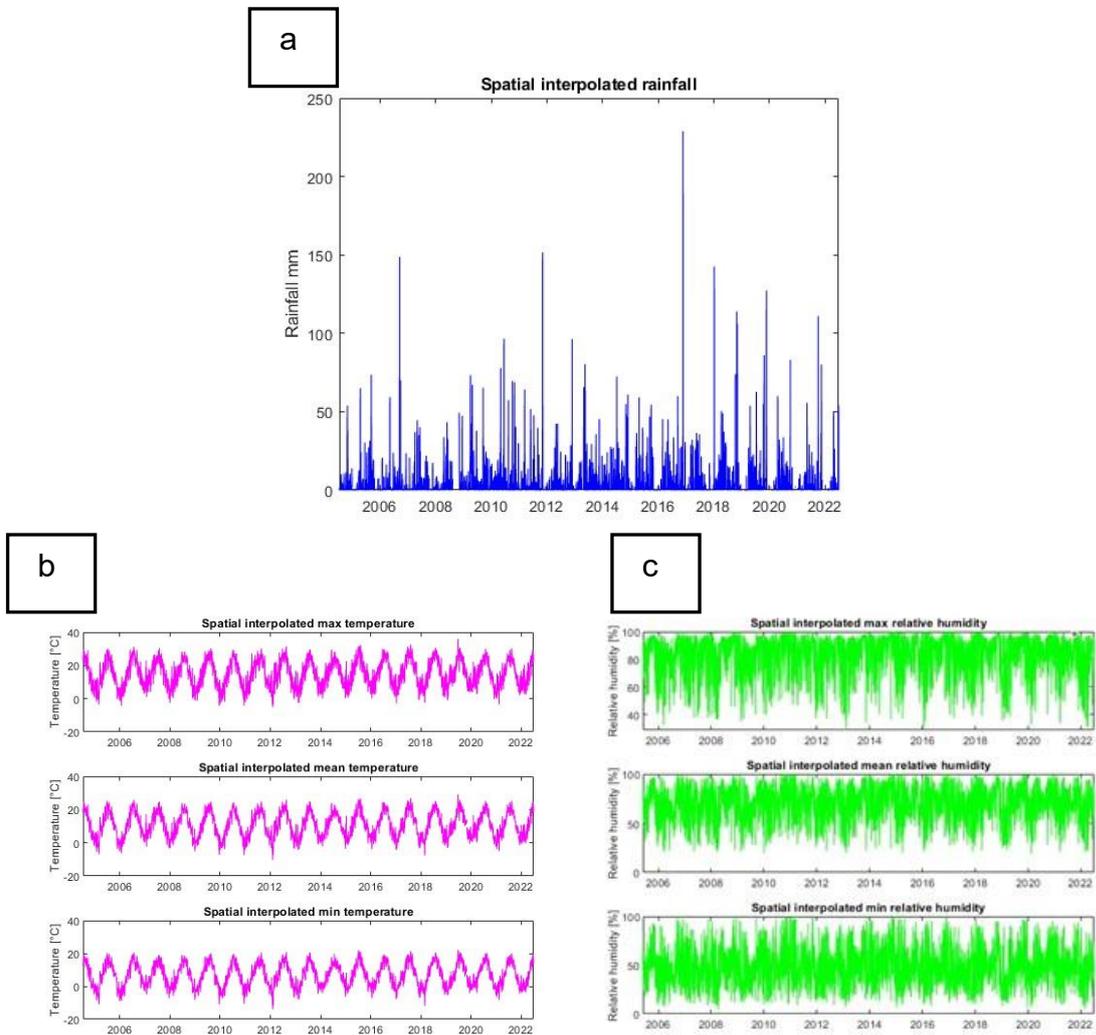


Figure 35. a) Trend of a spatial interpolated rainfall obtained from Thiessen polygons method. b) Trend of a spatial temperature (max, mean, min) obtained from Thiessen polygons method. c) Trend of a spatial interpolated relative humidity (max, mean, min) obtained from Thiessen polygons method.

4.2 Spectral signature analysis and false color images

The analysis of the classes' spectral signature was executed on five images: three are the images corresponding to the field surveys and the other two are the last before and the first after cloud-free images respecting the Google Earth image considered.

The following table 9 shows the characteristics of each image.

Table 9. Resume of ground truth source, date and related Sentinel-2 image.

Case	Ground truth source	Ground truth date	Sentinel-2 date
1	Google Earth Pro	3/2/2022	3/3/2022
2	Google Earth Pro	7/3/2022	3/3/2022
3	Field survey	9/11/2022	9/11/2022
4	Field survey	9/16/2022	9/16/2022
5	Field survey	10/3/2022	10/3/2022

Comparing ROIs of the class "general water bodies" with those representing water in the Sangone River channel (pools, complete pools, flow) shows immediately how different their spectral signatures are. The shallowness of water in Sangone defines a higher reflectance more affected by the reflectance of sediments. The selected water bodies are characterized by low reflectance along the entire spectrum, typically due to the high absorption coefficient of water. The spectral signatures obtained from ground truth data are visible in figures 36-37.

Spectral signatures for case 1-2

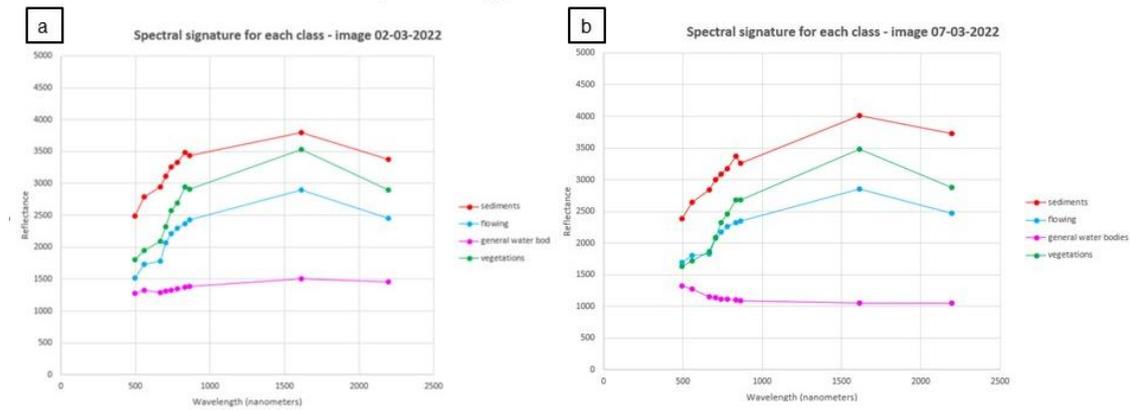


Figure 36. Graphs with the spectral signatures for the different classes. The signatures are related to the images corresponding to the Google Earth Pro source on late winter. a: spectral signatures for 02/03/2022 image. a: spectral signatures for 07/03/2022 image.

From the graphs in figure 37, related to the field survey, the different trend lines define an insignificant difference between the spectral signatures of the completed pools and pools. This significant result ensures non-dependence on supervised pixel selection to describe the polygons of the acquired pools.

Spectral signatures for case 3-4-5

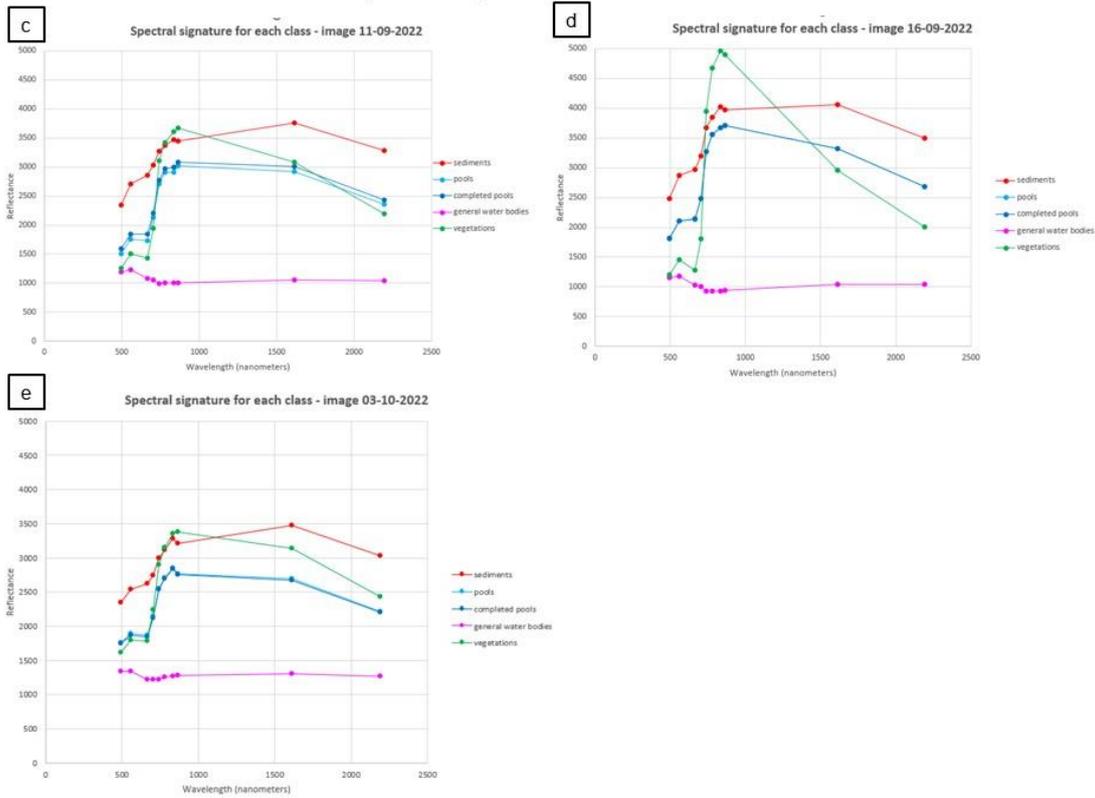


Figure 37. Graphs with the spectral signatures for the different classes. The signatures are related to the images corresponding to field surveys source on late summer. c: spectral signatures for 11/09/2022. d: spectral signatures for 16/09/2022 image. a: spectral signatures for 03/10/2022 image.

From the single charts (figure 36 and 37) and the average one (figure 38), it is possible to extrapolate some general assumptions:

- for visible bands (B2, B3, B4) the different flowing statuses have a spectral signature close, sometimes overlaid, with vegetations; sediments have a high separability due to the higher reflectance values corresponding to these bands,
- the vegetation red edge spectral region (B5, B6, B7) is the region with less separability between all signatures due to the rise of vegetations and flowing statuses reflectance,
- In the NIR (B8, B8A), the vegetation signature is almost the same as the sediments, the flowing statuses present general lower values and good separability,

- The SWIR spectral region (B11, B12) presents a good separability for all three bands for B11 and a condition like the visible region for B12.

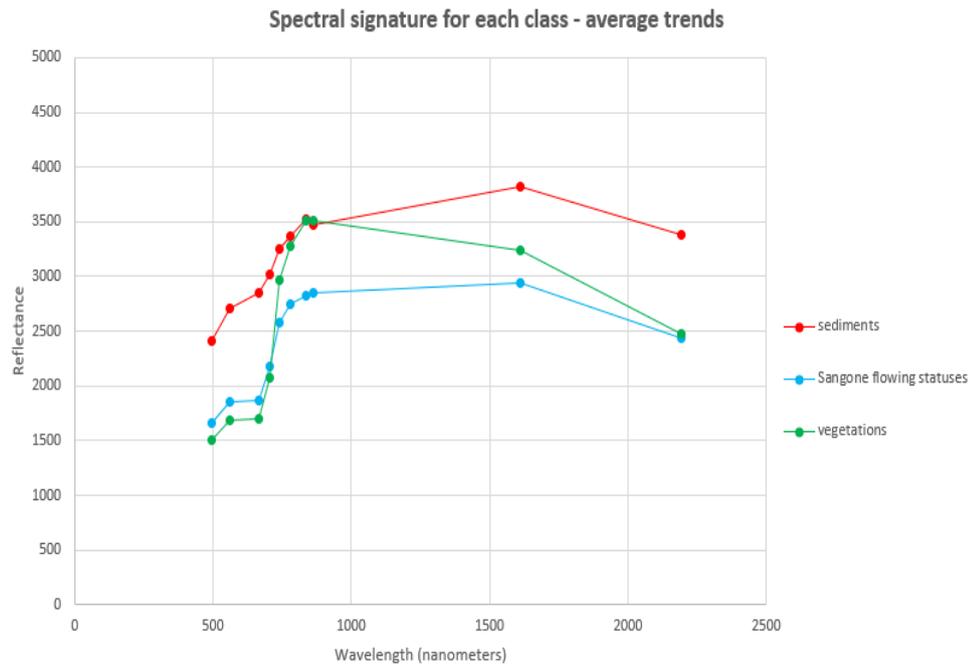


Figure 38. The final graph is the resultant from the average of the five spectral signatures of each class. For Sangone flowing statuses class, the classes used were flow from case 1 and 2 and pools from case 3-4-5.

These results are coherent with the result obtained by *Cavallo et al.*³¹ and, thus, the triplet chosen to create the FCI image with the best separability was the same one corresponding to R:B11, G:B8 and V:B4.

The B11 band provides good overall separability, the B8 allows the distinction of the water in the Sangone river from other classes, and, on the other hand, the B4 allows the possibility to recognize sediments from water and vegetation. For these last two, a visible band and B8 were preferred instead of the corresponding B12 and B8A bands due to their higher spatial resolution, 10m than 20m.

Also, the higher separability may be a consequence of an easier definition of the ROIs for each class. For example, the flowing status allows a simplified selection of a consistent number of pixels that define water in the river channel. The small and few pools, usually in a lateral position where lateral vegetation could cover them, that were present during the three field surveys to reach a valid number of pixels imposed a selection that could be

less precise in distinguishing only water pixels instead of only vegetation and sediment pixels.

In any case, the consistency of the assumptions for the ponding days returns a robust validation of the supervised classification. The FCI combination gives back images where the water presents a black/dark blue color, vegetation is in a bright green color, and sediments are white.

The comparison in figure 39 illustrates how visualizing the FCI image allows different classes to be differentiated to improve the accuracy of supervised classification.

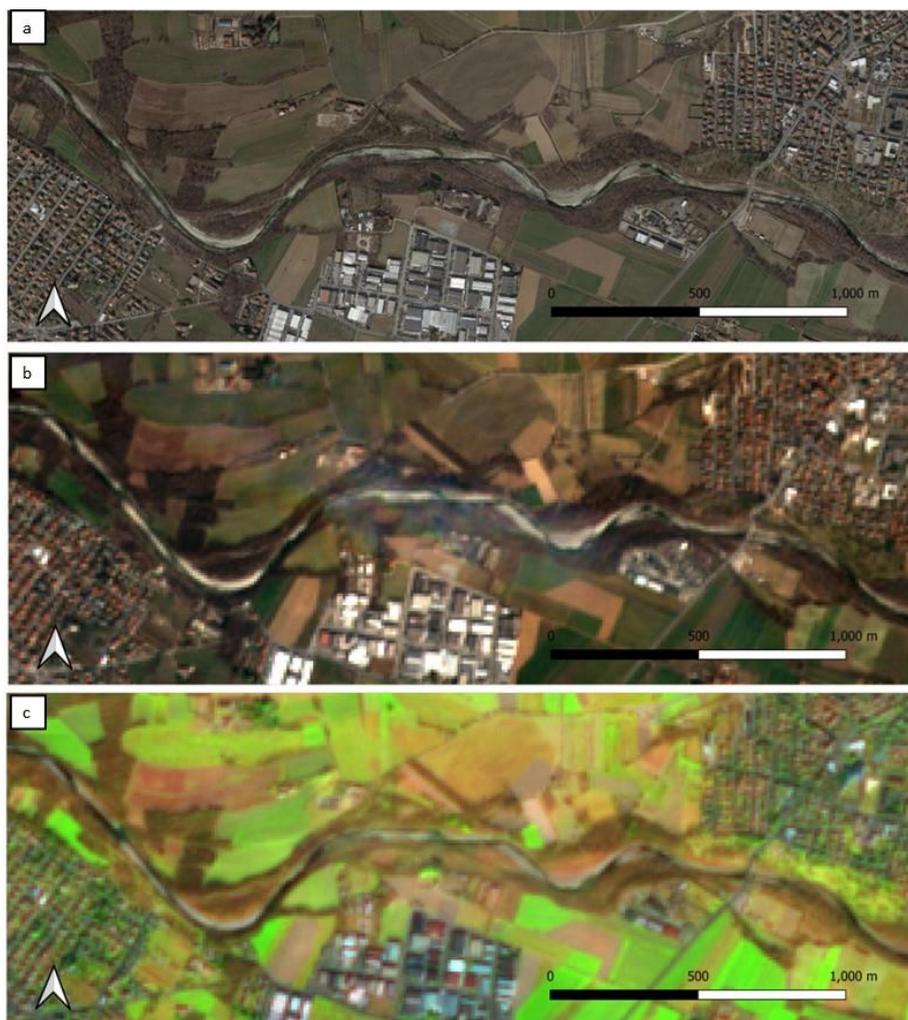


Figure 39. a: High resolution image from Google Earth Pro 03/03/2022. b: RGB Sentinel image 02/03/2022. c: FCI Sentinel image 02/03/2022.

4.3 Supervised classification of Sentinel 2 images

Comparison of false color images with ground truth data

The comparison between FCIs and ground truth data can exhibit potentialities and limits of the FCIs Sentinel-2 images in the visualization of water presence in the Sangone river channel. The first field survey on 11/09/2022 was carried out after days with moderate rain events or no rain at all. Between the second and first surveys, there weren't significant rainfall days and the comparison can highlight evapotranspiration phenomena and their influence on water presence and pool extensions.

The last field survey on 03/10/2022 was conducted two days after a significant meteorological event in the area.

In the following pages are reported several examples of the evolution in the same area between these days, considering both ground truth data and FCIs.

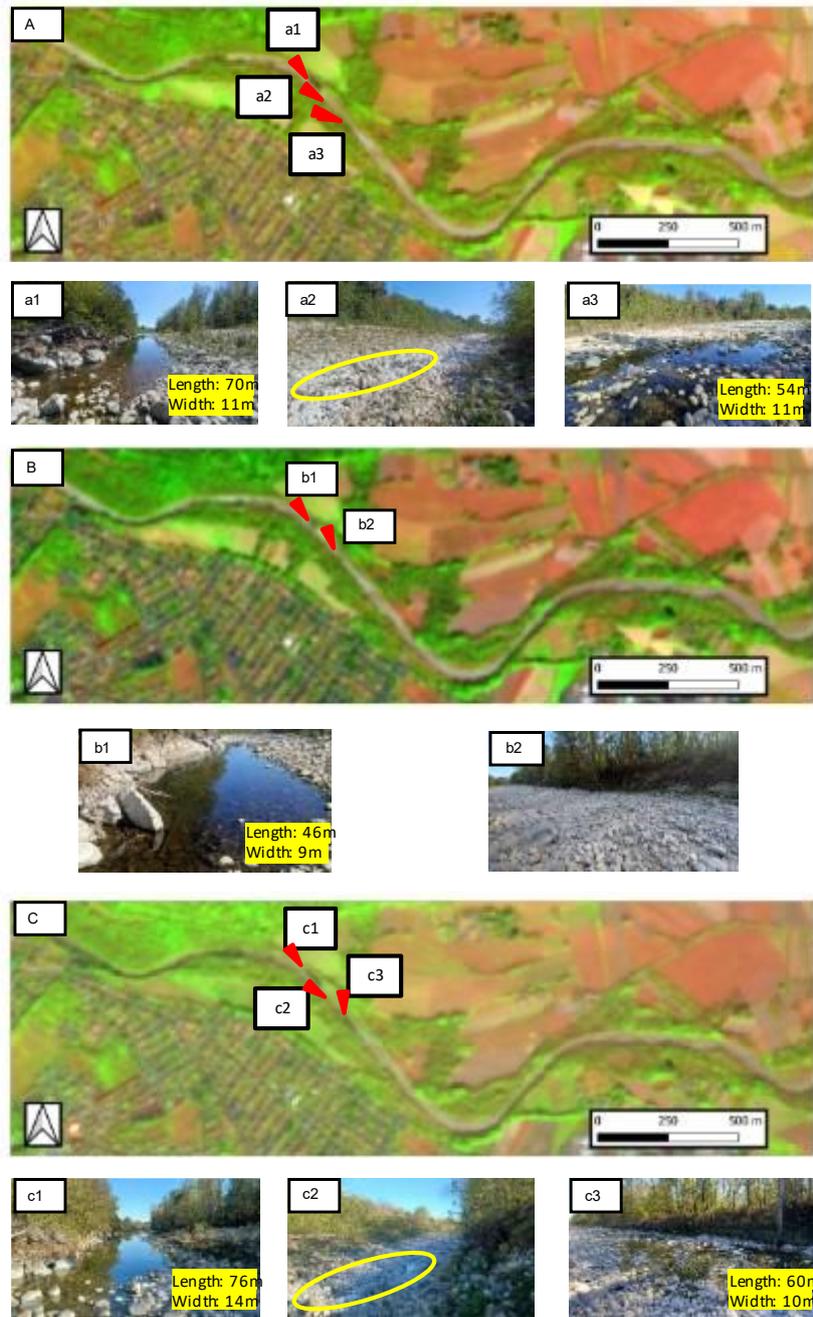


Figure 40. Comparison between FCIs images and field surveys for different date: A is the FCI on 11/09/2022 and a1, a2, a3 are photo taken on field on 11/09/2022; B is the FCI on 16/09/2022 and b1, b2 are photo taken on 16/09/2022; C is the FCI on 03/10/2022 and c1, c2, c3 are photo taken on 03/10/2022. The FCIs images are in scale 1:10000, the same used for the supervised classification. The portion of Sangone river investigated in this figure corresponds to the segment 1.

