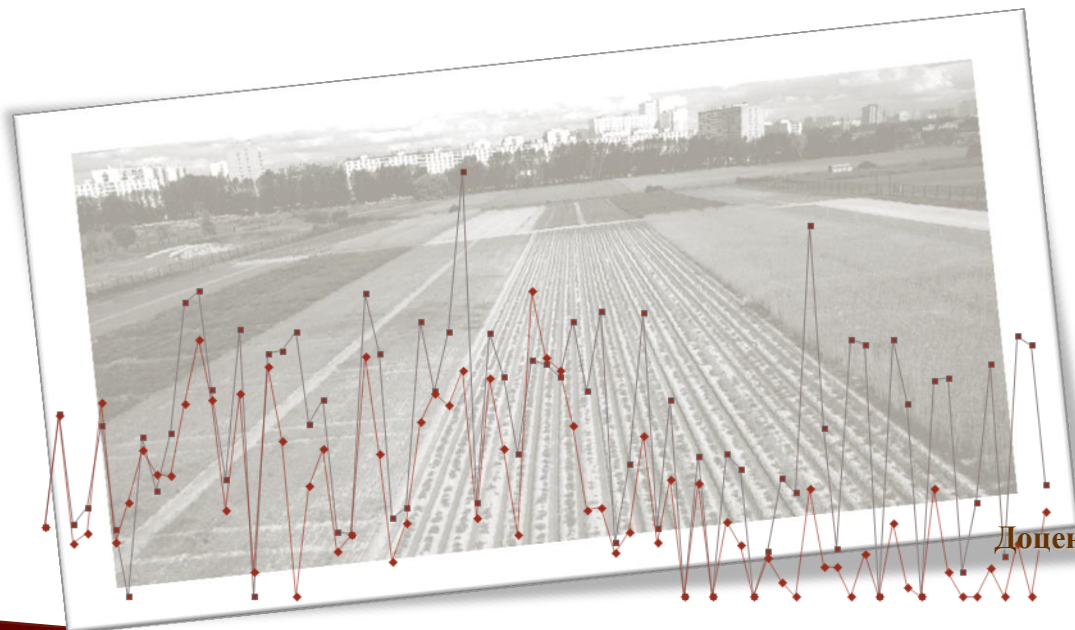


Использование Deductor для обработки данных 100-летнего полевого опыта РГАУ-МСХА

Deductor



Аналитическая
Платформа



Доценты кафедры экономической кибернетики

Карпузова В.И., Чернышева К.В.

Доцент кафедры земледелия и методики опытного дела

Усманов Р.Р.

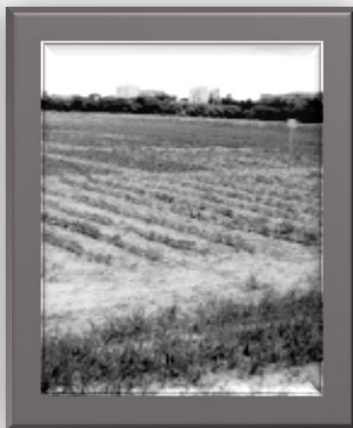
Москва, 2016

Патриарх русской агрономической науки профессор Стебут И.А. отмечал:

«Изучайте природу, вас окружающую, изучайте почву, от которой вы ожидаете себе урожая и надежной прибыли».

Длительный полевой опыт

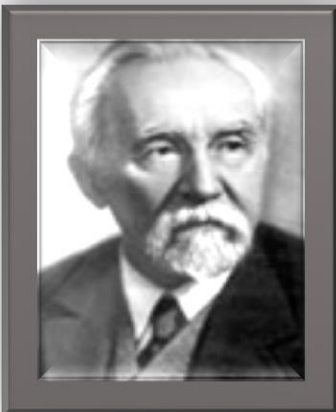
Историческая справка



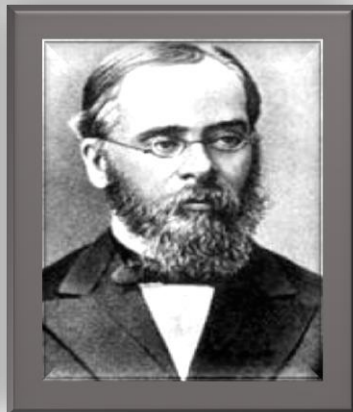
В 1867 г. профессором Стебутом И.А. основана полевая опытная станция.

В 1912 г. по инициативе Прянишникова Д.Н. заложен многофакторный опыт.

Профессором А.Г. Дояренко разработаны схема, методика опыта и программа исследований.



ПРЯНИШНИКОВ Д.Н.



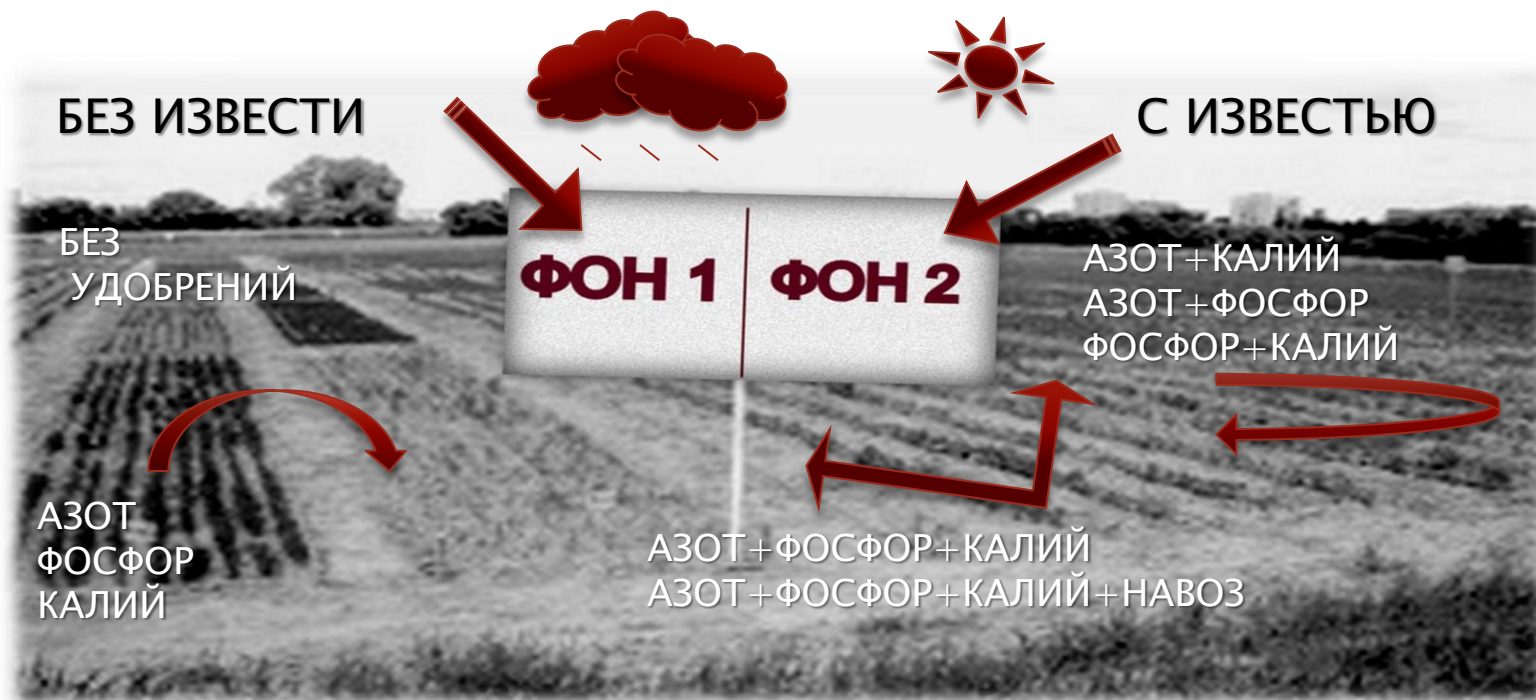
СТЕБУТ И.А.



