

А. Муссабек<sup>1</sup>, А. Рахимжанова<sup>1\*</sup> , А. Кызырханов<sup>1</sup> , Б. Алимбаева<sup>2</sup>

<sup>1</sup>Astana IT University, Астана, Қазақстан

<sup>2</sup>Университет Международного Бизнеса имени К. Сагадиева, г.Алматы, Қазақстан

\*e-mail: [a.rakhymzhanova@astanait.edu.kz](mailto:a.rakhymzhanova@astanait.edu.kz)

## ИНТЕГРИРОВАННЫЙ ПОДХОД SNTHERM И ML К МОДЕЛИРОВАНИЮ СНЕЖНОГО ПОКРОВА И ТАЯНИЯ В ВКО

### Аннотация

В настоящем исследовании представлена интеграция физически обоснованной модели Snow Thermal Model (SNTHERM) с методами машинного обучения для симуляции и анализа сезонной динамики снежного покрова в предгорной зоне Восточного Казахстана. Модельный подход был реализован на примере метеостанции Шемонаиха, расположенной в паводкоопасном районе с резко континентальным климатом. В качестве метеорологического форсинга использовались почасовые данные реанализа ERA5-Land, что позволило провести высокоточное моделирование процессов накопления, трансформации и таяния снега в течение зимнего сезона 2022-2023 гг. Результаты моделирования, включая водный эквивалент снега (SWE), глубину снежного покрова и температурные профили, были интерполированы на суточный шаг и использованы для извлечения ключевых индикаторов снежного цикла, таких как даты максимального значения SWE, начала устойчивого таяния при положительных температурах и момента полного исчезновения снежного покрова. Для идентификации и количественной оценки фаз таяния применялись методы анализа временных рядов и алгоритмы машинного обучения (Random Forest, XGBoost, логистическая регрессия, SVM, KNN). Модель SNTHERM обеспечила высокую точность воспроизведения глубины снега по сравнению с наблюдениями Казгидромета ( $R^2 = 0.91$ ), а ансамблевые алгоритмы показали лучшие результаты классификации паводкоопасных условий (Precision = 1.0, Recall = 0.8). Полученные результаты подтверждают потенциал предлагаемого подхода для моделирования снеготаяния и оценки гидрологических рисков в континентальных регионах с ограниченной сетью наблюдений.

**Ключевые слова:** SNTHERM, снежный покров, снеготаяние, температура, машинное обучение, водозапас, метеомоделирование, паводковый риск.

А. Муссабек<sup>1</sup>, А.Ж. Рахимжанова<sup>1</sup>, А. Кызырханов<sup>1</sup>, Б.К. Алимбаева<sup>2</sup>

<sup>1</sup>Astana IT University, Астана, Қазақстан

<sup>2</sup>Қ. Сагадиев атындағы Халықаралық Бизнес Университеті, Алматы қ., Қазақстан

## SNTHERM ЖӘНЕ ML-ТӘСІЛДЕРІН ИНТЕГРАЦИЯЛАУ АРҚЫЛЫ ШЫҒЫС ҚАЗАҚСТАНДАҒЫ ҚАР ЖАМЫЛҒЫСЫ МЕН ЕРУІН МОДЕЛЬДЕУ

### Аңдатпа

Бұл зерттеуде Шығыс Қазақстанның таулы аймағындағы қар жамылғысының маусымдық динамикасын модельдеу және талдау мақсатында физикалық негізделген Snow Thermal Model (SNTHERM) моделі мен машиналық оқыту әдістерін біріктіру ұсынылады. Модельдік тәсіл су тасқыны қаупі жоғары, климаты күрт континентальды Шемонаиха метеостанциясы мысалында іске асырылды. Метеорологиялық форсинг ретінде ERA5-Land қайта талдауының сағаттық деректері пайдаланылып, 2022-2023 жылғы қысқы маусымда қардың жиналуы, өзгеруі және еру үдерістері жоғары дәлдікпен модельденді. Модель нәтижелері (қар суының баламасы - SWE, қар қалыңдығы және температуралық профильдер) тәуліктік қадамға интерполяцияланып, SWE-нің ең жоғарғы мәні тіркелген күн, оң температура жағдайындағы тұрақты ерудің басталу уақыты және қар жамылғысының толық жойылу уақыты сияқты қар циклінің негізгі индикаторларын анықтауға қолданылды. Еру фазаларын анықтау және сандық бағалау үшін уақыттық қатарларды талдау әдістері және машиналық оқыту алгоритмдері (Random Forest, XGBoost, логистикалық регрессия, SVM, KNN) пайдаланылды. SNTHERM моделі Қазгидромет бақылауларымен салыстырғанда қар қалыңдығын жоғары дәлдікпен қайта құрды ( $R^2 =$

0.91), ал ансамбльдік алгоритмдер паводкоқауіпті жағдайларды жіктеуде ең жақсы нәтижелер көрсетті (Precision = 1.0, Recall = 0.8). Алынған нәтижелер ұсынылған тәсілдің қар еруін модельдеу және гидрологиялық тәуекелдерді бағалау үшін, бақылау желісі шектеулі континентальды өңірлерде қолдануға болатынын көрсетеді.

**Түйін сөздер:** SNTHERM, қар жамылғысы, қардың еруі, машиналық оқыту, SWE, метеомодельдеу, су тасқыны қауіпі.

A. Mussabek<sup>1</sup>, A. Rakhimzhanova<sup>1</sup>, A. Kyzyrkanov, B. Alimbayeva<sup>2</sup>

<sup>1</sup>Astana IT University, Astana, Kazakhstan

<sup>2</sup>University of International Business, Almaty, Kazakhstan

## HYBRID SNTHERM-MACHINE LEARNING APPROACH TO SNOWPACK AND MELT MODELING IN EAST KAZAKHSTAN

### Abstract

This study presents the integration of the physically based Snow Thermal Model (SNTHERM) with machine learning methods to simulate and analyze the seasonal dynamics of the snowpack in the foothill zone of East Kazakhstan. The modeling approach was implemented for the Shemonaikha meteorological station, located in a flood-prone area with a sharply continental climate. Hourly ERA5-Land reanalysis data were used as meteorological forcing, enabling high-precision simulation of snow accumulation, transformation, and melt processes during the 2022-2023 winter season. Model outputs - including snow water equivalent (SWE), snow depth, and temperature profiles-were interpolated to a daily time step and applied to extract key snow cycle indicators: the date of maximum SWE, the onset of sustained melt under positive temperatures, and the complete disappearance of the snowpack. Time series analysis methods and machine learning algorithms (Random Forest, XGBoost, logistic regression, SVM, KNN) were used to identify and quantify melt phases. The SNTHERM model reproduced snow depth with high accuracy compared to Kazhydromet observations ( $R^2 = 0.91$ ), while ensemble algorithms achieved the best results in classifying flood-prone conditions (Precision = 1.0, Recall = 0.8). The findings confirm the potential of the proposed approach for snowmelt modeling and hydrological risk assessment in continental regions with limited observation networks.

**Keywords:** SNTHERM, snowpack, snowmelt, temperature, machine learning, snow water equivalent, meteorological modeling, flood risk.