Figure 40 allows underlining of the visualization capacity of FCIs images compared to the ground through datasets. Table 10 reports the dimension of different pools.

Table 10. The table reports the dimension of the pools acquired during the field surveys related to figure 40.

Date	Pool	Length [m]	Width [m]
11/9/2022	a1	70	11
	a2	54	8
16/09/2022	b1	46	9
03/10/2022	c1	76	14
	c2	60	10

The evolution of the upper pool presented during all the field surveys (a1, b1, c1) is visible with the FCI images. The dark spots changes correspond to the pool shift extension due to the different dimensions detected on the three days. It is meaningful evidence of the high capacity to recognize presence of water and its changes over time.

Also, the second pool, present only during the first and the last field survey, is visible from the FCIs images even if the vegetation cover and its shallowness determine a less pronounced prominence concerning the first one.

On the other hand, the small and shallow water rivulet that connects these two pools on the date 11/09/2022 and 03/10/2022 are impossible to be detected visualizing only the FCIs image. It is evidence of the limits a supervised classification can encounter in detecting an Oligorheic aquatic state.



Figure 41. Comparison between FCIs images and field surveys for different date: D is the FCI on 11/09/2022 and d1, d2, d3 are photo taken on 11/09/2022; E is the FCI on 16/09/2022 and e1, e2, e3 are photo taken on 16/09/2022; F is the FCI on 03/10/2022 and f1, f2, f3 are photo taken on 03/10/2022. The FCIs images are in scale 1:10000, the same used for the supervised classification. The portion of Sangone river investigated in this figure corresponds to the portion excluded due to the anthropogenic alteration of the flows.

Figure 41 presents other examples of the comparison between FCIs images and field surveys. The Sangone portion considered in this case is the excluded one (cap. 2.2) due to the anthropogenic pressures that altered the natural flows. While a change in extent consistent with weather events prior to the days of the field surveys could be observed for the pools in figure 40, the interconnected pools acquired in this portion of the river show no consistent changes, and the length and width of the pools remain almost unchanged. The lengths and widths for each pool in figure 41 are reported in table 11.

Table 11. The table reports the dimension of the pools acquired during the field surveys related to figure 41.

Date	Pool	Length [m]		Width [m]	
11/9/2022	d1	80		17	
	d2	54		10	
	d3	20	25	3.5	4.5
16/09/2022	e1	80		16	
	e2	55		11	
	e3	22	25	3	4.5
11/9/2022	f1	79		17	
	f2	55		10	
	f3	19	22	3	4.5

Moreover, it is important to underline that the pools, considered in images d3, e3 and f3, do not present the corresponding black spots in the three FCIs. The width between 3 and 4.5 m, combined with the lateral vegetation cover, seems to determine a limit to the possible identification of the water presence only exploiting FCIs.

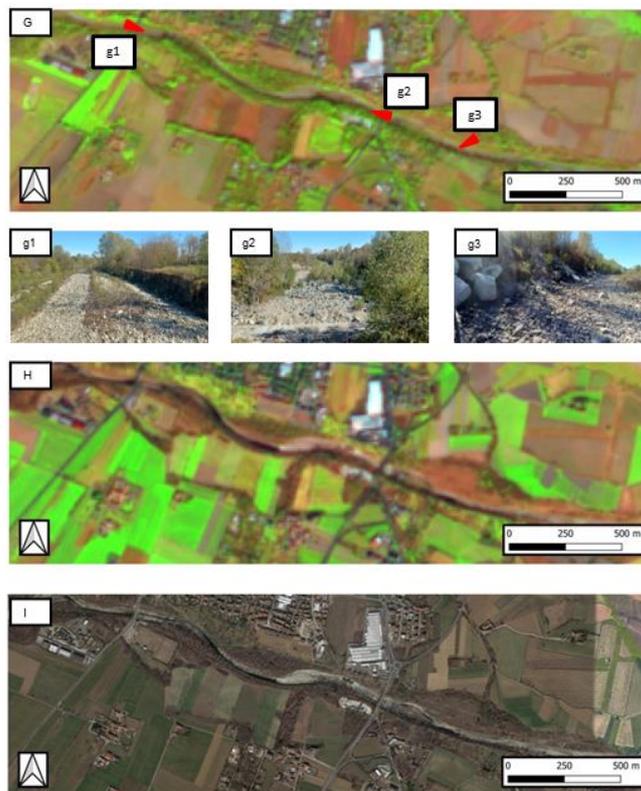


Figure 42. Comparison between FCIs images and field surveys that underlines the possible misclassification of water presence with shadows in the riverbed. G is the FCI on 03/10/2022 and g1, g2 and g3 are photo taken on 03/10/2022 during the field survey. H is the FCI image on 02/03/2022 and I the Google Earth Image on 02/03/2022.

Figure 42 shows the possible misclassification in the supervisor that can occur during the supervised classification. In the FCI image on 03/10/2022 in figure 42, there are dark spots similar to the ones that represent the pools acquired but the photos taken during the field survey demonstrate the dry riverbed in those locations. From the 02/03/2022 FCI and the Google Earth Image on 03/03/2022 in figure 41, it is possible to visualize the normal path of water in the river channel. This a priori information could be very important for the supervisor to avoid errors.

Dataset obtained from supervised classification

The supervised classification of the FCIs between 2015-2021 determined the dataset reported in table 12.

Table 12. Results of supervised classification.

	Segment 1	Segment 2	Segment 3	Average
dry	61	111	81	-
ponding	42	28	48	-
flowing	493	446	459	-
effective revisit time	8.26	8.58	8.49	8.44
percentage of dry days on clear image	10%	19%	14%	14.33%
percentage of ponding days on clear image	7%	5%	8%	6.67%
percentage of flowing days on clear image	83%	76%	78%	79.01%

The percentage of flowing, ponding and dry days are extremely unbalanced, with the flowing days more than 75% for every segment. Furthermore, considering cloudy days, the effective revisit time is between 8 and 9 days. However, before the launch of Sentinel-2B, the revisit time is higher in the first three years 2015-2017. Table 13 shows the trend of revisit time during the year.

Table 13. Trend of revisit time during the years.

year	revisit time [days]			
	Segment 1	Segment 2	Segment 3	Average
2015	13.36	13.36	13.36	13.36
2016	9.38	9.38	9.63	9.47
2017	5.62	5.98	5.70	5.77
2018	4.10	4.24	4.20	4.18
2019	3.72	3.72	3.69	3.71
2020	3.73	3.81	3.81	3.79
2021	4.10	4.10	4.06	4.09

Table 13 shows how the effective revisit time is lower in the last period. The effective revisit time during 2018-2021 is about 3.94, even lower during the dry season when cloudy days are rare. The effective revisit time of fewer than 4 days underscores the power of the Sentinel-2 archive for monitoring a phenomenon, such as changes in flowing status, that could evolve with a speed of a few days. In addition, the next years, which will have a revisit time similar to the 2018-2021 period, would significantly improve the capacity and the accuracy of exploiting Sentinel-2 as a tool for assessing and monitoring the intermittency of a river.

4.4 Random Forest results

Models F/NF and D/P considering flowrate and water table level series

The supervised classification of Sentinel-2 images is, then, used to evaluate the effective prediction capacities of the RF classification models through the statistical parameter of Accuracy, Sensitivity, Specificity and TSS.

The obtained statistical parameters defined a high capacity of RF models to predict the daily flowing statuses for each segment of the Sangone river. Especially, the F/NF models present meaningful results in which the accuracy is between 0.97-0.99 (average 0.98). The sensitivity scores are from 0.98 to 1 (average 0.99), and the specificity is slightly lower between 0.94-0.99 (average 0.965). The TSS parameter is always between 0.94-0.97 (average 0.955).

The second type of model, which classifies NF events between D and P, presents significant values of statistical parameters but is lower than the precedent ones. This model is probably affected by the challenge that the distinction between the ponding and dry phase could present due to the high speed of passing from pools to dry riverbeds and vice-versa. Moreover, these models depend on the first ones' uncertainties that enhance the final values of statistical parameters for these models. However, the classification results return these models with relative accuracy between 0.92-0.97 (average 0.945) and absolute between 0.89-0.95 (average 0.92). The relative sensitivity scores ranged from 0.91-0.94 (average 0.925) and absolute 0.89-0.92 (average 0.905). On the other hand, the relative specificity is always between 0.92-1 (average 0.96) and absolute between 0.86-

0.98 (average 0.92). The relative TSS is in the range of 0.83-0.94 (average 0.885), with an absolute between 0.78-0.89 (average 0.835).

The resultant significant variables and the statistical parameters for each final model are represented in the following table 14.

Table 14. Characterization of binary models, considering flowrate and water table level measures, for each segments through selected most significant variables (in order of importance from top-down direction) and statistical parameters (accuracy, sensitivity, specificity and TSS).

Segment	Model	Significant variables	Accuracy		Sensitivity		Specificity		TSS	
			rel.	abs.	rel.	abs.	rel.	abs.	rel.	abs.
1	F/NF	RIVALTA_EXSTAB_WTL RAINFALL_CUM_10 T_MAX_CUM_MEAN_90 RAINFALL_CUM_30 TRANA_FLOWRATE		0.97		1.00		0.94		0.94
		T_MAX_CUM_MEAN_90 TRANA_FLOWRATE RAINFALL_CUM_10 H_AVR_CUM_MEAN_30 RIVALTA_EXSTAB_WTL	rel.	0.92	rel.	0.92	rel.	0.92	rel.	0.83
	D/P	H_AVR_CUM_MEAN_30 RIVALTA_EXSTAB_WTL RAINFALL_CUM_30 T_MAX_CUM_MEAN_90	abs.	0.89	abs.	0.92	abs.	0.86	abs.	0.78
2	F/NF	RIVALTA_EXSTAB_WTL T_MAX_CUM_MEAN_90 TRANA_FLOWRATE RAINFALL_CUM_30 RAINFALL_CUM_90		0.98		0.98		0.98		0.96
		H_AVR_CUM_MEAN_30 RIVALTA_EXSTAB_WTL RAINFALL_CUM_30 T_MAX_CUM_MEAN_90	rel.	0.97	rel.	0.94	rel.	1.00	rel.	0.94
	D/P	RIVALTA_EXSTAB_WTL RAINFALL_CUM_30 T_MAX_CUM_MEAN_90	abs.	0.95	abs.	0.92	abs.	0.98	abs.	0.89
3	F/NF	RIVALTA_EXSTAB_WTL TRANA_FLOWRATE T_MAX_CUM_MEAN_90 RAINFALL_CUM_30 RAINFALL_CUM_90		0.99		0.99		0.99		0.97
		RIVALTA_EXSTAB_WTL RAINFALL_CUM_30 H_AVR_CUM_MEAN_30 T_MAX_CUM_MEAN_90	rel.	0.95	rel.	0.91	rel.	1.00	rel.	0.91
	D/P	RIVALTA_EXSTAB_WTL RAINFALL_CUM_30 H_AVR_CUM_MEAN_30 T_MAX_CUM_MEAN_90	abs.	0.94	abs.	0.89	abs.	0.99	abs.	0.88

The column with the most significant variables selected for the models highlights the great prediction power of water table level RIVALTA_EXSTAB_WTL that is present in every single model obtained. This result is coherent with literature where, often, groundwater was identified as one of the main drivers for non-perennial rivers.⁷

Also, the temperature, especially T_MAX_CUM_MEAN_90, demonstrates a general and meaningful prediction capacity for both models. This variable is probably the most linked to seasonality and this could be one of the reasons for its importance.

Moreover, as one might have imagined, TRANA_FLOWRATE is a crucial predictor for F/NF models, but due to the difficulties of discharge gauging stations to evaluate the ponding phase, almost no D/P models presents it as a significant variable.

On the other hand, if evapotranspiration effects are negligible for the distinction between F/NF, the phenomenon affects crucially the changes between ponding and dry statuses. Indeed, the relative humidity, in particular the series H_AVR_CUM_MEAN_30, is one of the most important meteorological variables in D/P models. For no F/NF models are meaningful.

One of the cumulative rainfalls is present in any model. RAINFALL_CUM_30 is present in every model, except for the D/P model in segment 1. For segments 2 and 3 F/NF model is the rainfall series with the greatest prediction capacities and this could be interpreted as this temporality, 30 days, is the one meant for the reactivation of the riverbed because contains inherently the information pertaining to the water these segments need to reactivate the flow of water.

Its importance for the D/P models of segments 2 and 3 could be linked to the relationship between the water table level and its recharging phenomena.

However, the two Boolean models for segment 1 present several differences and changes in the variable's importance. In the F/NF model, the RAINFALL CUM 10 becomes the most important between the cumulative rainfall. It is possible to interpret that if RAINFALL CUM 30 is the variable that account for the reactivation of flowing status for segment 2 and 3, for the other one is RAINFALL CUM 10. Indeed, this could be coherent with the field survey and the satellite's images where the first segment presents a faster answer to meteorological events and the less frequent dry status during the year.

Moreover, the D/P model of segment 1 underlines a shift of relevance from the RIVALTA_EXSTAB_WTL to TRANA_FLOWRATE. Also, this result could be seen as consistent due to the greater distance from the observation well in Rivalta concerning the other segments. Moreover, during the field survey, segment 1 presents often an Oligorheic aquatic state with water rivulets that connects pools. Due to the 10m resolution of the satellite's image, it is possible that the supervised classification determines the ponding status of this condition and that could explain the relevance of TRANA_FLOWRATE in this D/P model.

Models F/NF and D/P without flowrate and water table level series

Table 15 resumes the characteristics of the extra models obtained excluding flowrate and water table level series from the original datasets.

Table 15. Characterization of binary models, without flowrate and water table level measures, for segment 2 through selected most significant variables (in order of importance from top-down direction) and statistical parameters (accuracy, sensitivity, specificity and TSS).

Segment	Model	Significant variables	Accuracy		Sensitivity		Specificity		TSS	
2	F/NF	T_MAX_CUM_MEAN_90 RAINFALL_CUM_30 RAINFALL_CUM_90		0.95		0.98		0.93		0.91
	D/P	H_AVR_CUM_MEAN_30 RAINFALL_CUM_30 RAINFALL_CUM_90	rel.	0.96	rel.	0.92	rel.	1.00	rel.	0.92
			abs.	0.91	abs.	0.90	abs.	0.93	abs.	0.83

The results of both RF classifications highlight the modest decrease in statistical parameters compared to the corresponding original models. The difference between corresponding values goes from 0 (the sensitivity of the F/NF model and the relative specificity of the D/P model) to 0.6 (the absolute TSS for D/P).

The difference underlines the importance of enhancing the presence of gauging stations and observation wells along intermittent reaches of the river network to improve the prediction capacities of daily flowing statuses and their monitoring. On the other hand, the models still present a solid prediction capacity that allows the possibility to apply now even when only meteorological series are collected.

4.5 Threshold values for effective flowing status monitoring: results and analysis

The Partial Dependence Plots of `TRANA_FLOWRATE` and `RIVALTA_EXSTAB_WTL` for F/NF models in Annex I permit a visualization of how the flowing statuses classifications are affected by each single significant variables chosen.

The figure 43 shows the flowing and not-flowing observations with the corresponding values of flowrate and water table level for all three segments.

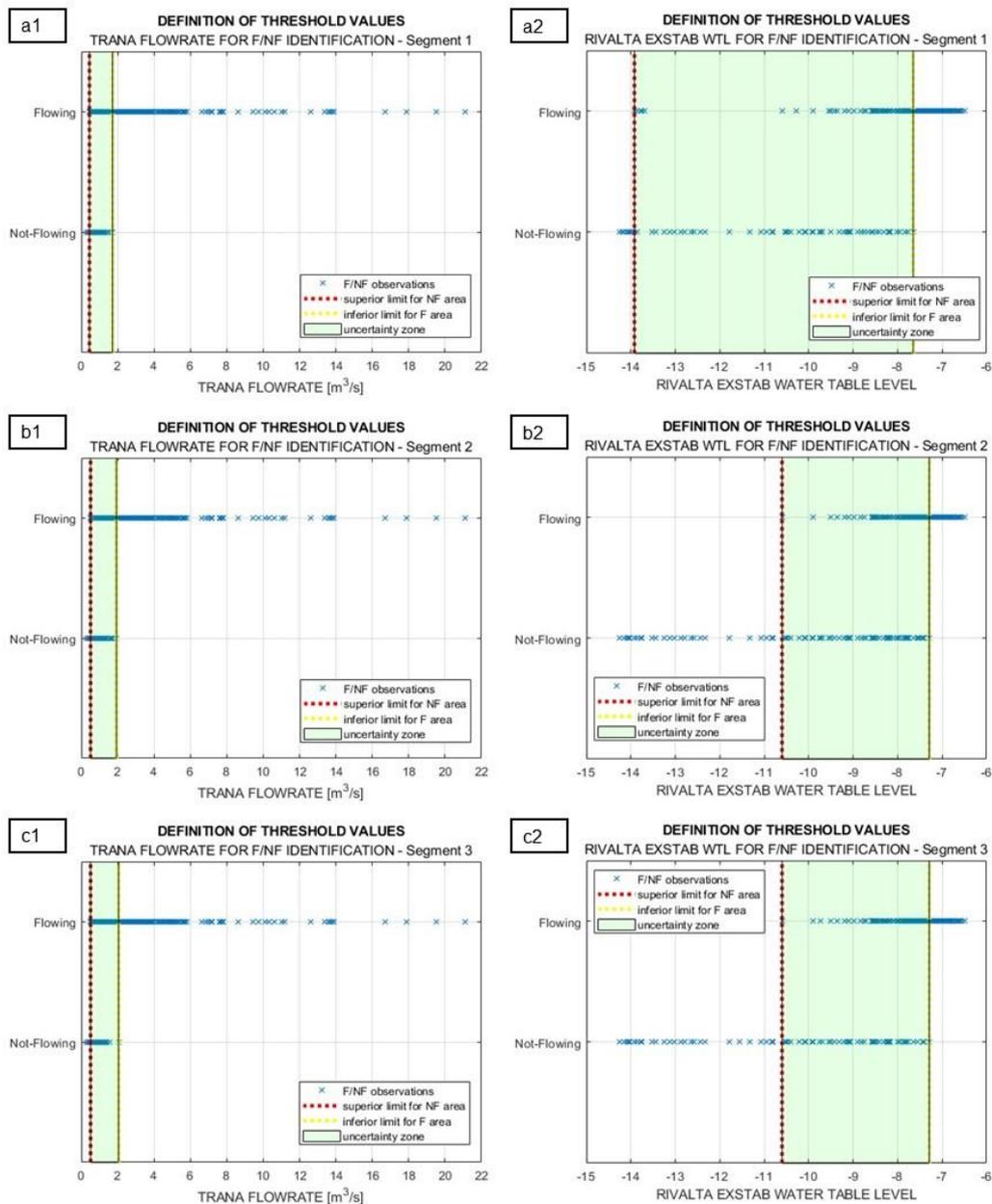


Figure 43. Using of `RIVALTA_EXSTAB_WTL` and `TRANA_FLOWRATE` to define threshold values for underline not flowing and flowing area. In green the uncertainty zone where there is the overlapping of statuses: a1 flowrate thresholds for segment 1, a2 water table level thresholds for segment 1, b1 flowrate thresholds for segment 2, b2 water table level thresholds for segment 2, c1 flowrate thresholds for segment 3 and c2 water table level thresholds for segment 3.

From graphs in figure 43, it is possible to underline the presence of two different regions in which it is possible to predict the flowing status from the value of the single variable (flowrate or water table level). Only with segment 1, the most far away from the observation wells considered, the definition determines a wider uncertainty zone in which there is overlapping of flowing and not-flowing presence. This result is coherent with the less importance of RIVALTA_EXSTAB_WTL seen for this segment with RF models. The threshold values for uncertainty zones are reported in table 16.

Table 16. The table reports the significant values of flowrate and water table level underlined in figure 43.