ДОЯРЕНКО А.Г.

Длительный полевой опыт

Стационарный полевой опыт РГАУ–МСХА является наглядным пособием, демонстрирующим роль севооборота, удобрений и известкования в повышении урожайности.



Первый фактор

**Бессменные культуры с 1912 года
(рожь, картофель, ячмень, клевер,
лен и поле «вечного» пара);**

**шестипольный севооборот с
ротацией тех же культур во
времени**

**(пар – рожь – картофель – ячмень –
клевер – лен).**

Второй фактор

Удобрения

(9 вариантов по схеме
Жоржа Вилля)

Рассматривались следующие варианты:

O – без удобрений,

N – азот,

P – фосфор,

K – калий,

NP – азот+фосфор,

NK – азот+калий,

PK – фосфор+калий,

NPK – азот+фосфор+калий,

NPK + H – азот+фосфор+калий+навоз.

Третий фактор

Известкование почвы с 1950 г.

Опыт проводился по двум фонам:

фон 1 – без извести,

фон 2 – по извести.

Исходные данные для анализа

1. Урожайность овса (до 1972 г.) и ячменя (с 1973 по 2011 гг.) с учетом:

- 9 вариантов удобрений;

- известкования почвы

(фон 1–без извести, фон 2– по извести).

2. Метеоусловия (t и осадки)

Исходные данные

Год	Фон	О	N	P	K	NP	NK	PK	NPK	NPK+.	Навоз, т/га	N, кг д.в. на 1 га	P205	K20	t Апрель	t Май	t Июнь	t Июль	t Средняя за год	o Апрель	o Май	o Июнь	o Июль	o Сумма за год
1950	1	21,8	31,5	27,5	19,5	28,9	19,5	28,8	40,9	38,1	20	75	60	90	9,1	11,9	15	15,9	4,1	22	83	102	60	706
1951	1	7,6	7	12,4	11,4	3,8	2,4	9,2	9	10,6	20	75	60	90	8,4	9,6	17,5	18,4	4,4	25	81	5	109	535
1952	1	21	23	24	24	26,2	13,2	29,2	25,4	23,6	20	75	60	90	5,2	10,2	17,1	17,6	4,6	17	52	57	155	896
1953	1	14,2	14,2	17,6	16,4	29,4	19,6	25,3	21,2	26,1	20	75	60	90	7,1	11,4	19,1	19	4,3	30	72	55	144	770
1954	1	6	5,1	6,3	6,8	6,4	7,8	11,9	13,8	11,4	20	75	60	90	3	12,7	18,9	20,9	4,5	30	46	69	63	557
1955	1	29,4	34,2	34,3	24,2	32,5	18,8	33,1	22,8	21,8	10	50	75	60	1,4	10,3	15	17,9	4,1	99	62	67	30	626
1956	1	23	10,5	26,2	24,4	17,5	20,8	27,5	22,4	25,7	10	50	75	60	4	10,7	20,8	15,2	3	52	100	88	60	698
1957	1	21,8	19,7	15,6	16,5	27,3	16,2	19,8	21,2	29,7	10	50	75	60	6,7	14,5	15,4	18,7	5,9	17	44	102	46	606
1958	1	16,5	20,5	17,9	16	20,4	26,4	16	26,5	30,2	10	50	75	60	4,2	13,2	15	18,4	4,4	48	53	131	39	712
1959	1	8,3	15,2	11,3	8,6	12,1	14,9	19,2	19,8	22,8	10	50	75	60	6,7	11,5	17	20,7	4,7	32	78	87	48	636
1960	1	8,6	11,6	10,4	7,7	7,2	7,3	15,3	27,5	13,9	10	50	75	60	5,3	11,8	18,6	21,5	5	12	78	44	48	700
1961	1	4,2	6	4,4	5,6	4,8	9,6	6,2	5,2	10,2	10	50	75	60	4,2	12	19,3	19,5	5,8	54	40	32	79	593
1962	1	6,2	5,9	7,7	6,3	9,6	10	9,6	12,7	15,2	10	50	75	60	7,6	13,2	13,4	16,4	5	35	112	74	110	767
1963	1	15,5	9,6	9,7	10,3	16,1	8,4	14,8	27,3	25,8	10	50	75	60	4	17,2	13,5	19,1	3,9	27	15	106	92	574
1964	1	5,2	8,2	6,6	7,1	6,9	7,2	8,9	6,6	9,3	10	50	75	60	4,3	11,4	19	20,2	5	28	94	26	26	461
1965	1	11,3	11,9	10	8,1	10,2	14,1	17,1	19	11	10	50	75	60	2,6	9,6	16	16,6	3,9	40	63	93	195	823
1966	1	0	0	0	0	0	0	0	0	0	10	50	75	60	8,8	15,4	16,7	19,2	5,2	40	48	50	101	810
1967	1	10,9	10,3	12,2	13	11,4	14,4	20	13,5	17	10	50	75	60	6,6	17	16,6	18,2	5,4	57	27	44	38	586
1968	1	0	0	0	0	0	0	0	0	0	10	50	75	60	5,9	12,6	18,7	15,9	4,3	44	47	21	158	678
1969	1	7,2	8	10,9	8,2	6,3	5,5	8,9	13,7	10,8	10	50	75	60	5,9	11,1	14,8	18,1	3,2	53	56	92	70	627
1970	1	4,9	4,3	4,2	3,4	5,3	6,5	8,3	12,3	11,7	10	50	75	60	5,8	12,7	16	19,7	4,8	104	18	85	39	753
1971	1	0	0	0	0	0	0	0	0	0	10	50	75	60	3,7	13	16,8	17,6	4,8	48	30	58	79	682
1972	1	3,7	1,5	2,9	3,7	2,1	2,2	4,9	4,3	7,2	10	50	75	60	6	12,7	19,4	23	6,1	53	80	63	16	529
1973	1	1,4	0,5	0,5	0,8	0,3	0,5	17,1	11,4	20,4	20	100	150	120	7,8	13,3	18,6	18,3	5,2	69	74	21	109	823
1975	1	0	0	5,6	0	0	0	13,9	10	19,4	20	100	150	120	10,1	16,1	17,9	18,9	6,7	36	24	66	78	559
1977	1	10,4	20,4	18,9	5,9	10,8	13,1	26,2	35,8	34,2	20	100	150	120	7,1	14,3	16,9	18,8	5	49	75	92	77	786
1979	1	2,9	4,1	6,2	7,9	5	4,3	34,2	16,2	28,8	20	100	150	120	3,3	17,3	17,5	16,7	5,1	12	19	40	161	708
1981	1	2,8	0	6	3,3	0	0	20,5	4,6	18,1	20	100	150	120	3,3	14,1	19,9	21,8	6,5	27	18	67	131	856
1983	1	0	5,2	0	0	0	0	0	24,8	28	20	100	150	120	9,3	15,6	14,7	18	6,3	89	9	100	65	725
1985	1	4,1	4,1	9,1	6,4	3,6	8,9	25,5	24,3	34,3	20	100	150	120	5,5	13,2	14,8	16,5	4,1	45	60	111	92	772
1987	1	0	0	0	0	0	0	14,5	0	32	20	100	150	120	2,9	13	17,9	16,9	3,3	23	97	113	53	587
1989	1	7,1	10,2	14,6	12,4	9	10,6	20,5	24,8	37,2	20	100	150	120	7,9	13,7	20,3	19,4	7,3	47	33	71	37	738
1991	1	0,9	6,2	2,9	1,1	1,1	3,7	12,1	18,6	37,1	20	100	150	120	7,3	13,6	19,1	18,4	6,3	51	48	122	101	851
1993	1	0	0	0	0	0	0	0	0	27	20	100	150	120	5,8	14,8	14,1	17,5	4,7	35	16	140	156	869
1995	1	10,4	9,7	8,2	10,2	9,5	13,7	11,1	20,8	26,2	20	100	150	120	9,3	14,8	19,8	17,7	6,7	48	27	68	56	565
1997	1	2,3	0	0	0	0	14	20,4	21,1	23,8	20	100	150	120	4,9	11,5	18,2	19,2	5,3	22	38	108	9	670
1999	1	0	0	0	0	0	0	0	2,4	2,6	20	100	150	120	10,2	9	21,9	22,2	6,8	28	33	7	67	575
2001	1	0	0	0	0	2,2	3,8	1,1	9,1	14	20	100	150	120	11,3	11,3	16,4	23,2	6,1	26	122	64	64	715
2003	1	2,7	6,9	2,9	2,7	5,1	5,6	20,2	22,4	32,2	20	100	150	120	4,9	15,8	12,9	20,8	5,9	45	42	67	91	710
2005	1	0	0	0	0	0	0	0	3,8	4,2	20	100	150	120	7,4	15	16,5	19,6	6,4	48	96	82	110	678
2007	1	5,1	10,7	8,9	10,2	10,2	10,4	16,8	25,1	25,1	20	100	150	120	6	16,1	17,7	19,1	7,3	19	31	24	64	586
2009	1	0	0	0	0	0	0	20,2	24,2	34,2	20	100	150	120	5,4	13,8	17,4	19	6,3	24	58	55	88	733
2011	1	8,2	7,2	8,9	6,8	9	8,5	8,9	10,8	11,4	20	100	150	120	6,7	14,7	19,4	23,8	6,8	43	29	38	75	590

