

АНАЛИЗ НА МЕТЕОРОЛОГИЧНИ СВОБОДНО ДОСТЪПНИ ДАННИ ЗА ПРОГНОЗИРАНЕ НА ВРЕМЕТО С ИЗКУСТВЕН ИНТЕЛЕКТ

Михаил Първанов

*Национален институт по метеорология и хидрология (НИМХ),
студент бакалавърска програма „Информатика“ в Нов български университет
e-mail: mparvanovm@gmail.com*

Ключови думи: метеорологични данни, анализ на данни, изкуствен интелект

Резюме: Докладът представя модел за автоматично калибриране на данни от свободни и частни метеорологични станции. Чрез сравнение с официалните измервания алгоритъмът открива и коригира систематичните грешки, произтичащи от различия в хардуера, монтажните условия и локалната среда. По този начин се повишава точността и пространствената резолюция на наблюденията, което създава по-надеждна основа за изграждане на прогностични и аналитични модели в метеорологията и съществено подобрява локалните краткосрочни прогнози. Това е от особена важност за земеделието, градската управа и координация на действия свързани с системите за защита от градушки.

ANALYSIS OF METEOROLOGICAL OPEN DATA FOR AI-BASED FORECASTING

Mihail Parvanov

*National Institute of Meteorology and Hydrology (NIMH), undergraduate student in the “Informatics”
Bachelor’s program at New Bulgarian University
e-mail: mparvanovm@gmail.com*

Keywords: meteorological data, data analysis, artificial intelligence

Abstract: The report presents a model for automatic calibration of data from open and private weather stations. By comparing sensor readings with official measurements, the algorithm detects and corrects systematic errors arising from differences in hardware, installation conditions, and local environments. This process increases the accuracy and spatial resolution of observations, providing a more reliable basis for building predictive and analytical models in meteorology and significantly improving local short-term forecasts. The approach is particularly important for agriculture, urban management, and the coordination of operations related to hail protection systems.

Въведение

Свободните и частни метеорологични станции предоставят ценна, но по правило некалибрирана информация, много от тях имат отворена архитектура, което позволява модификация на фърмуера, и интеграция със системи то по-високо ниво. В този процес техните собственици са отворени и биха оказали нужното съдействие, за да осигурят максимална калибрация и достоверност на получаваните данни. Те имат висока пространствена плътност и могат да допълнят официалната наблюдателна мрежа, но техните измервания се различават от тези на станциите на НИМХ поради различия в хардуера, височината на монтаж, градската среда и липсата на регулярна калибрация. Повечето създатели разчитат на заводската калибрация на сензорите, но нямат нито умения нито знания да развият този процес във времето. Целта на изследването е разработване на концептуален модел за автоматично калибриране, като по този начин се осигурява тяхната съпоставимост и надеждност.

Метеорологичните данни представляват ключов елемент в процеса на прогнозиране, управление на климатични рискове и на дългосрочни прогностични модели касаещи

земеделието и туризма. Наличните модули и модели позволяват изготвяне на краткосрочни прогнози, за анализ на дългосрочни тенденции, свързани с изменението на климата са необходими повече пространствени данни. Ефективното използване на тези данни дава възможност за изграждане на системи за ранно предупреждение при екстремни явления като наводнения, засушавания и бури.

Съвременните подходи в областта на метеорологията включват интегриране на данни от различни източници — наземни станции, сателити, радари и мобилни сензори. Комбинирането на тези потоци от данни позволява създаването на богати, многомерни информационни масиви, подходящи за обработка чрез алгоритми на изкуствения интелект и машинното обучение.

Настоящото изследване има за цел не само да представи статистически анализ на наличните данни, но и да демонстрира значението на предварителната им обработка и инженерството на признаци при изграждането на ефективни прогностични модели.

Климатичните промени и необходимостта от по-надеждни прогнози изискват разработването на устойчиви системи, базирани на висококачествени, дългосрочни и калибрирани данни. Това изисква огромно количество централизирани инвестиции в изграждане на инфраструктура, но и създава предпоставки за използването на отворени модели, които вече са инсталирани и интегрирани в множество системи. В България съществува добре развита мрежа от метеорологични станции, предоставящи данни с висока времева резолюция, която — в комбинация с данните от свободните сензори — може да формира основата на ново поколение AI-базирани прогностични системи.