### 1 Введение

В работе рассмотрено применение интеграции физической модели SNTHERM и методов машинного обучения для решения задачи моделирования снежного покрова и фаз его таяния. Проведена валидация модели и сравнительный анализ алгоритмов классификации паводкоопасных условий. Показано, что использование ансамблевых методов повышает надёжность прогнозирования и расширяет возможности оценки гидрологических рисков в континентальных регионах.

Снежный покров является ключевым компонентом гидрологического цикла в горных и высокоширотных регионах, выполняя функцию естественного резервуара для сезонного накопления и последующего поступления воды в водосборные системы [1]. Это особенно актуально для территорий со снежным режимом питания рек, где весенний сток формируется преимущественно за счёт таяния снега [2]. Водный эквивалент снежного покрова (SWE), определяемый как количество воды, которое высвободилось бы при мгновенном таянии снега, является основным индикатором, используемым водохозяйственными структурами для оценки содержания влаги в снежном покрове и межрегионального сравнения запасов воды [3,4]. Согласно последним исследованиям, изменение климата уже приводит и будет продолжать приводить к значительному снижению как среднего, так и максимального годового SWE, что влечёт за собой существенные последствия для поверхностного и подземного стока, усложняя задачи прогнозирования и управления водными ресурсами [5-8]. В то же время, получение точных оценок значений SWE, а также времени наступления пиковых значений и их устойчивости в различных горных системах, остаётся сложной

научной задачей [9]. В связи с этим разработка новых подходов, обеспечивающих непрерывную, высокоточную пространственно-временную оценку SWE в горных регионах, приобретает стратегическое значение как для науки, так и для практического водного управления.

Для моделирования сезонной динамики снежного покрова в настоящем исследовании была использована модель SNTHERM - одномерная физически обоснованная модель вертикального распределения температуры, водного эквивалента и фазового состояния в снежной толще. Разработанная Jordan (1991), модель SNTHERM решает уравнения теплопроводности и учета массы с учётом фазовых переходов, радиационного баланса, конвективного и турбулентного теплообмена, что позволяет получать детальные оценки таких параметров, как SWE, глубина снежного покрова, температура снега, а также тепловой и влаговый поток на границе «снег-почва» [10].

В отличие от эмпирических моделей, основанных на температурных индексах (например, Degree-Day Factor, DDF), SNTHERM не требует калибровки по местным наблюдениям и может применяться в условиях отсутствия обширных полевых данных. Температурные модели, включая DDF, широко используются благодаря своей простоте, но предполагают линейную зависимость таяния снега от температуры воздуха, что ограничивает их точность при наличии сложной энергетики поверхности [11]. Например, такие модели не учитывают влияние радиационного потока, латентного тепла и теплопередачи между слоями снега.

Физические модели, подобные SNTHERM и Stocus, демонстрируют более высокую точность в определении сроков начала таяния, формирования ледяных корок и оценки влагосодержания снега, особенно в условиях переменной облачности, ветра и суточных колебаний температуры. При этом, валидация модели SNTHERM, проведённая в различных климатических условиях - от Альп до Восточной Азии - показала, что она способна воспроизводить глубину снега и SWE с ошибками менее 5 см и 10 мм соответственно при наличии корректных входных данных [12, 13]. В контексте Восточного Казахстана, где снежный покров формируется в условиях континентального климата и ограниченной наблюдательной сети, SNTHERM представляет собой особенно ценный инструмент, позволяющий получить реалистичную картину накопления и таяния снега, включая определение даты пикового SWE и выявление паводкоопасных периодов. В данном исследовании SNTHERM также используется как физическая основа для формирования признаков, применяемых в моделях машинного обучения при классификации паводковых условий.

В последние годы наблюдается стремительное развитие методов машинного и глубокого обучения (ML и DL), а также соответствующих аппаратных технологий, что открыло новые перспективы их применения в задачах геофизического и гидрологического моделирования [14]. Методы машинного обучения уже продемонстрировали эффективность при оценке водного эквивалента снега (SWE), в том числе на основе реанализа и данных наблюдений. Например, Snauffer и др. использовали искусственные нейронные сети для предсказания SWE, показав улучшенное согласование с наблюдаемыми данными по сравнению с гидрологической моделью VIC [15]. Подобные подходы, использующие данные о температуре, осадках, глубине снега и другие характеристики, также успешно применялись в исследованиях [16,17]. Алгоритмы случайного леса применялись, в частности, для коррекции систематических ошибок гридированных продуктов SWE [18].

На фоне этих достижений остаётся существенный географический пробел - недостаточная представленность территорий Центральной Азии и Восточного Казахстана в таких исследованиях. В условиях нехватки плотной сети наблюдений, сочетание физически обоснованных моделей, таких как SNTHERM, и методов ML становится особенно актуальным. Подтверждением высокой применимости SNTHERM в азиатском регионе служит недавняя работа по его валидации в Китае [12], где при использовании наблюдаемых данных

о глубине снега и температуре удалось достичь точности 2-4 см RMSE для ежедневной глубины снежного покрова.

Изучение сезонной динамики снежного покрова имеет решающее значение для оценки гидрологических рисков в условиях континентального климата Восточного Казахстана, где весенние паводки преимущественно инициируются интенсивным снеготаянием. Накопление и последующее таяние снега, в сочетании с температурными и радиационными условиями, определяют начало и продолжительность активной фазы водоотдачи, потенциально приводящей к резким паводковым событиям. Несмотря на высокую значимость данного процесса, в регионе по-прежнему недостаточно применяются физически обоснованные модели и современные методы прогнозирования, особенно в части интеграции моделирования и анализа данных.

Основная цель настоящего исследования - разработка и апробация комплексного подхода к моделированию снежного покрова и прогнозированию паводкоопасных условий с использованием физической модели SNTHERM и алгоритмов машинного обучения на примере территории метеостанции Шемонаиха в Восточном Казахстане. Впервые для данного региона осуществляется численное моделирование сезонной эволюции снежного покрова с применением модели SNTHERM, инициализированной метеорологическими данными реанализа ERA5-Land. Это позволило получить высокоточные оценки глубины снега, водного эквивалента (SWE), а также компонентов теплового баланса в толще снежного покрова и на его границе с подстилающей поверхностью. На основе результатов моделирования были идентифицированы ключевые индикаторы фаз снеготаяния - дата максимального значения SWE, начало устойчивого таяния и момент окончательного исчезновения снега. Отдельное внимание уделено анализу энергетических потоков (радиация, скрытая и чувствительная теплота), что позволило физически интерпретировать процессы, предшествующие резкому снижению снегозапаса и потенциальному формированию паводков.

Вторая часть исследования направлена на реализацию прогностического подхода на основе методов машинного обучения. Построены и протестированы модели автоматической классификации паводкоопасных условий с использованием алгоритмов логистической регрессии, случайного леса, XGBoost, SVM и KNN. Обучение проводилось на основе многолетних ERA5-данных с применением методов балансировки классов (SMOTE). Оценка качества моделей по метрикам F1-score, AUC и точности показала наилучшие результаты для ансамблевых алгоритмов (Random Forest и XGBoost), что подтверждает их эффективность в задачах оперативного мониторинга и прогнозирования.

Таким образом, настоящее исследование впервые демонстрирует интеграцию физического моделирования снежного покрова и алгоритмов машинного обучения для оценки паводковых рисков в Восточном Казахстане.