Фон 1 – без извести Фон 2 – с известью

Аналитическая платформа Deductor Studio применяется в университете с 2006 г.

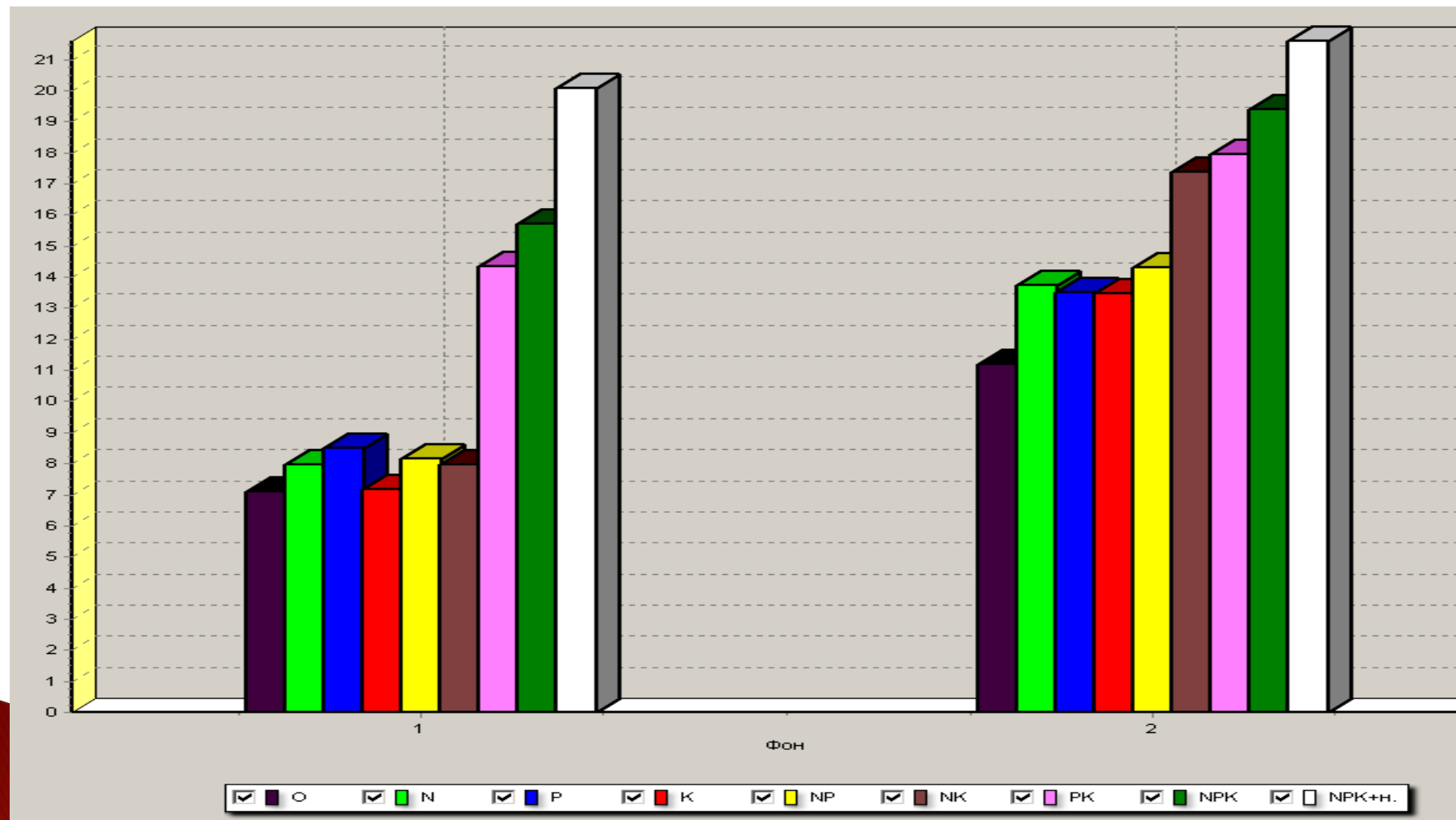
За истекший период прошли обучение более 1000 студентов, а также магистры, аспиранты и преподаватели сельскохозяйственных ВУЗов.

