

РАСПОЗНАВАНИЕ СОЛНЕЧНЫХ ВСПЫШЕК В НАБЛЮДЕНИЯХ МИКРОВОЛНОВОГО ИЗЛУЧЕНИЯ С ПОМОЩЬЮ МАШИННОГО ОБУЧЕНИЯ

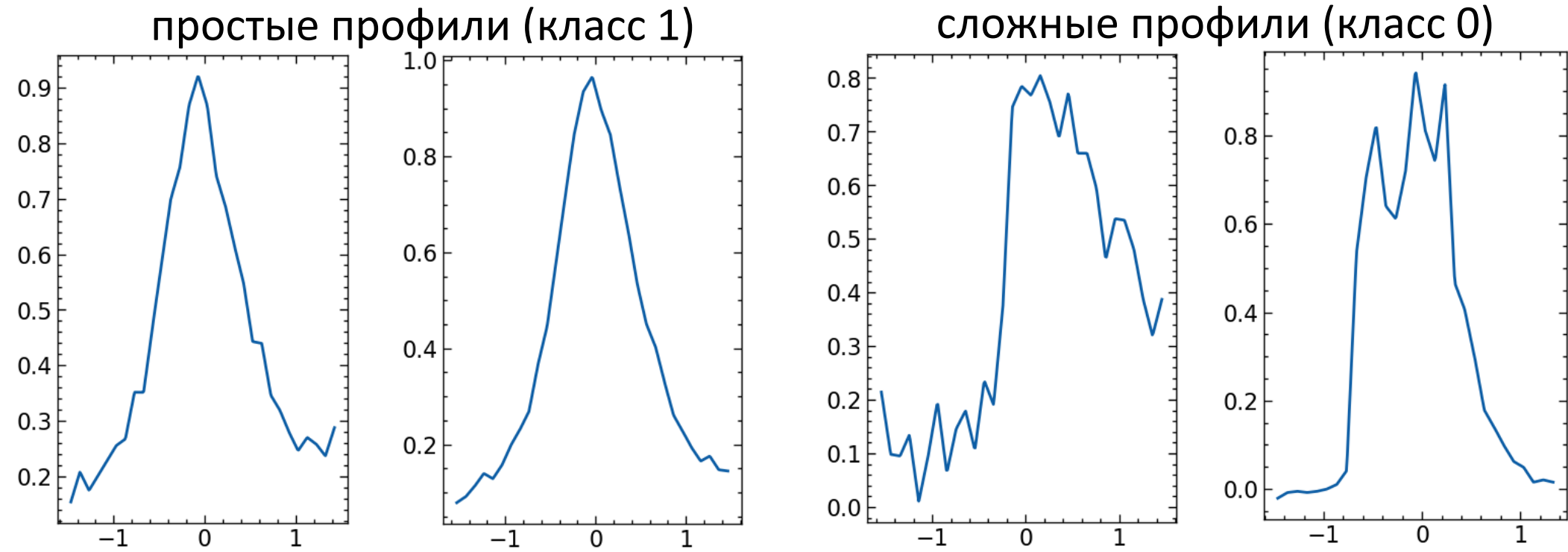


Шамсутдинова Ю. Н., Кашапова Л.К., Рожкова Д.В. ИСЗФ СО РАН, Иркутск yulia@iszf.irk.ru

