

doi.org/10.36073/1512-0902-2025-136-87-91

უაგ:111.111

გვალვის მონიტორინგის შესაძლებლობები მანქანური სწავლების საშუალებით ანა ფალავანდიშვილი

საქართველოს ტექნიკური უნივერსიტეტის ჰიდრომეტეოროლოგიის ინსტიტუტი
Palavandishvili.ana21@gtu.ge

რეზიუმე. ფიზიკის აქტუალურ პრობლემათა შორის ერთ-ერთი მნიშვნელოვანია ატმოსფეროში მიმდინარე ფიზიკური პროცესების კვლევა თანამედროვე მეთოდების და ახალი მონაცემების გამოყენებით. ბუნებრივ კატასტროფებს შორის გვალვა ერთ-ერთ საშიშროებას წარმოადგენს მსოფლიოს ბევრი რეგიონისთვის, მათ შორის საქართველოსთვის. კვლევა ჩატარდა საქართველოს ტერიტორიაზე, აღებულია 1960-2022 წლების პერიოდის მონაცემები გარემოს ეროვნული სააგენტოს მონაცემთა ბაზიდან, ასევე გამოყენებულია ნალექის 1960-1990 კომპენიკუსის ERA5-ის რეანალიზის მონაცემები, რომლებიც ვალიდაციის მიზნით შედარებულია საქართველოს ტერიტორიაზე არსებული მიწისპირა სადგურების მონაცემებთან. საქართველოს მიწისპირა და სატელიტური წყაროების მონაცემების ანალიზით დადგინდა, რომ მანქანური სწავლების რეგრესიის მეთოდის გამოყენებით საკმარისია შეფასდეს 1960 - 2000 წლების მონაცემები სწავლებისთვის და 2001 - 2022 წლების მონაცემები ტრენინგისთვის. სატრენინგო ობიექტად შეირჩა SPI-3 სამ თვიანი ნალექის სტანდარტიზებული ინდექსი. მანქანური სწავლების მოდელის და ალგორითმის გამოყენება ნალექის განაწილების ანალიზის, გვალვის მონიტორინგის და პროგნოზირებისთვის მიზანშეწონილია როგორც საქართველოს, ასევე სხვა რეგიონებშიც დაკვირვების მონაცემებთან სათანადო ბაზის (60 წლიანი) პირობებში. მიღებული შედეგების გამოყენება რეკომენდებულია ადრეული გაფრთხილების სისტემაში გვალვის მონიტორინგისთვის.

საკვანძო სიტყვები: მანქანური სწავლება, კლიმატის ცვლილება, გვალვის მონიტორინგი

შესავალი. ნალექის ცვალებადობის შეფასება სივრცესა და დროში გვალვის ადრეული გაფრთხილებისა და გარემოს მონიტორინგის მთავარი ასპექტია. თუმცა, სატელიტური მონაცემებიდან მიღებული შეფასებები იძლევა ადგილის საშუალო მაჩვენებლებს, რომლებიც გაფანტულია რთული რელიეფის გამო, რის გამოც ხშირად ექსტრემალური ნალექების ინტენსივობა ზუსტად არ ფასდება. ამის საპირისპიროდ, სადგურების მონაცემებით წარმოებული ნალექის ბადე ხარვეზიანია უფრო სოფლად, სადაც ნაკლები სადგურები და საგუშაგოებია.

კვლევისათვის საჭირო მონაცემები აღებულია გარემოს ეროვნული სააგენტოს ბაზიდან. ეროვნული სააგენტოს მონაცემებით დაკვირვებები წარმოებს 1930 წლიდან, საქართველოს ტერიტორიაზე სადგურების მაქსიმალურმა რაოდენობამ მიაღწია 1960 წელს 170 ერთეულს.

გამომდინარე იქიდან, რომ საქართველოში გარკვეული სადგურები დაიკეტა 90 - იანი წლების პერიოდში, რადგან ხელსაწყოები არ იყო დაკალიბრებული ან იყო დაზიანებული, შესაბამისი პერიოდის მონაცემები არის ნაკლები სანდოობის. მონაცემებს გააჩნია წყვეტები, რაც წარმოადგენს დიდ პრობლემას ინდექსის გამოთვლისას და მანქანური სწავლებისთვის. ამიტომ ისტორიული მონაცემებისთვის შეირჩა ERA5 - ის რეანალიზის მონაცემები. ბადის უჯრედის ზომის სიდიდიდან გამომდინარე ვერ მოხერხდა სანაპირო ზოლის ახლოს მდებარე სადგურების შედარება.