Segment	Variable	Uncertainty zone	
		Inf. limit	Sup. Lim
1	Flowrate [m ³ /s]	0.45	1.72
	Water Table Level [m]	-13.92	-7.65
2	Flowrate [m ³ /s]	0.51	1.94
	Water Table Level [m]	-10.6	-7.29
3	Flowrate [m ³ /s]	0.51	2.05
	Water Table Level [m]	-10.6	-7.29

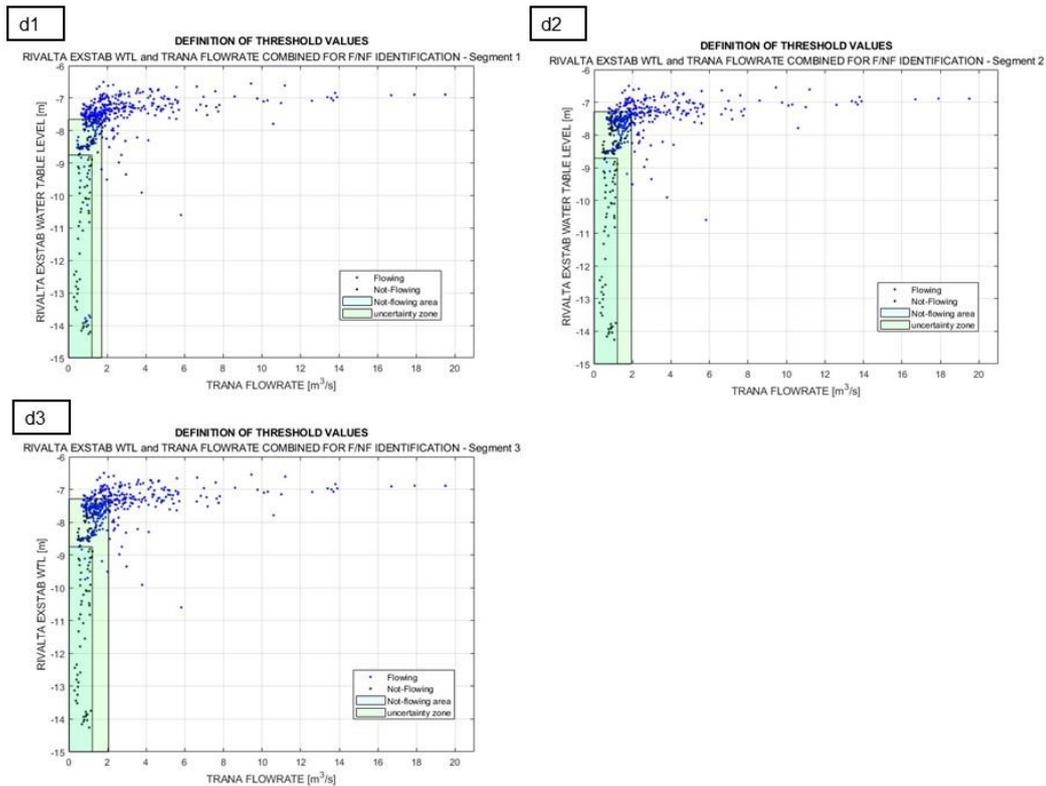


Figure 44. Scatterplots of RIVALTA_EXSTAB_WTL and TRANA_FLOWRATE to visualize the not-flowing area and the uncertainty zones for each segment: d1 segment 1, d2 segment2 and d3 segment 3. In cyan the not flowing area and in green the uncertainty zones.

Anyway, all graphs present a wide range of uncertainty where the two statuses can coexist considering only one of each variable. Due to decrease the uncertainty zone, it was built up a scatterplot considering both variables at the same time.

The charts in figure 44 determines a wider area that is characterized by general not-flowing status where only few flowing observations are present events. For segment 2 and segment 3 the percentage of not-flowing events is between 0.95-0.97, the first segment always presents a lower capacity in distinction with a percentage of 0.84. The following table 17 highlight the limits that characterizes the Not-flowing areas and Uncertainty zones for each segment.

Table 17. The table reports the threshold values that define the Not-flowing area and the uncertainty zones.

Segment	Variable	Limits for Not-flowing area	Limits for Uncertainty zone
1	Flowrate [m³/s]	1.21	1.72
	Water Table Level [m]	-8.75	-7.65
2	Flowrate [m³/s]	1.21	1.94
	Water Table Level [m]	-8.71	-7.29
3	Flowrate [m³/s]	1.21	2.05
	Water Table Level [m]	-8.75	-7.29

Nevertheless, the threshold values that defines the uncertainty zones, that could be seen as the conditions for which the not-flowing statuses starts to compare in the segments, and the threshold values for the not-flowing areas, that could be seen as the conditions for which is possible to define in not-flowing conditions the segments, could be use from the body in charge to define a possible scale of levels, similar to the one used for managing flood events, through which regulate on time the withdrawals and pumping wells, adapting to the flowing statuses present on time in the riverbed.

4.6 Duration of flowing status

Predictions of flowing statuses

The accuracy reached with the RF algorithm permits the predictions of flowing statuses on a daily scale. For each segment, the duration of each flow state is shown in the histogram in the figure 45.

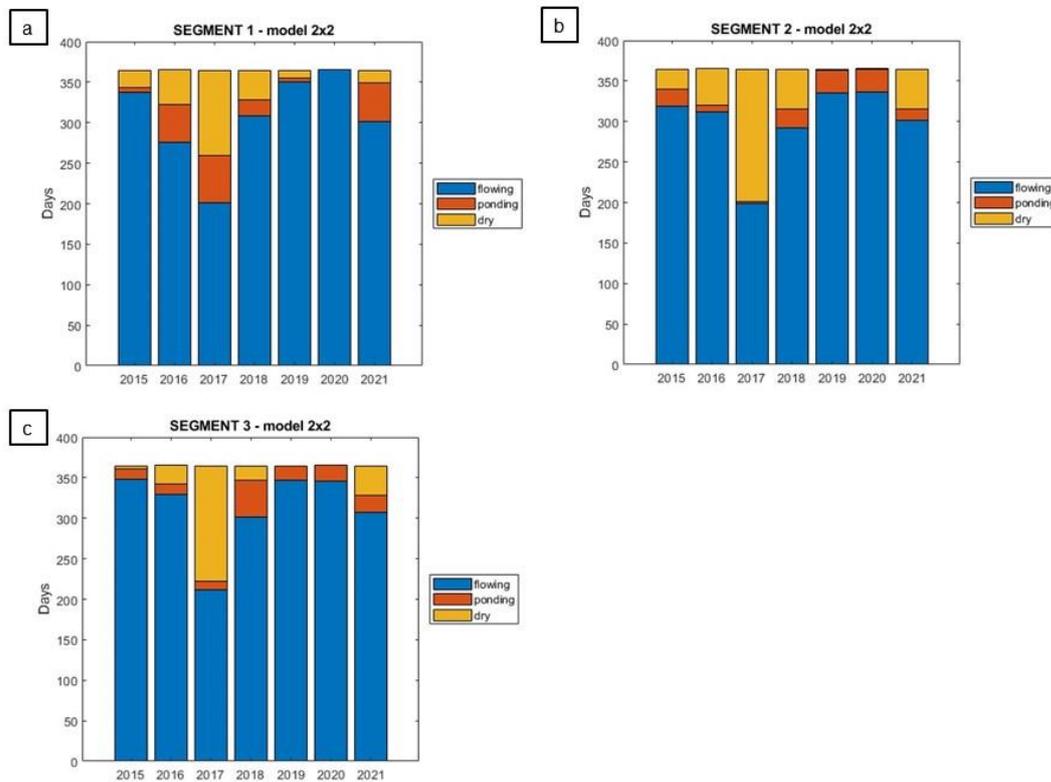


Figure 45. Duration of flowing statuses per year for each segment. a: segment 1, b: segment 2, c: segment 3.

The results underline immediately how the intermittency of a river is variable for a different year and, in addition, the great spatial variability comparing the daily occurrences of each status for these three contiguous segments.

The difference gives back the evidence about how, even if the meteorological events and conditions are equal, the evolution of flowing status is strictly correlated to characteristic of each segment.

Segment 1 is characterized by an amount of non-flow statuses days between 0 and 164. This segment presents a year (2020) in which the RF model does not predict a not-flowing status. On the other hand, the D status has a daily occurrence between 0 to 105 days and the P one between 0 to 59 days. The average daily events per year of each flowing status

are 306 F and 55 NF (22 P,33 D). Although there is a trend toward more dry days rather than days with disconnected pools, compared to the other segments, segment 1 shows a general trend toward a higher frequency of ponding events.

Segment 2 presents a higher amount for each day of not flowing statuses. The average values are 299 F and 66 NF (18 P, 48 D). The variability of NF days goes from 199 to 336 in which the P status varies between 2 to 29 and the D condition between 1 to 164. It is more prone to ceasing water flow events with an average of two months per year of not flowing days and a significant preference to stabilize to dry conditions of riverbed rather than ponding one. It is meaningful how in 2017, with 164 events of dry riverbeds, only two days are predicted in ponding status. Compared with segment 1, there is no year with 0 NF events.

The last segment tends to have a higher duration of flowing statuses. The mean per year is around 313 F and 52 NF (20 P, 32 D). Such as segment 2, there is no year with 0 NF status even if two consequence years (2019 and 2021) present 0 days of D. The variability of P is between 10 to 46, and for D is between 0 to 143.

From a general point of view, it is possible to highlight some general characteristics from daily occurrences. First, the histograms show how 2017 is the year with an extraordinary duration of NF statuses rather than the other years. On the other hand, the biennium 2019-2020 presents the NF days for less than a month. The mean duration of NF for the period 2015-2021 is two months.

The graphs in Annex 2 show the frequency and duration of the flowing statuses during the year. These graphs distinguish between predicted and observed (classified from Sentinel-2 archives) flowing statuses.

For the whole set of graphs, it is possible to define late August and September as the typical NF periods. For years with a longer duration of NF days, the period usually is extended between the end of July and the first half of October with, sometimes, a short number of flowing days in it. Rarely, there are NF days outside this period. The two exceptions are:

- in 2017/2018, the D condition is stable for all three segments from late July 2017 until the first days of January 2018,
- Segment 1 presents a ponding status in February 2016.

Some graphs, like for segment 2, presents in February 2016 an observed ponding phase for a day. The absence other days of NF in that period for this segment could be linked to the limited accuracy of RF's predictions. Especially, the prediction could be more affected by misclassification when the changes between flowing statuses are rapid and when the observed dataset is less frequent. Indeed, 2015 and 2016 are the years before Sentinel-2B launched and, thus, present a much higher revisit time comparing with the last five year of 2015-2021 period.

Classification of regime type with TREHS metrics

The duration of flowing, ponding and dry statuses obtained with the RF allows classifying the regime of each segment with the TREHS classification, comparing the permanence of flow, pools and dry riverbed with the threshold values defined by *Gallart et al. (2017)*.²⁷

The following table 18-20 presents the classification for a single year and the general one considering the whole period between 2015 and 2021.

Table 18. Regime classification of segment 1 through TREHS classification.

	Md	Mp	Mf	Regime
2015	0.06	0.01	0.93	Qp
2016	0.12	0.13	0.75	AF
2017	0.29	0.16	0.55	AF
2018	0.10	0.05	0.84	AF
2019	0.03	0.01	0.96	Qp
2020	0.00	0.00	1.00	Pe
2021	0.04	0.13	0.82	FS
General	0.09	0.07	0.84	FS

Table 19. Regime classification of segment 2 through TREHS classification.

	Md	Mp	Mf	Regime
2015	0.07	0.06	0.87	FS
2016	0.13	0.02	0.85	AF
2017	0.45	0.01	0.55	AF
2018	0.13	0.07	0.80	AF
2019	0.01	0.08	0.92	Qp
2020	0.00	0.08	0.92	Qp
2021	0.13	0.04	0.82	AF
General	0.13	0.05	0.82	AF

Table 20. Regime classification of segment 3 through TREHS classification.

	Md	Mp	Mf	Regime
2015	0.01	0.04	0.95	FS
2016	0.07	0.03	0.90	AF
2017	0.39	0.03	0.58	AF
2018	0.05	0.13	0.82	AF
2019	0.00	0.05	0.95	Qp
2020	0.00	0.05	0.95	Qp
2021	0.10	0.06	0.84	AF
General	0.09	0.05	0.86	FS

As the histograms in the figure 45, the TREHS classification highlights the variability between different years.

Segment 1 presents four different regimes in the 2015-2021 period: from the perennial (Pe) regime in 2020 to the Alternant-Fluent (AF) from 2016 to 2018. The general classification defines a Fluent-Stagnant regime for this segment.

As defined before, the second segment is more prone to NF conditions. There is no perennial regime between these years but at most a quasi-perennial (Qp) in the 2019-2020 biennium. Due to its higher NF duration, this segment presents a general Alternant-Fluent regime classification.

The last segment does not present a perennial regime during the investigated period. Nonetheless, as the first segment, the less frequent presence of NF statuses determines a general Fluent-Stagnant regime classification.

Comparison of models' prediction and classification for segments 2

The graph with the duration of each status for each year underlines the coherence between the model with and without flowrate and water table level as predictive variables.

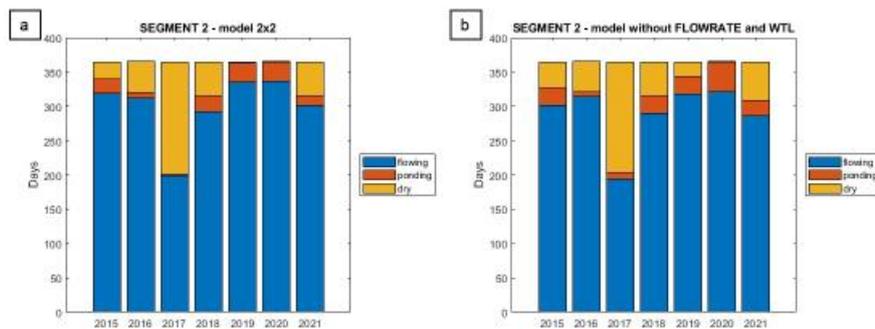


Figure 46. Comparison between resultant flowing statuses durations of segment 2 using flowrate and water table level as predictor variables or not. a: model with flowrate and water table level as predictors, b: model without flowrate and water table level as predictors.

Moreover, it is possible to see a trend to enhance the days that present NF statuses. In addition, this trend is evidenced by the average values of F, P, and D increasing from 299, 18, and 48 with the original models to 289, 23, and 53 with the new. There are ten days of flowing status less and five more for each not flowing status.

From the graph in Annex 2, there is no significant evidence beyond an early beginning of the not-flowing season and a later end. Only the chart for 2019 shows an evident difference in October: the original model predicts an isolated event with pools during the beginning of October; however, the second model presents numerous predictions of ponding (4 days) and dry (10 days). The presence of observed flowing statuses in that period could suggest the misclassification of the second model.

Table 21. Comparison TREHS classification between models on segment 2 with and without flowrate and water table level as predictive variables. The last column is the precedent classification shown in table 13 from the model with the original datasets. In yellow are underlined year with different classification.

	Md	Mp	Mf	Regime	Regime with original datasets
2015	0.10	0.07	0.82	AF	FS
2016	0.12	0.02	0.86	AF	AF
2017	0.44	0.03	0.53	AF	AF
2018	0.14	0.07	0.79	AF	AF
2019	0.06	0.07	0.87	Qp	Qp
2020	0.00	0.12	0.88	FS	Qp
2021	0.15	0.06	0.78	AF	AF
General	0.15	0.06	0.79	AF	AF

Comparison between the classification regime, reported in table 21, gives back a coherent result for General classification on the entire investigation period 2015-2021. Considering the classification for each year, the regimes are coincidental for 5 out of 7 years. This table underlines how little difference in statistical parameters could affect noticeably the investigation of the intermittency regime for a year.

It is important to underline the fact that this kind of classification is no geomorphologic dependent, considering only the presence of water in the riverbed. Moreover, the definition of thresholds could be more consistent updating the critical values after studies that considers regional and local information on morphology, sediments, flooding events and biological knowledge on local river ecosystem.

5. Conclusion

The study has shown the results of exploiting multispectral Sentinel-2 images to assess the flowing status of three consecutive segments of Sangone river during the period 2015-2021. The Sentinel-2 mission was suitable for this purpose because of its high spatial resolution and revisit time of 5 days for each satellite. Due to cloudy days, for the period 2015-2021, the effective revisit time was 9 days, however in the period 2018-2021, after the launch of the twin satellite in 2017, the effective revisit time is reduced to less than 4 days, which is extremely useful for monitoring temporal phenomena for which changes could occur in a few days.

The results of spectral signature analysis of water pools acquired during field surveys have determined a triplet of bands with B11, B8, and B4 that enhance the distinction of water from sediments and river vegetation. Especially, B4 and generally visible bands permit the contrast between sediments and the other two classes. On the other hand, the NIR B8 band is the wavelength through which the distinction between sediments and vegetation. In the end, the SWIR bands are the ones for which there was a greater separability between all the spectral signature but their resolution of 20 m instead of 10 m determine the selection of only one of them, B11.

Comparing the FCIs and ground-truth data revealed the ability of satellite imagery to define the flowing status of segments among three classes: flowing, ponding, and dry. In addition, the comparison highlighted some limitations in visualizing rivulets, typical of the Oligorheic phase, and emphasizing pools with a width less than 10 m wide, especially when lateral vegetation covers them.

The supervised classification applied to FCIs determined a training dataset to evaluate the accuracy of RF classification models to predict the daily status of each segment investigated. The machine learning algorithm used TRANA_FLOWRATE, RIVALTA_EXSTAB_WTL and the spatial interpolated, with Thiessen polygons method, meteorological series of rainfall, temperature (max, mean, min) and relative humidity (max, mean, min) and their own cumulative at 3,5,7,10,30 and 90 days.

Comparison between the predictions of RF models and supervised image training datasets returns significant results. The F/NF models show high accuracy in the range of 0.97-0.99. The sensitivity scores ranging 0.98-1.00, while the specificity is 0.94-0.99. These results determine a TSS value that goes from 0.94 to 0.97. Even if the values of statistical parameters are robust, the values are slightly lower because of the significant difficulties in distinguishing dry and ponding status. The relative accuracy of these models is between 0.92-0.95 (the absolute 0.89-0.95) while the relative sensitivity and the relative specificity scores range respectively 0.91-0.94 (the absolute 0.89-0.92) and 0.92-1.00 (the absolute 0.86-0.99). The final TSS values are between 0.83-0.94 (the absolute 0.78-0.89).

The oversampled double binary models used to identify and predict flowing status from not-flowing status (F/NF models), and then between dry and ponding status (D/P models), determines the water table level as the generally most important predictor variables for all models. Also, the mean of the previous 90-day maximum air temperature has significant predictive power for all models. Moreover, the flowrate presents a significant importance for F/NF models and the 30-days cumulative and, on the other hand, the mean of the previous 30-days-average relative humidity presents for D/P models. The cumulative 30-days rainfall and 90-days rainfall, for Segments 2 and 3, and 10-days rainfall for Segment 1 proved to have good predictive capabilities. Due to the different importance of rainfall series, it is possible to assess for segment 1 the 10-day rainfall is the one responsible for the reactivation of the flowing statuses while 30-days rainfall has the same role for segments 2 and 3.

Furthermore, the model for segment 2, which does not consider point measurements such as flowrate and water table level, gives back a modest decrease in accuracy and coherent results with the precedent. This model presents a slight tendency to overestimate not-flowing days. Nonetheless, due to the typical ungauged condition of non-perennial rivers, these results are crucial to enable a more general use of this method.

Even if the model without flowrate and water table level series as predictive variables shows solid results, the partial plot underlines the flowing status dependency on water table level and flowrate. Exploiting these variables that give an on-time measure concerning the cumulative series, it is possible to determine threshold values that could

be a significant indicator for better management in real time of withdrawals from the river to limit the anthropogenic pressure on modified Sangone intermittency.

The daily flowing status prediction determined high temporal variability in the permanence of the non-flowing status over several years and how consecutive segments may present different flowing statuses on the same day, even if the weather conditions are identical. In addition, not-flowing days can range from 0 to 166 days per year, with segment 1 having a greater tendency toward runoff conditions and segment 2 in the dry status. Using the TREHS classification metric, based on the permanence of different flow states, it was possible to evaluate the flow regime for each segment determining segment 1-3 as Fluent-Stagnant (FS), segment 2 as Alternate-Flowing (AF).

Because of the solid and meaningful results obtained from this model and its easy and inexpensive reproducibility, it is a valid candidate to contribute to enhancing the mapping of non-perennial rivers and to give back solid instruments to evaluate the evolution during time of the temporariness of a river and to improve the management starting in considering the daily flowing status to regulate on-time withdrawals and pumping wells.

ANNEX I – RF Model Interpretation: Key Charts and Metrics for Understanding Model Output

Segment 1

F/NF model Segment 1

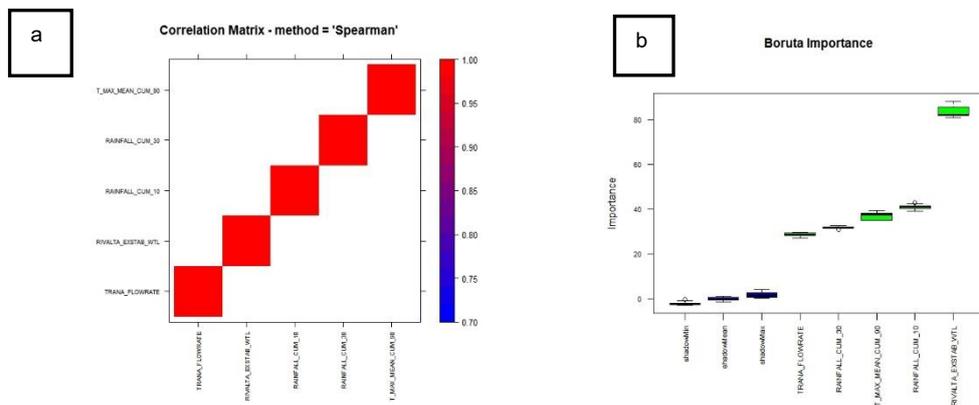


Figure 47. a: Results for Spearman analysis to defines the evaluate the correlation between variables. b: definition of variable importance through Boruta algorithm.