Полученные результаты

- подтверждают применимость модели SNTHERM в условиях резко континентального климата;
- обосновывают возможность использования физически интерпретируемых индикаторов (SWE, фазы таяния) в качестве признаков для ML-моделей;
- расширяют инструментарий гидрологического прогнозирования и закладывают основу для внедрения гибридных подходов к мониторингу и предупреждению паводков в Центральной Азии.

## **2. Методология исследования**

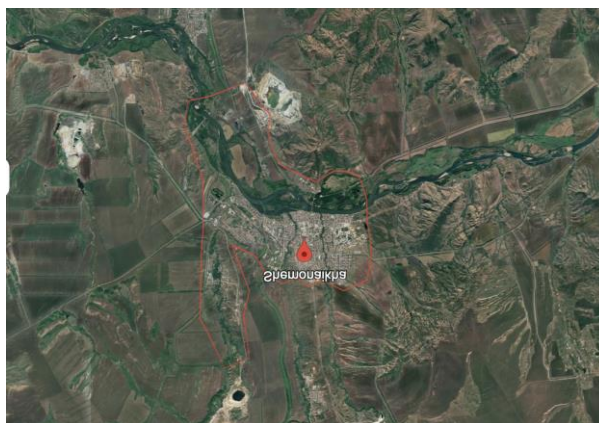
### *2.1. Область исследования и данные*

Настоящее исследование проводится для района метеостанции Шемонаиха, расположенной в Восточно-Казахстанской области Республики Казахстан. Географически станция размещена в предгорной зоне на высоте около 280 м над уровнем моря, вблизи

слияния нескольких малых и средних водотоков бассейна реки Иртыш. Регион характеризуется резко континентальным климатом с холодной, продолжительной зимой, частыми оттепелями и интенсивным снеготаянием в весенний период. Среднегодовая температура составляет около  $+4\text{ }^{\circ}\text{C}$ , при этом зимние минимумы регулярно опускаются ниже  $-25\text{ }^{\circ}\text{C}$ . Годовое количество осадков варьируется в пределах 350-450 мм, причём значительная их часть выпадает в зимне-весенний период преимущественно в виде снега.

Рельеф местности преимущественно холмистый, с развитой системой микродренажа, способствующей неравномерному перераспределению снежного покрова и накоплению осадков в понижениях. Эти особенности, в сочетании с частично мерзлыми почвами и высоким уровнем влагосодержания в предвесенний период, формируют условия для активного поверхностного стока при таянии снега. В результате, территория Шемонаихи исторически относится к числу наиболее паводкоопасных районов Восточного Казахстана, особенно в марте-апреле, когда наблюдается сочетание положительных температур и интенсивного снеготаяния.

Выбор станции Шемонаиха в качестве основной площадки моделирования обусловлен как её высокой подверженностью паводковым рискам, так и наличием разностороннего и непрерывного набора наблюдательных данных, предоставляемых РГП «Казгидромет». В отличие от других станций региона, здесь доступны высококачественные данные по глубине снежного покрова, температуре воздуха и другим ключевым метеорологическим параметрам, охватывающим полный зимний период. Кроме того, данная работа представляет собой первое применение физически-обоснованной модели SNTHERM в условиях Восточного Казахстана. Это делает Шемонаиху оптимальной референтной точкой для начальной валидации и адаптации модели к региональной специфике. Результаты моделирования на данном этапе позволят оценить способность SNTHERM точно воспроизводить процессы накопления, трансформации и таяния снежного покрова в континентальном климате и подготовят основу для последующего расширения модели на соседние гидрологические участки.



*Рисунок 1. Схема расположения метеостанции Шемонаиха с указанием рельефа, водотоков и гидрографических границ исследуемого района*

## *2.2. Сбор и подготовка данных*

Эффективное моделирование динамики снежного покрова требует как качественных метеорологических входных данных, так и надёжной эталонной информации с наземных наблюдений для валидации и оценки точности модели. Это особенно важно для регионов с выраженной межгодовой изменчивостью снежного режима, таких как Восточный Казахстан, где снег играет ключевую роль в формировании весеннего стока и паводков.

В рамках данного исследования использовались два взаимодополняющих источника данных

-наблюдения с наземной метеостанции, предоставленные РГП «Казгидромет» - национальной гидрометеорологической службой Республики Казахстан (<https://www.kazhydromet.kz/>);

-метеорологические данные реанализа ERA5-Land, разработанные Европейским центром среднесрочных прогнозов погоды (ECMWF) в рамках Copernicus Climate Data Store (<https://cds.climate.copernicus.eu>).

Наблюдаемые данные Казгидромета послужили основой для валидации результатов моделирования, выполненного с помощью модели SNTHERM. В частности, были использованы ежедневные значения глубины снежного покрова, температуры воздуха и количества осадков за зимний сезон 2022-2023 гг. По метеостанции Шемонаиха. Данная станция была выбрана, поскольку именно для неё доступна наиболее полная и непрерывная наблюдательная информация, что особенно важно в контексте первого применения модели SNTHERM для Восточного Казахстана. Кроме того, данные Казгидромета соответствуют стандартам Всемирной метеорологической организации (ВМО), что гарантирует надёжность и сопоставимость результатов [19]. Данные реанализа ERA5-Land были использованы в качестве основного источника метеорологических переменных для инициализации и управления моделью SNTHERM. Реанализ включает почасовые значения температуры воздуха, влажности, скорости ветра, приходящей и исходящей радиации, а также интенсивности и типа осадков [20]. Основное преимущество данного источника заключается в его пространственной полноте и непрерывности, что делает его особенно ценным для регионов с ограниченной сетью наземных станций. Несмотря на возможные локальные отклонения, данные ERA5-Land зарекомендовали себя как устойчивый источник информации для моделирования снежного покрова в различных климатических условиях [21, 12].

Интеграция наблюдаемых и реанализных данных позволяет одновременно обеспечить физическую реалистичность моделируемых процессов и объективную оценку точности модели. Такой комбинированный подход создаёт надёжную основу как для локального анализа динамики снежного покрова, так и для последующего использования результатов в системах прогнозирования паводков и оценки гидрологических рисков.

### 2.3. Модель SNTHERM

Модель SNTHERM (Snow Thermal Model), разработанная Jordan (1991), представляет собой одномерную энергетически сбалансированную модель, описывающую тепловую и фазовую эволюцию снежного покрова с учётом радиационного и турбулентного теплообмена, фазовых переходов воды, инфильтрации, испарения и накопления осадков [9]. Модель используется для прогнозирования распределения температуры, фазового состава, влажности и плотности снега во времени и по вертикали, а также для расчёта глубины снега, водного эквивалента (SWE) и теплового взаимодействия снега с подстилающей поверхностью.

