

УДК 551.578:519.246:004.942

UDC 551.578:519.246:004.942

4.1.3. Агрохимия, агропочвоведение, защита и карантин растений (биологические науки, сельскохозяйственные науки)

4.1.3. Agrochemistry, agro-soil science, plant protection and quarantine (biological sciences, agricultural sciences)

**МОДЕЛЬ ПРОГНОЗИРОВАНИЯ ЗАСУХ НА
ОСНОВЕ ВРЕМЕННОГО РЯДА ИНДЕКСА
SPI: ПОДХОД С ИСПОЛЬЗОВАНИЕМ
ГАУССОВСКИХ ПРОЦЕССОВ И
АДАПТИВНОЙ НАСТРОЙКОЙ ЯДРА**

**A DROUGHT FORECASTING MODEL BASED
ON SPI TIME SERIES: A GAUSSIAN
PROCESS APPROACH WITH ADAPTIVE
KERNEL TUNING**

Аврахам Елена Михайловна
канд. физико-математических наук
ФГБОУ ВО Кубанский государственный
аграрный университет имени И.Т. Трубилина,
Краснодар, Россия

Avraham Elena Mikhailovna
Candidate of Physical and Mathematical Sciences
Kuban State Agrarian University named after I.T.
Trubilin, Krasnodar, Russia

В статье представлено исследование, направленное на разработку модели предсказания метеорологической засухи с использованием данных, полученных методом дистанционного зондирования Земли. В работе представлен подход к прогнозированию индекса SPI с использованием моделей гауссовских процессов (Gaussian Process Regression, GPR). Данные о засухе в масштабах SPI-1 месяц рассчитаны на основе ежемесячных данных об осадках, полученных со спутников дистанционного зондирования земли в период с 1981 по 2024 год, были взяты в качестве входных данных в модели GPR. Данные о засухе SPI использовались в период с 1981 по 2022 год в качестве обучающих данных, а в период с 2022 по 2024 год — в качестве тестовых данных. Для точного моделирования временных рядов SPI была выбрана комбинация ядер, учитывающих разные аспекты данных (сезонность, периодичность и шумовые характеристики временного ряда SPI): WhiteKernel, Matern, ExpSineSquared. Особое внимание уделено адаптивной настройке параметров ядра для учета с использованием байесовского оптимизатора (Bayesian Optimization)

The article presents a study aimed at developing a meteorological drought prediction model using remote sensing data. The paper presents an approach to forecasting the SPI index using Gaussian Process Regression (GPR) models. SPI-1 month drought data calculated from monthly precipitation data obtained from remote sensing satellites for the period from 1981 to 2024 were taken as input to the GPR models. SPI drought data for the period from 1981 to 2022 were used as training data, and for the period from 2022 to 2024 - as test data. To accurately model the SPI time series, a combination of kernels taking into account different aspects of the data (seasonality, periodicity and noise characteristics of the SPI time series) was chosen: WhiteKernel, Matern, ExpSineSquared. Particular attention is paid to adaptive tuning of kernel parameters to account for the use of the Bayesian Optimizer (Bayesian Optimization)

Ключевые слова: ПРОГНОЗИРОВАНИЕ
ЗАСУХИ, ИНДЕКС SPI, ДИСТАНЦИОННОЕ
ЗОНДИРОВАНИЕ ЗЕМЛИ,
ГАУССОВСКИЕ ПРОЦЕССЫ (GAUSSIAN
PROCESS REGRESSION, GPR),
ЕЖЕМЕСЯЧНЫЕ ДАННЫЕ ОБ ОСАДКАХ,
СПУТНИКОВЫЕ ДАННЫЕ, ВРЕМЕННЫЕ
РЯДЫ, МЕТЕОРОЛОГИЧЕСКАЯ ЗАСУХА,
АНАЛИЗ ДАННЫХ

Keywords: DROUGHT FORECASTING, SPI
INDEX, EARTH REMOTE SENSING,
GAUSSIAN PROCESS REGRESSION (GPR),
MONTHLY PRECIPITATION DATA, SATELLITE
DATA, TIME SERIES, METEOROLOGICAL
DROUGHT, DATA ANALYSIS

<http://dx.doi.org/10.21515/1990-4665-208-026>

<http://ej.kubagro.ru/2025/04/pdf/26.pdf>

Обоснование актуальности исследования и обозначение проблемы.