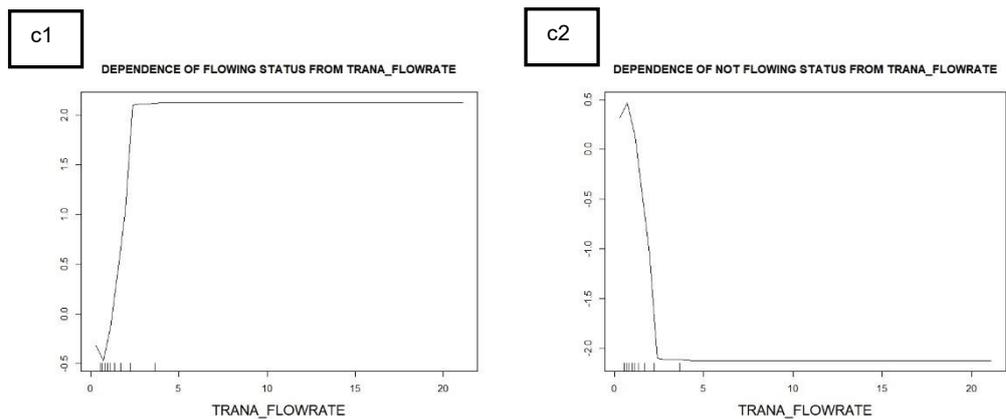
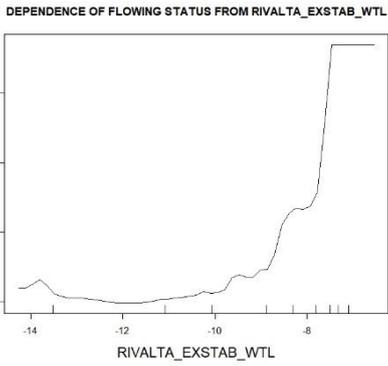


Figure 48. Partial dependence plots for TRANA_FLOWRATE variable: c1, flowing status; c2, not-flowing status.

d1



d2

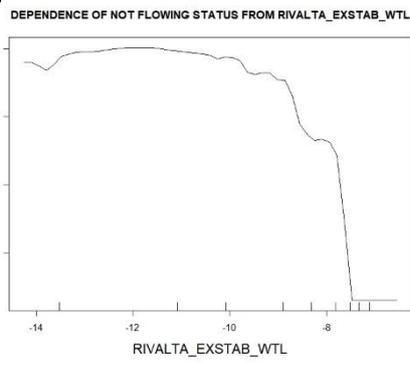
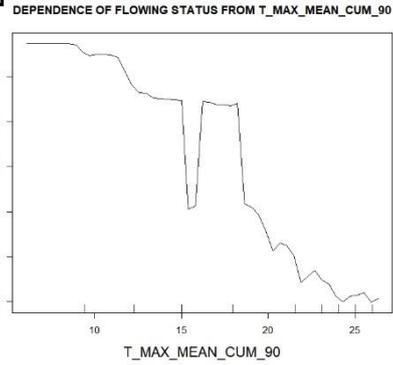


Figure 49. Partial dependence plots for RIVALTA EXSTAB WTL variable: d1, flowing status; d2, not-flowing status.

e1



e2

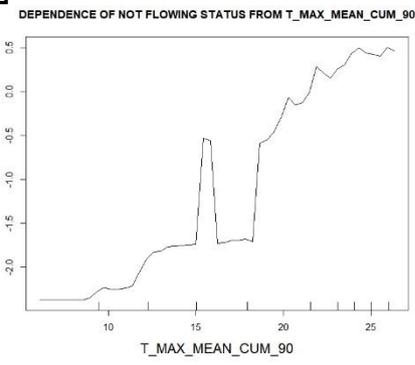
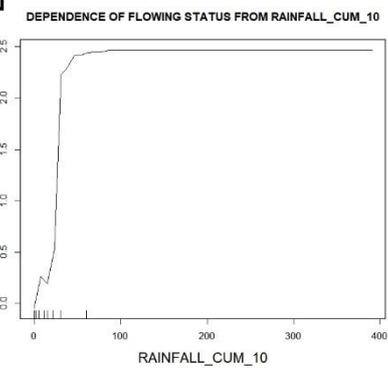


Figure 50. Partial dependence plots for T MAX MEAN CUM 90 variable: e1, flowing status; e2, not-flowing status.

f1



f2

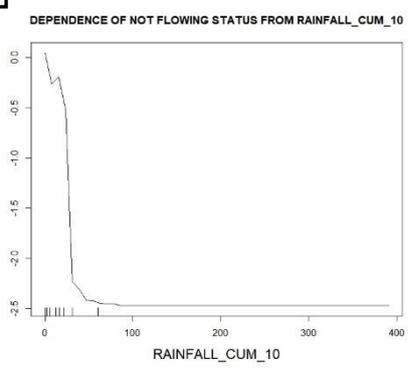
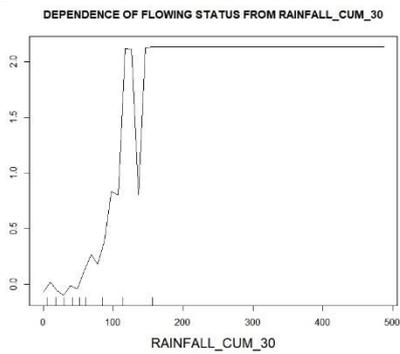


Figure 51. Partial dependence plots for RAINFALL CUM 10 variable: f1, flowing status; f2, not-flowing status.

g1



g2

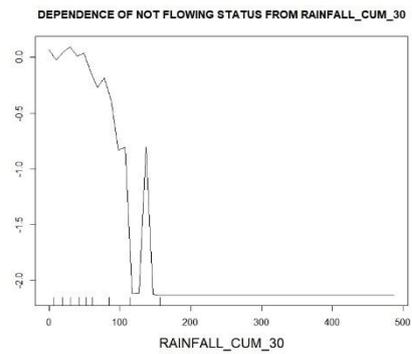


Figure 52. Partial dependence plots for RAINFALL CUM 30 variable: g1, flowing status; g2, not-flowing status.

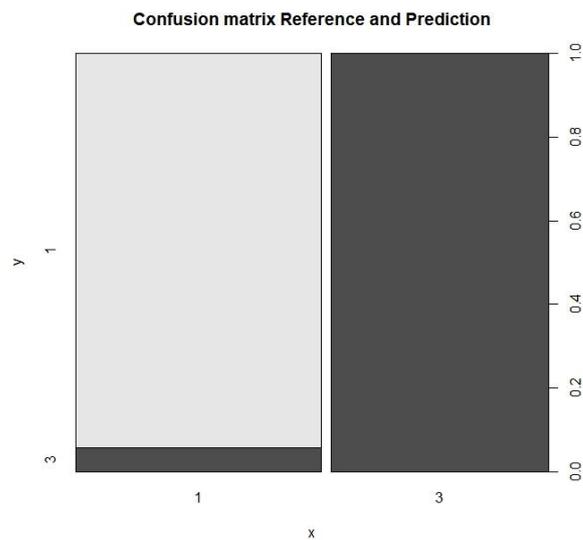


Figure 53. Confusion matrix of F/NF model for segment 1 through which is possible to visualize the accuracy, the sensitivity and the specificity of the classification. 1 is the not flowing status and 3 is the flowing one.

D/P model Segment 1

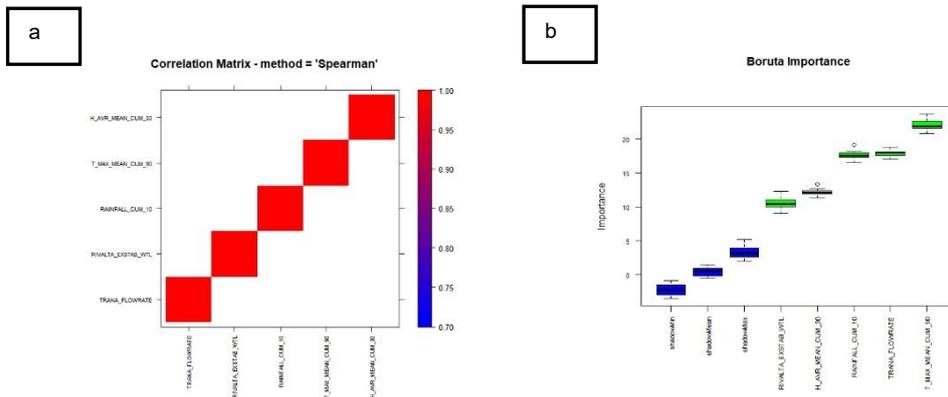


Figure 54. a: Results for Spearman analysis to defines the evaluate the correlation between variables. b: definition of variable importance through Boruta algorithm.

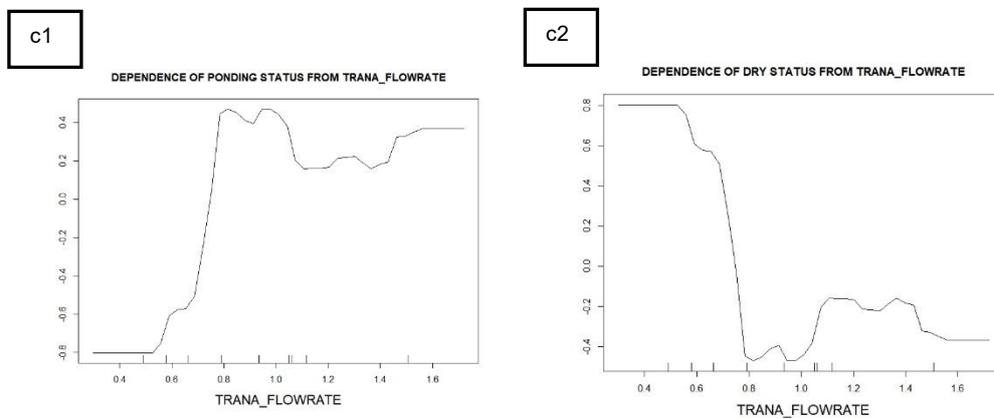


Figure 55. Partial dependence plots for TRANA_FLOWRATE variable: c1, ponding status; c2, dry status.

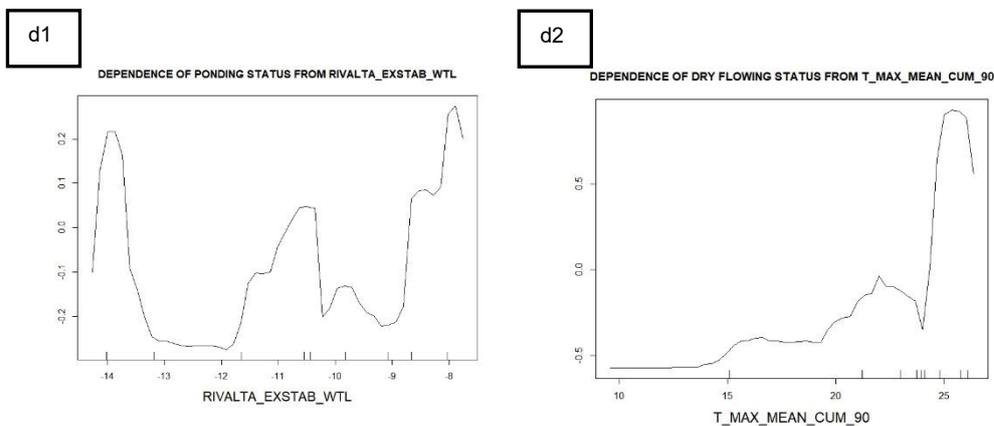
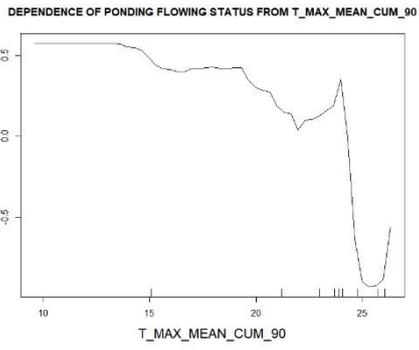


Figure 56. Partial dependence plots for RIVALTA_EXSTAB_WTL variable: d1, ponding status; d2, dry status.

e1



e2

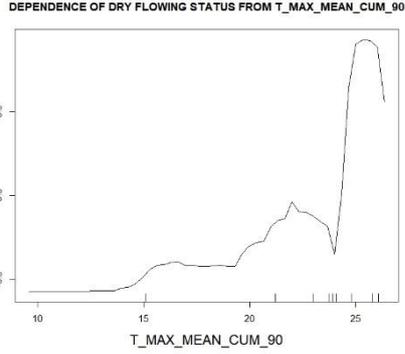
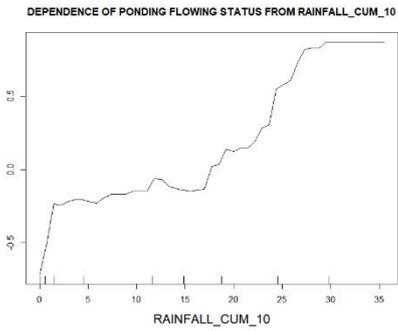


Figure 57. Partial dependence plots for T MAX MEAN CUM 90 variable: e1, ponding status; e2, dry status.

f1



f2

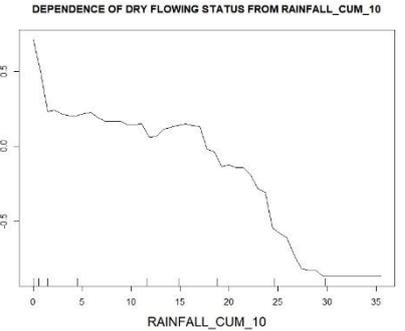
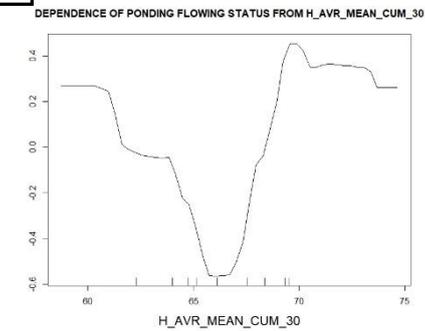


Figure 58. Partial dependence plots for RAINFALL CUM 10 variable: f1, ponding status; f2, dry status.

g1



g2

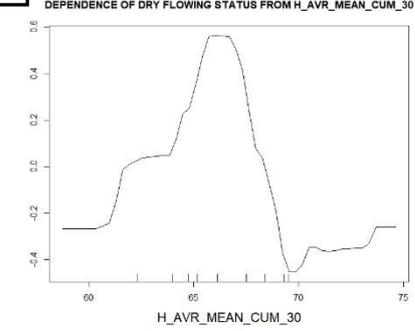


Figure 59. Partial dependence plots for H AVR MEAN CUM 30 variable: g1, ponding status; g2, dry status.

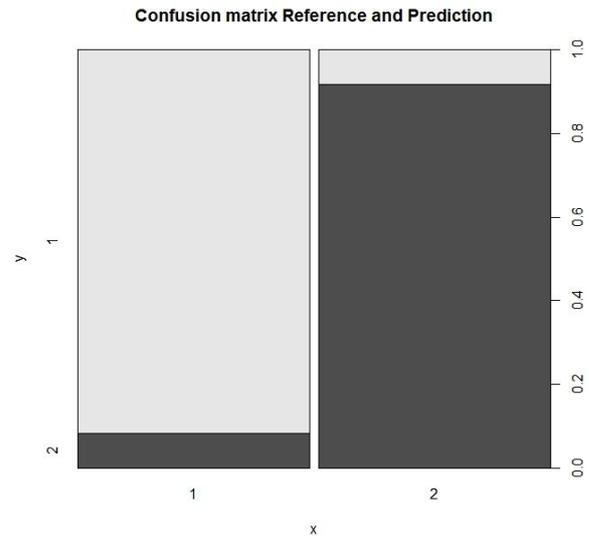


Figure 60. Confusion matrix of F/NF model for segment 1 through which is possible to visualize the accuracy, the sensitivity and the specificity of the classification. 1 is the dry status and 2 is the ponding one.

Segment 2

F/NF model Segment 2

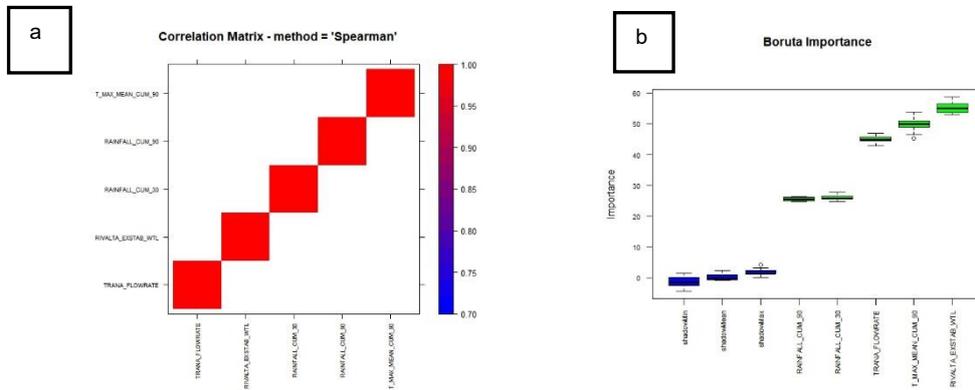


Figure 61. a: Results for Spearman analysis to defines the evaluate the correlation between variables. b: definition of variable importance through Boruta algorithm.

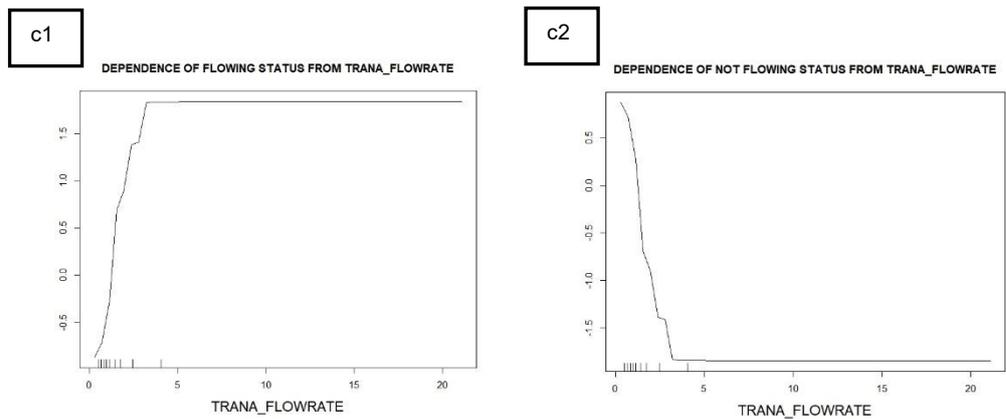
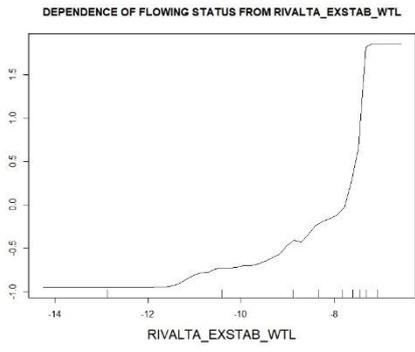


Figure 62. Partial dependence plots for TRANA_FLOWRATE variable: c1, flowing status; c2, not-flowing status.

d1



d2

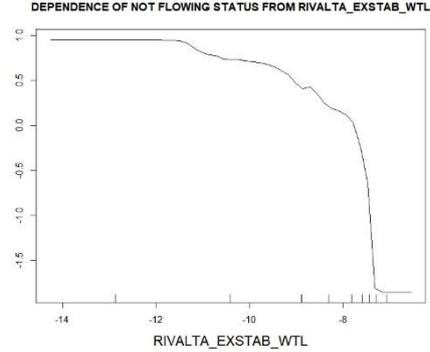
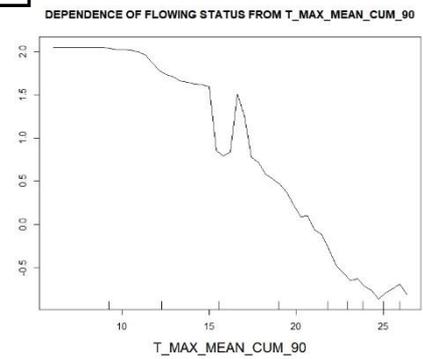


Figure 63. Partial dependence plots for RIVALTA EXSTAB WTL variable: d1, flowing status; d2, not-flowing status.

e1



e2

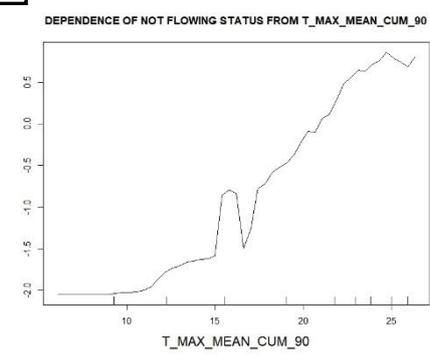
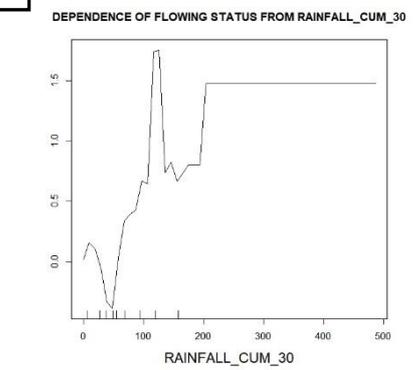


Figure 64. Partial dependence plots for T MAX MEAN CUM 90 variable: e1, flowing status; e2, not-flowing status.

f1



f2

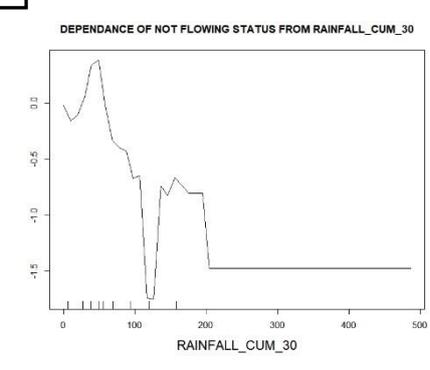
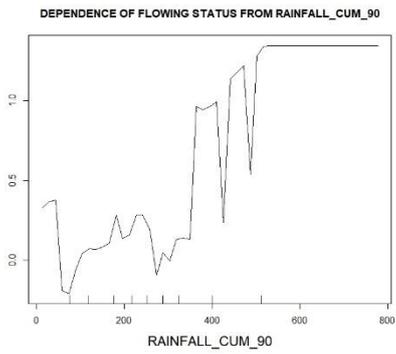


Figure 65. Partial dependence plots for RAINFALL CUM 30 variable: f1, flowing status; f2, not-flowing status.

g1



g2

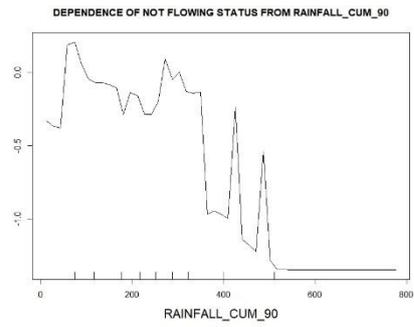


Figure 66. Partial dependence plots for RAINFALL CUM 90 variable: g1, flowing status; g2, not-flowing status.

