

# *Выбор алгоритма классификации*

*Для распределения пикселей снимка по классам можно использовать разные методы, причем выбор того или иного классифицирующего правила зависит от типа исходных данных и решаемой задачи. Все методы классификации можно разделить на параметрические и непараметрические. При использовании параметрических методов предполагается, что векторные данные  $X_s$ , полученные на этапе обучения для каждого класса и каждого спектрального диапазона, имеют нормальное распределение. При использовании непараметрических методов такого предположения не требуется. Полезно рассмотреть несколько различных алгоритмов классификации, среди которых чаще всего используют алгоритм наименьшего расстояния, алгоритм параллелепипеда и алгоритм максимального*

# Алгоритм классификации на основе определения наименьшего расстояния

- Этот алгоритм является одним из самых простых и наиболее часто используемых. Входными данными служат средние векторы, полученные на этапе обучения для каждого класса и каждого спектрального диапазона. Для каждого значения яркости пиксела  $(BV\#)$ , не отнесенного к классу, вычисляется эвклидово расстояние  $D$  от пиксела до среднего вектора класса  $\mu_{ck}$  в соответствии с формулой:

$$D = \sqrt{(BV_{\alpha} - \mu_{\alpha})^2 + (BV_{\beta} - \mu_{\beta})^2}.$$

- Пиксел приписывается тому классу, для которого это расстояние оказывается наименьшим. Интересно отметить, что точность

# Алгоритм параллелепипеда

- Этот алгоритм классификации основан на обычной булевой логике и статистических показателях обучающей выборки в  $n$  спектральных диапазонах. Сначала для каждого класса  $c$  и диапазона  $k$  вычисляется среднее значение яркости в обучающей выборке и  $\sigma_{ck}$ .

$$BV_{jk}$$

- После этого  $\mu_{ck} - \sigma_{ck} \leq BV_{jk} \leq \mu_{ck} + \sigma_{ck}$  пикселей снимка применяют следующее правило. Пиксел принадлежит классу  $c$  тогда и только тогда, когда его яркость ВУ/Д удовлетворяет следующему условию:

$$L_{ck} = \mu_{ck} - \sigma_{ck}$$

и

$$H_{ck} = \mu_{ck} + \sigma_{ck},$$

то условие (6.15) запишется в виде:

$$L_{ck} \leq BV_{jk} \leq H_{ck}.$$

- где  $c = 1, 2, 3,$

$m$  —

- Множество точек, подчиняющихся этому условию, образует параллелепипед в  $n$ -мерном пространстве спектральных признаков. Если значения яркости пиксела лежат внутри этого параллелепипеда, пиксел приписывается данному классу. В противном случае пиксел относят к категории неклассифицируемых.



# Алгоритм максимального правдоподобия

- В рассмотренных методах классификации не учитывались возможные вариации спектральных признаков и проблемы, возникающие из-за пересечения классов. Последние часто возникают в тех случаях, когда требуется классифицировать пиксели, близкие по своим спектральным характеристикам.
- В методе максимального правдоподобия пиксел приписывается к тому классу, который максимизирует функцию правдоподобия классификации. Данные из обучающей выборки используются для вычисления среднего вектора измерений  $M_c$  и ковариационной

$$p_c > p_i, i=1, 2...m,$$

где  $m$  — число классов, вероятность  $p_c$  определяется формулой

$$p_c = [-0,5 \log \{\det (V_c)\}] - 0,5[(X-M_c)^T(V_c)^{-1}(X-M_c)],$$

а  $p_i$  обозначает вероятность того, что данный класс существует.