В современных условиях изменение климата и увеличение частоты экстремальных погодных явлений, таких как засухи, представляют серьезную угрозу для сельского хозяйства, водных ресурсов и экономики в целом [1]. Прогнозирование засух на основе индекса стандартизированных осадков (SPI) становится актуальным инструментом для своевременного реагирования и принятия мер по минимизации их последствий [2]. Однако традиционные методы прогнозирования, такие как линейные модели или классические алгоритмы машинного обучения, часто сталкиваются с ограничениями при работе с нелинейными зависимостями, сезонными колебаниями и шумовыми компонентами данных. Эти ограничения снижают точность прогнозов и затрудняют выявление экстремальных значений, характеризующих засушливые периоды [3].

Гауссовские процессы (Gaussian Process Regression, GPR) представляют собой перспективный подход к моделированию временных рядов, который позволяет учитывать сложные нелинейные зависимости и обеспечивает гибкость за счет точной настройки параметров ядра. Благодаря своей способности работать с неопределенностью и адаптироваться к различным типам данных, GPR особенно подходит для задач, связанных с прогнозированием экстремальных явлений, таких как засухи.

Цель данной работы — разработка и оптимизация модели гауссовских процессов для прогнозирования SPI с акцентом на выявление экстремальных значений, характеризующих засушливые периоды.

Описание данных.

В исследовании использовался набор данных CHIRPS — это набор данных о квазиглобальных осадках за более чем 30 лет. CHIRPS объединяет

спутниковые снимки с разрешением $0,05^\circ$ и данные станций для создания сеточных временных рядов осадков для анализа тенденций и мониторинга сезонной засухи. Данные предоставлены платформой Earth Engine, которая предоставляет бесплатный доступ к данным для исследователей. Разрешение данных составило 5566 метров. Диапазоны данных представлены в таблице 1.

Имя	Единицы измерения	Мин.	Макс	Описание
precipitation	мм / сут	0*	1444.34*	Осадки

Таблица 1 — описание диапазонов данных (* обозначено расчетное минимальное или максимальное значение)

Данные осадков для исследования взяты для географии территории Краснодарского края Российской Федерации. На основе данных осадков был рассчитан стандартизированный индекс засухи SPI (Standardized Precipitation Index). Стандартизированный индекс осадков (SPI) — это широко используемый индекс для характеристики метеорологической засухи в различных временных масштабах. Для анализа использовались временные ряды SPI с месячным разрешением, охватывающие период в 30 лет. Данные включают значения SPI, рассчитанные на основе исторических осадков некоторых районов территории Краснодарского края.

Для проверки предположения о нормальности распределения значений SPI были использованы статистические тесты Шапиро-Уилка и Колмогорова-Смирнова. Эти методы позволяют оценить, насколько распределение данных соответствует гауссовскому, так как регрессия гауссовских процессов (GPR) предполагает гауссовское распределение данных, что делает эти тесты необходимыми для обоснования корректности применения модели.