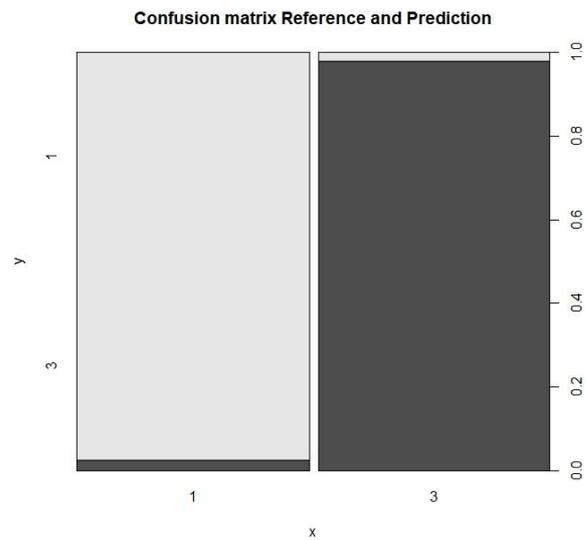


Figure 67. Confusion matrix of F/NF model for segment 2 through which is possible to visualize the accuracy, the sensitivity and the specificity of the classification. 1 is the not flowing status and 3 is the flowing one.

D/P model Segment 2

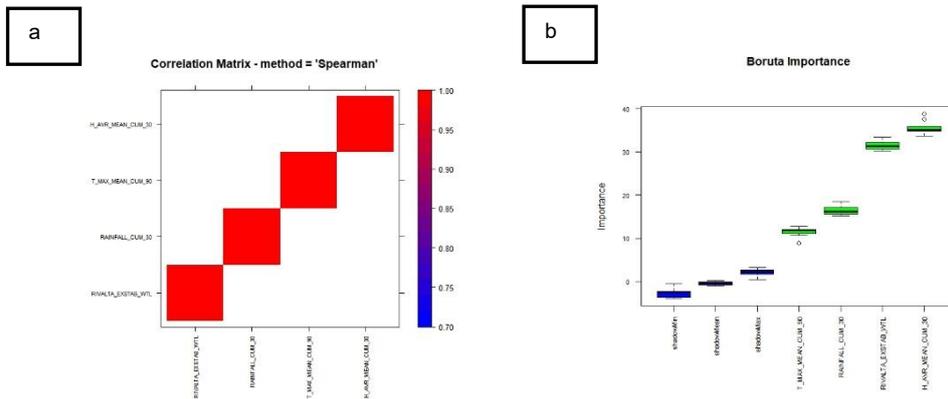


Figure 68. a: Results for Spearman analysis to defines the evaluate the correlation between variables. b: definition of variable importance through Boruta algorithm.

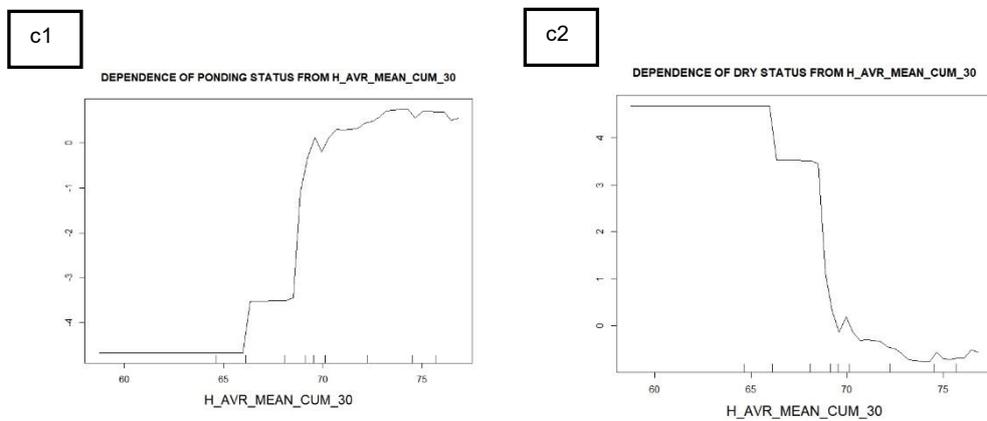


Figure 69. Partial dependence plots for $H_AVR_MEAN_CUM_30$ variable: c1, ponding status; c2, dry status.

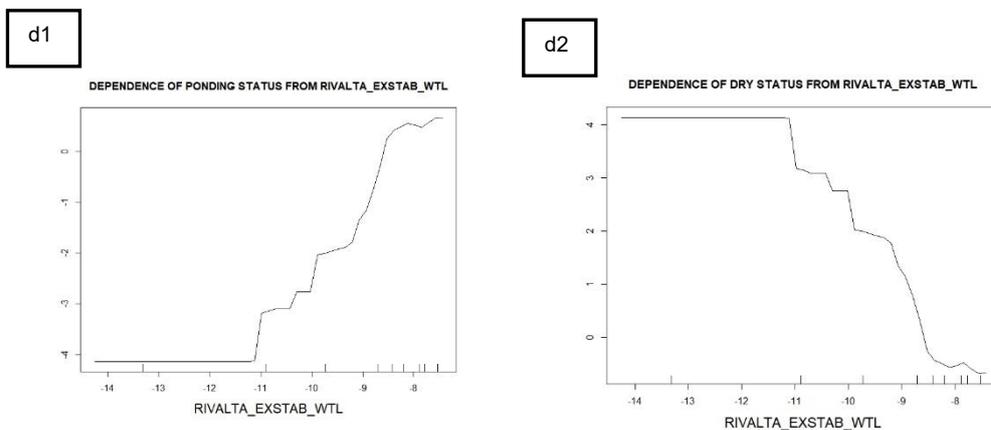
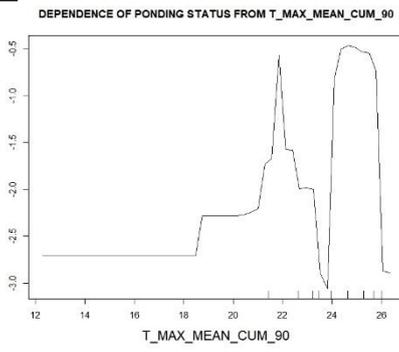


Figure 70. Partial dependence plots for $RIVALTA_EXSTAB_WTL$ variable: d1, ponding status; d2, dry status.

e1



e2

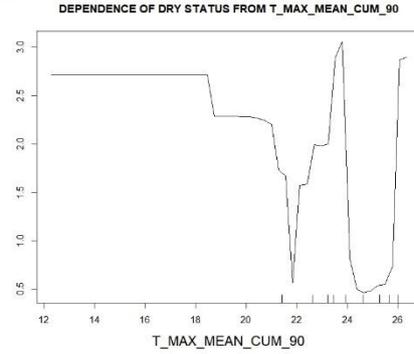
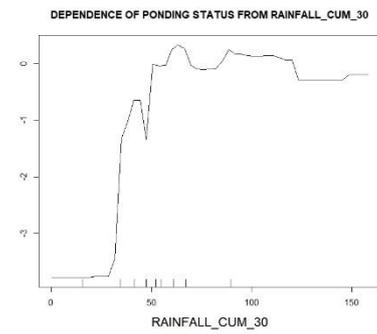


Figure 71. Partial dependence plots for T MAX MEAN CUM 90 variable: e1, ponding status; e2, dry status.

f1



f2

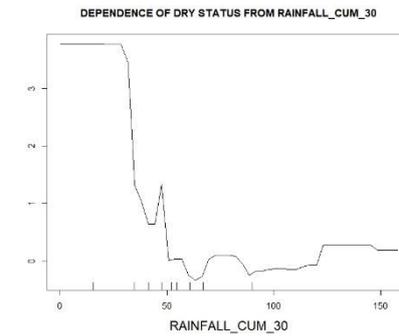


Figure 72. Partial dependence plots for RAINFALL CUM 30 variable: f1, ponding status; f2, dry status.

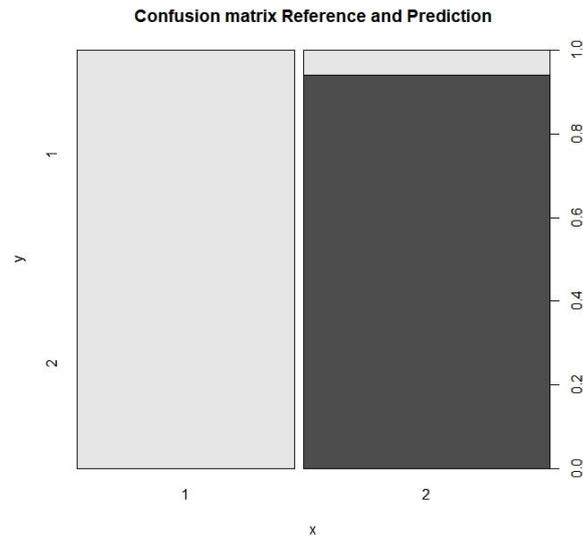


Figure 73. Confusion matrix of D/P model for segment 2 through which is possible to visualize the accuracy, the sensitivity and the specificity of the classification. 1 is the dry status and 2 is the ponding one.

Segment 3

F/NF model Segment 3

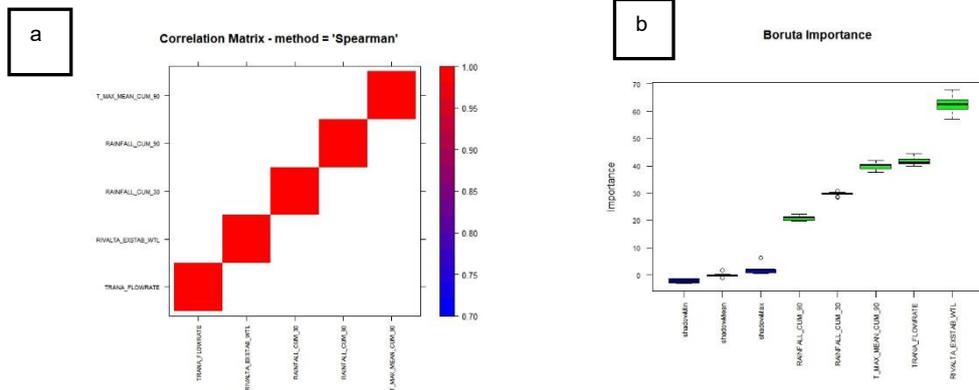


Figure 74. a: Results for Spearman analysis to defines the evaluate the correlation between variables. b: definition of variable importance through Boruta algorithm.

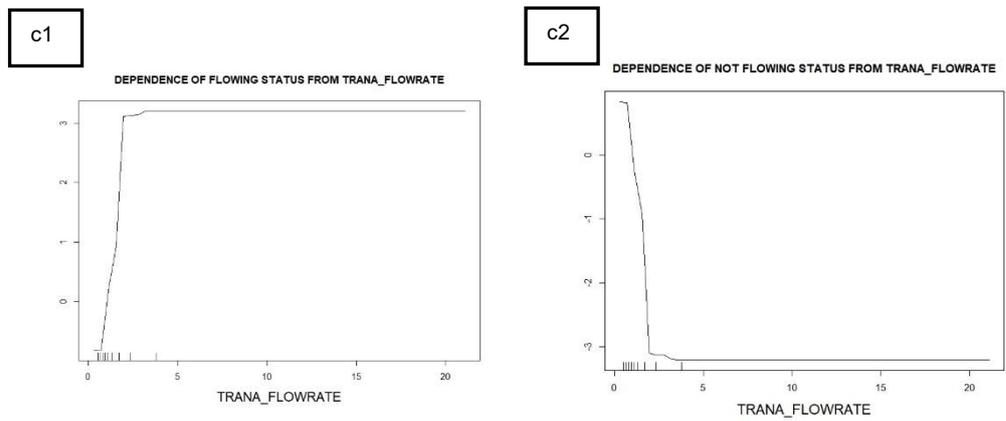
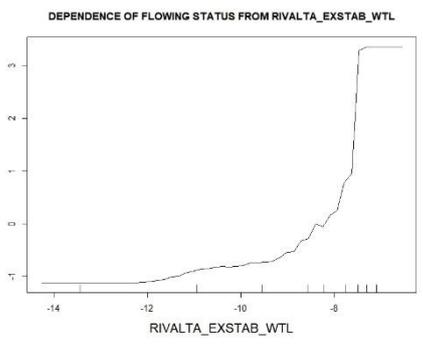


Figure 75. Partial dependence plots for TRANA_FLOWRATE variable: c1, flowing status; c2, not-flowing status.

d1



d2

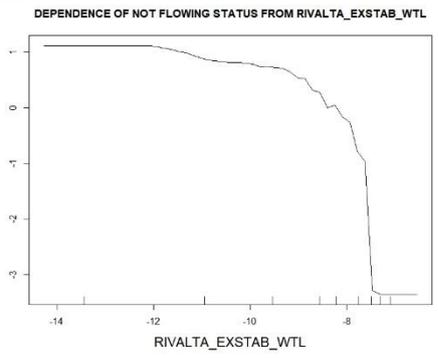
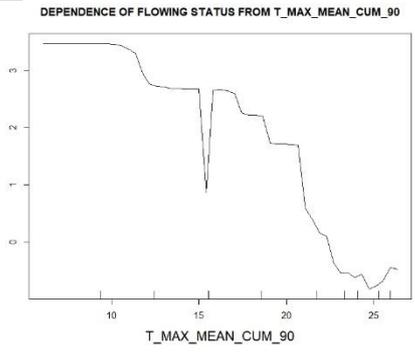


Figure 76. Partial dependence plots for RIVALTA EXSTAB WTL variable: d1, flowing status; d2, not-flowing status.

e1



e2

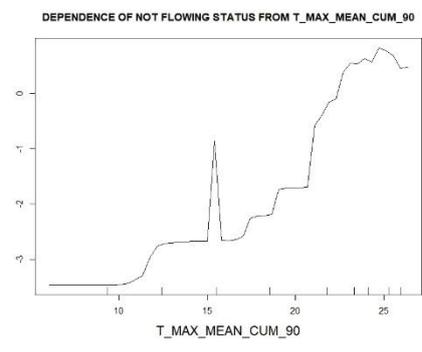
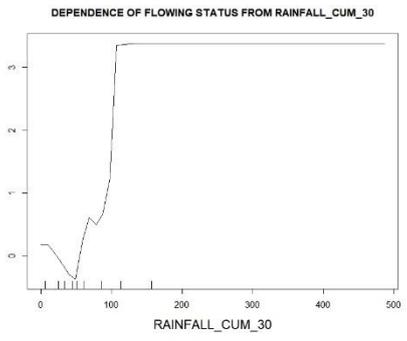


Figure 77. Partial dependence plots for T MAX MEAN CUM 90 variable: e1, flowing status; e2, not-flowing status.

f1



f2

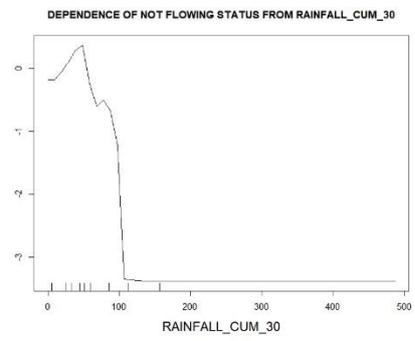
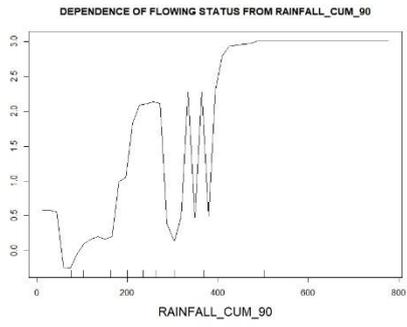


Figure 78. Partial dependence plots for RAINFALL CUM 30 variable: f1, flowing status; f2, not-flowing status.

g1



g2

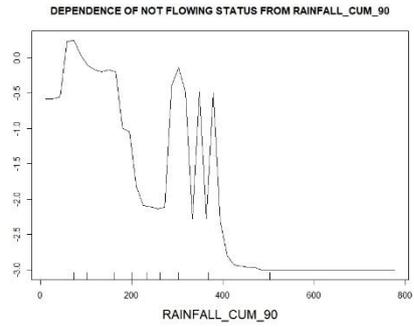


Figure 79. Partial dependence plots for RAINFALL CUM 90 variable: g1, flowing status; g2, not-flowing status.

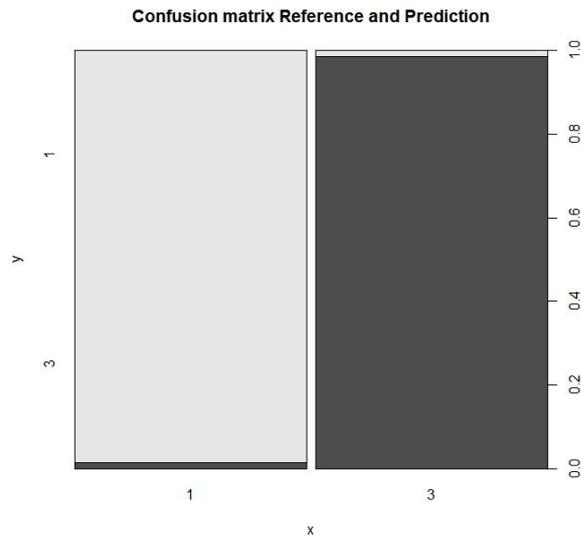
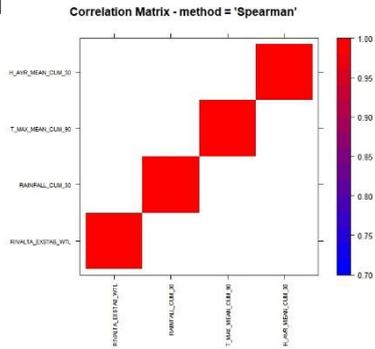


Figure 80. Confusion matrix of F/NF model for segment 3 through which is possible to visualize the accuracy, the sensitivity and the specificity of the classification. 1 is the not flowing status and 3 is the flowing one.