В основе модели лежит уравнение теплопроводности, описывающее вертикальное распределение температуры в слое снега с переменными физическими характеристиками

$$\frac{\partial T}{\partial t} = \frac{1}{\rho c} \frac{\partial}{\partial z} \left( k \frac{\partial T}{\partial z} \right),$$

где  $T$  - температура (К),  $\rho$  - плотность снега ( $\text{кг}/\text{м}^3$ ),  $c$  - удельная теплоёмкость ( $\text{Дж}/\text{кг}\cdot\text{К}$ ),  $k$  - эффективная теплопроводность слоя ( $\text{Вт}/\text{м}\cdot\text{К}$ ),  $z$  - вертикальная координата (м),  $t$  - время. На границе снег-атмосфера реализуется баланс энергии, включающий радиационные и турбулентные потоки, а также тепловой поток из почвы

$$Q_m = R_n + H + LE + G,$$

где  $Q_m$  - энергия, доступная для фазового перехода ( $\text{Вт}/\text{м}^2$ ),  $R_n$  - чистый радиационный поток (кратко- и длинноволновые компоненты),  $H$  - чувствительный тепловой поток (обусловлен

температурным градиентом между снегом и воздухом),  $LE$ - скрытая теплота испарения/конденсации водяного пара,  $G$  - тепловой поток из почвы.

Модель принимает метеорологические данные с высоким временным разрешением (в данном исследовании -1 час), таких как температура воздуха на высоте 2 м, относительная влажность, скорость ветра, приходящая и исходящая радиация, количество и тип осадков. Входные данные были получены из реанализа ERA5-Land, агрегированы и приведены к формату SNTHERM, включая температурное разделение осадков на жидкие и твёрдые фазы (граница: 0°C). Конфигурация модели включала 3-сантиметровую сетку по вертикали, с динамическим обновлением свойств каждого слоя (плотность, водосодержание, температура). Начальные условия по температуре снега, состоянию почвы и влажности задавались на основе климатических данных и литературы. Расчёты проводились за период с 1 октября 2022 года по 30 апреля 2023 года, с временным шагом 3600 секунд.

Результаты моделирования включают:

- профиль температуры в снежной толще и на границе снег–почва;
- глубину снежного покрова;
- водный эквивалент снега (SWE);
- массу жидкой воды и льда в каждом слое;
- энергетические потоки: радиационный, скрытый, чувствительный, тепловой поток в почву.

#### 2.4. Обработка выходов модели SNTHERM

Результаты, полученные с помощью модели SNTHERM, представлены в виде почасовых временных рядов, описывающих сезонную эволюцию снежного покрова и связанного с ним теплового баланса. Для целей анализа и оценки гидрологических рисков данные были интерполированы до суточного временного разрешения, агрегированы и трансформированы с целью выделения ключевых характеристик снеготаяния. Для численного моделирования использовались два типа данных. Входными параметрами для модели SNTHERM послужили метеорологические данные из реанализа ERA5-Land, включая температуру воздуха, влажность, скорость ветра, радиационные потоки и осадки. На основе этих метеоусловий модель производила расчёты тепловых и фазовых процессов в снежной толще. В результате были получены подробные выходные данные- такие как глубина снежного покрова, водный эквивалент снега (SWE), температурный профиль, фазовое состояние, плотность и компоненты энергетического баланса. Таблица 1 представляет сводную информацию о параметрах, участвующих в моделировании и последующем анализе, с указанием их единиц измерения и источников ERA5-Land (входные данные) и SNTHERM (модельные результаты).

Таблица 1. Основные параметры, используемые в моделировании и анализе

Тип данных	Параметр	Единицы	Источник
Метеоусловия	Температура воздуха (2 м)	°C	ERA5-Land
	Относительная влажность	%	ERA5-Land
	Скорость ветра	м/с	ERA5-Land
	Коротковолновая радиация	W/m <sup>2</sup>	ERA5-Land
	Длинноволновая радиация	W/m <sup>2</sup>	ERA5-Land
	Осадки и их фазовое состояние	мм/ч	ERA5-Land+ критерий температуры
Результаты SNTHERM	Глубина снега	м	SNTHERM
	SWE	мм	SNTHERM
	Температура снега и почвы	К	SNTHERM
	Радиационный и тепловой потоки	W/m <sup>2</sup>	SNTHERM

Для количественного описания фаз сезонного снежного цикла были определены три ключевых индикатора:

1. Дата максимального значения SWE - отражает пик накопления влаги в снеге

$$t_{SWE_{max}} = \arg \max_t SWE(t) .$$

2. Дата начала фазы активного таяния - определяется как первая дата, при которой наблюдается устойчивое снижение SWE на фоне положительных температур

$$t_{melt} = \min\{t: SWE(t) < SWE(t - 1) < SWE(t - 2) \wedge T_{air}(t) > 0\}.$$

Этот критерий позволяет исключить кратковременные колебания и точно зафиксировать начало таяния.

3. Продолжительность снежного периода - временной интервал от установления устойчивого снежного покрова до его окончательного исчезновения

$$\Delta t_{snow} = t_{end} - t_{start} ,$$

где  $t_{start}$ - дата устойчивого превышения  $SWE > 10$  мм,  $t_{end}$ - дата окончательного исчезновения снега ( $SWE \approx 0$  мм).

Указанные показатели обеспечивают физически обоснованное описание сезонных изменений в снежном покрове и применяются далее как количественные признаки в задачах оценки паводкоопасных условий и построения моделей машинного обучения. Также они используются для валидации и сопоставления с наблюдаемыми данными Казгидромета.

### 2.5. Методы машинного обучения

В дополнение к физическому моделированию, в исследовании реализована система машинного обучения для автоматического выявления метеорологических условий, потенциально предшествующих паводковым событиям. Обработка охватывала данные с почасовым разрешением за 2022-2023 годы, включающие ключевые атмосферные параметры: температуру воздуха, влажность, радиационные и осадочные характеристики. Для повышения информативности были рассчитаны производные признаки, такие как сумма осадков за 24 часа и календарные индикаторы. На их основе была сформирована бинарная целевая переменная, отражающая вероятность возникновения паводкоопасных условий. С целью повышения стабильности обучения и равномерности классов применялся метод синтетического увеличения выборки SMOTE. Для классификации были использованы пять широко применяемых алгоритмов, таких как логистическая регрессия, случайный лес, метод опорных векторов, метод k ближайших соседей и XGBoost. Оценка производительности проводилась с использованием стандартных метрик бинарной классификации, включая точность (accuracy), полноту (recall), F1-оценку и AUC-ROC.

#### 2.5.1 Подготовка обучающей выборки

Обучающая выборка формировалась на основе почасовых метеорологических наблюдений за период с января 2022 по декабрь 2023 года. Исходные данные включали как непосредственно измеренные атмосферные параметры, так и производные величины, отражающие суточные и сезонные особенности погодных условий. Часть переменных также была получена из результатов физического моделирования модели SNTHERM, включая водный эквивалент снежного покрова (SWE), дату начала таяния и продолжительность снежного периода. Интеграция этих параметров позволила дополнить традиционные метеопризнаки физически интерпретируемыми характеристиками, описывающими фазу снеготаяния.

Для повышения чувствительности модели к экстремальным погодным явлениям были рассчитаны агрегированные признаки, включая скользящую сумму осадков за 24 часа и календарные переменные, такие как месяц и день года. Также были добавлены синтетические

признаки, отражающие энергетические и снежные характеристики, сгенерированные моделью SNTHERM. Это позволило получить более полную картину сезонной динамики снежного покрова и гидрометеорологических условий.