Данни и методология

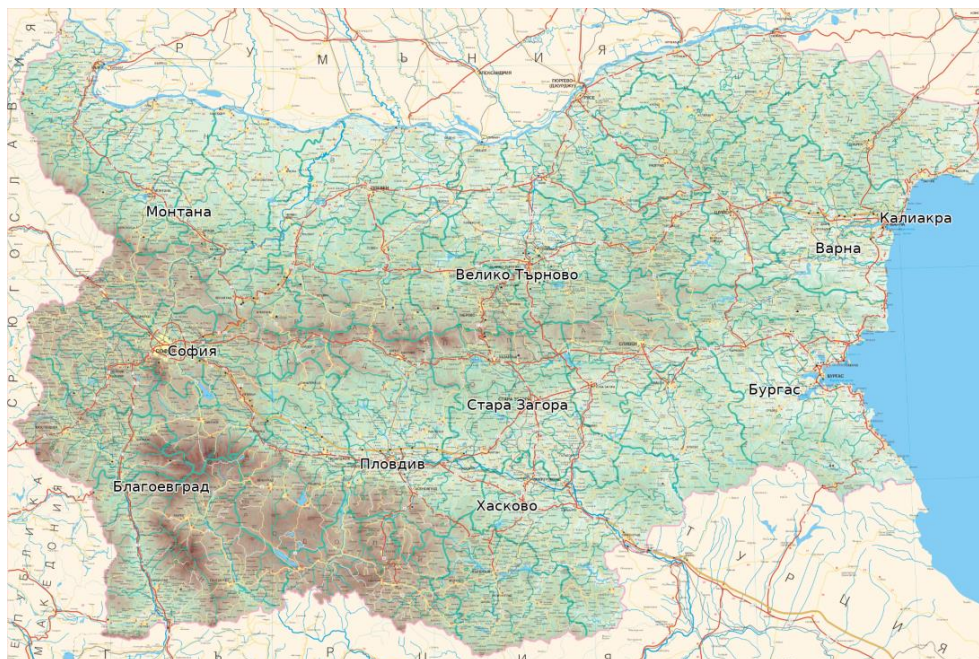
Изследването използва предварително обучен модел с данни от националната мрежа метеорологични станции на територията на България, обхващащи периода 1991–2025 г. Измерванията са извършвани на тричасови интервали, което позволява детайлно проследяване на динамиката на основните атмосферни параметри.

Основните използвани променливи са:

- Температура на въздуха (°C)
- Валежи (mm)
- Скорост и посока на вятъра (m/s, °)
- Относителна влажност (%)
- Атмосферно налягане (hPa)

Преди анализа данните преминават през процес на предварителна обработка, включващ:

- премахване на липсващи и некоректни стойности;
- хармонизиране на времевите формати;
- филтриране на шум и аномални отклонения;
- пространствено обвързване с географските координати и надморската височина на станциите.



Фиг. 1. Разпределение на станции

Мрежата от станции (фиг. 1) е разположена равномерно в различни климатични и географски зони, което осигурява представителност на данните. Това пространствено покритие е от съществено значение за моделиране на локални особености и микроклиматични ефекти. Планинските станции предоставят важна информация за вертикалните температурни градиенти и влиянието на надморската височина върху валежите и влажността — фактори, които са решаващи при прогнозиране на явления като интензивни снеговалежи и мъгли.

Времеви зависимости и нелинейности

Метеорологичните данни са силно зависими както от времето, така и от пространството, което позволява да се представят във времеви редове с нелинейни взаимовръзки. За да се улови тази динамика, в набора от данни се добавят лагови (закъснели) стойности и сезонни компоненти, които позволяват на алгоритмите да разпознават периодични цикли и краткосрочни тенденции. Подобен подход е особено ефективен при използване на невронни мрежи и модели за последователности като *LSTM* и *GRU*, които могат да обработват зависимости във времето, но и се хранят постоянно с множество реални данни.

Нелинейната природа на атмосферните процеси изисква прилагането на гъвкави модели, които да улавят сложните взаимодействия между температура, налягане, вятър и влажност. Комбинирането на тези параметри в общи предсказващи функции и включването на калибрани данни от различни източници води до по-точни и устойчиви прогнози, особено в локален мащаб при липса на достатъчно данни, каквито са радарни и сателитни изображения.