Визуализатор «Статистика»

Таблица X Статистика X Диаграмма X Куб X									
	Метка столбца	Статистика: Кол-во значений = 86							
		Минимум	Максимум	Среднее	Стандартное откл.	Σ Сумма	Σ ² Сумма квадратов	s Кол-во уникальных значений	0 Кол-во пустых значений
1	ab Год	4	4	4	0	344	1376	43	0
2	ab Фон	1	1	1	0	86	86	2	0
3	9.0 O	0,0	30,3	9,1	8,1	786,6	12741,8		0
4	9.0 N	0,0	43,6	10,9	9,7	933,6	18071,9		0
5	9.0 P	0,0	36,4	11,0	9,3	945,8	17706,6		0
6	9.0 K	0,0	37,2	10,3	8,8	887,9	15815,8		0
7	9.0 NP	0,0	41,8	11,2	10,2	966,0	19734,0		0
8	9.0 NK	0,0	36,1	12,7	10,4	1089,4	23045,6		0
9	9.0 PK	0,0	42,7	16,2	10,3	1389,6	31433,5		0
10	9.0 NPK	0,0	44,5	17,5	11,3	1508,6	37343,2		0
11	9.0 NPK+н.	0,0	54,5	20,8	12,0	1790,6	49449,3		0
12	9.0 Навоз, т/га	10,0	20,0	15,8	5,0	1360,0	23600,0		0
13	9.0 N, кг д.в. на 1 га	50,0	100,0	76,2	23,6	6550,0	546250,0		0
14	9.0 P205	60,0	150,0	108,1	39,5	9300,0	1138500,0		0
15	9.0 K20	60,0	120,0	91,4	28,3	7860,0	786600,0		0
16	9.0 t Апрель	1,4	11,3	6,1	2,3	525,8	3648,9		0
17	9.0 t Май	9,0	17,3	13,2	2,1	1133,4	15317,7		0
18	9.0 t Июнь	12,9	21,9	17,2	2,1	1483,0	25955,9		0
19	9.0 t Июль	15,2	23,8	18,9	2,0	1627,0	31117,8		0
20	9.0 t Средняя за год	3,0	7,3	5,2	1,1	449,0	2449,0		0
21	9.0 o Апрель	12,0	104,0	40,8	20,6	3506,0	178890,0		0
22	9.0 o Май	9,0	122,0	54,1	28,6	4656,0	321560,0		0
23	9.0 o Июнь	5,0	140,0	69,9	32,9	6014,0	512766,0		0
24	9.0 o Июль	9,0	195,0	81,1	42,5	6978,0	720034,0		0
25	9.0 o Сумма за год	461,0	896,0	684,2	104,7	58842,0	41192526,0		0

 Кросс-таблица (Строки: Фон; Факты: O, N, P, K, NP, NK, PK, NPK, NPK+H.)

Фон	O	N	P	K	NP	NK	PK	NPK	NPK+H.
1	7,1	8,0	8,5	7,2	8,1	8,0	14,4	15,7	20,1
2	11,2	13,7	13,5	13,5	14,3	17,4	18,0	19,4	21,6



На кислых дерново-
подзолистых почвах
известкование почвы
является мощным фактором
повышения урожайности.

Фон	O	N	P	K	NP	NK	PK	NPK	NPK+H.
1	7,1	8,0	8,5	7,2	8,1	8,0	14,4	15,7	20,1
2	11,2	13,7	13,5	13,5	14,3	17,4	18,0	19,4	21,6

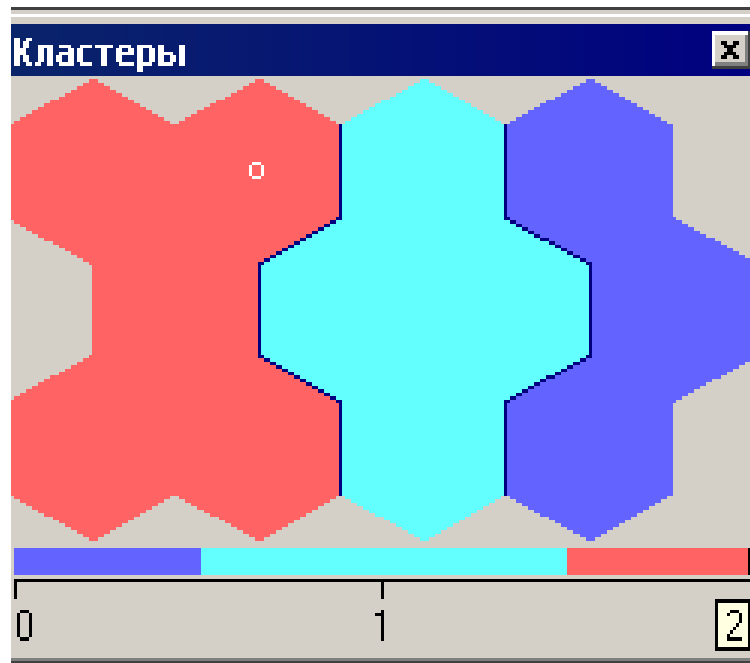
Кластеризация по годам (фон 1)



Номер кластера	Год
0	1959
0	1960
0	1963
0	1965
0	1967
0	1977
0	1979
0	1985
0	1989
0	1995
0	2007
1	1950
1	1952
1	1953
1	1955
1	1956
1	1957
1	1958
2	1951
2	1954
2	1961
2	1962
2	1964
2	1969
2	1970
2	1972
2	1973
2	1975
2	1981
2	1983
2	1987
2	1991
2	1993
2	1997
2	1999
2	2001
2	2003
2	2005
2	2009