щим

- Если нет дополнительных сведений о пространственных объектах, вероятность  $p$  для всех классов будет одинаковой. Если же известно, что вероятность существования одних классов больше, чем других, оператор может задать набор априорных значений вероятности для соответствующих спектральных признаков. Критерий принадлежности пиксела к классу  $c_i$  определяется следующим образом

$$p_i(a_j) > p_l(a_j), \quad i=1, 2, \dots, m,$$

где

$$p_i(a_j) = \log(a_j) - [-0,5 \log \{\det(V_i)\}] - 0,5[(X-M_i)^T(V_i)^{-1}(X-M_i)]. \quad (6.20)$$

Пиксел  $x$  принадлежит классу  $c$  в том и только в том случае, если

# *Методы неконтролируемой классификации*

- *При использовании методов неконтролируемой классификации от оператора практически не требуется вводить каких-либо входных данных. Все операции выполняются автоматически, при этом программа анализирует пространство спектральных параметров и на основании определенных критериев разделяет пиксели на классы, рассчитывая для каждого из них средние значения признаков и ковариационные матрицы. После того как все данные распределены по спектральным классам, оператор старается сопоставить их с известными информационными классами пространственных объектов. Эта задача зачастую оказывается весьма непростой. Некоторые спектральные классы (кластеры) могут не соответствовать каким-либо классам пространственных объектов или соответствовать сразу нескольким из них. Указанная неоднозначность может быть устранена только оператором на основе дополнительной информации о спектральных характеристиках изучаемых объектов. Алгоритмы кластерного анализа, применяемые для неконтролируемой классификации данных ДЗ, различаются, прежде всего, своей*



# Этап 1. Выделение кластеров

- $\Delta$  1.  $R$  — радиус кластера. ся следующие данные:  
2.  $C$  — пороговое расстояние, при котором два кластера объединяются.  
3.  $N$  — количество анализируемых пикселей.  
4.  $C_{\max}$  — максимально допустимое количество кластеров, которое может быть выделено в результате анализа.

- *Процесс анализа начинается с того, что в качестве центра первого кластера выбирается первый пиксел снимка. Затем в пространстве спектральных признаков вычисляется расстояние между этой точкой и вторым пикселом. Если это расстояние меньше или равно  $R$ , пиксел присваивается классу 1, а новый центр класса определяют, рассчитывая среднее значение по двум пикселям. Если же расстояние между пикселями больше  $R$ , тот новый пиксел становится центром класса 2. После этого все перечисленные действия повторяются для всех пикселей вплоть до  $N$ , при этом проверяется их принадлежность ко всем ранее выделенным классам.*

- *Следующий шаг состоит в проверке разделимости выделенных классов. На этом этапе рассматриваются расстояния между центрами классов, которые не должны быть меньше заданного значения  $C$ . Если расстояние между двумя центрами кластеров меньше  $C$ , они объединяются, а центр нового кластера вычисляется как взвешенное среднее центров исходных кластеров. После проверки разделимости процесс выделения кластеров*

## Этап 2. Классификация пикселов снимка

- *После выделения центров кластеров каждый пиксел снимка присваивается одному из выделенных классов в соответствии с критерием минимального расстояния. Затем наступает очередь оператора. Поскольку выделение кластеров и классификация пикселов выполнялись в пространстве спектральных признаков, оператор должен сопоставить спектральные классы информационным. Однако может оказаться, что для получения одного информационного класса необходимо объединить два или даже более спектральных классов. При этом размер результирующего кластера может оказаться слишком большим и потребуются разделить этот кластер на два. К сожалению, в алгоритме CLUSTER предусмотрена возможность объединения кластеров, но не их разделения.*
- *Обе эти возможности присутствуют в другом алгоритме, который называется ISODATA (Iterative self Organizing Data Analysis Technique A). Еще одной особенностью этого алгоритма является то, что в нем операции выделения*

# Оценка точности классификации

- *Классификацию данных дистанционного зондирования нельзя считать завершенной, пока не получена оценка ее точности. Оператор должен определить, насколько точно классы объектов на снимке соответствуют классам объектов на земной поверхности. В области обработки изображений термин **точность** означает меру согласованности стандартной информации в некоторой пространственной точке с информацией в соответствующей точке классифицированного снимка. Обычно проверка точности классификации основана на сравнении двух карт — той, которая получена в результате анализа данных ДЗ, и второй, контрольной карты. Для построения контрольной карты обычно используют несколько источников данных, поэтому она считается более точной. Обычно на контрольной карте нанесено несколько отдельных участков земной поверхности, для каждого из которых используется отдельное условное обозначение.*
- *Простейшим методом определения точности классификации является сравнение тех участков двух карт, которые относятся к выделенным классам объектов. Результатом является отчет, в котором указывается общая точность классификации для всей карты или точность классификации отдельных ее участков.*
- *Оценку общей точности классификации формулируют в единицах площади объектов каждого класса, не обращая внимания на несовпадение объектов в отдельных точках сравниваемых карт. При оценке точности классификации определенного участка карты расхождение между классами объектов анализируется более подробно.*
- *При использовании метода контролируемой классификации, наиболее простой способ оценки ее*