სატელიტური CHIRPS - ისა და IMERG - ის ინფორმაციის შედარების საფუძველზე დადგინდა [8], რომ CHIRPS - ს საქართველოს ტერიტორიაზე არსებულ სადგურების მონაცემებთან აქვს კარგი კორელაცია. ის სადგურები, სადაც მონაცემების 50 % აკლდა ან არ შეესაბამებოდა რეალობას, არ დაექვემდებარა ანალიზს. აღსანიშნავია ისიც, რომ სატელიტი ვერ აღიქვამს მაღალმთიან რეგიონს, დაკვირვებას არ ექვემდებარება ზღვა, რადგან დადებული აქვს ფილტრი (mask), ამიტომ ხშირ შემთხვევაში სანაპირო ზოლზე ვერ ხერხდება სატელიტური მონაცემის მოპოვება, ამ შემთხვევაში ხდება სხვა ბადის ახლომდებარე ბადის უჯრის არჩევა.

განხორციელდა CHIRPS - ისა და სადგურის მონაცემების შედარება, რისთვისაც გამოთვლილ იქნა, სისტემური ცდომილება - BIAS, რომელიც გულისხმობს თანამგზავრიდან და მიწისპირა სადგურიდან გაზომილი ნალექების ყოველთვიური ჯამების სხვაობის შეფასებას. დათვლილ იქნა ასევე, კორელაცია დაკვირვების მთელი პერიოდისთვის, ყველა თვის და წლისთვის, საშუალო აბსოლუტური ცდომილებები და სტანდარტული გადახრა იმავე პრინციპით. გამოთვლებმა აჩვენა, რომ იმ სადგურებში, სადაც დაკვირვების პერიოდიდან არ არსებობს მონაცემთა 50%-ზე მეტი მაგალითად, მანგლისში, შეუძლებელია სტატისტიკური

ანალიზის ჩატარება. გამოვლინდა სულ ცხრა ასეთი სადგური. სტატისტიკური მახასიათებლების გამოთვლის ანალიზიდან ირკვევა, რომ ყველაზე დაბალი კორელაციური მნიშვნელობები 0.33 არის მთა-საბურთის სადგურზე და მაქსიმალური 0.72 - შოვის სადგურზე, მინიმალური საშუალო აბსოლუტური ცდომილება არის 18.8 გორის სადგურზე, და მაქსიმალური - 83.2 ბათუმის სადგურზე, ყველაზე მცირე სტანდარტული გადახრა არის 23.4 გორის სადგურზე, უდიდესი - 118.46 მირვეთის სადგურზე, რაც ახსნადია ზღვისპირა და მთაგორიანი რელიეფის გათვალისწინებით. მაღალი კორელაციის კოეფიციენტი უკეთ გამოხატავს სატელიტისა და სადგურის მონაცემების თანხვედრას.

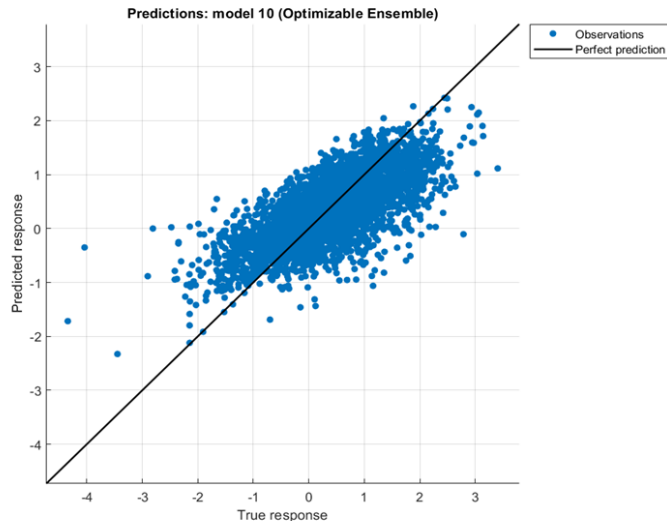
R - instat პროგრამული უზრუნველყოფა გამოიყენებოდა პირსონის კორელაციის და სხვა სტატისტიკური პარამეტრების გამოსათვლელად. სადგურები, რომლებიც გამოტოვებულია, შეუძლებელი იყო ანალიზის გაკეთება. შედარებულ იქნა ისტორიული ნალექის მონაცემები 1961 - 1991 წლებში ERA5 - ის ნალექის რეანალიზის მონაცემებთან. კახეთის რეგიონზე შეირჩა 19 სადგური. რა თქმა უნდა, ანალიზს არ დაექვემდებარა ის პუნქტები, სადაც არ იყო საკმარისი მონაცემი.