Номер кластера	О Δ	N	P	K	NP	NK	PK	NPK	NPK+н.
2	21,1	21,9	23,3	20,1	26,0	19,2	25,7	25,8	27,9
1	8,6	10,7	10,9	9,2	9,6	10,9	20,1	23,1	25,1
0	2,9	3,3	3,7	3,1	2,9	4,0	9,9	10,9	17,8
Итого:	7,6	8,6	9,1	7,7	8,8	8,5	15,4	16,9	21,6

Кластеризация по годам (фон 2)



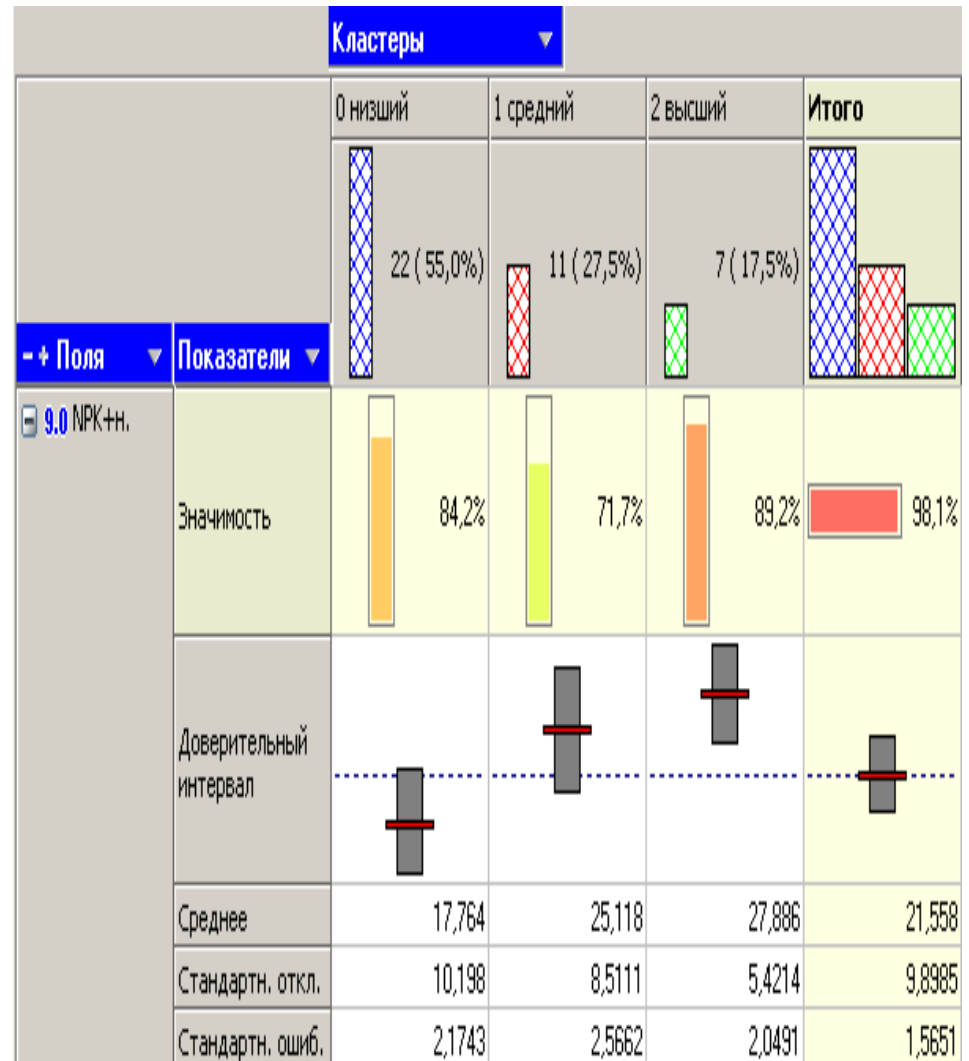
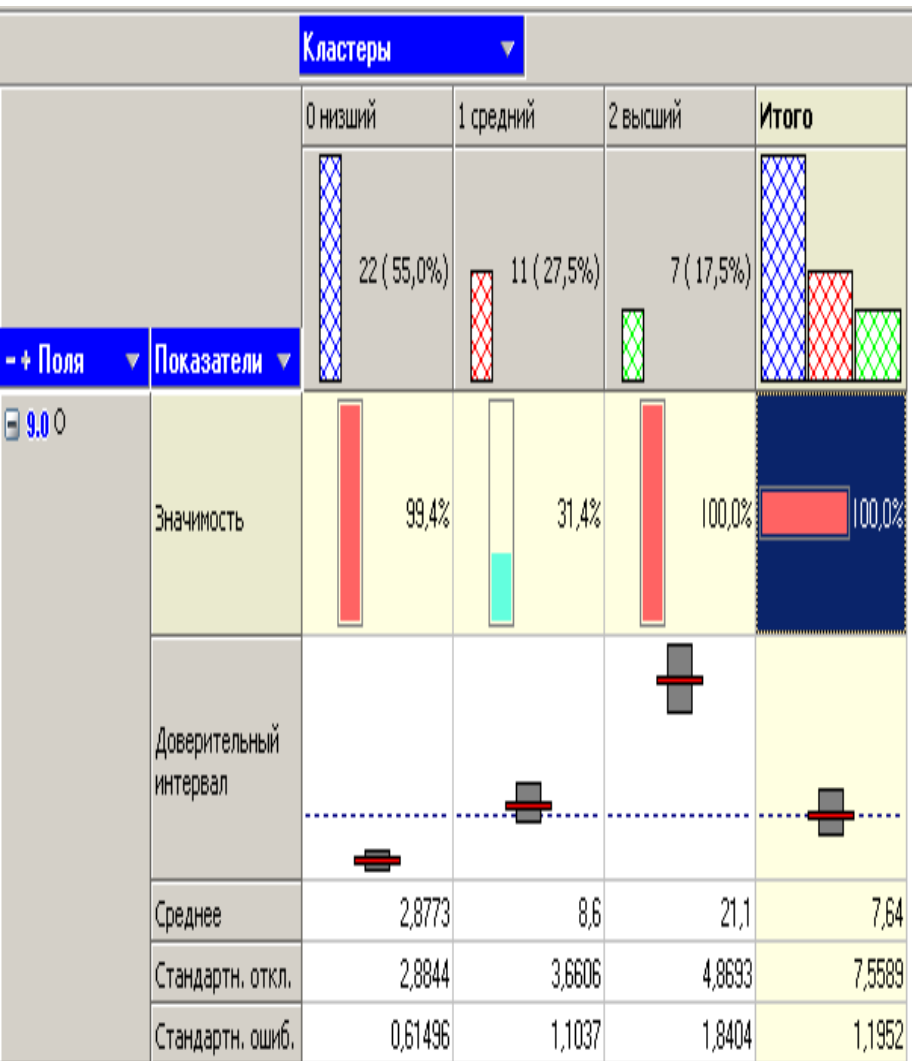
Номер кластера	Год
0	1974
0	1976
0	1977
0	1978
0	1984
0	1985
0	1987
0	1993
1	1979
1	1980
1	1983
1	1986
1	1989
1	1990
1	1991
1	1992
1	2002
1	2003
1	2004
1	2006
2	1973
2	1975
2	1981
2	1982
2	1988
2	1994
2	1995
2	1996
2	1997
2	1998
2	1999
2	2000
2	2001
2	2005
2	2007
2	2008
2	2009