# Матрица ошибок

Стандартной формой представления оценки точности классификации для определенного местоположения является **матрица ошибок**, которая характеризует не только погрешность классификации для каждого класса, но и ошибки, связанные с неверной классификацией. Матрица ошибок состоит из  $y$  столбцов и  $l$  строк, где  $y$  — количество классов объектов на контрольной карте. Строки матрицы — это **истинные классы**, представленные на контрольной карте, а столбцы — классы, выделенные на анализируемой карте. Схематичное представление матрицы ошибок показано на рис. 6.12, а пример действительной матрицы приведен в табл. 6.7. В последнем столбце представлено количество точек в каждом информационном классе контрольной карты, а в последней строке матрицы — количество точек соответствующих классов на анализируемой карте. Диагональные элементы матрицы — это количество точек, принадлежащих одинаковым классам на обеих картах. Значение в нижнем правом углу таблицы представляет собой количество точек, правильно классифицированных на обеих картах, за вычетом пропусков классификации с

$$p = p' - \left[ 1,645 \sqrt{\frac{(p')(q)}{n} + \frac{50}{n}} \right]. \quad (6.21)$$

Здесь  $p$  — общая точность классификации, соответствующая доверительному уровню 95%,  $p'$  — общая точность классификации,  $q = 100 - p'$ ,  $n$  — размер выборки.

В табл. 6.7 для оценки точности классификации использовали данные, принадлежащие пяти



Класс	Классифицируемый снимок						Всего
	Город	Поселы	Пастбище	Водоем	Лес	Пустошь	
Город							
Поселы							
Пастбище							
Водоем							
Лес							
Пустошь							
Всего							
Граница столбцов							

Строки характеризуют пропуски классификации, столбцы – ошибки, связанные с ложной классификацией.

Правильно определенные пиксели.

Общая сумма правильно определенных пикселей.

Рис. 6.12. Схематическое представление матрицы ошибок [11]