Приложения на изкуствения интелект в метеорологията

Изкуственият интелект (AI) намира все по-широко приложение в метеорологията, като допълва традиционните числени методи за прогнозиране. Алгоритмите за машинно обучение могат да извличат закономерности от големи обеми данни и да улавят сложни зависимости, които често остават скрити за класическите статистически модели. Важно е да се отбележи, че те не бива директно да бъдат замествани, но могат да се ползват локално, като матрици, които да бъдат ползвани от множество устройства разположени на голяма географска територия.

AI се прилага успешно в следните направления:

- Краткосрочни прогнози с висока пространствена и времева резолюция;
- Идентифициране на екстремни явления, като бури, градушки и интензивни валежи;
- Дългосрочно климатично моделиране и анализ на тенденции;
- Интеграция на хетерогенни източници на данни – сателити, радарни станции и мобилни сензори;
- Оценка на риска при природни бедствия, свързани с времето.

Комбинирането на свободни и частни наблюдения, след корекция чрез модел за калибрация, разширява пространственото покритие и обогатява входните данни за AI алгоритмите. Това позволява изграждането на прогностични системи, които са едновременно по-гъвкави, по-точни и по-адаптивни към локалните условия. Това ги прави подходящи за нуждите на земеделието.

За повишаване на точността на прогностичните модели се извършва инженеринг на признаци — създаване на нови, информативни параметри от наличните данни. Такива признаци включват:

- Закъснели стойности (lag features) – за улавяне на динамиката и инерцията на атмосферните процеси;
- Сезонни компоненти и индекси – за моделиране на годишните цикли и климатичната периодичност;
- Комбинирани фактори – като температурно-влажностни и налягане-вятър зависимости, които улавят нелинейните взаимодействия между параметрите.

Инженерство на признаци и адаптивност на модела

Класическото представяне на сезонността чрез адитивни хармонични функции предполага стабилна и симетрична годишна цикличност. В условията на съвременния климат се наблюдават размествания по фаза и промени в амплитудата и продължителността на сезонните преходи. Това води до нарушаване на традиционната периодичност, поради което стандартните хармонични модели вече не описват адекватно реалните климатични процеси.

За по-реалистично улавяне на сезонните и краткосрочните колебания се прилагат времезакъснителни и изпреварващи модели, при които даден параметър (напр. температура или влажност) се изразява не само чрез своите минали стойности, но и чрез прогнози за бъдещи тенденции, извлечени от локални зависимости. По този начин моделът става чувствителен към несинхронни сезонни сигнали — като аномално топли зими или ранни пролетни затопляния.

За откриване на непериодични и скокови промени се използва уейвлет анализ, който позволява разлагане на времеви ред по честоти и засичане на локални изменения в сезонната структура. За разлика от класическите хармонични редове, уейвлетите описват динамични и променливи цикли, което ги прави подходящ инструмент за анализ на нестационарни атмосферни процеси.

В контекста на изменящия се климат е възможно да се проявяват бифуркационни поведения – резки преходи в режима на атмосферната система при достигане на определен праг (например внезапно застудяване или активиране на конвективни процеси). За описване на подобни нелинейни преходи се използват ансамблови модели на градиентно буустване, като Gradient Boosting Regressor, XGBoost и LightGBM, които са способни да моделират сложни зависимости и взаимодействия между входните променливи.

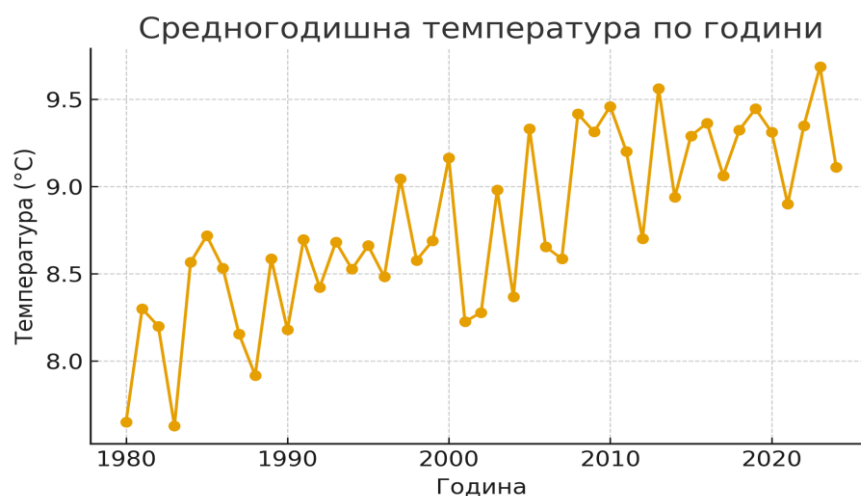
Бифуркационните индикатори се въвеждат като допълнителни управляващи входове, чрез които моделът адаптивно регулира своите вътрешни коефициенти и тегла в зависимост от степента на динамична нестабилност в системата, детектирана за реален предишен времеви ред, касаещ точната прогноза сега. По този начин се реализира контролирано преминаване между различни прогностични режими, което повишава стабилността и точността на модела при екстремни събития.