Номер кластера	O	N	P	K	NP	NK	PK	NPK	NPK+n	Расстояние до центра ячейки	Расстояние до центра кластера
1	9,67	12,36	11,78	11,06	9,22	12,74	17,95	19,28	25,68	0,92	2,53
2	11,77	13,50	13,08	12,89	12,06	12,95	27,80	26,25	36,48	1,03	2,62
0	5,85	9,37	9,14	8,00	6,74	7,26	11,16	10,09	12,26	1,26	4,60
	9,00	11,67	11,27	10,57	9,27	10,84	18,80	18,31	24,46	1,08	3,31

Столь существенные колебания урожайности овса и ячменя объясняются, прежде всего, **неустойчивыми погодными условиями.**

Так, урожайность яровых культур в неблагоприятные годы (очень сухие, чрезмерно увлажненные): **1951, 1961, 1964, 1970, 1972, 1988, 1998, 1999, 2005, 2010, 2011 гг.** по всем вариантам опыта была в 2–4 раза ниже по сравнению урожайностью в нормальные по температуре и увлажнению годы.

Кластеризация K – means (фон 1)



Кластеризация K – means (фон 2)

		Кластеры			
		0	1	2	Итого
+ Поля	Показатели				
9.0 NPK+H,	Значимость	99,9%	79,1%	100,0%	100,0%
	Доверительный интервал				
	Среднее	15,374	32,183	46,7	26,972
	Стандартн. откл.	7,0611	3,6979	6,8383	13,927
	Стандартн. ошиб.	1,6199	1,0675	2,4177	2,2301

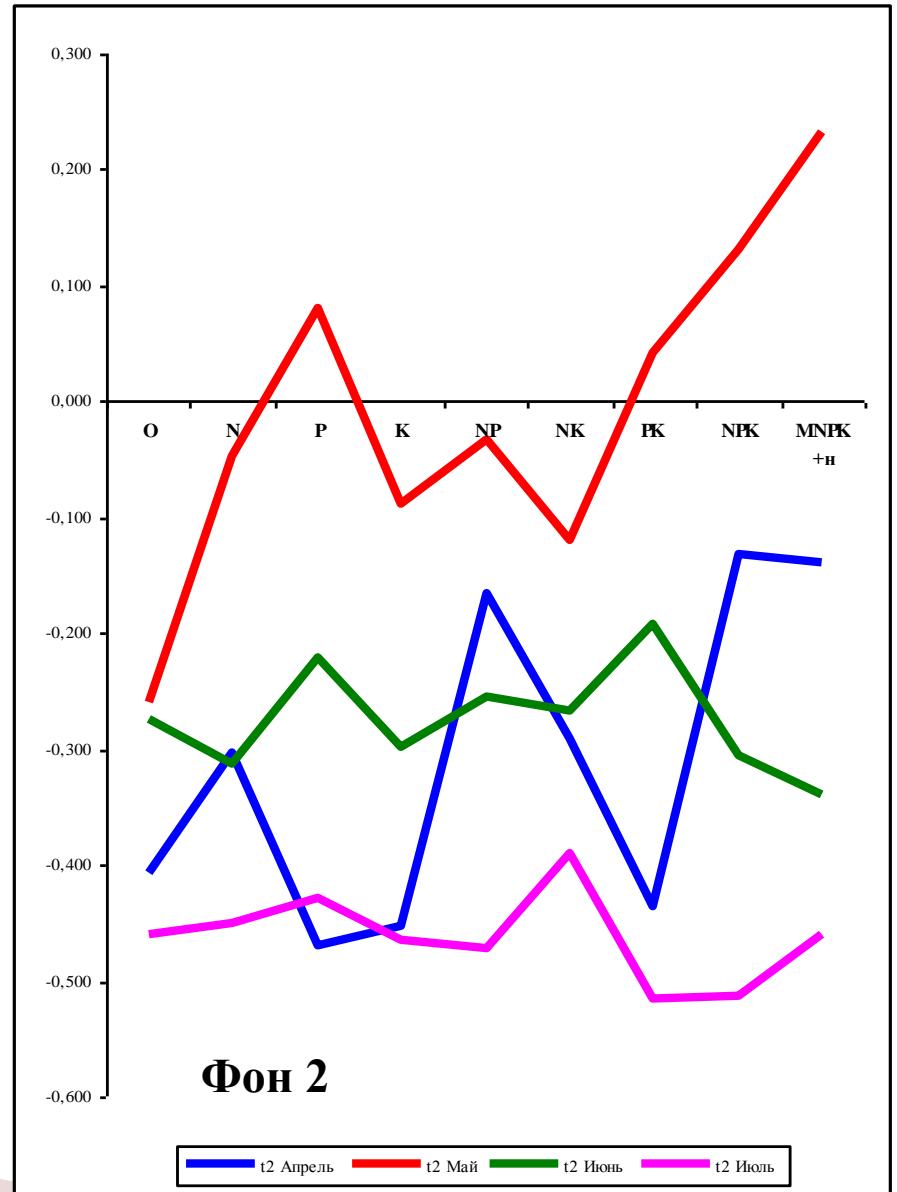
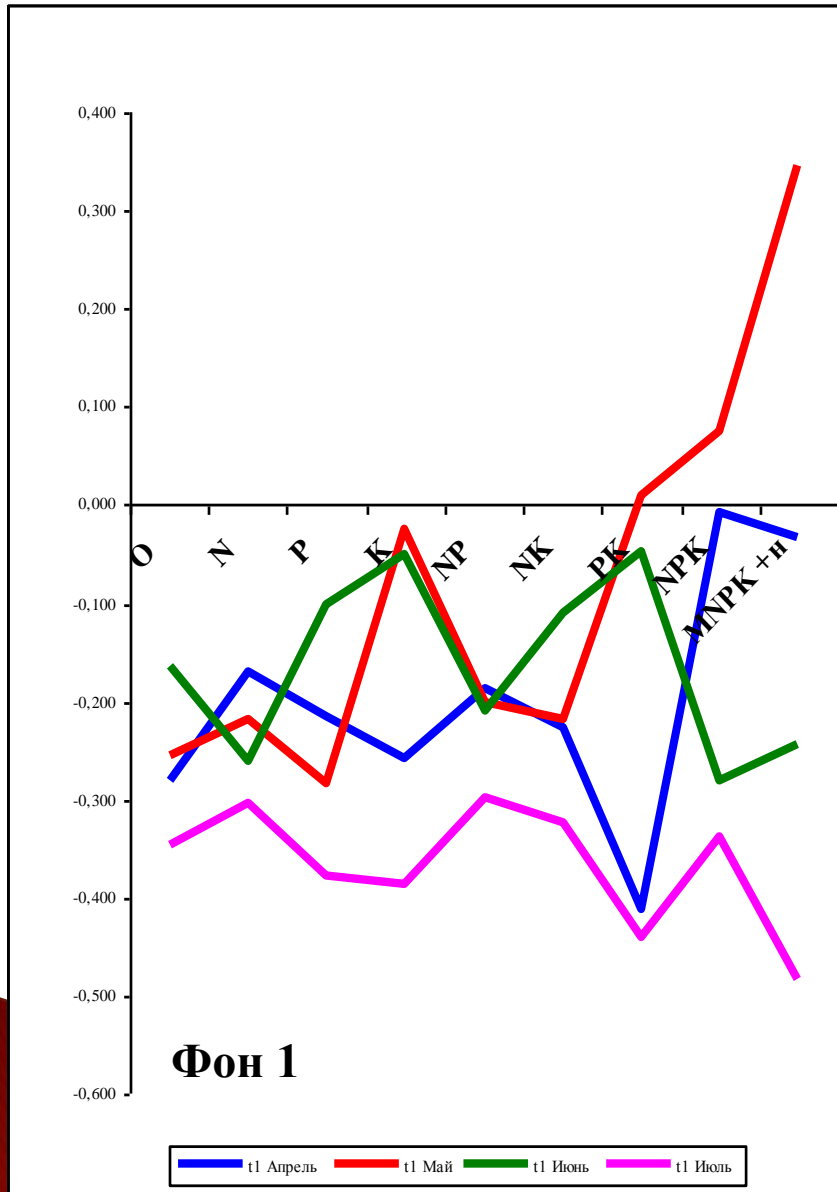
		Кластеры			
		0	1	2	Итого
+ Поля	Показатели				
9.0 O	Значимость	98,2%	77,2%	96,2%	100,0%
	Доверительный интервал				
	Среднее	9,0737	21,75	27,038	16,659
	Стандартн. откл.	7,1831	12,784	12,075	12,579
	Стандартн. ошиб.	1,6479	3,6905	4,2692	2,0143