# Характеристики снимков Landsat 5 TM



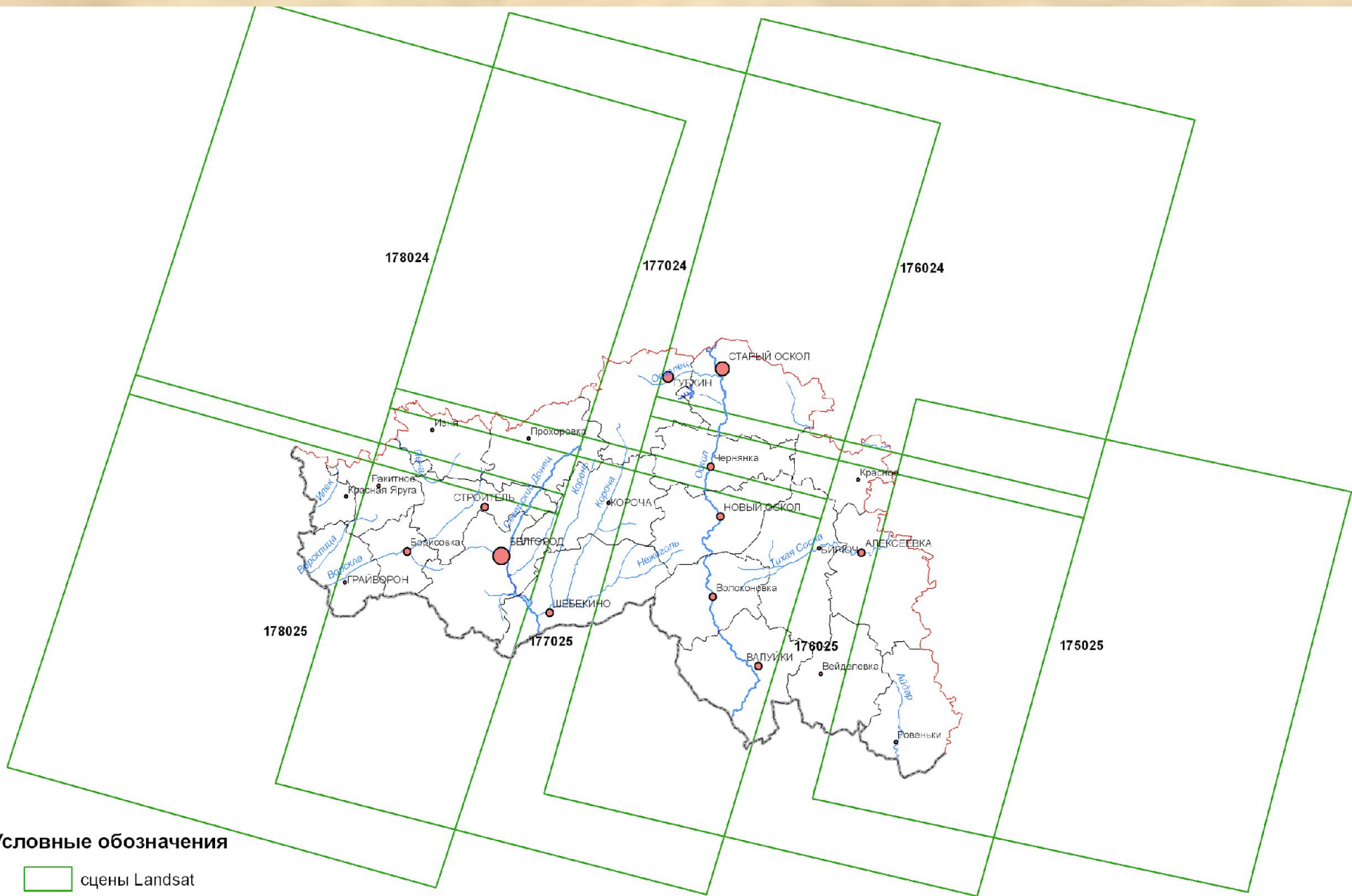
Запущен весной 1984 .  
Прекратил работу  
осенью 2011.

Формат данных	GeoTIFF с метаданными
Пространственное разрешение	30 м/пиксель
Размер кадра	185×185 км
Периодичность съемки	16 дней
Количество каналов	7
Области спектра	Видимая (1-3), ближняя инфракрасная (4), средняя инфракрасная (5,7), тепловая (6).
Радиометрическое разрешение	8 бит

Табл. 6.7. Пример матрицы ошибок классификации [30]

Фактический класс	Расчетный класс						Всего
	Вереск	Водоем	Лес 1	Лес 2	Обна- женная почва	Паст- бище	
Вереск	826	0	0	5	27	0	858
Водоем	0	878	0	0	0	0	878
Лес 1	0	0	720	183	0	7	910
Лес 2	33	0	21	878	2	0	934
Открытая почва	61	0	0	0	560	0	621
Пастбище	0	0	0	1	0	219	220
Всего	920	878	741	1067	589	226	4081

# Покрытие снимками Landsat 5 TM Белгородской области





## Характеристики данных Landsat 8



Запущен в феврале 2013 .