D/P model Segment 3

a



b

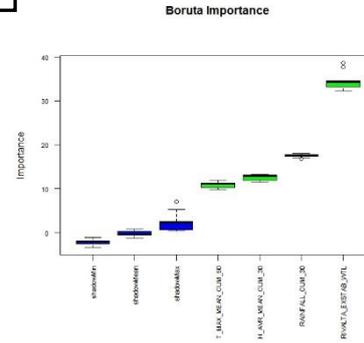
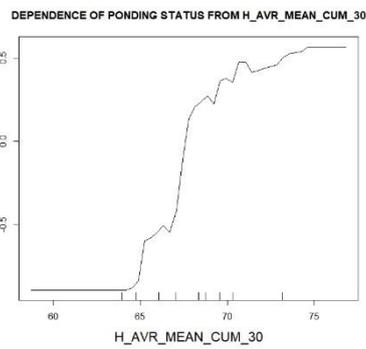


Figure 81. a: Results for Spearman analysis to defines the evaluate the correlation between variables. b: definition of variable importance through Boruta algorithm.

c1



c2

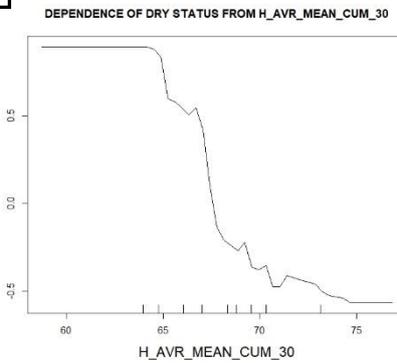
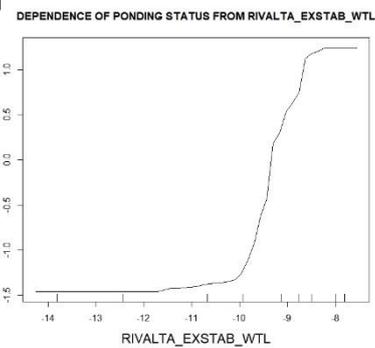


Figure 82. Partial dependence plots for H_AVR_MEAN_CUM_30 variable: c1, ponding status; c2, dry status.

d1



d2

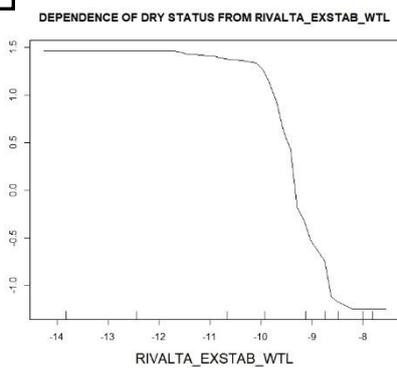
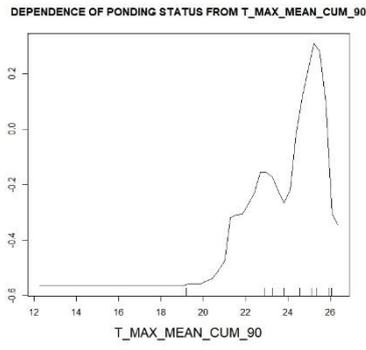


Figure 83. Partial dependence plots for RIVALTA_EXSTAB_WTL variable: d1, ponding status; d2, dry status.

e1



e2

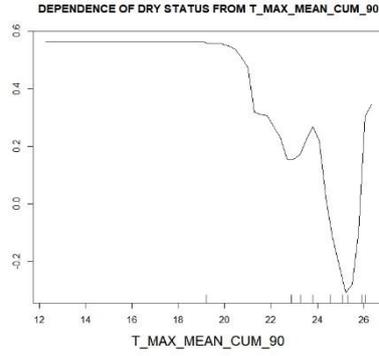
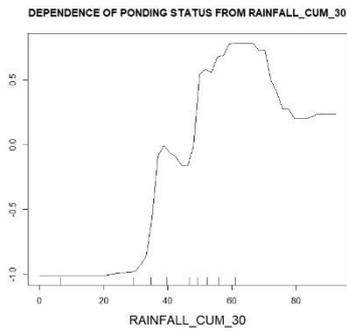


Figure 84. Partial dependence plots for T MAX MEAN CUM 90 variable: e1, ponding status; e2, dry status.

f1



f2

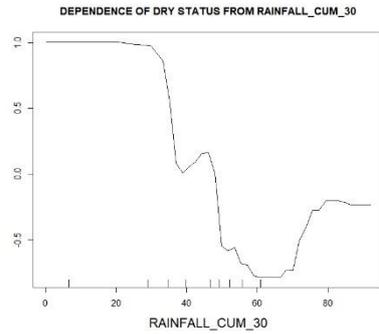


Figure 85. Partial dependence plots for RAINFALL CUM 30 variable: f1, ponding status; f2, dry status.

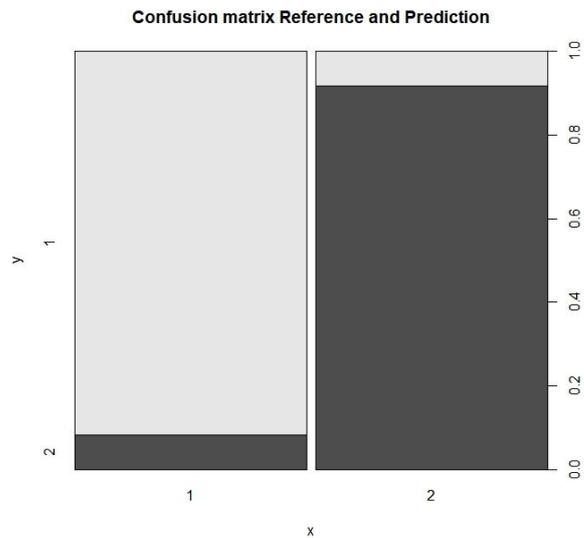


Figure 86. Confusion matrix of D/P model for segment 3 through which is possible to visualize the accuracy, the sensitivity and the specificity of the classification. 1 is the dry status and 2 is the ponding one.

Segment 2 without flowrate and water table level

F/NF model Segment 2 without flowrate and water table level

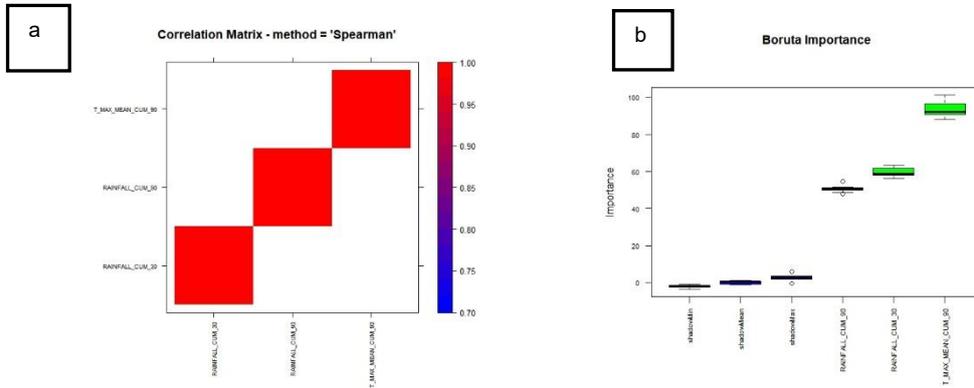


Figure 87. a: Results for Spearman analysis to define the evaluate the correlation between variables. b: definition of variable importance through Boruta algorithm.

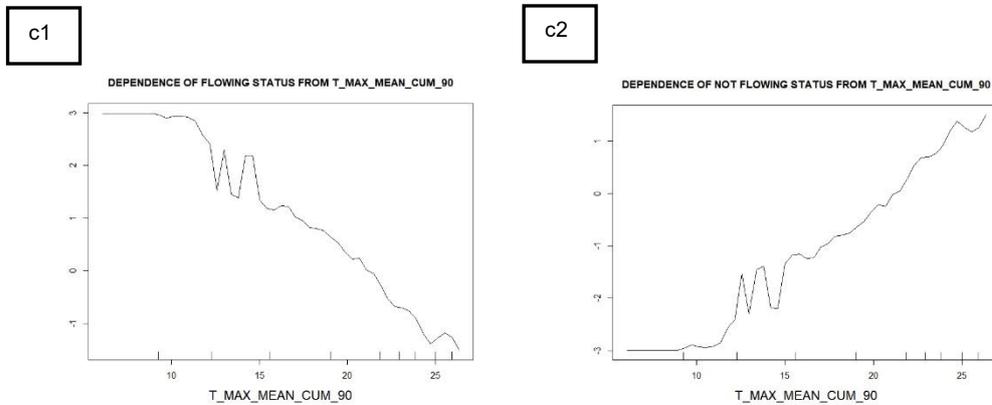


Figure 88. Partial dependence plots for T MAX MEAN CUM 90 variable: c1, flowing status; c2, not-flowing status.

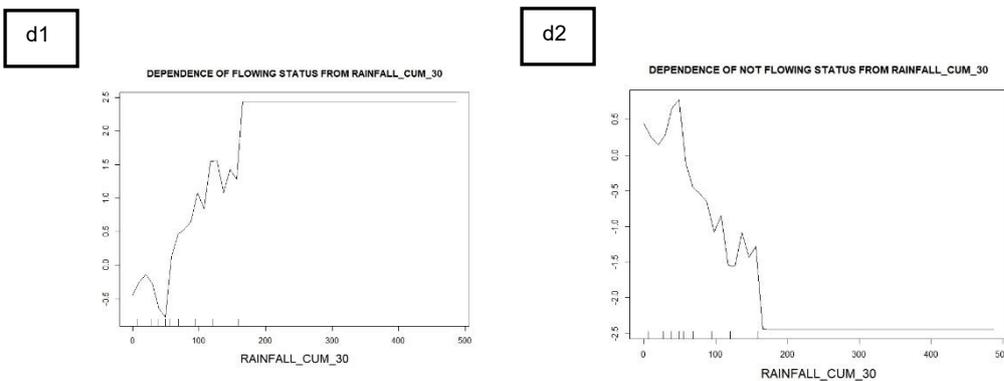
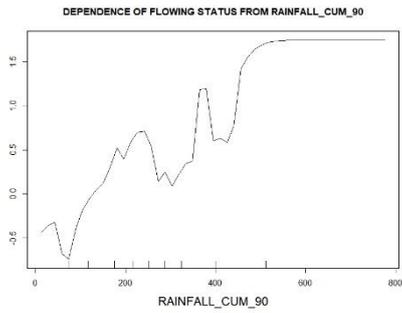


Figure 89. Partial dependence plots for RAINFALL CUM 30 variable: d1, flowing status; d2, not-flowing status.

e1



e2

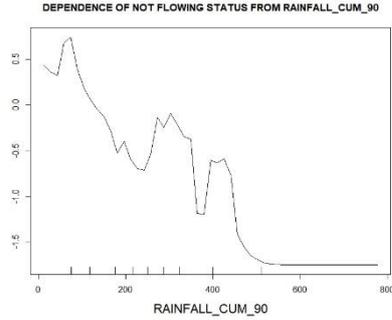


Figure 90. Partial dependence plots for RAINFALL CUM 90 variable: e1, flowing status; e2, not-flowing status.

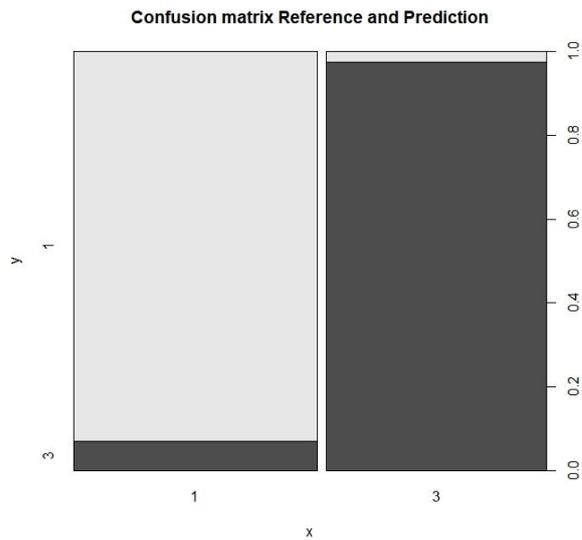


Figure 91. Confusion matrix of the second F/NF model for segment 3 through which is possible to visualize the accuracy, the sensitivity and the specificity of the classification. 1 is the not flowing status and 3 is the flowing one.

D/P model Segment 2 without flowrate and water table level

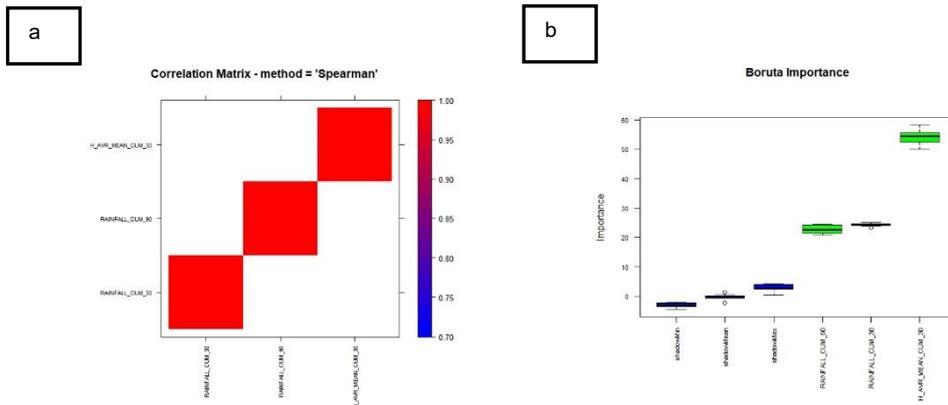


Figure 92. a: Results for Spearman analysis to defines the evaluate the correlation between variables. b: definition of variable importance through Boruta algorithm.

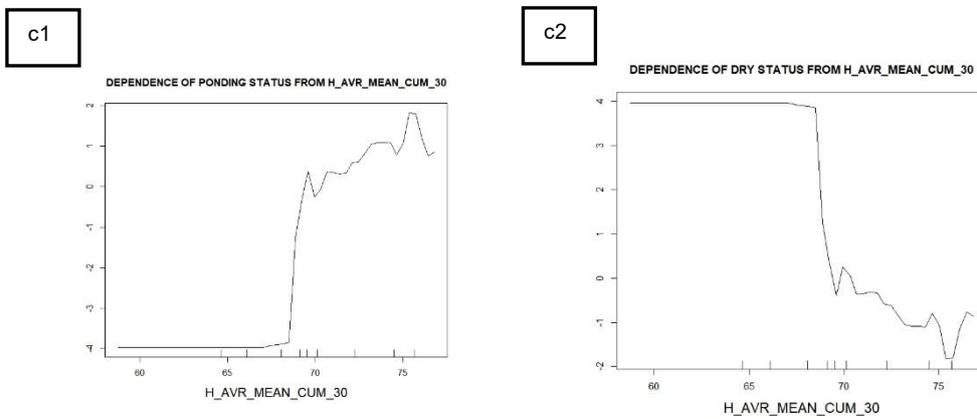


Figure 93. Partial dependence plots for $H_AVR_MEAN_CUM_30$ variable: c1, ponding status; c2, dry status.

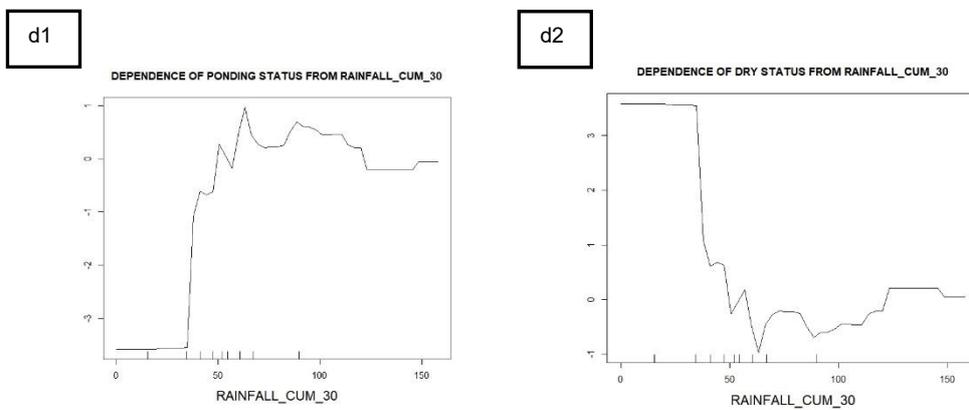
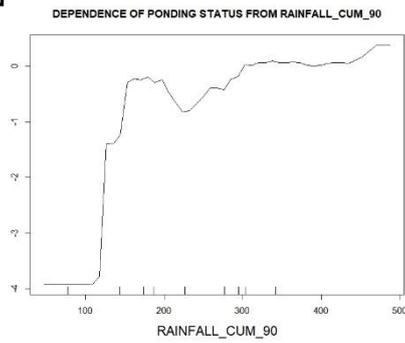


Figure 94. Partial dependence plots for $RAINFALL_CUM_30$ variable: d1, ponding status; d2, dry status.

e1



e2

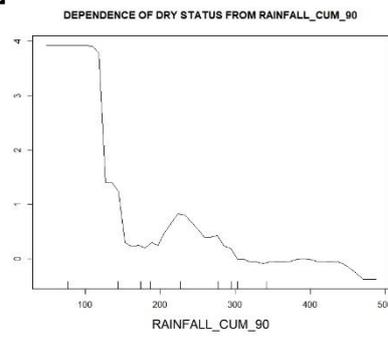


Figure 95. Partial dependence plots for RAINFALL CUM 90 variable: e1, ponding status; e2, dry status.

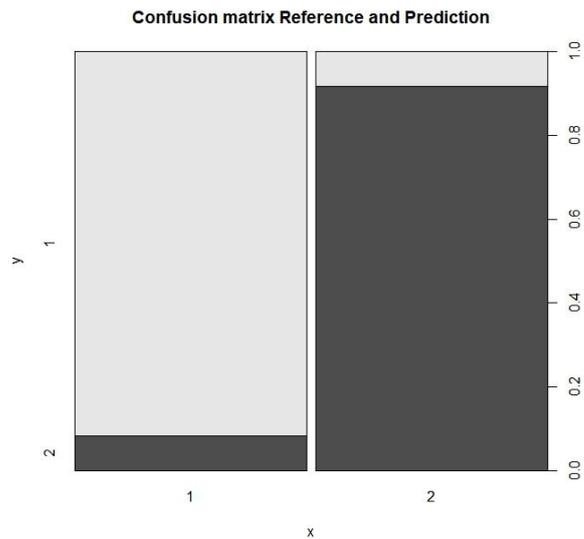


Figure 96. Confusion matrix of the second D/P model for segment 2 through which is possible to visualize the accuracy, the sensitivity and the specificity of the classification. 1 is the dry status and 2 is the ponding one.

ANNEX II – Daily Flowing Status: Predicted vs Observed for 2015-2021 years

Segment 1

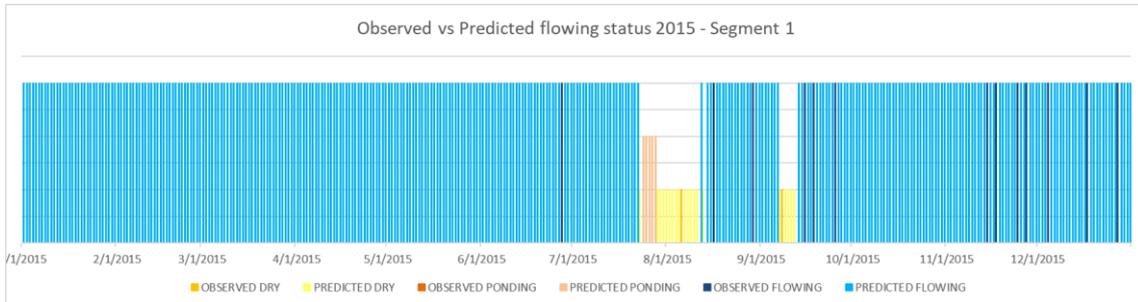


Figure 97. Evolution of observed and predicted flowing status for segment 1 during 2015.

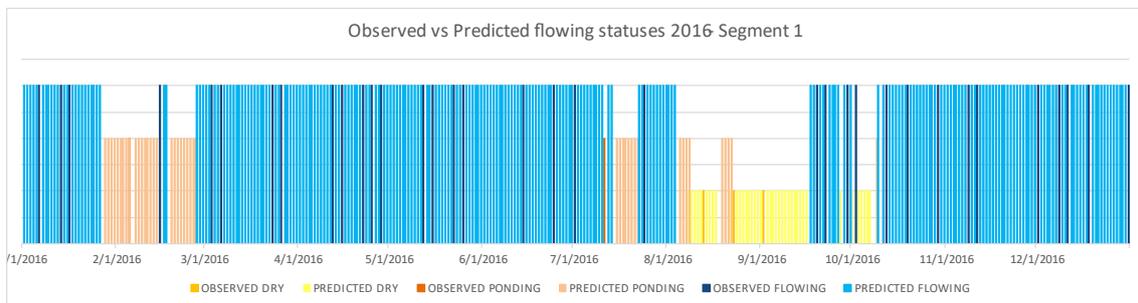


Figure 98. Evolution of observed and predicted flowing status for segment 1 during 2016.

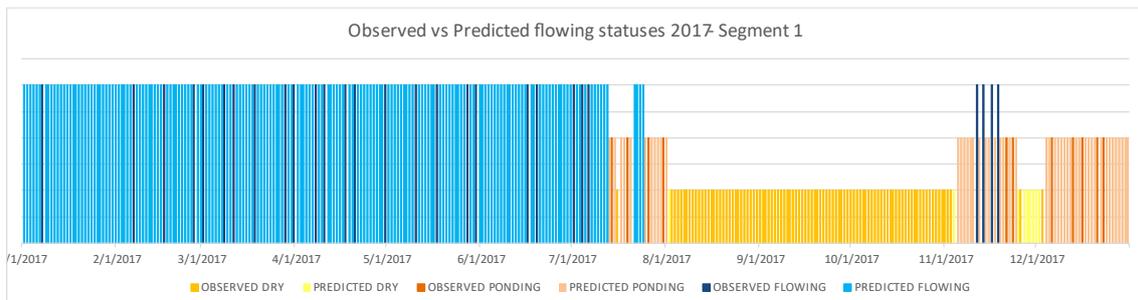


Figure 99. Evolution of observed and predicted flowing status for segment 1 during 2017.

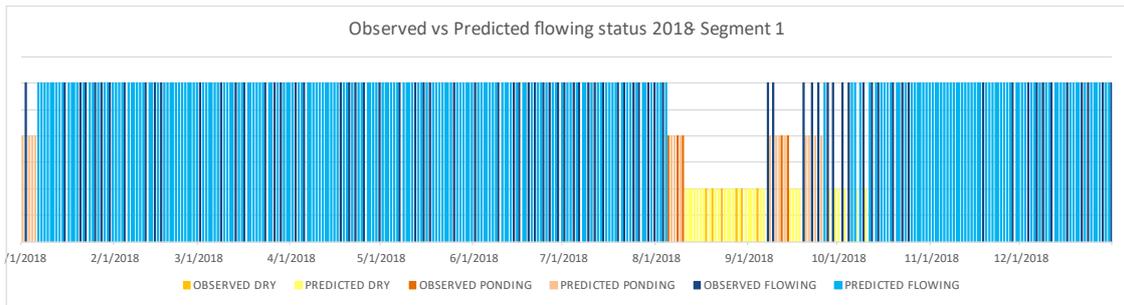


Figure 100. Evolution of observed and predicted flowing status for segment 1 during 2018.