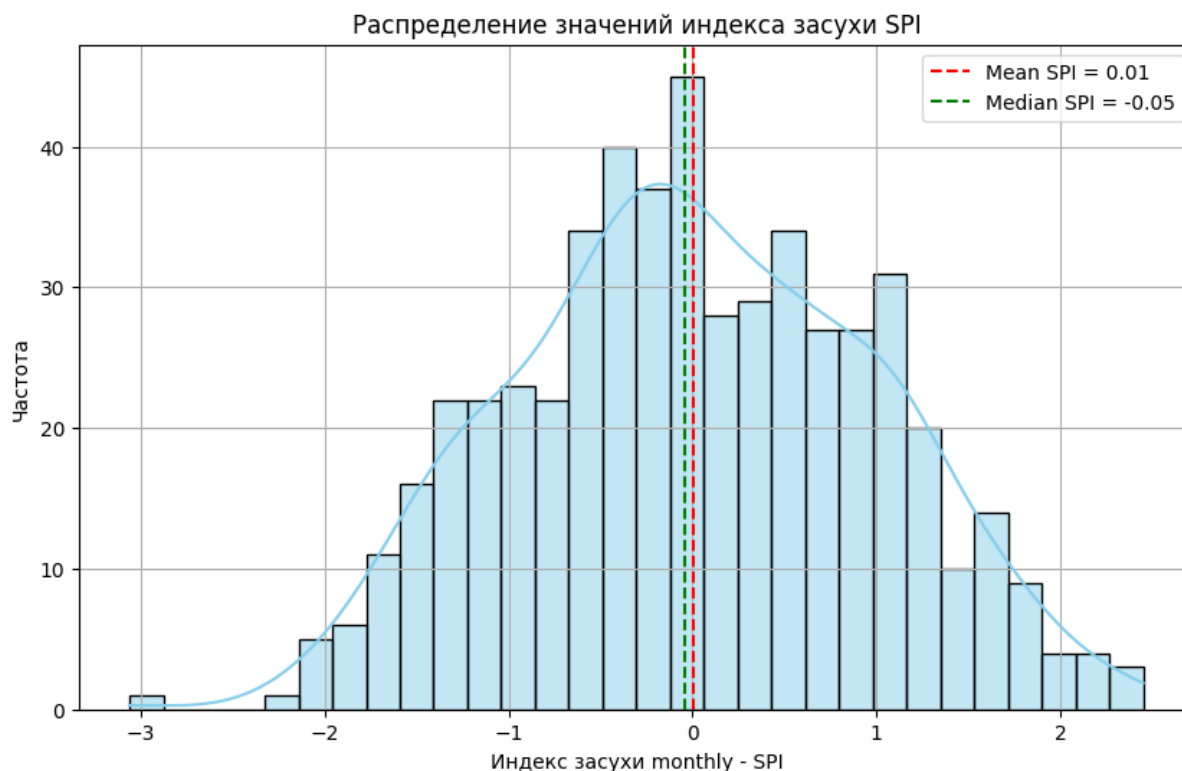


Рисунок 1 – распределение значений индекса засухи SPI

Методика исследования.

Гауссовские процессы представляют собой вероятностный подход к регрессии, где предполагается, что наблюдения распределены согласно многомерному гауссовскому распределению. Модель предполагает, что отклик y на каждом наблюдении состоит из истинного сигнала $f(\mathbf{x})$ и аддитивного гауссовского шума:

$$y = f(\mathbf{x}) + \epsilon,$$

где $f(\mathbf{x}) \sim \mathcal{GP}(\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$ — гауссовский процесс с априорным средним $\mu(\mathbf{x})$ и ковариационной функцией $k(\mathbf{x}, \mathbf{x}')$, $\epsilon \sim \mathcal{N}(0, \sigma_{\text{noise}}^2)$ — гауссовский шум с дисперсией σ^2 . Для всех точек входного пространства $X = \{x_1, x_2, \dots, x_n\}$, значения $f(x_i)$ априори распределены согласно многомерному нормальному закону:

$$f = [f(x_1), f(x_2), \dots, f(x_n)]^T \sim \mathcal{N}(\mu(X), K),$$

где $\mu(X) = [\mu(x_1), \mu(x_2), \dots, \mu(x_n)]^T$ — вектор среднего, $K = \{k(x_i, x_j)\}_{i,j=1}^n$ — ковариационная матрица, построенная с использованием ядра $k(x, x')$. Мы используем комбинированное ядро $k(\mathbf{x}, \mathbf{x}')$, составленное из трёх компонентов: шумового ядра, ядра Матерна и ядра экспоненциально затухающего синусоида:

$$k(\mathbf{x}, \mathbf{x}') = k_0(\mathbf{x}, \mathbf{x}') + k_1(\mathbf{x}, \mathbf{x}') + k_3(\mathbf{x}, \mathbf{x}'),$$

Шумовое ядро k_0 учитывает случайные шумы и записывается как:

$$k_0(\mathbf{x}, \mathbf{x}') = \sigma_{\text{noise}}^2 \cdot \delta(\mathbf{x}, \mathbf{x}').$$

Ядро Матерна k_1 моделирует долгосрочные тренды и плавные изменения. Оно задаётся как:

$$k_1(\mathbf{x}, \mathbf{x}') = \sigma_{\text{const}}^2 \cdot \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu} |\mathbf{x} - \mathbf{x}'|}{l} \right)^\nu K_\nu \left(\frac{\sqrt{2\nu} |\mathbf{x} - \mathbf{x}'|}{l} \right).$$