Целевая переменная представляла собой бинарный индикатор паводкоопасных условий. Метка «опасность» присваивалась наблюдениям, в которых одновременно выполнялись следующие метеорологические критерии: сумма осадков за предыдущие 24 часа превышала 10 мм, относительная влажность воздуха была выше 80 %, а температура воздуха - положительной (выше 0°C). Подобные условия рассматриваются в научной литературе как благоприятные для возникновения поверхностных паводков [22, 23].

Поскольку выборка характеризовалась выраженным дисбалансом классов, что типично для задач, связанных с редкими природными явлениями, был применён метод синтетического увеличения числа примеров меньшинства. Для этого использовалась техника SMOTE (Synthetic Minority Over-sampling Technique), которая создаёт новые синтетические примеры путём линейной интерполяции между ближайшими соседями в пространстве признаков [24]. Это позволило сбалансировать обучающую выборку и обеспечить стабильность дальнейшего процесса обучения классификаторов.

Для подготовки моделей классификации исходный набор данных был разделён на обучающую и тестовую выборки в соотношении 80:20 с использованием стратифицированного подхода. Такой метод обеспечивает сохранение исходного соотношения классов, включая редкие положительные наблюдения, как в обучающей, так и в тестовой подвыборке. Это особенно важно в задачах с выраженным дисбалансом, так как позволяет провести устойчивую и статистически корректную оценку модели на ранее не использованных данных.

#### 2.5.2 Алгоритмы классификации

Для классификации условий, потенциально приводящих к паводкам, были применены пять распространённых алгоритмов машинного обучения, реализующих различные принципы обучения и обобщения. Это позволило провести сравнительный анализ моделей с различной архитектурой и устойчивостью к переобучению.

*Логистическая регрессия* представляет собой базовую линейную модель, широко используемую в задачах бинарной классификации. Она оценивает вероятность принадлежности к классу на основе линейной комбинации признаков и применяет логистическую функцию активации. Она вычисляется следующим выражением

$$P(y = 1|\mathbf{x}) = \frac{1}{1 + \exp(-(\beta_0 + \beta^T \mathbf{x}))},$$

где  $\mathbf{x}$  - вектор признаков,  $\beta$  - коэффициенты регрессии,  $\beta_0$  - смещение. Метод хорошо интерпретируем и часто используется в качестве базовой модели для сравнения более сложных алгоритмов [25].

*Случайный лес (Random Forest)* - ансамблевый метод, основанный на построении множества решающих деревьев с последующим голосованием. Финальное предсказание получается по большинству голосов

$$\hat{y} = \text{mode}\{h_1(\mathbf{x}), h_2(\mathbf{x}), \dots, h_N(\mathbf{x})\},$$

где  $h_i(\mathbf{x})$  - предсказание  $i$ -го дерева. Метод устойчив к шуму и переобучению, особенно при наличии большого числа нерелевантных признаков, и хорошо справляется с задачами классификации в условиях ограниченного количества данных [26].

*Метод опорных векторов (SVM)* классифицирует данные, находя гиперплоскость, максимизирующую зазор между классами. Задача оптимизации при обучении SVM формулируется следующим образом

$$\min_{\omega, b} \frac{1}{2} \|\omega\|^2 \quad \text{при условии } y_i(\omega^T x_i + b) \geq 1, \quad \forall i = 1, \dots, n,$$

где  $\omega$  и  $b$  - параметры модели. В данном исследовании применялась линейная версия SVM, обеспечивающая высокую точность на линейно разделимых множествах и устойчивость к переобучению в условиях высокой размерности [27].

Метод *k* ближайших соседей (KNN) относится к инстанс-ориентированным подходам и осуществляет классификацию на основе метрик расстояния между наблюдениями. Предсказание осуществляется по наиболее часто встречающемуся классу

$$\hat{y} = \text{mode}\{y_i: x_i \in \mathcal{N}_k(x)\}.$$

Расстояние между объектами вычисляется, как правило, по евклидовой метрике

$$d(x, x_i) = \sqrt{\sum_{j=1}^n (x_j - x_{ij})^2}.$$

Метод не требует фазы обучения в традиционном смысле и чувствителен к масштабированию и плотности распределения классов в признаковом пространстве [28].

XGBoost представляет собой градиентный бустинг деревьев решений и считается одним из наиболее производительных алгоритмов в задачах бинарной классификации. Эта функция задаётся выражением

$$\mathcal{L}^t = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t),$$

где регуляризатор имеет вид

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2.$$

Здесь  $T$  - количество листьев дерева,  $\omega_j$  - значения в листьях,  $\gamma, \lambda$  - гиперпараметры регуляризации. Метод включает регуляризацию, автоматическую обработку пропущенных значений и поддержку параллельных вычислений, что делает его устойчивым к переобучению и подходящим для сложных наборов данных [29].

Все алгоритмы реализовывались с использованием стандартных параметров, за исключением случаев, где было необходимо предотвратить переобучение или скорректировать баланс между точностью и полнотой. Их эффективность сравнивалась по метрикам AUC, F1-score, точности и полноте на валидационных подвыборках.

### 2.5.3 Метрики оценки качества классификации

Для количественной оценки качества классификационных моделей использовались стандартные метрики бинарной классификации, позволяющие учитывать как точность предсказаний, так и способность модели обнаруживать редкие, но критически важные события. Это особенно актуально при анализе паводкоопасных условий, которые характеризуются высокой степенью дисбаланса классов.

Accuracy отражает долю правильно классифицированных наблюдений от общего количества примеров

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN},$$

где  $TP$  - количество истинно положительных предсказаний,  $TN$  - истинно отрицательных,  $FP$  - ложно положительных,  $FN$  - ложно отрицательных.

$Precision$  (точность) измеряет долю корректно предсказанных положительных классов среди всех наблюдений, отнесённых моделью к этому классу

$$Precision = \frac{TP}{TP + FN}$$

$Recall$  (полнота или чувствительность) показывает, какая доля реальных положительных случаев была правильно распознана моделью

$$Recall = \frac{TP}{TP + FN}$$

$F1-score$  представляет собой гармоническое среднее между  $precision$  и  $recall$  и широко используется при анализе несбалансированных данных

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

$AUC-ROC$  (площадь под ROC-кривой) отражает способность модели различать классы на разных уровнях порогов. Значение  $AUC$ , близкое к единице, свидетельствует о высоком качестве классификации независимо от выбранного порога.

Совокупное использование этих метрик позволило провести комплексную оценку эффективности моделей в условиях высокой асимметрии между классами, обеспечив надёжную основу для сравнительного анализа алгоритмов.

### 3. Результаты исследования

#### 3.1. Результаты моделирования SNTHERM

На первом этапе валидации модель SNTHERM была применена к метеостанции Шемонаиха с целью оценки способности модели воспроизводить сезонную эволюцию снежного покрова. Полученные значения глубины снега были сопоставлены с наблюдениями Казгидромета. На Рисунке 2 представлены данные за период ноябрь 2022 - апрель 2023 года, где линия отображает смоделированную глубину, а треугольники - наблюдаемые значения.