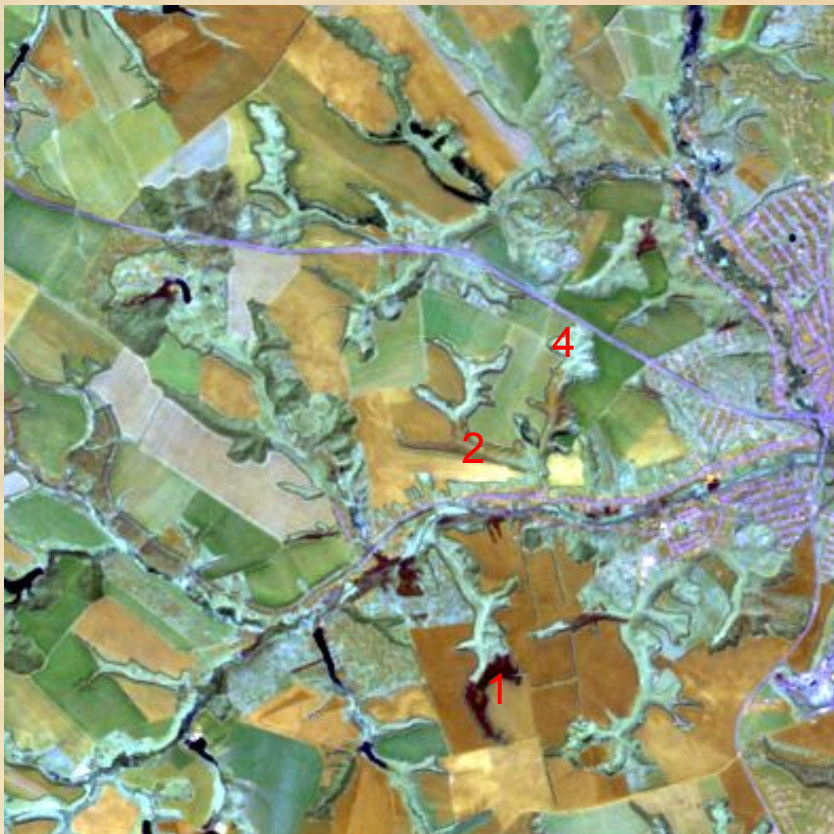
Доступен для скачивания с конца мая 2013

Формат данных	GeoTIFF с метаданными
Пространственное разрешение	30 м/пиксель (15 м/пиксель в панхроматическом канале)
Размер кадра	185×185 км
Периодичность съемки	16 дней
Количество каналов	11
Области спектра	Видимая (1-4), ближняя инфракрасная (5), средняя инфракрасная (6,7,9), панхроматический диапазон(8), тепловой (10,11)
Радиометрическое разрешение	12 бит

# Дешифрование гарей разного возраста

22 апреля 2011

6 мая 2007



1 – свежие, 2 – среднего возраста, 3 – старые, 4 – невыгоревшие участки

Свежие гари имеют темно-бордовый (или вишневый) или темно-коричневый цвет. По мере зарастания травой он сменяется на бурый и палевый, а затем – на желтовато-зеленый. Постепенно цвет гарей и нетронутых огнем участков выравнивается, и к лету они уже неразличимы.