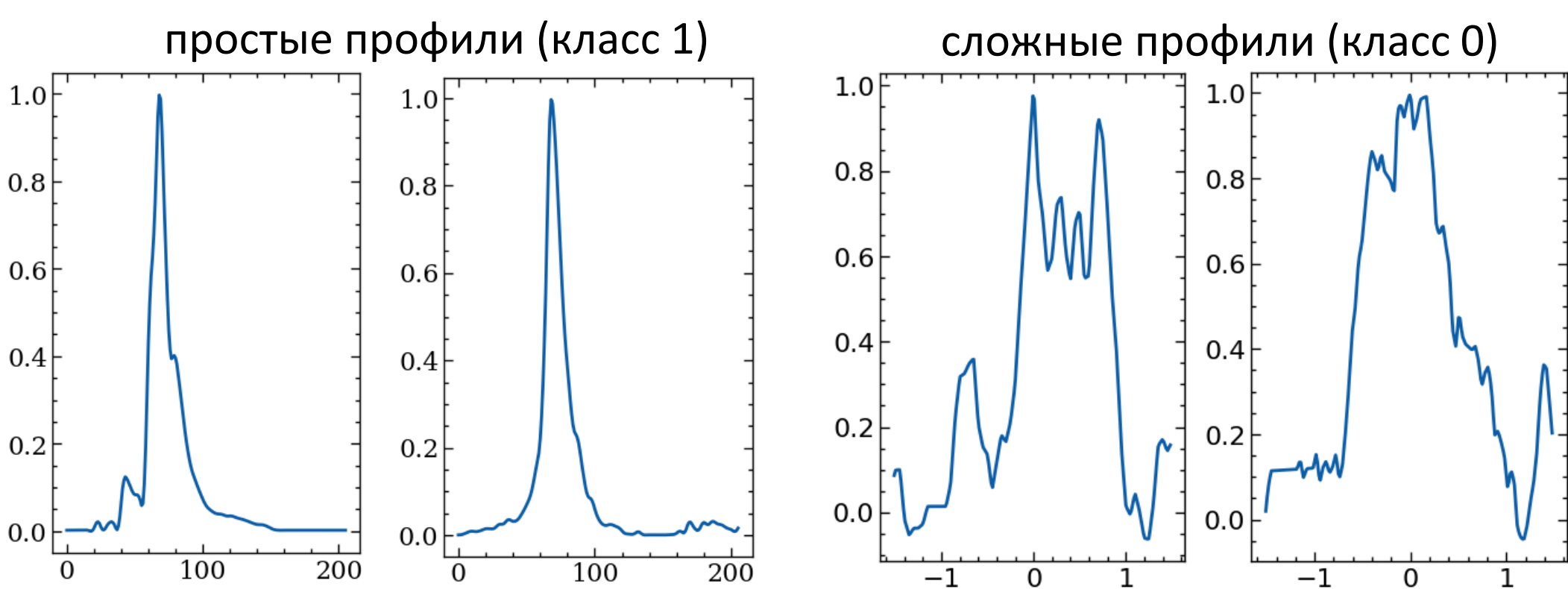
Мы представляем результаты тестирования методов машинного обучения для распознавания солнечных вспышек, наблюдавшихся Сибирским Радиогелиографом (СРГ) в диапазоне частот 3-24 ГГц. В качестве входных данных были использованы временные профили потоков на частоте 9.4 ГГц. На выходе определялась форма вспышки: «классический профиль» и «сложный профиль». Для временных профилей был применен Метод Опорных Векторов (SVM) с использованием модели, предварительно обученной и протестированной на временных профилях Nobeyama Spectropolarimeter (NoRP) на аналогичной частоте 9.4 ГГц, но для временных профилей так называемых корреляционных кривых. Показано, что результат кросс-корреляции значительно хуже, чем SVM метода. Были подобраны модели позволяющие наилучшим образом распознавать тип события. Обсуждаются и сравниваются модели, обученные на данных различных инструментов.