Комбинирането на уейвлет анализ (за детекция на локални промени), градиентно буустване (за нелинейно обучение) и бифуркационно управление (за адаптация в реално време) създава мощна основа за изграждане на динамични AI модели, способни да реагират на променящите се атмосферни режими и да поддържат висока прогностична надеждност дори при нестабилни климатични условия. Поради тази причина анализът на локалните аномалии е от съществено значение, което налага агрегацията на множество данни, дори и такива от некалибрирани сензори, тъй като отчитаме локалните градиенти на параметрите, а не точните стойности.

Комбинацията от уейвлет анализ, времеви лагове и бифуркационно управление формира адаптивен подход за извличане и актуализиране на признаци, който не разчита на фиксирана сезонност, а следва реалната еволюция на атмосферните процеси. Така се изграждат по-гъвкави и устойчиви AI модели, способни да реагират на преходи между климатични режими и да поддържат висока прогностична точност при променяща се среда.

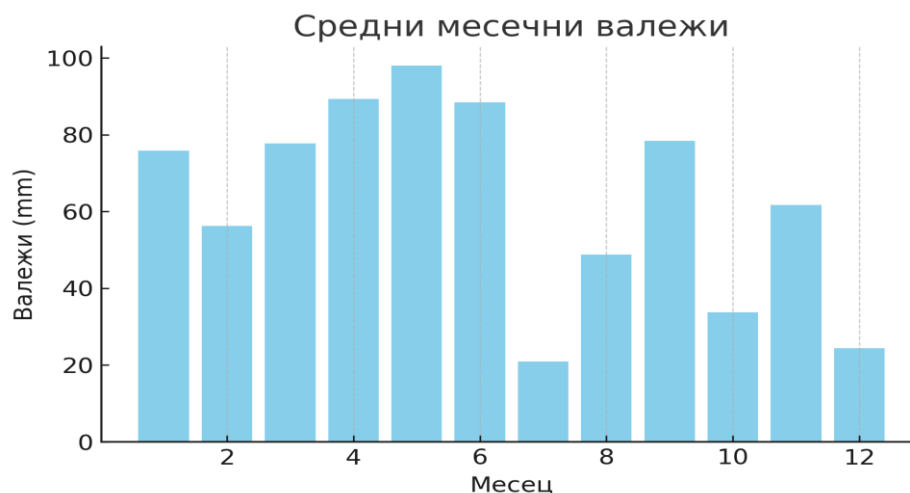
Анализ и резултати

Анализът показва ясно изразена сезонност в температурата и валежите. Средногодишната температура варира от 6 °C в планинските райони до 14 °C в южните. Валежите са най-високи през месеците май и юни. Наблюдава се и обща тенденция на затопляне от 1.2 °C за последните 40 години.



Фиг. 2. Средногодишна температура по години

Наблюдава се отчетлив възходящ тренд в средногодишната температура (фиг. 2) за периода 1980–2024 г., който е в синхрон с глобалните климатични тенденции. Тази информация е изключително ценна при създаването на модели за прогнозиране на климатични промени. Вижда се, че вариациите в температурата не са линейни — има периоди на по-бързо покачване и стабилизация. Такива характеристики трябва да се вземат предвид при създаването на модели за времеви редове.