4:3:2



7:4:2



5:4:3



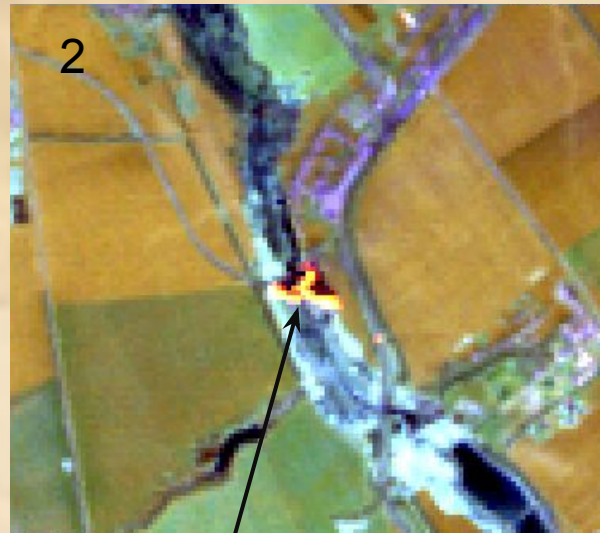
3:2:1



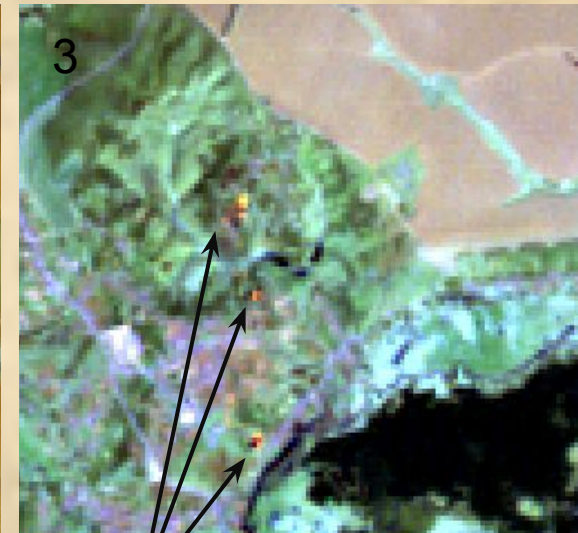
# Дешифрование пламени



пламя



пламя



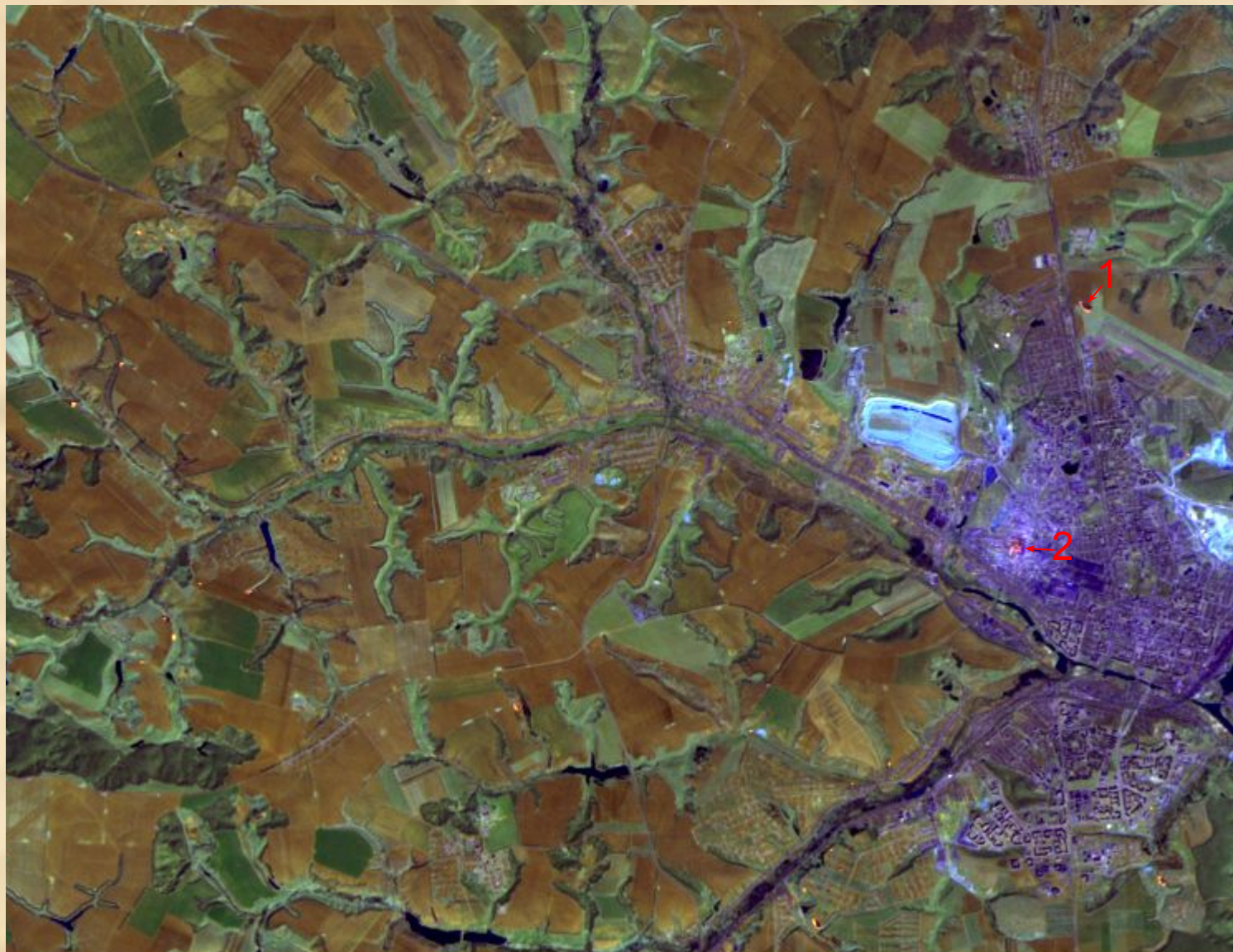
пламя

1,2 – крупные очаги пламени на склонах балок; 3 – мелкие очаги пламени в населенном пункте

В комбинации каналов Landsat 7:5:3 пламя выглядит как небольшой по площади объект, площадью от нескольких до нескольких десятков пикселей. Для него характерно сочетание ярко-желтого, оранжевого и красного цвета.