Ядро k_3 используется для моделирования периодических процессов и имеет вид:

$$k_3(\mathbf{x}, \mathbf{x}') = \sigma_{\text{const}_k3}^2 \cdot \exp \left(- \frac{2 \sin^2 \left(\pi \frac{|\mathbf{x} - \mathbf{x}'|}{p} \right)}{l^2} \right).$$

Оптимизация параметров ядра проводилась с использованием метода семейства Байесовских методов оптимизации, где замещающая модель (surrogate-модель) используется для приближенного поиска минимума целевой функции. Для построения замещающей модели используется Gradient Boosted Trees (GBT).

Обозначим целевую функцию, которую нужно минимизировать:

$$f(\mathbf{x}): \mathbb{R}^d \rightarrow \mathbb{R},$$

где $\mathbf{x}$ — точка из d -мерного пространства.

Для приближенного представления $f(\mathbf{x})$ используется регрессионная модель, построенная на основе градиентного бустинга над решающими деревьями. Эта замещающая модель $\hat{f}(\mathbf{x})$ аппроксимирует поведение $f(\mathbf{x})$ в известных точках. На каждой итерации метод обучает замещающую модель на наборах данных:

$$\mathcal{D} = \{(\mathbf{x}_i, f(\mathbf{x}_i))\}_{i=1}^n,$$

где $\mathbf{x}_i$ — точки из пространства поиска, а $f(\mathbf{x}_i)$ — значения целевой функции в этих точках. Целевая функция $f(x)$ представляет собой метрику качества модели:

$$f(x) = \text{MAE}(y_{\text{test}}, y_{\text{predicted}}).$$

Каждая итерация минимизации выглядит следующим образом:

$$x_{n+1} = \operatorname{argmax} \alpha(x),$$

где $\mathcal{X}$ — пространство поиска.

После получения x_{n+1} обновляется замещающая модель:

$f^n + 1(\mathbf{x}) \approx f(\mathbf{x})$ на основе нового набора данных $\mathcal{D}$.

Результаты исследования.

Полученная в результате исследования модель гауссовских процессов регрессии (GPR), продемонстрировала высокую прогностическую точность, достигнув средней ошибки MAE на уровне 0,11 при кросс-валидации. Коэффициент детерминации (R^2) составил 0,89, а корреляция между наблюдаемыми и прогнозируемыми значениями достигла 78%.

Результаты тестирования на независимой выборке показали значительную временную согласованность предсказаний и успешное воспроизведение сезонных колебаний индекса SPI. Модель проявила устойчивость к шумовым компонентам данных, способность учитывать долгосрочные тренды и адаптироваться к периодическим сезонным изменениям. Благодаря своей гибкости, данный подход может быть применен для различных климатических условий и временных масштабов.

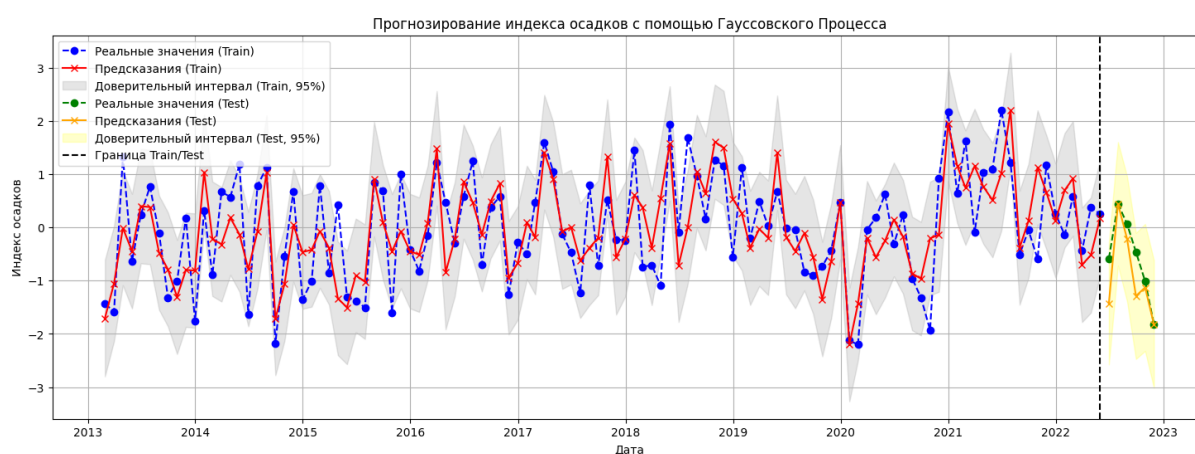


Рисунок 2 – Прогнозируемые и фактические значения SPI.