Уравнение регрессии

Выходное поле: NPK+н.	
Фон 1	
Атрибут	Коэффициент
9.0 <Константа>	27,559
9.0 Навоз, т/га	-4,3692
9.0 N, кг д.в. на 1 га	0,33477
9.0 P205	-0,57232
9.0 K20	1,7884
9.0 t Апрель	-0,8674
9.0 t Май	1,2582
9.0 t Июнь	0,1153
9.0 t Июль	-1,8191
9.0 o Апрель	-0,04314
9.0 o Май	-0,0064736
9.0 o Июнь	0,11465
9.0 o Июль	-0,076473

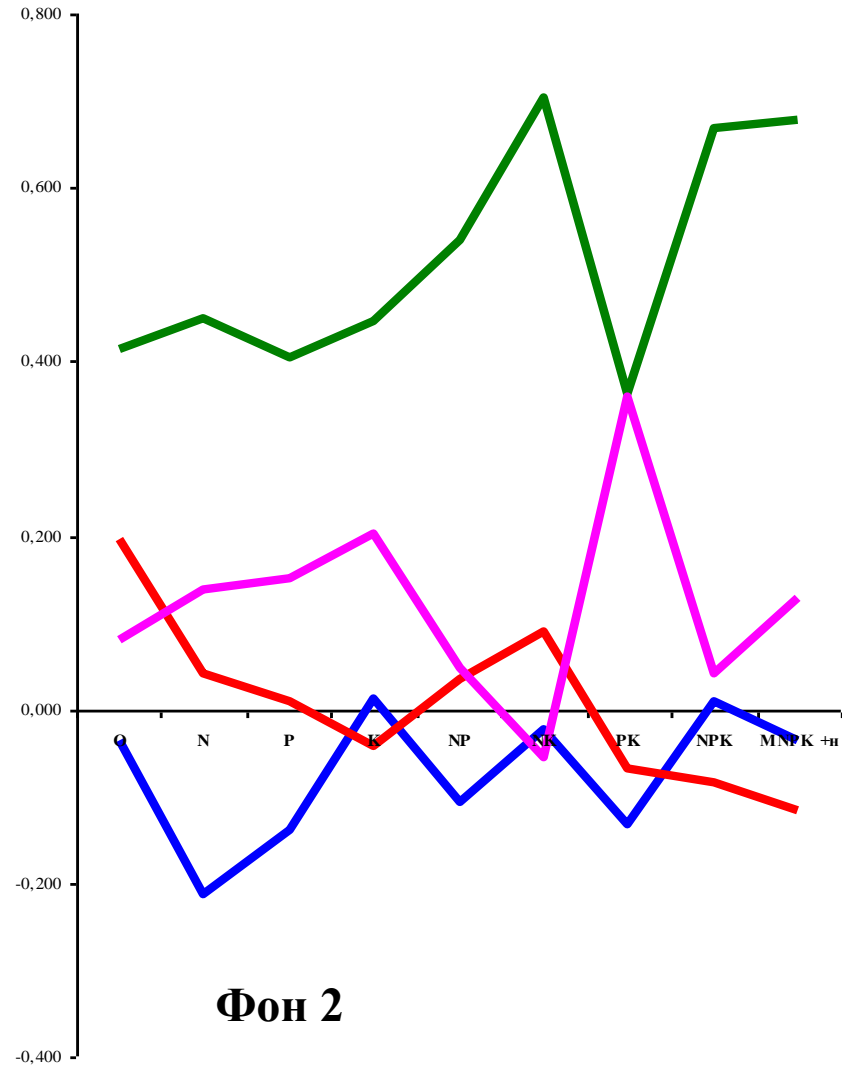
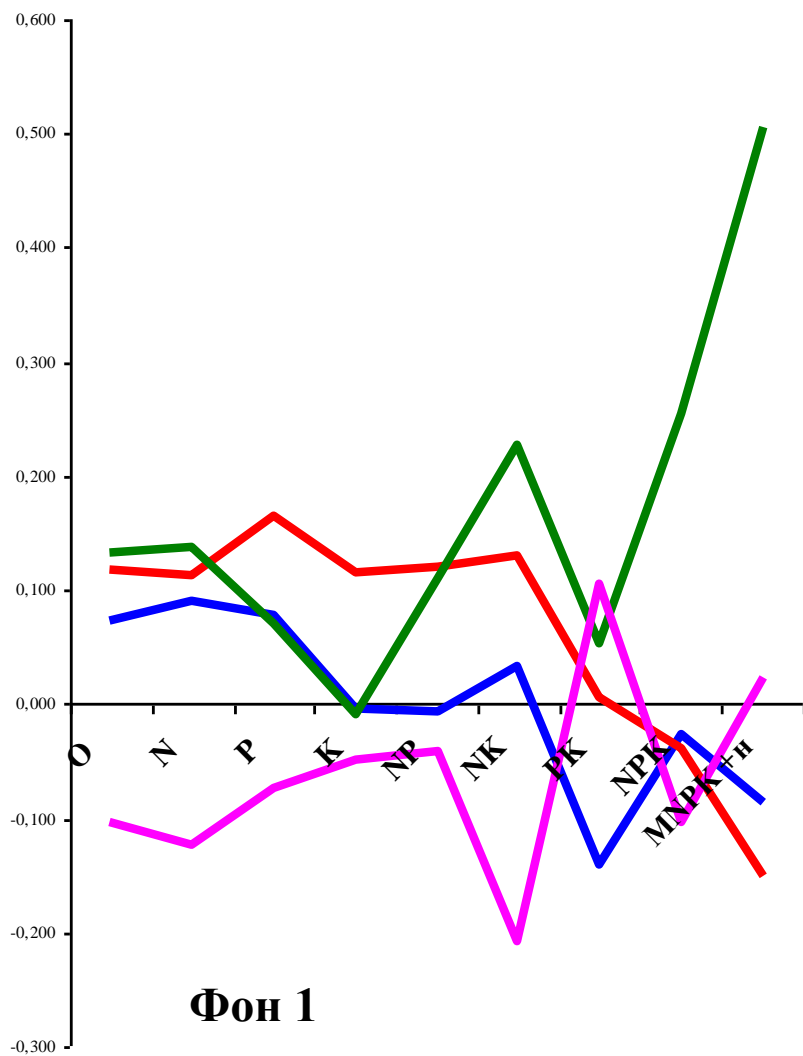
Выходное поле: NPK+н.	
Фон 2	
Атрибут	Коэффициент
9.0 <Константа>	30,331
9.0 Навоз, т/га	0,55152
9.0 N, кг д.в. на 1 га	1,2896
9.0 P205	-0,11872
9.0 K20	-2,3162
9.0 t Апрель	-0,99215
9.0 t Май	0,013469
9.0 t Июнь	-0,19562
9.0 t Июль	-1,3915
9.0 o Апрель	-0,012893
9.0 o Май	-0,012242
9.0 o Июнь	0,21851
9.0 o Июль	0,013225