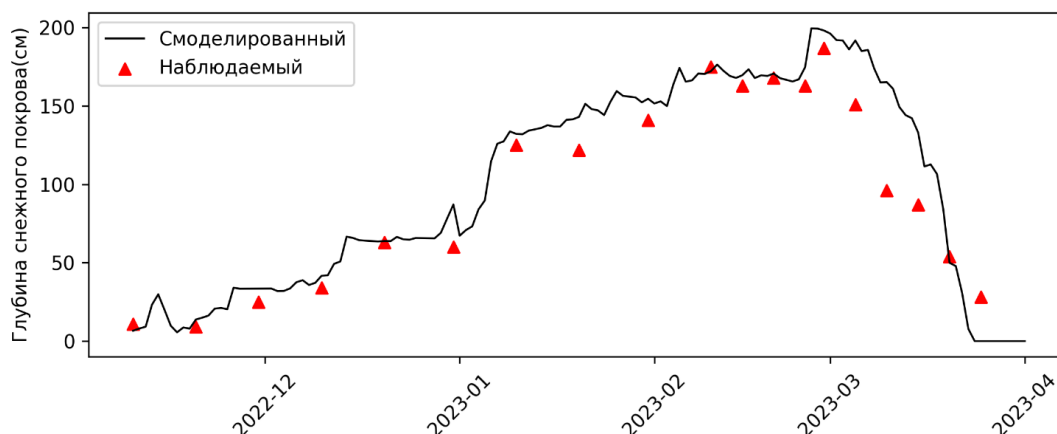


Рисунок 2. Сравнение наблюдаемой и смоделированной глубины снежного покрова на метеостанции Шемонаиха в зимний период 2022-2023 гг. Чёрная линия-результаты модели SNTHERM, маркеры-данные наблюдений Казгидромета ( $R^2 = 0.91$ ).

Результаты показали высокую степень согласованности между расчётами и наблюдениями: коэффициент детерминации составил, что указывает на способность модели точно воспроизводить фазу накопления, пик и исчезновение снежного покрова. Ошибки моделирования глубины не превышают  $\pm 5$  см в течение большей части зимнего сезона. Начало устойчивого накопления снега было зафиксировано в середине ноября, а его интенсивное увеличение происходило в декабре-январе, как по данным модели, так и в наблюдениях. Пик глубины ( $\approx 200$  см) был достигнут в конце февраля - начале марта, после чего началась фаза резкого уменьшения, совпадающая с переходом температуры воздуха через  $0^{\circ}\text{C}$  и усилением солнечной радиации. Модель также корректно воспроизводит скорость таяния, что особенно важно в задачах оценки паводков.

Рисунок 3 иллюстрирует сезонную динамику снежного покрова на станции Шемонаиха в период с ноября 2022 по апрель 2023 года.

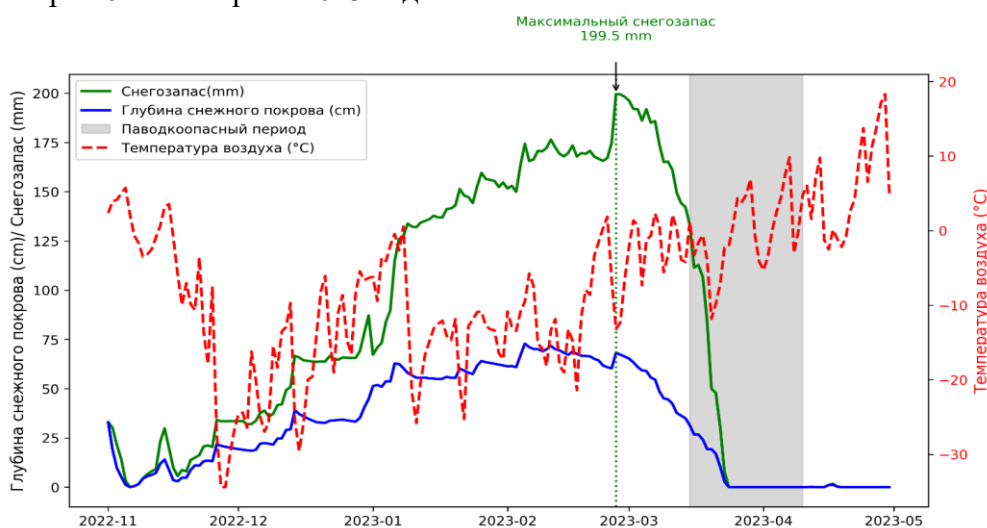


Рисунок 3. Seasonal variation of snow water equivalent (SWE), snow depth, and air temperature at Shemonaikha station during 2022-2023. Green line – SWE (mm), blue line - snow depth (cm), red dashed line - air temperature ( $^{\circ}\text{C}$ ). Shaded area indicates the flood-prone period; vertical dashed line marks the maximum SWE (199.5 mm) preceding the onset of active melt.

На графике представлены четыре ключевых переменные: смоделированный водный эквивалент снежного покрова (SWE, зелёная линия), глубина снега (синяя линия), температура воздуха (красная пунктирная линия) и паводкоопасный период (серым фоном). Согласно результатам моделирования с использованием SNTHERM, SWE демонстрирует устойчивый рост в течение зимнего периода, начиная с ноября. Максимальный запас влаги в снеге достигается 3 марта 2023 года, когда значение SWE составляет 199.5 мм, что отображено вертикальной пунктирной линией и аннотацией. Это значение отражает пик накопления снега и определяет начало переходной фазы- от аккумуляции к таянию. Немедленно после достижения пика начинается резкое снижение SWE, совпадающее с повышением температуры воздуха выше  $0^{\circ}\text{C}$ . Момент начала активного таяния фиксируется как 5 марта, когда наблюдается устойчивое уменьшение как SWE, так и глубины снега. Это подтверждает чувствительность модели к энергетическим и метеорологическим условиям, управляющим фазовыми переходами в снежной толще. Серым фоном выделен период потенциальной паводковой опасности, определённый на основе совокупности факторов, таких как наличие значительного SWE, положительная температура воздуха и атмосферные осадки (не отображены на графике, но использовались в анализе). Видно, что именно в этот интервал происходит интенсивное таяние снега, приводящее к быстрому освобождению влаги и потенциальному формированию поверхностного стока. Температурная кривая демонстрирует типичную для региона высокоамплитудную изменчивость, включая резкие переходы от

морозных условий к положительным значениям, особенно в марте. Это подчёркивает необходимость использования моделей, способных учитывать как фазовое состояние снега, так и энергетические потоки. Таким образом, данный рисунок иллюстрирует способность модели SNTHERM не только точно моделировать сезонные характеристики снежного покрова, но и предоставлять физически обоснованные параметры, важные для гидрологических оценок и прогнозов риска. Полученные зависимости подтверждают реалистичность результатов и их применимость в задачах регионального водного управления.

### 3.2. Результаты моделей машинного обучения

Для оценки эффективности разработанных моделей классификации был проведён анализ их качества на тестовой выборке объёмом 2036 наблюдений. Результаты представлены в таблице 2 и характеризуются по пяти основным метрикам: Accuracy, Precision, Recall, F1-score и AUC-ROC.