# Дешифрование пламени



1 - пожар у  
взлетной  
полосы  
белгородского  
аэропорта (!!!)

2 - тепловой  
выброс  
цементного  
завода (не  
пожар)

16 апреля 1994 года. Видно 26 очагов пламени.

# Дешифрование дыма



3:2:1

Сгоревшая трава

Дымовой шлейф



7:5:3

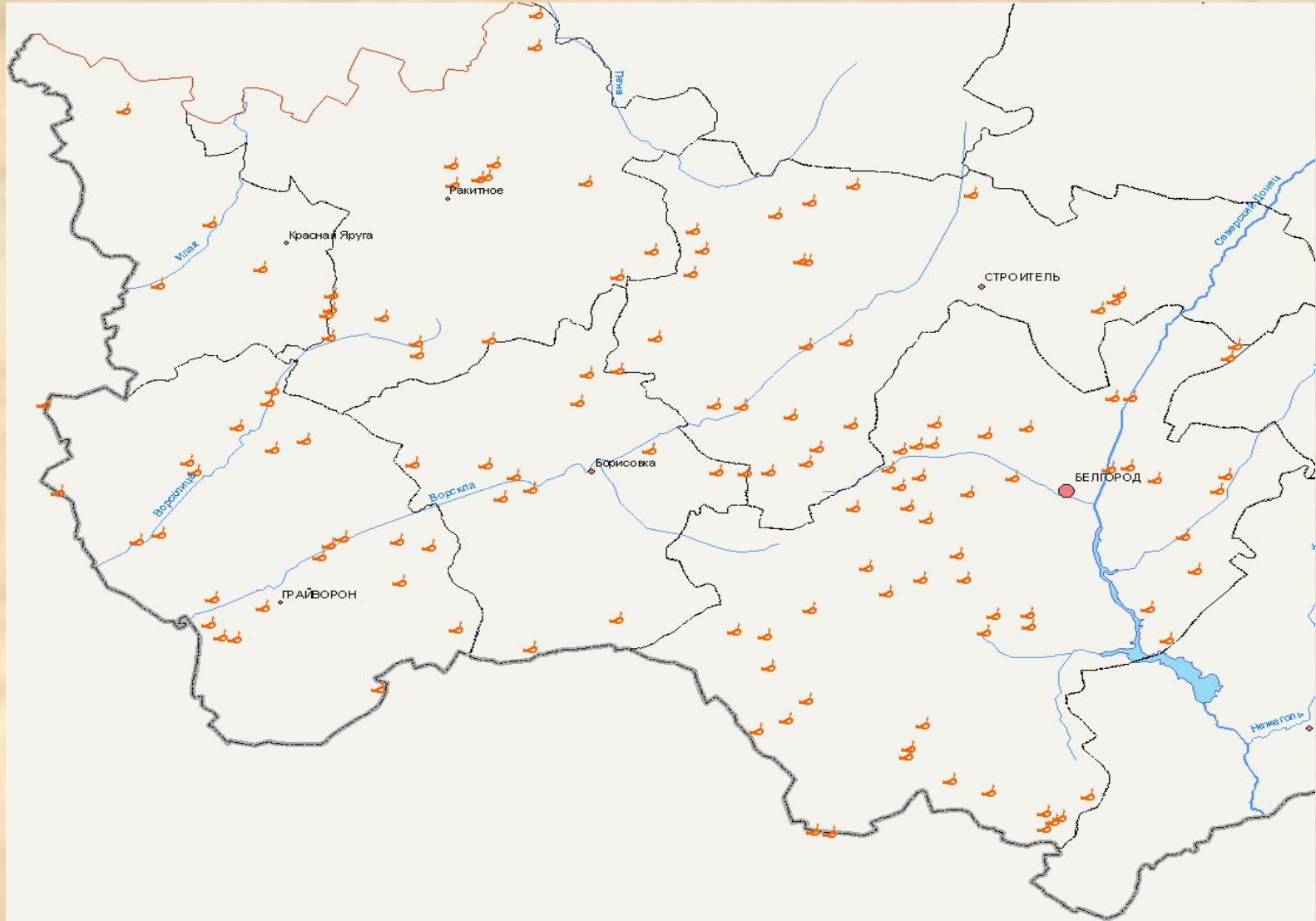
Пламя

Дым лучше всего виден в комбинации каналов Landsat 3:2:1 (каналы видимой части спектра). Он представляет собой белую пелену разной степени прозрачности, под которой просматриваются объекты местности.

В инфракрасной части спектра (каналы Landsat 4, 5 и 7 дым, в отличие от облаков, становится «невидимым»).



# Травяные палы на западе Белгородской области



22 апреля 2011 года.