Фиг. 3. Средни месечни валежи

Видно е ясно изразено увеличение на валежите (фиг. 3) през пролетните и началото на летните месеци, което е характерно за умерения климат в България. Тази сезонност е от особено значение при моделиране на водни ресурси, селско стопанство и хидрологични процеси. По-високите стойности на валежите през определени месеци могат да бъдат комбинирани с информация за температура и влажност, за да се създадат комплексни индикатори, които подобряват точността на прогнозните модели.

Заклучение

Изследването подчертава значението на висококачествените, дългосрочни метеорологични данни като основа за надеждни AI прогнози. В бъдеща работа се предвижда разширяване на набора от станции, използване на дълбоки невронни мрежи и климатични сценарии до 2050 г. Настоящото изследване представя детайлен анализ на пространственото разпределение на ключови български градове върху политическа карта на страната, използвайки реални географски координати за тяхната точна локализация. Чрез визуализацията на градове като София, Пловдив, Варна, Бургас, Велико Търново, Благоевград, Хасково, Монтана, Калиакра и Стара Загора се постига ясно разбиране за тяхното географско разположение спрямо основните геофизични и инфраструктурни особености на България. Подобен подход предоставя прецизна и лесна за интерпретация информация, която е особено ценна в контекста на метеорологични изследвания, климатично моделиране и създаване на AI-базирани системи за прогнозиране.

Точната географска привързаност на тези локации има съществено значение за коректното интерпретиране на наблюдателни данни от метеорологични станции и за оптималното им интегриране в модели за числено прогнозиране на времето. Например, местоположенията на крайбрежните градове (като Варна, Бургас) са ключови при анализ на морски и климатични влияния, докато вътрешните градове (София, Пловдив и Велико Търново) позволяват по-добро разбиране на континенталните ефекти и орографското въздействие. Освен това, визуалното позициониране на тези градове върху карта с политически и географски контур е важен елемент при разработването на обучителни и прогнозни модели, които разчитат на точни входни параметри. Това има пряко приложение не само в метеорологията, но и в области като управление на природни ресурси, градско планиране, анализ на климатични рискове и кризисен мениджмънт.

В заключение, изготвянето на географски точни карти с ключови градски точки не е просто визуална помощ, а съществен компонент от изграждането на надеждни и прецизни научни и практически системи. В бъдещи разработки подобни визуализации могат да бъдат обогатени с допълнителни слоеве данни — като топографски, климатични и инфраструктурни — което ще позволи още по-задълбочени анализи и моделиране на сложни природни процеси.

Литература:

1. Warner, T. T. (2010). Numerical Weather and Climate Prediction. Cambridge University Press. ISBN 978-0-521-51389-0
2. Santurette, P., & Georgiev, C. (2005). Weather Analysis and Forecasting: Applying Satellite Water Vapor Imagery and Potential Vorticity Analysis. Academic Press. ISBN 978-0-12-619262-9
3. Camps-Valls, G., Tuia, D., Zhu, X. X., & Reichstein, M. (2021). Deep Learning for the Earth Sciences: A Comprehensive Approach to Remote Sensing, Climate Science and Geosciences. Wiley-Blackwell. ISBN 978-1-119-64618-6
4. De Smith, M. J., Longley, P. A., & Goodchild, M. F. (2023). Geospatial Analysis (7th ed.). Winchelsea Press. ISBN 978-1-908-44213-3
5. Jensen, J. R. (2019). Remote Sensing and GIS for Meteorology. CRC Press. ISBN 978-1-138-35764-0
6. Westra, E. (2015). Python Geospatial Development (2nd ed.). Packt Publishing. ISBN 978-1-78328-113-8
7. European Centre for Medium-Range Weather Forecasts (ECMWF). (2024). Reanalysis and Forecast Datasets. Retrieved from <https://www.ecmwf.int>
8. World Meteorological Organization (WMO). (2024). Global Observing System and Guidelines. Retrieved from <https://public.wmo.int>
9. NOAA National Centers for Environmental Information (NCEI). (2024). Global Climate Data. Retrieved from <https://www.ncei.noaa.gov>
10. European Environment Agency (EEA). (2024). Geospatial Climate Data for Europe. Retrieved from <https://www.eea.europa.eu>
11. National Institute of Meteorology and Hydrology (NIMH). (2024). Национални климатични архиви. Retrieved from <https://www.meteo.bg>
12. НИМХ – БАН. (2015). Климатичен атлас на България. София: Национален институт по метеорология и хидрология – БАН. ISBN 978-954-8889-54-3.