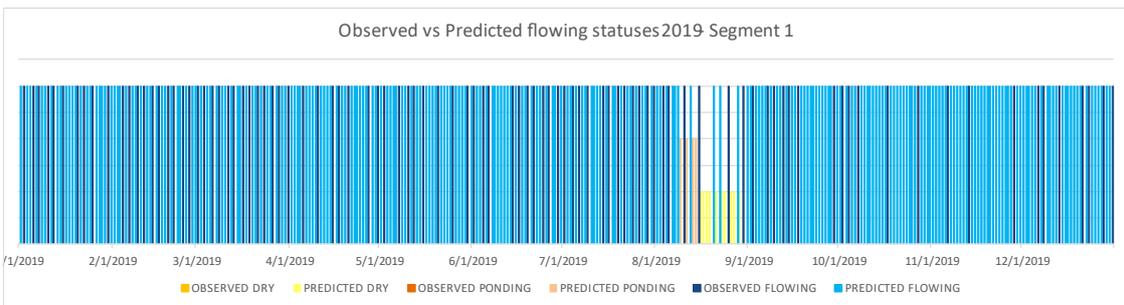


Figure 101. Evolution of observed and predicted flowing status for segment 1 during 2019.

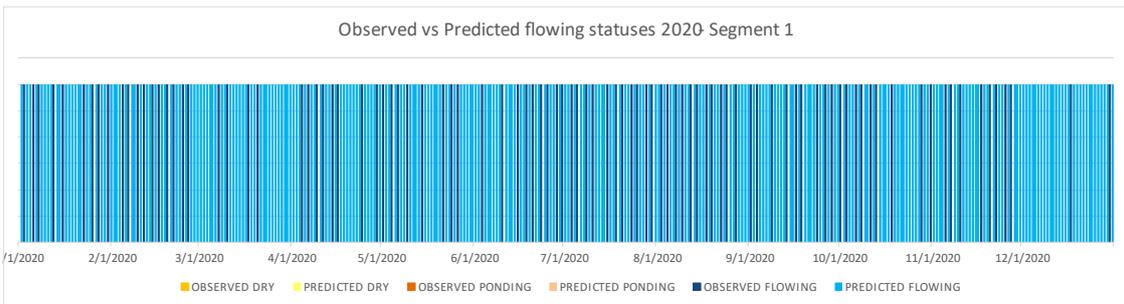


Figure 102. Evolution of observed and predicted flowing status for segment 1 during 2020.

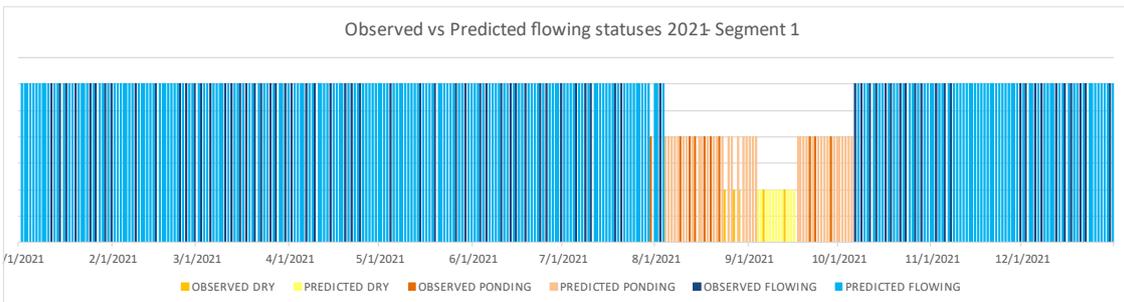


Figure 103. Evolution of observed and predicted flowing status for segment 1 during 2021.

Segment 2

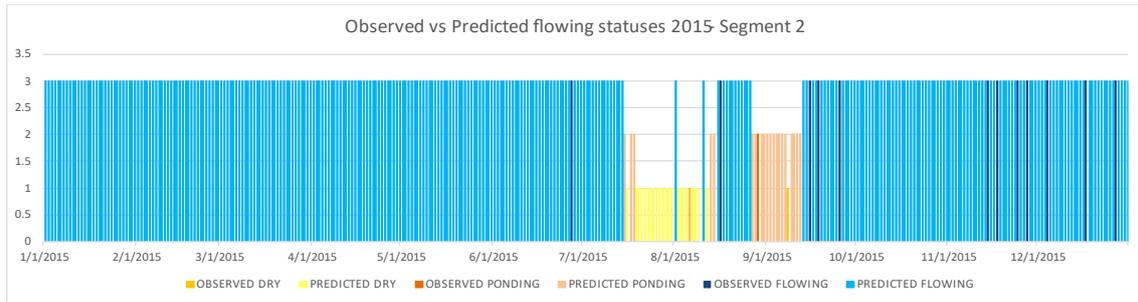


Figure 104. Evolution of observed and predicted flowing status for segment 2 during 2015.

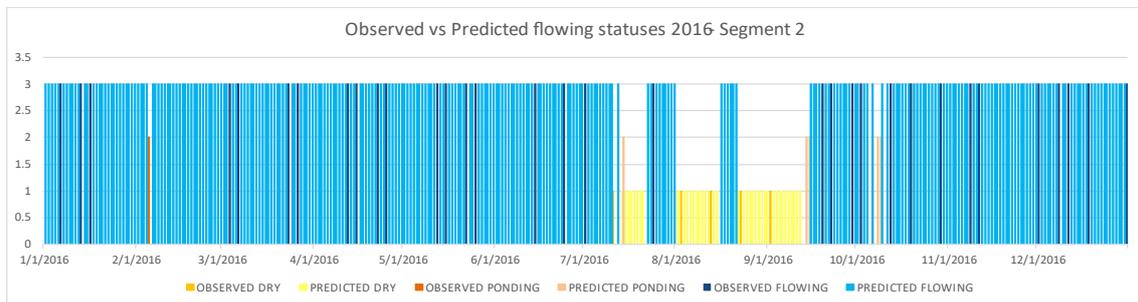


Figure 105. Evolution of observed and predicted flowing status for segment 2 during 2016.

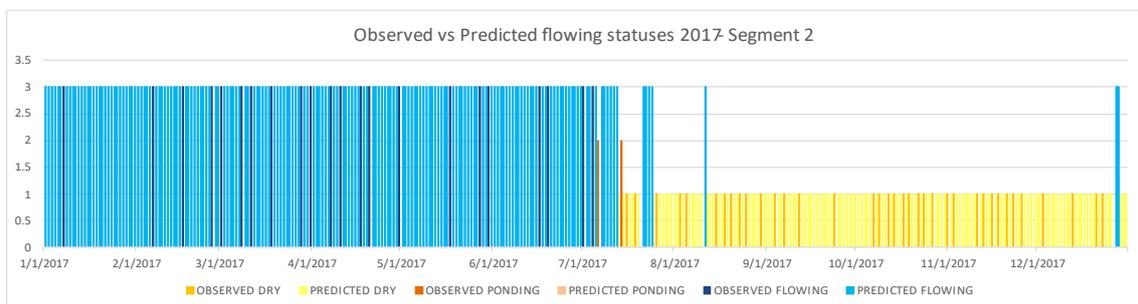


Figure 106. Evolution of observed and predicted flowing status for segment 2 during 2017.

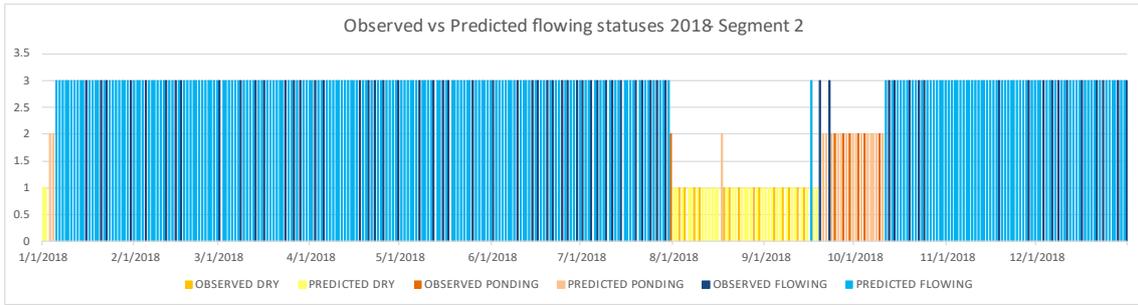


Figure 107. Evolution of observed and predicted flowing status for segment 2 during 2018.

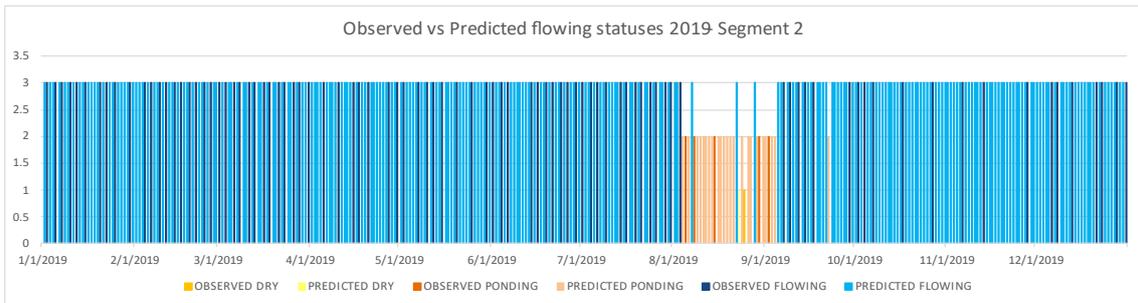


Figure 108. Evolution of observed and predicted flowing status for segment 2 during 2019.

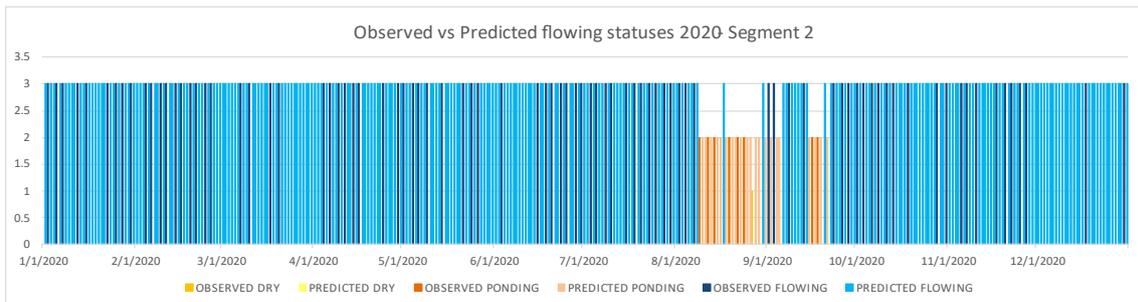


Figure 109. Evolution of observed and predicted flowing status for segment 2 during 2020.

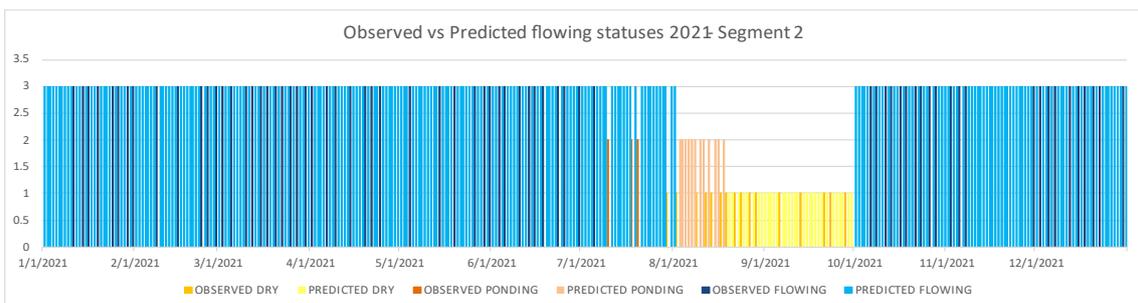


Figure 110. Evolution of observed and predicted flowing status for segment 2 during 2021.

Segment 3

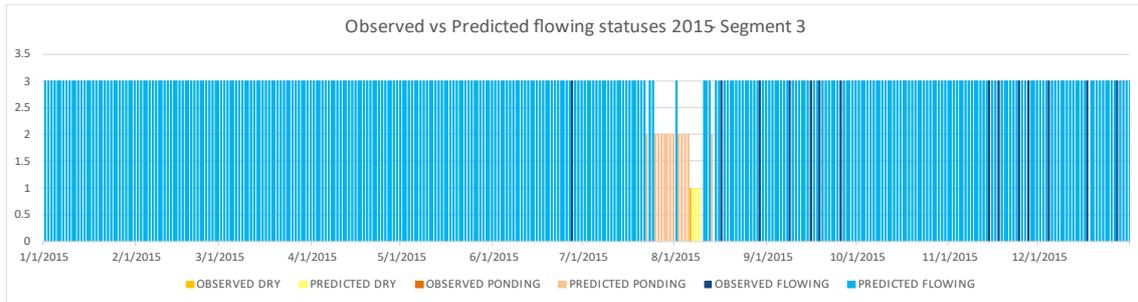


Figure 111. Evolution of observed and predicted flowing status for segment 3 during 2015.

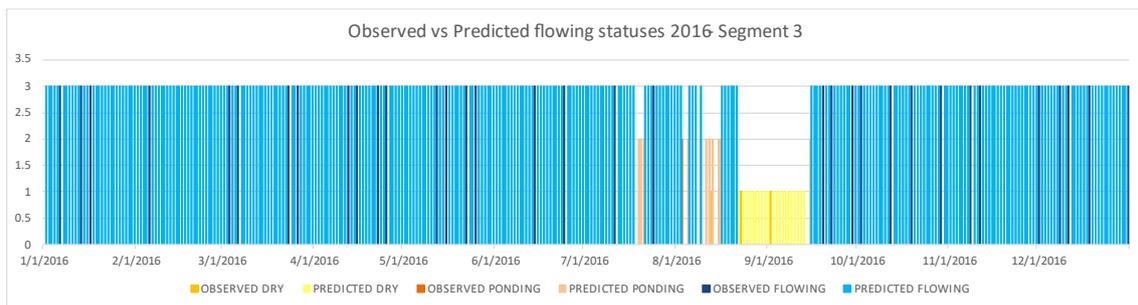


Figure 112. Evolution of observed and predicted flowing status for segment 3 during 2016.

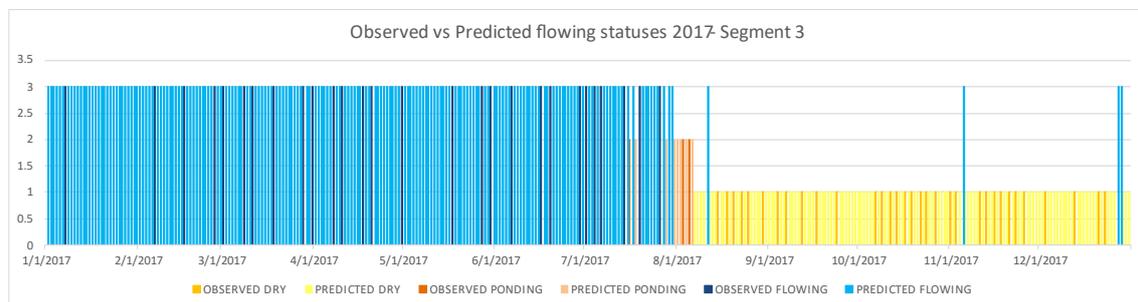


Figure 113. Evolution of observed and predicted flowing status for segment 3 during 2017.

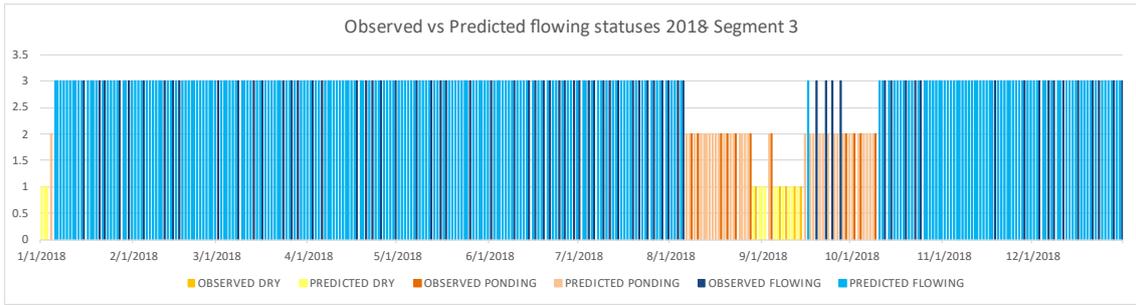


Figure 114. Evolution of observed and predicted flowing status for segment 3 during 2018.

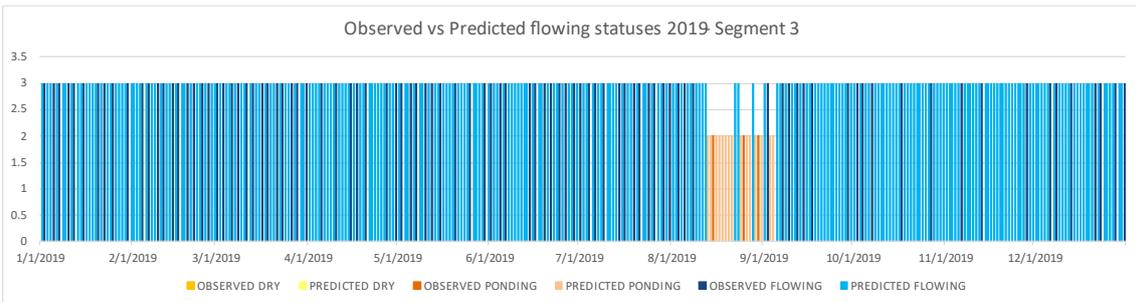


Figure 115. Evolution of observed and predicted flowing status for segment 3 during 2019.

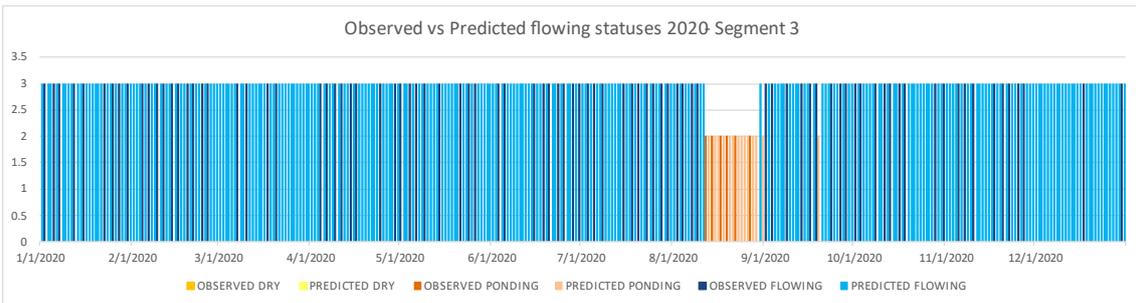


Figure 116. Evolution of observed and predicted flowing status for segment 3 during 2020.

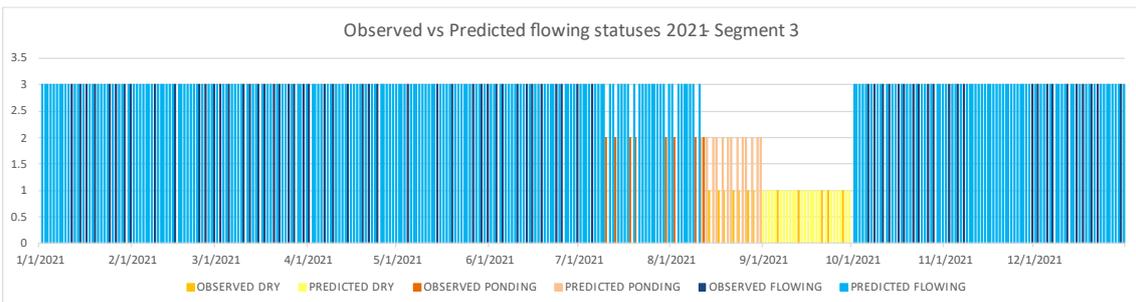


Figure 117. Evolution of observed and predicted flowing status for segment 3 during 2021.

Segment 2 – model no Flowrate and Water Table Level

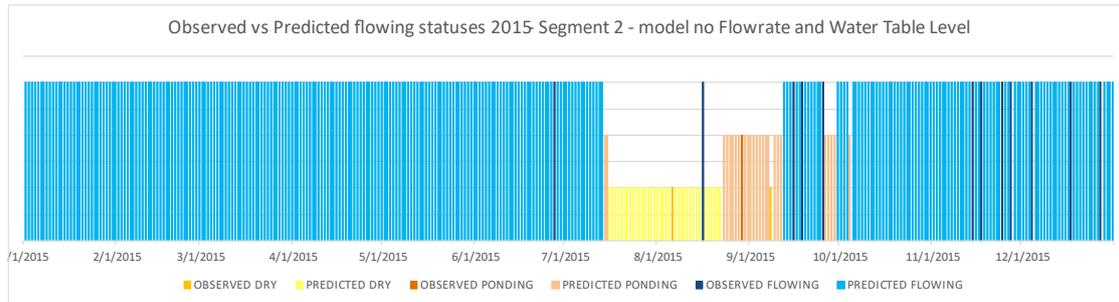


Figure 118. Evolution of observed and predicted flowing status for segment 2 during 2015, second model.