Травяные палы тяготеют к окраинам крупных населенных пунктов и автомагистралям



## Палы на территории ООПТ



6 мая 2007

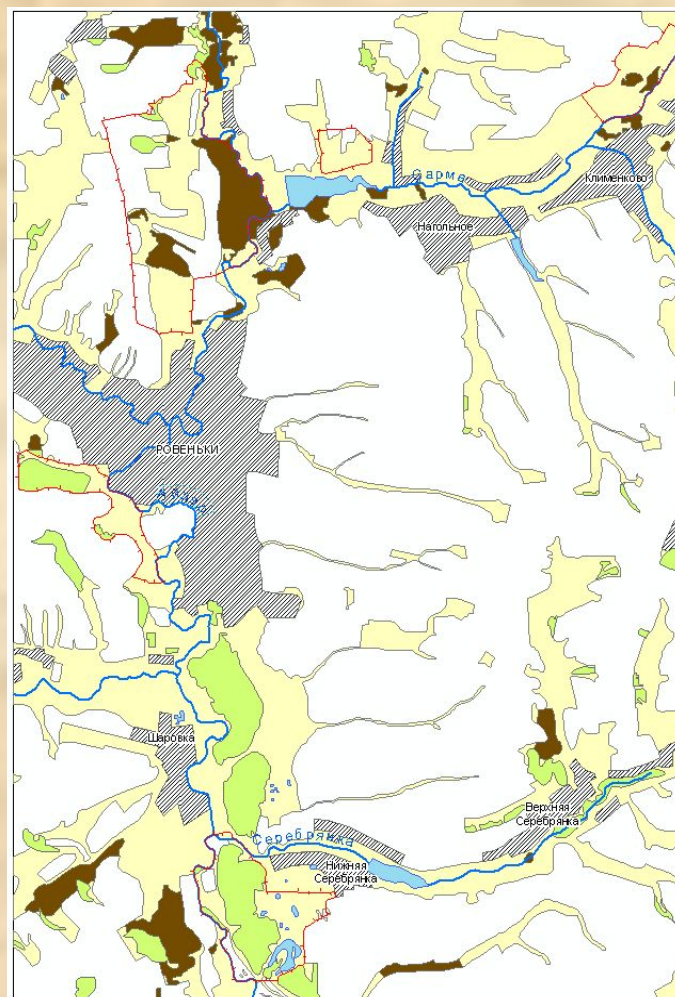
На снимке видна выгоревшая пойма в охранной зоне Леса на Ворскле, выгоревшие участки Острасьевых Яров (справа) и выгоревший заказник Борки (слева)



# Весенние травяные палы в региональном природном парке «Ровеньский» и его окрестностях в 2011 году



Снимок Landsat 5 TM, 3 мая 2011  
Комбинация каналов 7:5:3



## Условные обозначения

- граница природного парка
- травянистая растительность
- выгоревшие участки
- населенные пункты
- леса

Результат картографирования  
выгоревших территорий

**СПАСИБО ЗА ВНИМАНИЕ!**