კვლევის ობიექტი და მეთოდოლოგია. მონაცემთა შედარების შედეგად გამოიკვეთა ყველაზე დაბალი კორელაციის მაჩვენებელი (0,309) სადგურ შირაქზე, მაქსიმალური (0,657) სადგურ ომალოზე. მინიმალური საშუალო აბსოლუტური ცდომილება (1,662) დაფიქსირდა სადგურ უდაბნოზე, მაქსიმალური (3,041) - სადგურ შილდაზე. ყველაზე მცირე სტანდარტული გადახრა (4,047) არის სადგურ უდაბნოზე, უდიდესი (7,624) - სადგურ ლაგოდეხზე. 1961 - 1991 წლების სადგურის ისტორიული და ERA5-ის რეანალიზის მონაცემების შედარებით ჩანს, რომ მოდელისთვის გამოდგება ალტერნატიულ მონაცემთა ბაზად, როგორც კახეთის, ასევე საქართველოს მთელი ტერიტორიაზე. დათვლილ იქნა climact2 R - პროგრამული ენის პაკეტის დახმარებით SPI და SPEI ინდექსები. ამავე დროს, განხორციელდა მონაცემთა ხარისხის კონტროლი, გაანალიზდა მაქსიმალური და მინიმალური ტემპერატურის ტენდენციები, შეფასდა გვალვის ინდექსები 3, 6, 9 და ერთ წლიანი დროითი ბიჯებით. გამოთვლილ იქნა 3 თვიანი პერიოდისთვის SPI და SPEI. ანალიზმა აჩვენა, რომ 3 თვიანი ნალ სრულყოფილად ასახავს მეტეოროლოგიური, სასოფლო-სამეურნეო, სოციალურ-ეკონომიური გვალვის სიმკაცრეს.

განხილულ იქნა გვალვის განვითარების ტენდენციის შეფასების შესაძლებლობა. ქვემოთ მოყვანილია (ნახაზები 8 - 10) რამდენიმე ლოკაციისთვის climact2 პაკეტის გამოყენებით დათვლილი და ვიზუალიზირებული SPI 24 (2 წლიანი) ნალექის სტანდარტიზებული ინდექსი, რომელიც კარგად გვაჩვენებს გვალვის განვითარების ტენდენციას ბოლო 60 წლის განმავლობაში შერჩეულ რეგიონებში. SPI და SPEI ინდექსების ეფექტურობის ანალიზისთვის, ერთმანეთს შედარდა სამი სადგურის მონაცემები. სადგურ თელავში დაფიქსირდა SPI -ის მინიმალური მნიშვნელობა -3,62 და მაქსიმალური 2,43, ხოლო SPEI - ის მინიმალური მნიშვნელობა 2,52 და მაქსიმალური 2,43. ამ სადგურზე უფრო გამოკვეთილია გვალვის ტრენდები SPI - ის გამოყენებით.

ექსპერიმენტული შედეგები და განსჯა. პირველი მოდელის დამხმარე ვექტორული მანქანის (SVM) მიერ გენერირებული მოდელის შეფასებისთვის პარამეტრებია : RMSE - 0,58848, R2 _0.64, MSE_0.34631, MAE_0.45479, პროგნოზის სიჩქარე ~430000 obs/sec (დაკვირვება წამში), სწავლების დრო - 30,296 წმ, ფოთლის მინიმალური ზომა - 45 (ოპტიმიზებული პარამეტრი), ჰიპერპარამეტრული ძებნის რანჟირება: მინიმალური ფოთლის ზომა_1-3420. ყველა ეს პარამეტრი საჭიროა რათა შეფასდეს მოდელი.