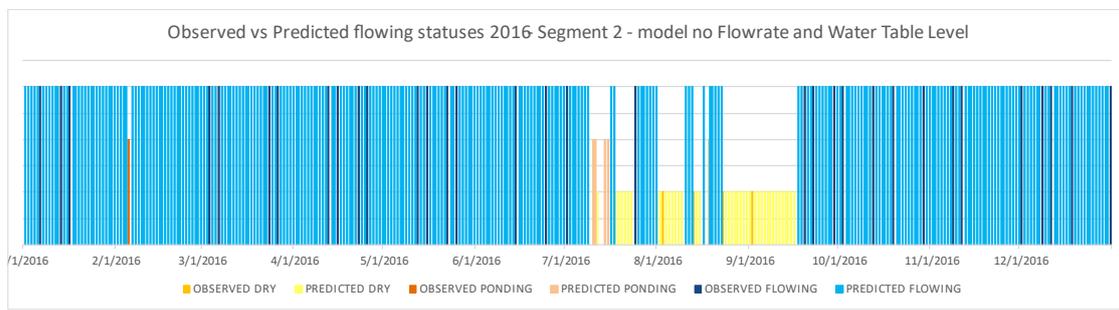


Figure 119. Evolution of observed and predicted flowing status for segment 2 during 2016, second model.

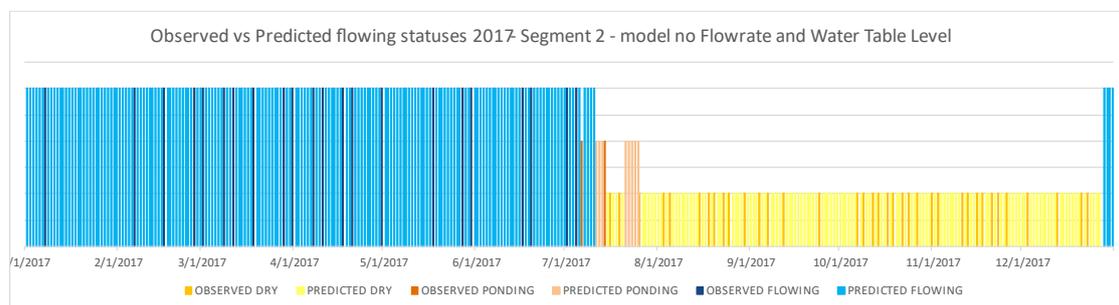


Figure 120. Evolution of observed and predicted flowing status for segment 2 during 2017, second model.

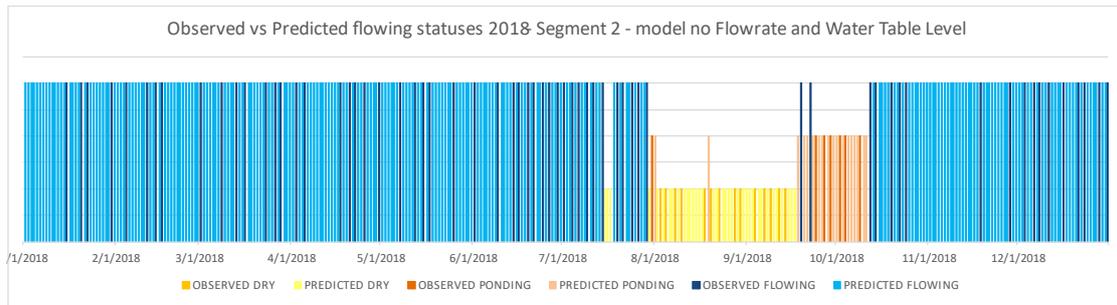


Figure 121. Evolution of observed and predicted flowing status for segment 2 during 2018, second model.

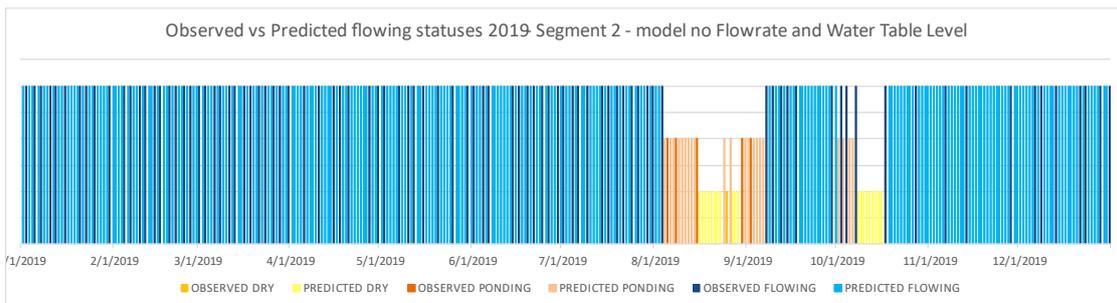


Figure 122. Evolution of observed and predicted flowing status for segment 2 during 2019, second model.

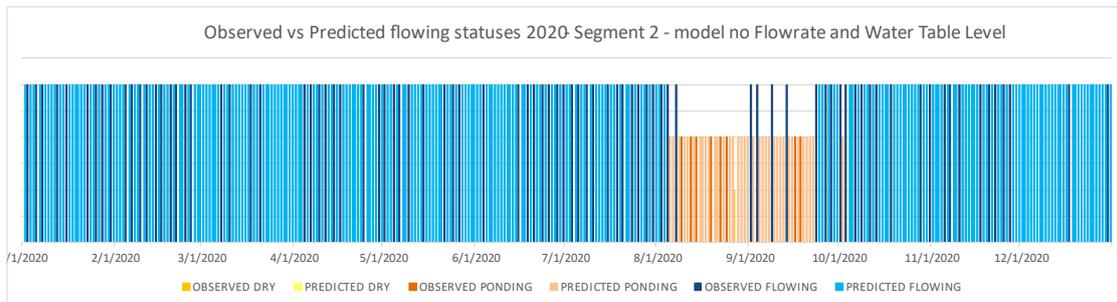


Figure 123. Evolution of observed and predicted flowing status for segment 2 during 2020, second model.

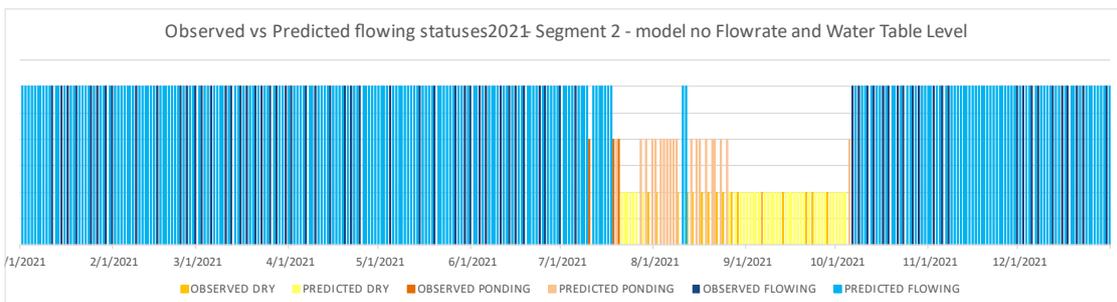


Figure 124. Evolution of observed and predicted flowing status for segment 2 during 2021, second model.

Bibliography

1. Coordinator, C. M. *et al.* Intermittent Rivers and Ephemeral streams : What water managers need to know Edited by : Contributing authors.
2. Busch, M. H. *et al.* What's in a name? Patterns, trends, and suggestions for defining non-perennial rivers and streams. *Water (Switzerland)* **12**, (2020).
3. Shanafield, M., Bourke, S. A., Zimmer, M. A. & Costigan, K. H. An overview of the hydrology of non-perennial rivers and streams. *Wiley Interdiscip. Rev. Water* **8**, 1–25 (2021).
4. Messenger, M. L. *et al.* Global prevalence of non-perennial rivers and streams. *Nature* **594**, 391–397 (2021).
5. Döll, P. & Schmied, H. M. How is the impact of climate change on river flow regimes related to the impact on mean annual runoff? A global-scale analysis. *Environ. Res. Lett.* **7**, (2012).
6. Pumo, D., Caracciolo, D., Viola, F. & Noto, L. V. Climate change effects on the hydrological regime of small non-perennial river basins. *Sci. Total Environ.* **542**, 76–92 (2016).
7. Zipper, S., Popescu, I., Compare, K., Zhang, C. & Seybold, E. C. Alternative stable states and hydrological regime shifts in a large intermittent river. *Environ. Res. Lett.* **17**, (2022).
8. Datry, T., Larned, S. T. & Tockner, K. Intermittent rivers: A challenge for freshwater ecology. *Bioscience* **64**, 229–235 (2014).
9. Skoulikidis, N. T. *et al.* Non-perennial Mediterranean rivers in Europe: Status, pressures, and challenges for research and management. *Sci. Total Environ.* **577**, 1–18 (2017).
10. Datry, T. *et al.* Causes, Responses, and Implications of Anthropogenic versus Natural Flow Intermittence in River Networks. *Bioscience* **73**, 9–22 (2023).
11. Llanos-Paez, O. & Acuña, V. Analysis of the socio-ecological drivers of the recreational use of temporary streams and rivers. *Sci. Total Environ.* **807**, (2022).
12. Cottet, M., Robert, A., Tronchère-Cottet, H. & Datry, T. “It’s dry, it has fewer charms!”: Do perceptions and values of intermittent rivers interact with their management? *Environ. Sci. Policy* **139**, 139–148 (2023).
13. Pastor, A. V. *et al.* Rethinking ecosystem service indicators for their application to intermittent rivers. *Ecol. Indic.* **137**, (2022).

14. Acuña, V. *et al.* Why Should We Care About Temporary Waterways ? 6–7.
15. Fritz, K., Cid, N. & Autrey, B. *Governance, Legislation, and Protection of Intermittent Rivers and Ephemeral Streams. Intermittent Rivers and Ephemeral Streams: Ecology and Management* (Elsevier Inc., 2017). doi:10.1016/B978-0-12-803835-2.00019-X.
16. European Parliament. *WFD DIRECTIVE 2000/60/EC*.
17. Ufficiale, G. *DECRETO 131-2008 - Regolamento recante i criteri tecnici per la carat- terizzazione dei corpi idrici (tipizzazione, individua- zione dei corpi idrici, analisi delle pressioni) per la modifica delle norme tecniche del decreto legislativo 3 aprile 2006, n. .* (2008).
18. Skoulikidis, N. T. *et al.* Science of the Total Environment Non-perennial Mediterranean rivers in Europe : Status , pressures , and challenges for research and management. **577**, 1–18 (2017).
19. Hassan, M.A., Egozi, R. Impact of wastewater discharge on the channel morphology of ephemeral streams. *Earth Surf. Process. Landforms* (2001) doi:10.1002/esp.273.
20. POFF, N. L. & ZIMMERMAN, J. K. H. Ecological responses to altered flow regimes: a literature review to inform the science and management of environmental flows. *Freshw. Biol.* **55**, 194–205 (2010).
21. Zimmer, M. A. *et al.* Zero or not? Causes and consequences of zero-flow stream gage readings. *WIREs Water* **7**, e1436 (2020).
22. Oueslati, O., De Girolamo, A. M., Abouabdillah, A., Kjeldsen, T. R. & Lo Porto, A. Classifying the flow regimes of Mediterranean streams using multivariate analysis. *Hydrol. Process.* **29**, 4666–4682 (2015).
23. University of Zurich. CrowdWater. <https://crowdwater.ch/en/temporary-stream/>.
24. Johnson, M. F. *et al.* Network environmentalism: Citizen scientists as agents for environmental advocacy. *Glob. Environ. Chang.* **29**, 235–245 (2014).
25. Conrad, C. C. & Hilchey, K. G. A review of citizen science and community-based environmental monitoring: Issues and opportunities. *Environ. Monit. Assess.* **176**, 273–291 (2011).
26. Gallart, F. *et al.* A novel approach to analysing the regimes of temporary streams in relation to their controls on the composition and structure of aquatic biota. *Hydrol. Earth Syst. Sci.* **16**, 3165–3182 (2012).
27. Gallart, F. *et al.* TREHS: An open-access software tool for investigating and

- evaluating temporary river regimes as a first step for their ecological status assessment. *Sci. Total Environ.* **607–608**, 519–540 (2017).
28. Chapin, T. P., Todd, A. S. & Zeigler, M. P. Robust, low-cost data loggers for stream temperature, flow intermittency, and relative conductivity monitoring. *Water Resour. Res.* **50**, 6542–6548 (2014).
 29. Bhamjee, R. & Lindsay, J. B. Ephemeral stream sensor design using state loggers. *Hydrol. Earth Syst. Sci.* **15**, 1009–1021 (2011).
 30. Gao, S. *et al.* Mapping dynamic non-perennial stream networks using high-resolution distributed hydrologic simulation: A case study in the upper blue river basin. *J. Hydrol.* **600**, 126522 (2021).
 31. Cavallo, C. *et al.* Exploiting Sentinel-2 dataset to assess flow intermittency in non-perennial rivers. *Sci. Rep.* **12**, 1–17 (2022).
 32. Cavallo, C. *et al.* Continuous monitoring of the flooding dynamics in the albufera wetland (Spain) by landsat-8 and sentinel-2 datasets. *Remote Sens.* **13**, (2021).
 33. Costigan, K. H. *et al.* Chapter 2.2 - Flow Regimes in Intermittent Rivers and Ephemeral Streams. in *Intermittent Rivers and Ephemeral Streams* (eds. Datry, T., Bonada, N. & Boulton, A.) 51–78 (Academic Press, 2017). doi:<https://doi.org/10.1016/B978-0-12-803835-2.00003-6>.
 34. Gallo, E., Ganora, D., Laio, F., Masoero, A. & Claps, P. Atlante dei bacini imbriferi piemontesi. (2013).
 35. Diamanti, E., Pierluigi, C. & Graziadei, M. *Relazione di tirocinio: Caratterizzazione fisiografica, climatica e del suolo del bacino del Sangone per applicazione di modelli idrologico distribuiti.* (2014).
 36. Regione Piemonte - Direzione Pianificazione delle Risorse Idriche. *MONOGRAFIA AREA IDROGRAFICA A110: PIANO DI TUTELA DELLE ACQUE (D.C.R. n. 117-10731 del 13 marzo 2007).* (2007).
 37. Autorità di Bacino del fiume Po. *Monografia Sangone e Chisola-Lemina. Obiettivi di qualità ambientale e principali misure per il sottobacino Sangone e Chisola-Lemina, PdGPo.* (2010).
 38. Regione Piemonte - Direzione Ambiente Governo e Tutela del Territorio. *Piano di Tutela delle Acque - Relazione generale.* (2018).
 39. ARPA Piemonte. *PROCESSO DI IMPLEMENTAZIONE DELLA DIRETTIVA 2000/60/CE (WFD) IN PIEMONTE.* (2009).
 40. ARPA Piemonte. Idroweb. Annali idrologici della Regione Piemonte. Banca dati

- idrologica. https://www.arpa.piemonte.it/rischinaturali/accesso-ai-dati/annali_meteoidrologici/annali-meteo-idro/banca-dati-idrologica.html (2022).
41. Biddoccu, M., Turconi, L., Tropeano, D. & De, S. K. Slope Instability and Flood Events in the Sangone Valley, Northwest Italian Alps. *Stud. Geomorphol. Carpatho-Balcanica XLIV*, 77–112 (2010).
 42. ARPA Piemonte. *IMPLEMENTAZIONE DELLA DIRETTIVA 2000/60/CE ANALISI E VALUTAZIONE DEGLI ASPETTI MORFOLOGICI ATTIVITA' 2017-2018 TORRENTE SANGONE CI 04SS2N704PI*. (2018).
 43. Beck, H. E. *et al.* Present and future köppen-geiger climate classification maps at 1-km resolution. *Sci. Data* **5**, 1–12 (2018).
 44. Peel, M. C., Finlayson, B. L. & McMahon, T. A. Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.* (2007).
 45. Climate Change Knowledge Portal for development practitioner and policy makers. Climate knowledge portal. 2022 <https://climateknowledgeportal.worldbank.org/country/italy/climate-data-historical>.
 46. ARPA Piemonte, SNPA & Regione Piemonte. STATO DELL'AMBIENTE IN PIEMONTE. Relazione 2020: temperature. <http://relazione.ambiente.piemonte.it/2020/it/clima/stato/temperature> (2020).
 47. ARPA Piemonte, SNPA & Regione Piemonte. STATO DELL'AMBIENTE IN PIEMONTE. Relazione 2020: precipitazioni. <http://relazione.ambiente.piemonte.it/2020/it/clima/stato/precipitazioni> (2020).
 48. Arpa Piemonte & Regione Piemonte. Analisi degli scenari di clima regionale del periodo 2011- 2100. (2020).
 49. ARPA Piemonte. Portale sul clima in piemonte. (2022).
 50. ARPA Piemonte & SNPA. *IMPLEMENTAZIONE DELLA DIRETTIVA 2000/60/CE: ANALISI E VALUTAZIONE DEGLI ASPETTI IDROMORFOLOGICI II SESSENNIO RELAZIONE SUI CORPI IDRICI ANALIZZATI NEL 2017-2018*. (2018).
 51. Pinca, M. Italy's longest river affected by worst drought in 70 years. *Reuters* <https://www.reuters.com/news/picture/italys-longest-river-affected-by-worst-d-idUSRTS8WTIV> (2022).
 52. Guccione, G. Troppi prelievi, in sofferenza quattro fiumi su dieci in Piemonte. *la Repubblica, Torino* (2017).

53. Currà, O. NICHELINO - Il Sangone in secca rivela una discarica nel suo letto. *www.torinosud.it/cronaca/* (2019).
54. Bacci, S. Piemonte, allarme siccità Il Sangone in secca al 4% della portata. *www.rainews.it/tgr/piemonte* (2022).
55. ARPA Piemonte. *IMPLEMENTAZIONE DELLA DIRETTIVA 2000/60/CE ANALISI E VALUTAZIONE DEGLI ASPETTI MORFOLOGICI ATTIVITA' 2017-2018 Torrente sangone ci 06SS3F705PI.* (2018).
56. ARPA Piemonte. Meteoweb. Annali meteorologici della Regione Piemonte. Banca dati meteorologica. https://www.arpa.piemonte.it/rischinaturali/accesso-ai-dati/annali_meteoidrologici/annali-meteo-idro/banca-dati-meteorologica.html (2022).
57. Regione Piemonte. Monitoraggio acque sotterranee. <http://www.regione.piemonte.it/monitgis/jsp/cartografia/mappa.do>.
58. Città Metropolitana di Torino. CED ACQUA. Rete di monitoraggio. <http://www.cittametropolitana.torino.it/cms/ambiente/risorse-idriche/progetti-ris-idriche/rete-monitoraggio/cedacqua>.
59. Città Metropolitana di Torino. Risorse idriche. Rete di monitoraggio acque sotterranee. <http://www.cittametropolitana.torino.it/cms/ambiente/risorse-idriche/progetti-ris-idriche/rete-monitoraggio/acque-sotterranee>.
60. Casu, F., Manunta, M., Agram, P. S. & Crippen, R. E. Big Remotely Sensed Data: tools, applications and experiences. *Remote Sens. Environ.* **202**, 1–2 (2017).
61. Isikdogan, F., Bovik, A. & Passalacqua, P. RivaMap: An automated river analysis and mapping engine. *Remote Sens. Environ.* **202**, 88–97 (2017).
62. Jiang, H. *et al.* An automated method for extracting rivers and lakes from Landsat imagery. *Remote Sens.* **6**, 5067–5089 (2014).
63. Seaton, D., Dube, T. & Mazvimavi, D. Use of multi-temporal satellite data for monitoring pool surface areas occurring in non-perennial rivers in semi-arid environments of the Western Cape, South Africa. *ISPRS J. Photogramm. Remote Sens.* **167**, 375–384 (2020).
64. G. Kantharajan, Arur Anand, P. Krishnan, Rajeev K. Singh, Kundan Kumar, Ajay Kumar Yadav, Vindhya Mohindra, S. P. S. & K. K. L. Applications of Sentinel-2 satellite data for spatio-temporal mapping of deep pools for monitoring the riverine connectivity and assessment of ecological dynamics: a case from Godavari, a tropical river in India (2016–2021). *Environ. Monit. Assess.* (2022).

65. Paolo Vezza, Andrea Zanin, P. P. *Manuale tecnico-operativo per la modellazione e la valutazione dell'integrità dell'habitat fluviale*. ISPRA. vol. 13 (2017).
66. *MAP STREAM, User Manual*. vol. 3304 (2021).
67. Gorelick, N. *et al.* Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **202**, 18–27 (2017).
68. Boothroyd, R. J., Williams, R. D., Hoey, T. B., Barrett, B. & Prasojo, O. A. Applications of Google Earth Engine in fluvial geomorphology for detecting river channel change. *Wiley Interdiscip. Rev. Water* **8**, 1–27 (2021).
69. Genuer Robin & Poggi Jean-Michel. *Random Forests with R. Use R* (2020).
70. Probst, P., Wright, M. N. & Boulesteix, A. L. Hyperparameters and tuning strategies for random forest. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **9**, 1–15 (2019).
71. Prasad, A. M., Iverson, L. R. & Liaw, A. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems* **9**, 181–199 (2006).
72. Vezza, P., Muñoz-Mas, R., Martínez-Capel, F. & Mouton, A. Random forests to evaluate biotic interactions in fish distribution models. *Environ. Model. Softw.* **67**, 173–183 (2015).
73. Somodi, I., Lepesi, N. & Botta-dukát, Z. Prevalence dependence in model goodness measures with special emphasis on true skill statistics. 863–872 (2017) doi:10.1002/ece3.2654.
74. Zhang, W. Y., Wei, Z. W., Wang, B. H. & Han, X. P. Measuring mixing patterns in complex networks by Spearman rank correlation coefficient. *Phys. A Stat. Mech. its Appl.* **451**, 440–450 (2016).