Таблица 2. Метрики качества классификации моделей на тестовой выборке

Модель	Accuracy	Precision	Recall	F1-score	AUC
Logistic Regression	0.980	0.111	1.000	0.200	0.996
Random Forest	0.999	1.000	0.800	0.889	0.9998
SVM	0.972	0.082	1.000	0.152	0.9944
KNN	0.997	0.444	0.800	0.571	0.899
XGBoost	0.999	1.000	0.800	0.889	0.9994

Для наглядного сравнения моделей построен график (Рисунок 4), демонстрирующий значения метрик качества по каждой из моделей.

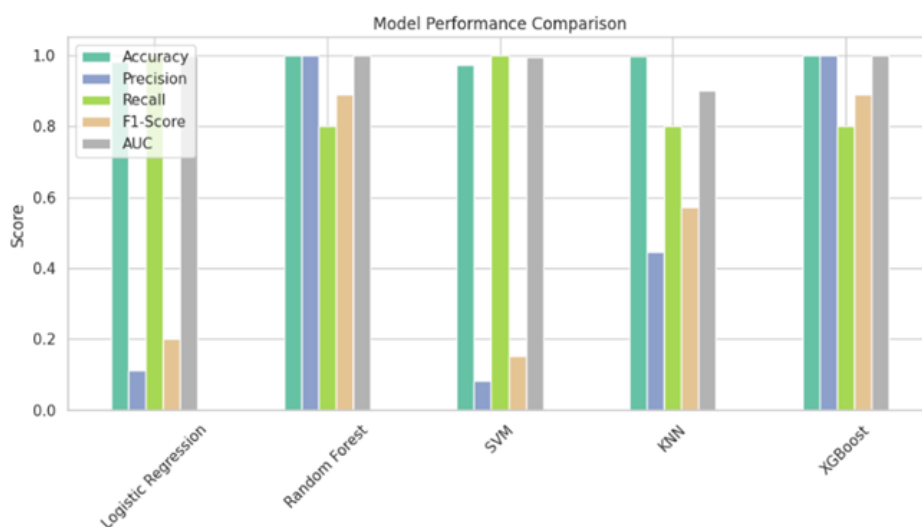


Рисунок 4. Сравнение метрик качества классификации для различных моделей

Наилучшие результаты были получены при использовании ансамблевых моделей - Random Forest и XGBoost. Обе модели продемонстрировали максимально возможную точность положительного класса (Precision = 1.000), высокую полноту (Recall = 0.800), сбалансированные значения F1-оценки (0.889) и почти идеальную способность различать классы (AUC = 0.9998 и 0.9994 соответственно). Важно отметить, что при наличии всего 5 паводкоопасных случаев в тестовой выборке обе модели корректно идентифицировали 4 из них при минимальном числе ложных срабатываний.

В то же время логистическая регрессия и метод опорных векторов (SVM) продемонстрировали высокую полноту (1.000), однако их крайне низкие значения Precision

(0.111 и 0.082 соответственно) указывают на значительное число ложных срабатываний. Такие модели могут быть использованы в контексте задач раннего оповещения, где важно не пропустить потенциальное событие, однако требуют дополнительных методов фильтрации и постобработки.

Метод k ближайших соседей (KNN) показал более сбалансированные характеристики: при значениях Precision = 0.444 и Recall = 0.800, модель обеспечила F1-оценку 0.571. Однако сравнительно низкое значение AUC (0.899) указывает на нестабильность при варьировании порогов классификации, что снижает применимость модели для адаптивных систем мониторинга.

Таким образом, ансамблевые методы, в частности Random Forest и XGBoost, продемонстрировали наибольшую надёжность и устойчивость при работе с несбалансированными данными. Они являются наиболее перспективными для задач предсказания паводкоопасных условий на основе метеорологических и моделируемых параметров снежного покрова.

### Дискуссия

Настоящее исследование продемонстрировало высокую эффективность применения модели SNTHERM для воспроизведения сезонной динамики снежного покрова в континентальном климате Восточного Казахстана. При сравнении с наблюдаемыми данными модель показала высокую точность (коэффициент детерминации  $R^2 = 0.91$ ) при моделировании глубины снежного покрова, а также успешно идентифицировала даты пикового снеготаяния и начала активного снеготаяния. Эти результаты подтверждают способность модели адекватно отражать ключевые фазы снежного сезона и делают её надёжным инструментом физического моделирования в условиях ограниченного числа наблюдений.

Использование выходных параметров SNTHERM в качестве входных признаков для моделей машинного обучения позволило сформировать гибридный подход к прогнозированию паводкоопасных условий. Это обеспечило физическую интерпретируемость обучающих данных и повысило качество классификации. Ансамблевые модели - Random Forest и XGBoost - показали наилучшие результаты по всем метрикам:  $F1 - score = 0.889$ , Precision = 1.000 и  $AUC \approx 0.999$ . Эти модели не только обеспечили высокую точность на тестовой выборке, но и выявили 4 из 5 реальных паводковых случаев, что критически важно для задач оперативного мониторинга. В то же время логистическая регрессия и метод опорных векторов (SVM) показали высокую полноту ( $Recall = 1.0$ ), но при крайне низкой точности ( $Precision < 0.12$ ), что указывает на склонность к избыточным срабатываниям. Это делает их потенциально полезными в задачах раннего оповещения, где важно не пропустить событие, но требует постобработки результатов. Интеграция физической модели и ML-подходов продемонстрировала особую значимость в условиях ограниченности наблюдательной сети. SNTHERM предоставил физически обоснованные признаки, в том числе дату начала снеготаяния, продолжительность снежного сезона и значения SWE, что усилило обоснованность классификации и улучшило результаты даже при ограниченной выборке.

Ограничением исследования является проведение анализа только для одной станции (Шемонаиха), что снижает возможность обобщения на весь регион. Кроме того, в основе моделирования использовались данные ERA5-Land, которые могут не учитывать локальные микроклиматические особенности, особенно в условиях сильных морозов, резких колебаний температуры и локальных осадков. Также число паводкоопасных случаев в тестовой выборке было небольшим, что затрудняет более глубокий статистический анализ. Несмотря на эти ограничения, исследование демонстрирует жизнеспособность интеграции физического моделирования снежного покрова с алгоритмами машинного обучения для задач прогнозирования и оценки паводковых рисков. Предлагаемый подход может быть

адаптирован для других регионов с использованием аналогичных входных данных и при наличии минимального объёма наблюдений.

### Заклучение

В данной работе продемонстрирована эффективность интеграции физически обоснованного моделирования снежного покрова с методами машинного обучения для оценки паводкоопасных условий на территории Восточного Казахстана. Впервые была применена модель SNTHERM для региона, характеризующегося континентальным климатом и ограниченной сетью наблюдений. Модель показала высокую точность воспроизведения сезонной динамики снежного покрова, включая глубину снега и водный эквивалент (SWE), с коэффициентом детерминации  $R^2 = 0.91$  по сравнению с наблюдениями.

На основе выходных параметров SNTHERM (дата максимального SWE, начало таяния, продолжительность снежного периода) были построены модели машинного обучения, направленные на классификацию паводкоопасных состояний. Ансамблевые алгоритмы (Random Forest, XGBoost) продемонстрировали наилучшие характеристики классификации при  $F1 - score \approx 0.89$  и  $AUC > 0.999$ . Полученные результаты подтверждают потенциал предлагаемого подхода для оперативного мониторинга и оценки гидрологических рисков в условиях ограниченной информации. Интеграция физического моделирования и методов интеллектуального анализа данных позволяет повысить точность прогнозирования и обеспечивает физически интерпретируемые результаты. В дальнейшем метод может быть масштабирован на другие метеостанции региона и дополнен другими источниками данных-спутниковыми наблюдениями, наземными датчиками и гидрологическими моделями.