გადაჭარბებული მორგებისაგან თავის აცილების მიზნით დამატებულ იქნა კიდევ ქართლის რეგიონის დამატებით 30 სადგური. ამ შემთხვევაშიც სწავლებაში შევიდა კოორდინატები, წელი, თვე, თვის ნალექის ჯამი, მაგრამ მიუხედავად დამატებული მონაცემებისა დაფიქსირდა მსგავსი შემთხვევა. ნახაზიდან 15 - ჩანს, რომ მონაცემები იმეორებს საუკეთესო პროგნოზის მრუდის ფორმას, რაც მეტყველებს გადაჭარბებულ მორგებაზე. ცნობილია, რომ 2022 წლისთვის მიწისპირა სადამკვირვებლო ქსელიდან აქტიური დარჩა 3 სადგური, ამიტომ დაგჭირდება მათი სატელიტის ან კოპერნიკუსის მონაცემთა ბაზის ინფორმაციით ჩანაცვლება. რეგრესიის ალგორითმის ტრენინგისას ყველაზე კარგი შედეგი აჩვენა Bagged trees ანსამბლირებულმა ოპტიმიზირებულმა ალგორითმმა. სწავლის დრო 326,21 წმ - ია, პროგნოზირების სისწრაფე - ~ 7900 დაკვირვება/წამში, მახასიათებელი პარამეტრები RMSE - 0,5046, R2- 0,64, MSE - 0,25466, MAE - 0,38065.



ნახ. 1. გამოსახულია დაკვირვებების წერტილები და ყველაზე კარგი პროგნოზი.

ამ ნახაზის ანალიზის მიხედვით შეგვიძლია ვიმსჯელოთ თუ როგორი მოდელია. თუ მონაცემი (წერტილები) არის ძალიან გაბნეული ეს მიუთითებს იმაზე რომ არის არასაკმარისი მორგება ანუ Underfitting, ხოლო წერტილები, რომლებიც ახლოსაა პროგნოზის მრუდთან (წრფესთან) მიუთითებს გადაჭარბებულ მორგებაზე ანუ Overfitting - ზე. როცა წერტილთა დიდი ნაწილი თავმოყრილია ერთ არეში შეიძლება დავასკვნათ, რომ მოდელი არის კარგი მორგების, ე.ი. დამაკმაყოფილებელი მინიმალური ფოთლის ზომა არის 19, ოპტიმიზაციის მიზნით მინიჭებულია 40 იტერაცია, ნაცვლად 30 იტერაციისა. ცისფრად აღნიშნულია სავარაუდო მინიმალური საშუალო კვადრატული გადახრა, ლურჯით დაკვირვებადი მინიმალური საშუალო კვადრატული გადახრა. სწავლების სისწრაფეა 0,27325. სწავლების ეტაპის შემდეგ მოდელი დაექვემდებარა 2001-2022 წლების მონაცემებზე მორგებას (fitting - ს) შემდგომი პროცესინგისთვის. აღსანიშნავია ისიც, რომ მორგებისა და პროგნოზის პროცესის ჩასატარებლად შეირჩა ის სადგურები, რომლებიც ოპერირებდნენ 2001 - 2022 წლების მანძილზე, რამდენიმე წლის განმავლობაში მაინც. ამის შემდეგ ისინი შეივსო CHIRPS - ის სატელიტური მონაცემით. პროგნოზის განსახორციელებლად საჭირო გახდა წრფივი რეგრესიის განტოლების გამოთვლა თითოეული სადგურისთვის, MatLab სივრცეში polyfit ფუნქციის გამოყენებით. კოეფიციენტების გამოთვლის შემდეგ, MatLab -ის სივრცეში polyval ფუნქციით გამოითვალა პოლინომები. საპროგნოზო სცენარის პირველ შემთხვევაში ნალექის რაოდენობა განისაზღვრა 0 სმ - დან 10 სმ - მდე, რადგან ჩვენთვის საინტერესოა მცირე ნალექის რაოდენობის შემთხვევაში გვალვის განვითარების პროცესი.

ცხრილი 1. ნალექის სტანდარტიზებული ინდექსი (SPI), 0 სმ - დან 10 - სმ - ის ჩათვლით ნალექის პროგნოზი კახეთის ზოგიერთი რეგიონისთვის.