Набор данных

Примеры данных Nobeyama Spectropolarimeter (NoRP) *Потоки в S.F.U.*

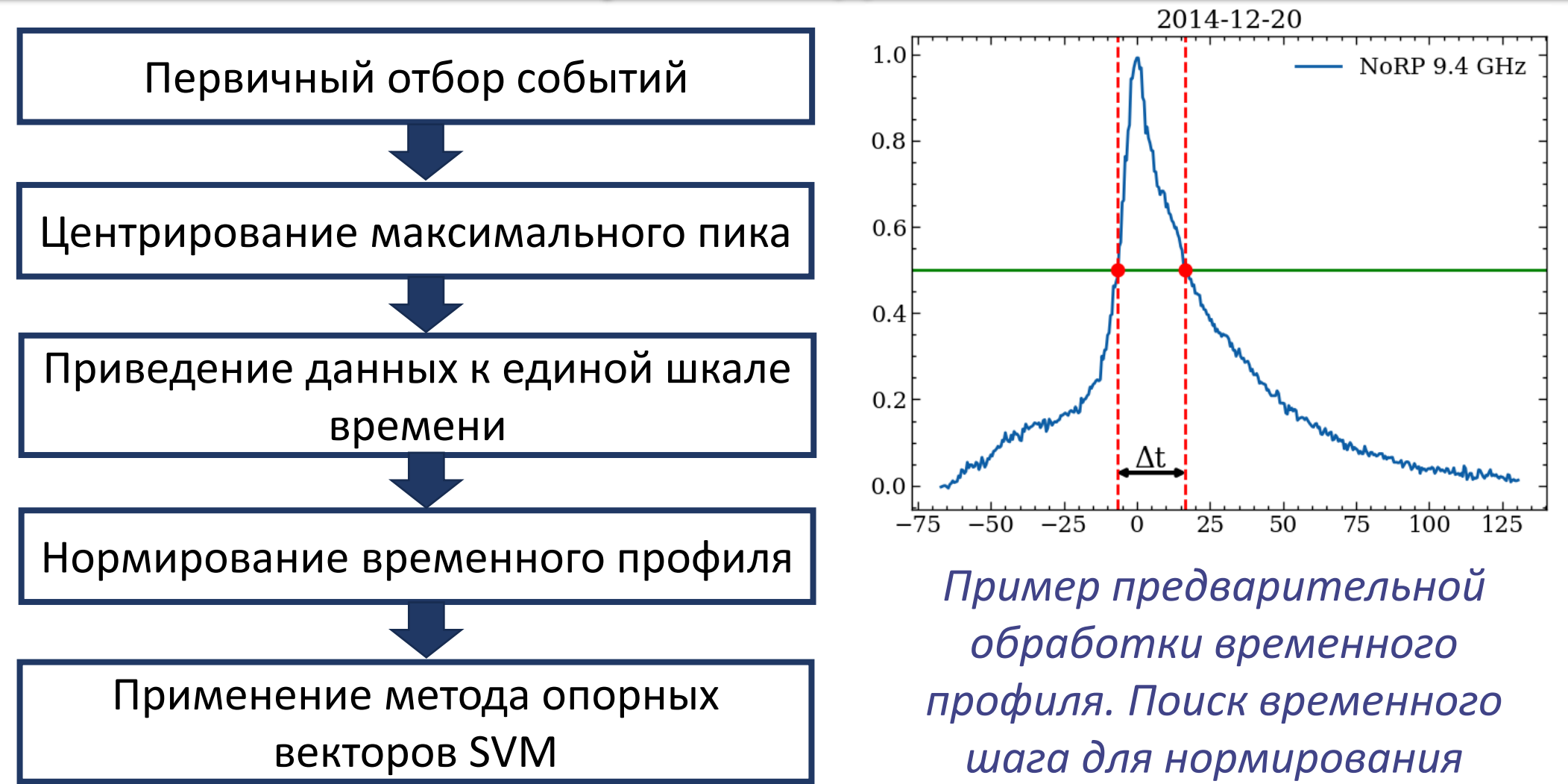


Примеры данных Сибирского Радиогелиографа (СРГ) *Корреляционные кривые*



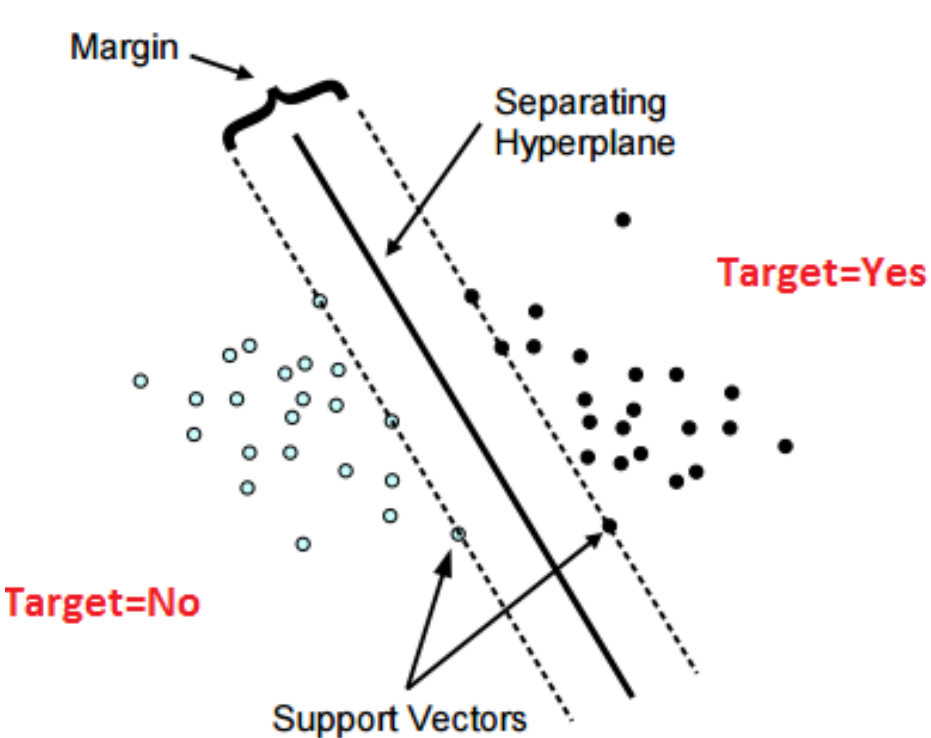
	класс 0	класс 1	всего
NoRP (2001, 2002, 2003, 2011, 2012, 2013, 2014)	155	80	235
СРГ (июнь-август 2023 года)	84	116	200

Обработка данных



Метод опорных векторов (Support Vector Machine, SVM)

Главная цель SVM как классификатора – найти уравнение разделяющей гиперплоскости, которая наилучшим образом разделяет набор данных на классы.