Рисунок 3 – Прогнозируемые и фактические значения SPI.

Заключение.

Проведенное исследование подтверждает высокую эффективность применения регрессии гауссовских процессов (Gaussian

Process Regression, GPR) с оптимизированными ядрами для прогнозирования индекса стандартизированных осадков (SPI), который является важным инструментом оценки метеорологических засух. Использование байесовской оптимизации для настройки гиперпараметров ядер гауссовских процессов позволило достичь значительного улучшения точности модели, демонстрируя её способность адаптироваться к сложным паттернам данных, таким как сезонные колебания, долгосрочные тренды и шумовые компоненты. Предложенная модель успешно воспроизводит временные зависимости и сохраняет согласованность прогнозов на тестовых выборках, что делает её эффективной в прогнозировании экстремальных значений SPI, характеризующих засушливые периоды.

Дальнейшие перспективы исследования включают расширение набора входных переменных за счет интеграции дополнительных климатических факторов, таких как температура, влажность почвы и скорость ветра, что может повысить детализацию и точность прогнозов. Также представляется перспективным внедрение гибридных моделей, сочетающих физически обоснованные уравнения с методами машинного обучения, для создания комплексных систем мониторинга засух.

Список использованных источников

1. Kazemi Garajeh M., Abdoli N., Seyedebrahimi E., Naboureh A., Kurdpour I., Bakhshi Lomer A.R., Sadeqi A., Mirzaei S. Impact of Long-Term Drought on Surface Water and Water Balance Variations in Iran: Insights from Highland and Lowland Regions // Remote Sensing. – 2024. – Vol. 16, No. 19. – P. 3636. – DOI: <https://doi.org/10.3390/rs16193636>.
2. Funk Chris, Peterson Pete, Landsfeld Martin, Pedreros Diego, Verdin James, Shukla Shraddhanand, Husak Gregory, Rowland James, Harrison Laura, Hoell Andrew, Michaelsen Joel. The climate hazards infrared precipitation with stations — a new environmental record for monitoring extremes // Scientific Data. – 2015. – Vol. 2. – P. 150066. – DOI: 10.1038/sdata.2015.66.3.

3. Mishra A.K., Singh V.P. A review of drought concepts // Journal of Hydrology. – 2010. – Vol. 391, No. 1–2. – P. 202–216. – DOI: <https://doi.org/10.1016/j.jhydrol.2010.07.012> .

References

1. Kazemi Garajeh M., Abdoli N., Seyedebrahimi E., Naboureh A., Kurdpour I., Bakhshi Lomer A.R., Sadeqi A., Mirzaei S. Impact of Long-Term Drought on Surface Water and Water Balance Variations in Iran: Insights from Highland and Lowland Regions // Remote Sensing. – 2024. – Vol. 16, No. 19. – P. 3636. – DOI: <https://doi.org/10.3390/rs16193636> .

2. Funk Chris, Peterson Pete, Landsfeld Martin, Pedreros Diego, Verdin James, Shukla Shraddhanand, Husak Gregory, Rowland James, Harrison Laura, Hoell Andrew, Michaelsen Joel. The climate hazards infrared precipitation with stations — a new environmental record for monitoring extremes // Scientific Data. – 2015. – Vol. 2. – P. 150066. – DOI: [10.1038/sdata.2015.66.3](https://doi.org/10.1038/sdata.2015.66.3).

3. Mishra A.K., Singh V.P. A review of drought concepts // Journal of Hydrology. – 2010. – Vol. 391, No. 1–2. – P. 202–216. – DOI: <https://doi.org/10.1016/j.jhydrol.2010.07.012> .