საპროგნოზო ნალექი (სმ)	დედოფლის წყარო	საგარეჯო	გურჯაანი	ყვარელი	ლაგოდეხი	ომალო	თელავი
0	-0.270	-0.336	-0.008	- 0.002	-0.524	0.270	-0.568
1	-0.021	-0.057	-0.097	-0.012	-0.269	0.429	-0.325
2	0.227	0.220	-0.186	-0.026	-0.014	0.587	-0.083
3	0.475	0.498	-0.275	-0.041	0.240	0.746	0.159
4	0.724	0.776	-0.365	-0.054	0.495	0.905	0.402
5	0.973	1.054	-0.454	-0.068	0.750	1.064	0.645
6	1.222	1.333	-0.543	-0.082	1.005	1.222	0.888
7	1.470	1.611	-0.633	-0.096	1.260	1.381	1.131
8	1.719	1.889	-0.722	-0.110	1.514	1.540	1.374
9	1.968	2.167	-0.811	-0.124	1.769	1.698	1.617
10	2.217	2.445	-0.901	-0.138	2.024	1.857	1.860

ანალიზმა აჩვენა, რომ გურჯაანი და ყვარელი ნალექის მცირე რაოდენობის მიმართ უფრო მოწყვლადია, ასევე გამოიკვეთა ნაკლებად მოწყვლადი რეგიონებიც - დედოფლისწყარო, საგარეჯო, ლაგოდეხი. ამ შედეგების გაანალიზებისას, გასათვალისწინებელია სადგურის სიმაღლე ზღვის დონიდან და ტენიანობა, ასევე, ის ფაქტი, რომ სატელიტური მონაცემის გარკვეული ნაწილი არ მოიძებნა ართანასა და მელაანის სადგურის შემთხვევაში, რის გამოც მოდელმა ვერ მოიცვა კახეთის სრული რეგიონი.

დასკვნა

- 1) ატმოსფერული პროცესების, კერძოდ ნალექწარმოქმნის პროცესების ამსახველი განტოლებების და მათემატიკური მოდელების შესწავლის შედეგად დადგინდა, რომ ამოცანის კომპლექსური და მრავალპარამეტრიანი ხასიათიდან გამომდინარე, არსებული მოდელები არ იძლევა პროცესების სრულყოფილ აღწერას და ზუსტი ანალიზური ამოხსნების ჩაწერის შესაძლებლობას, არის მიახლოებითი და შრომატევადი.
- 2) კვლევის ალტერნატიული მეთოდის სახით, შერჩეულ იქნა სიმულაციური მეთოდი მანქანური სწავლების გამოყენებით. მანქანური სწავლების ზედამხედველური მეთოდის გამოყენება ოპტიმალურია ანალოგიური კომპლექსური ამოცანების შემთხვევაში. მანქანური სწავლების ზედამხედველობითი რეგრესიის ალგორითმებიდან ყველაზე კარგი შედეგი მიღებულ იქნა Bagged Trees - ს ოპტიმიზირებული მოდელით.
- 3) მანქანური სწავლების წარმოდგენილ მოდელისთვის ოპტიმალურია შემდეგი შემავალი პარამეტრების გამოყენება: სადგურის დასახელება, წელი, თვე, თვის ნალექის ჯამი, მოდელის სწავლების საპასუხო ცვლადის სახით - სტანდარტიზებული ნალექის ინდექსი - SPI, არ არის აუცილებელი დაკვირვების სადგურის კოორდინატები.
- 4) სატელიტური დაკვირვების ბაზების CHIRPS - ის და IMERG - ის მონაცემთა შედარების შედეგად დადგინდა, რომ CHIRPS - ის მონაცემები კარგ თანხვედრაშია საქართველოს ტერიტორიაზე მიწისპირა სადგურების 2000 – 2020 წლების დაკვირვებების მონაცემებთან.
- 5) სტანდარტიზებული ნალექის ინდექსის (SPI) გამოყენება უკეთესად ასახავს გვალვის მზარდ ტრენდს კახეთის რეგიონში, ვიდრე ნალექის ევაპოტრანსპირაციის სტანდარტიზებული ინდექსი (SPEI).
- 6) საქართველოს მიწისპირა და სატელიტური წყაროების მონაცემების ანალიზით დადგინდა, რომ მანქანური სწავლების რეგრესიის მეთოდის გამოყენებით საკმარისია SPI ინდექსი შეფასდეს 1960 - 2000 წლების მონაცემებით სწავლებისთვის და 2001 - 2022 წლების მონაცემებით ტრენინგისთვის.
- 7) მიწისპირა სადგურების ისტორიული და ERA5 - ის რეანალიზის მონაცემების (1961 - 1991 წწ.) შედარებით (კახეთის რეგიონში), დადასტურდა ERA5 - ის მონაცემების გამოყენების შესაძლებლობა ალტერნატიულ მონაცემთა ბაზად მოდელისთვის, როგორც კახეთის, ასევე საქართველოს მთელი ტერიტორიისთვის.
- 8) პროგნოზირების ალგორითმით გამოიკვეთა კახეთის რეგიონში გვალვისადმი მოწყვლადი რაიონები.
- 9) მანქანური სწავლების წარმოდგენილი მოდელის და ალგორითმის გამოყენება ნალექის განაწილების ანალიზის, გვალვის მონიტორინგის და პროგნოზირებისთვის შესაძლებელია როგორც საქართველოს, ასევე სხვა რეგიონებშიც დაკვირვების მონაცემთა სათანადო ბაზის (60 წლიანი) პირობებში.
- 10) მიღებული შედეგების გამოყენება რეკომენდირებულია ადრეული გაფრთხილების სისტემებში გვალვის მონიტორინგის და პროგნოზირებისთვის.