Гиперплоскость (гиперплоскость)
Плоскость принятия решения. Разделяет набор объектов на классы

Support vectors (опорные вектора)
Точки данных, находящиеся ближе всего к гиперплоскости. Вычисляют границы (Margin).

Ядро (kernel) преобразует пространство входных данных в требуемую форму. С помощью ядра входное пространство низкой размерности преобразуется в пространство более высокой размерности.

В работе была использована **радиальная базовая функция (RBF)**. RBF может отображать входное пространство в бесконечном пространстве, что удобно при большом количестве признаков.

Параметр C – параметр регуляризации, который уравнивает компромисс между максимизацией границ (margin) и минимизацией ошибки классификации.

Параметр gamma определяет, какое влияние имеет один опорный вектор на данные.

RBF: $\exp(-\gamma|x - x'|^2)$ [<https://scikit-learn.ru/1-4-support-vector-machines/>]

Оценка метрик

Матрица ошибок

		предсказание	
		0	1
ИСТИНА	0	TN	FP
	1	FN	TP

TP (True positive) – истинно-положительный
FP (False positive) – ложно-положительный
TN (True negative) – истинно-отрицательный
FN (False negative) – ложно-отрицательный

$$\text{Precision} = \frac{TP}{TP + FP} \quad \text{Recall} = \frac{TP}{TP + FN} \quad F_1 \text{ score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Для поиска оптимальных гиперпараметров модели используется метод *GridSearchCV*.

Обучающая выборка – 50%. Тестовая выборка – 50%

Обучение и тестирование на данных СРГ

Классификатор SVC (kernel='rbf', C=1, gamma=1)

	Precision	Recall	F1-score
сложный профиль	0.93	1.00	0.96
простой профиль	1.00	0.95	0.97

		предсказание	
		0	1
ИСТИНА	0	41	0
	1	3	56

Классификатор SVC (kernel='rbf', C=0.1, gamma=0.1)

	Precision	Recall	F1-score
сложный профиль	1.00	1.00	1.00
простой профиль	1.00	1.00	1.00

		предсказание	
		0	1
ИСТИНА	0	41	0
	1	0	59

Обучение и тестирование на данных NoRP

Классификатор SVC (kernel='rbf', C=10, gamma=1)

	Precision	Recall	F1-score
сложный профиль	0.86	0.94	0.90
простой профиль	0.80	0.61	0.69

		предсказание	
		0	1
ИСТИНА	0	80	5
	1	13	20

Обучение на данных NoRP. Тестирование на данных СРГ

Классификатор SVC (kernel='rbf', C=10, gamma=1)

	Precision	Recall	F1-score
сложный профиль	0.41	0.95	0.57
простой профиль	0.00	0.00	0.00

		предсказание	
		0	1
ИСТИНА	0	80	4
	1	116	0

Модель не распознала простые профили СРГ

Применение кросс-корреляции на данных NoRP

Медиана с коэффициентом 0.85

	Precision	Recall	F1-score
сложный профиль	0.77	0.77	0.77
простой профиль	0.82	0.82	0.82

		предсказание	
		0	1
ИСТИНА	0	33	10
	1	10	47

Среднее значение с коэффициентом 0.85

	Precision	Recall	F1-score
сложный профиль	0.73	0.81	0.77
простой профиль	0.85	0.77	0.81

		предсказание	
		0	1
ИСТИНА	0	33	10
	1	10	47

Резюме

- 1) Проведено тестирование метода опорных векторов для распознавания сложных и простых временных профилей микроволнового излучения на частоте 9.4 ГГц на основе данных каталога спектрополяриметров Нобейма и временных профилей корреляционных кривых по наблюдениям Сибирского Радиогелиографа.
- 2) На данных спектрополяриметра Нобейма показано преимущество метода опорных векторов перед кросс-корреляцией. Однако обучение и тестирование на данных различных инструментов показал плохой результат, что можно объяснить различными способами получения временных профилей.
- 3) Подобраны параметры, позволяющие добиваться наилучшего результата распознавания для временных профилей солнечных вспышек в микроволнах.

Исследование выполнено за счет гранта Российского научного фонда № 24-22-00315, <https://rscf.ru/project/24-22-00315/>