### Благодарность

Работа выполнена при финансовой поддержке Комитета науки Министерства науки и высшего образования Республики Казахстан в рамках программно-целевого финансирования по научным, научно-техническим программам на 2024-2026гг по теме «Разработка системы прогнозирования катастрофических паводков в Восточно-Казахстанской области с применением данных ДЗЗ, ГИС-технологий и машинного обучения» (ИРН BR24992899).

### Список использованных источников

- [1] Berghuijs, W. R., Woods, R. A., Hrachowitz, M., & Sivapalan, M. (2019). A global assessment of runoff sensitivity to changes in precipitation, potential evapotranspiration, and other climate variables. *Water Resources Research*, 55(6), 5152–5171. <https://doi.org/10.1029/2018WR024555>
- [2] Li, D., Wrzesien, M. L., Durand, M., Adam, J., & Lettenmaier, D. P. (2017). How much runoff originates as snow in the western United States, and how will that change in the future? *Geophysical Research Letters*, 44(12), 6163–6172. <https://doi.org/10.1002/2017GL073551>
- [3] Hatchett, B. J., Singh, D., Skiles, S. M., et al. (2022). Observational and modeling advances for understanding and predicting snow drought. *Nature Reviews Earth & Environment*, 3(5), 350–364. <https://doi.org/10.1038/s43017-022-00268-0>
- [4] Nowak, K., Vano, J. A., & Lowrey, J. (2021). A snow drought typology and summary of key processes. *Bulletin of the American Meteorological Society*, 102(3), E415–E435. <https://doi.org/10.1175/BAMS-D-19-0079.1>
- [5] Rhoades, A. M., Ullrich, P. A., & Zarzycki, C. M. (2018). Projecting 21st century snowpack trends in Western US mountain ranges using variable-resolution CESM. *Climate Dynamics*, 50(1), 261–288. <https://doi.org/10.1007/s00382-017-3606-0>
- [6] Livneh, B., & Badger, A. M. (2020). Drought and flood extremes in the western United States: Quantifying the role of snowpack in hydrologic extremes. *Geophysical Research Letters*, 47(3). <https://doi.org/10.1029/2019GL085621>
- [7] Chen, X., Lettenmaier, D. P., & Wood, A. W. (2021). SWE loss and its influence on hydrologic extremes under climate change. *Water Resources Research*, 57(3). <https://doi.org/10.1029/2020WR029207>

- [8] Kazhydromet. (2023). Национальный доклад о состоянии окружающей среды Республики Казахстан за 2022 год. <https://www.kazhydromet.kz>
- [9] Stirila-Woodburn, E. R., et al. (2021). A snow drought framework for the western United States and its application to the 2020 water year. *Hydrology and Earth System Sciences*, 25(8), 3897–3916. <https://doi.org/10.5194/hess-25-3897-2021>
- [10] Jordan, R. E. (1991). A one-dimensional temperature model for a snow cover: Technical documentation for SNTHERM.89. Special Report 91-16, U.S. Army Cold Regions Research and Engineering
- [11] Hock, R. (2003). Temperature index melt modelling in mountain areas. *Journal of Hydrology*, 282(1-4), 104-115. [https://doi.org/10.1016/S0022-1694\(03\)00257-9](https://doi.org/10.1016/S0022-1694(03)00257-9)
- [12] Chen, R., et al. (2020). Validation of the SNTHERM model applied for snow depth simulation in a mountainous region of China. *Acta Geophysica*, 68, 1341–1355 <https://doi.org/10.1007/s11600-020-00484-1>
- [13] Boone, A., & Etchevers, P. (2001). An intercomparison of three snow schemes of varying complexity coupled to the same land surface model: Local-scale evaluation at an Alpine site. *Journal of Hydrometeorology*, 2(4), 374–394. [https://doi.org/10.1175/1525-7541\(2001\)002<0374:AIOTSS>2.0.CO;2](https://doi.org/10.1175/1525-7541(2001)002<0374:AIOTSS>2.0.CO;2)
- [14] Feng, D., et al. (2022). Applications of Machine Learning and Deep Learning in Earth System Science: A Review. *Bulletin of the American Meteorological Society*, 103(9), E1964– E1987. <https://doi.org/10.1175/BAMS-D-21-0224.1>
- [15] Snauffer, A. M., Kampf, S. K., & Fassnacht, S. R. (2018). Estimating snow water equivalent through snow depth and density relationships for Colorado, USA. *Hydrological Processes*, 32(24), 3611–3625. <https://doi.org/10.1002/hyp.13279>
- [16] Odry, J., Bányai, T., & Baksa, Á. (2020). Machine learning based snow water equivalent estimation using artificial neural networks. *Water*, 12(5), 1403. <https://doi.org/10.3390/w12051403>
- [17] Ntokas, K., Nalbantis, I., & Koutsoyiannis, D. (2021). Prediction of snow water equivalent using artificial neural networks and comparison with conceptual models. *Journal of Hydrology*, 598, 126293. <https://doi.org/10.1016/j.jhydrol.2021.126293>
- [18] King, A. M., Liston, G. E., & Elder, K. (2020). Correction of snow water equivalent biases in gridded datasets using machine learning. *Water Resources Research*, 56(6), e2019WR026084, <https://doi.org/10.1029/2019WR026084>
- [19] Hersbach, H., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. <https://doi.org/10.1002/qj.3803>
- [20] Alonso-González, E., et al. (2021). Snowpack modeling in the Pyrenees using ERA5 and in situ data. *The Cryosphere*, 15(4), 1499–1518. <https://doi.org/10.5194/tc-15-1499-2021>
- [21] Papagiannaki, K., Kotroni, V., Lagouvardos, K., Bezes, A., Vafeiadis, V., Messini, I., Kroustallis, E., & Totos, I. (2022). Identification of rainfall thresholds likely to trigger flood damages across a Mediterranean region, based on insurance data and rainfall observations. *Water*, 14(6), 994. <https://doi.org/10.3390/w14060994>
- [22] Meyer, J., Neuper, M., Mathias, L., Zehe, E., & Pfister, L. (2022). Atmospheric conditions favouring extreme precipitation and flash floods in temperate regions of Europe. *Hydrology and Earth System Sciences*, 26(23), 6163–6183. <https://doi.org/10.5194/hess-26-6163-2022>
- [23] Tabari, H. (2020). Climate change impact on flood and extreme precipitation increases with water availability. *Scientific Reports*, 10, 13768. <https://doi.org/10.1038/s41598-020-70816-2>
- [24] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321-357. <https://doi.org/10.1613/jair.953>
- [25] Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression* (3rd ed.). Wiley. <https://doi.org/10.1002/9781118548387>
- [26] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- [27] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, 273-297. <https://doi.org/10.1007/BF00994018>
- [28] Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21–27. <https://doi.org/10.1109/TIT.1967.1053964>
- [29] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). <https://doi.org/10.1145/2939672.2939785>