ლიტერატურა - References – Литература

1. A. Palavandishvili. Structural data set in environmental issued. The Regional Student Scientific and Practical Conference Digital Transformation in Engineering Human - Computer Interaction, Georgian Technical University (GTU), Faculty of Informatics and Control Systems. 2021.
2. M. Tatishvili, Z. Khvedelidze, I. Samkharadze, A. Palavandishvili. Atmosphere processes and climate parameters variation in River Mtkvari basin. საერთაშორისო სამეცნიერო კონფერენცია “ზუნებრივი კატასტროფები საქართველოში: მონიტორინგი, პრევენცია, შედეგების შერბილება”. შრომათა კრებული. 2019. ISBN 978-9941-13-899-7.
3. Mathwork - Matlab Statistics and Machine Learning Toolbox documentation (2016). stats.pdf (mathworks.com) გადამოწმებულია 07.06.2023.
4. Mathwork - Statistics and Machine Learning Toolbox Release notes rn.pdf (mathworks.com) გადამოწმებულია 07.06.2023.

5. M. Tatishvili, T. Tsintsadze, I. Samkharadze, A. Palavandishvili. Dependence of meteorological parameters on geomagnetic storms in Georgia. International Scientific Journal. Environmental Studies. 2019, v.8. issue 1, ISBN-9781671503274.

UDC:111.111

Drought monitoring capabilities through machine learning /Ana Palavandishvili/ Transactions IHM, GTU. -2025. - vol.136. -pp.87-91. - Georg., Summ. Georg., Eng.Rus.

One of the key current problems in physics is studying atmospheric physical processes using modern methods and new data. Drought is among the most serious natural threats for many regions worldwide, including Georgia. This study focused on Georgia, using data from 1960–2022 obtained from the National Environmental Agency, as well as Copernicus ERA5 precipitation reanalysis data from 1960–1990, which were compared with ground station data in Georgia for validation. The analysis of ground and satellite data showed that using machine learning regression methods, data from 1960–2000 are sufficient for training, while 2001–2022 data can be used for testing. The SPI-3 (three-month Standardized Precipitation Index) was chosen as the target variable. Applying machine learning models and algorithms for precipitation analysis, drought monitoring, and forecasting is appropriate for Georgia and other regions with a 60-year observational record. The results are recommended for use in early warning systems for drought monitoring..

УДК: 111.111

Возможности мониторинга засухи с помощью машинного обучения /Ана Палавандишвили / Сб. Трудов ИГМ ГТУ. - 2025. – том 136. - с. 87-91. - Груз.; Рез: Груз., Англ., Рус.

Одной из важнейших современных проблем физики является изучение физических процессов, происходящих в атмосфере, с использованием современных методов и новых данных. Среди стихийных бедствий засуха является одной из угроз для многих регионов мира, включая Грузию. Исследование проводилось на территории Грузии с использованием данных базы данных Национального агентства по охране окружающей среды за период 1960-2022 гг., а также данных реанализа осадков Copernicus ERA5 за период 1960-1990 гг., которые для проверки сравнивались с данными наземных станций на территории Грузии. Анализ данных наземных и спутниковых источников в Грузии показал, что при использовании метода регрессии машинного обучения для обучения достаточно оценить данные за 1960-2000 годы, а для тренировки — данные за 2001-2022 годы. В качестве объекта обучения был выбран трехмесячный стандартизированный индекс осадков SPI-3. Использование моделей и алгоритмов машинного обучения для анализа распределения осадков, мониторинга и прогнозирования засух целесообразно как в Грузии, так и в других регионах при наличии соответствующей базы данных наблюдений (60 лет). Полученные результаты рекомендуется использовать в системах раннего оповещения при мониторинге засух.