Коэффициенты корреляции (температура за апрель – июль)



Отмечаются слабые отрицательные корреляции между средней температурой и урожайностью овса и ячменя по всем вариантам удобрений, как на фоне извести, так и без извести.

Коэффициенты корреляции (осадки за апрель – июль)



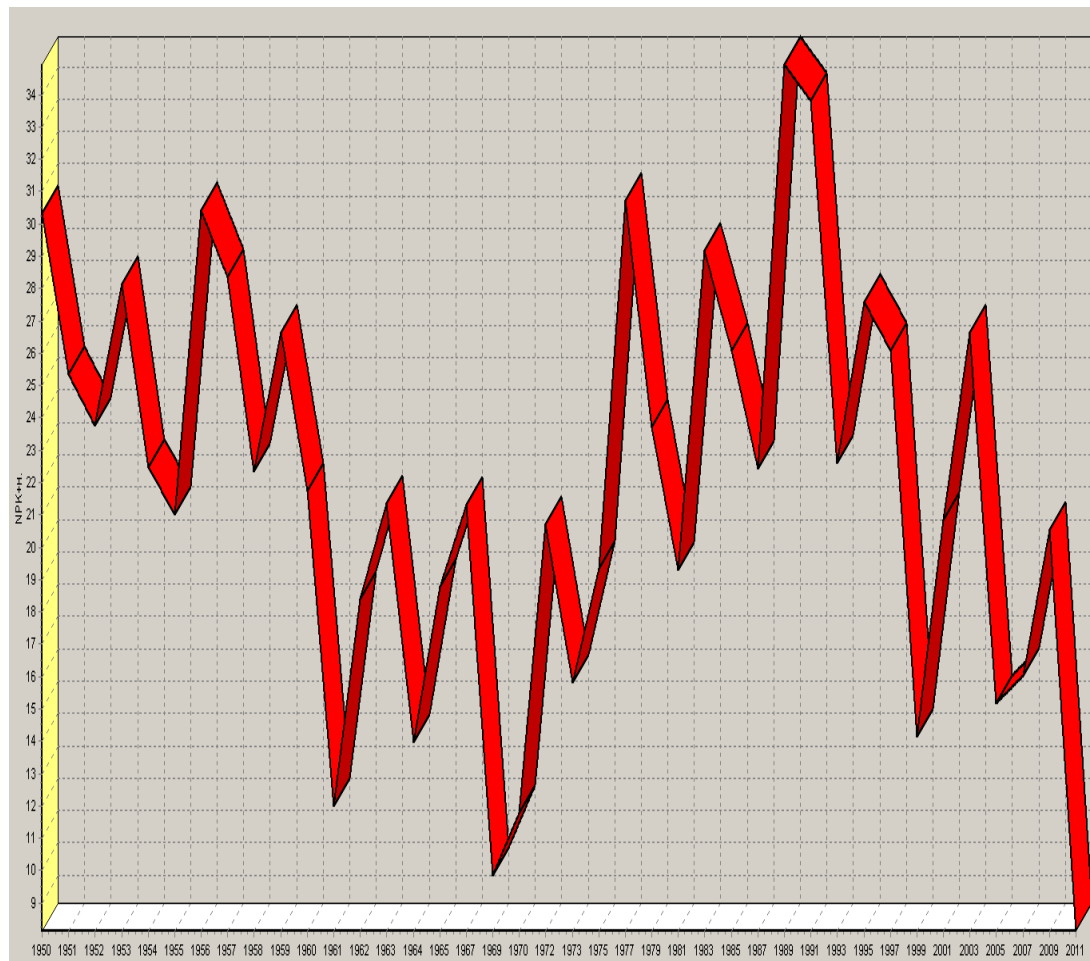
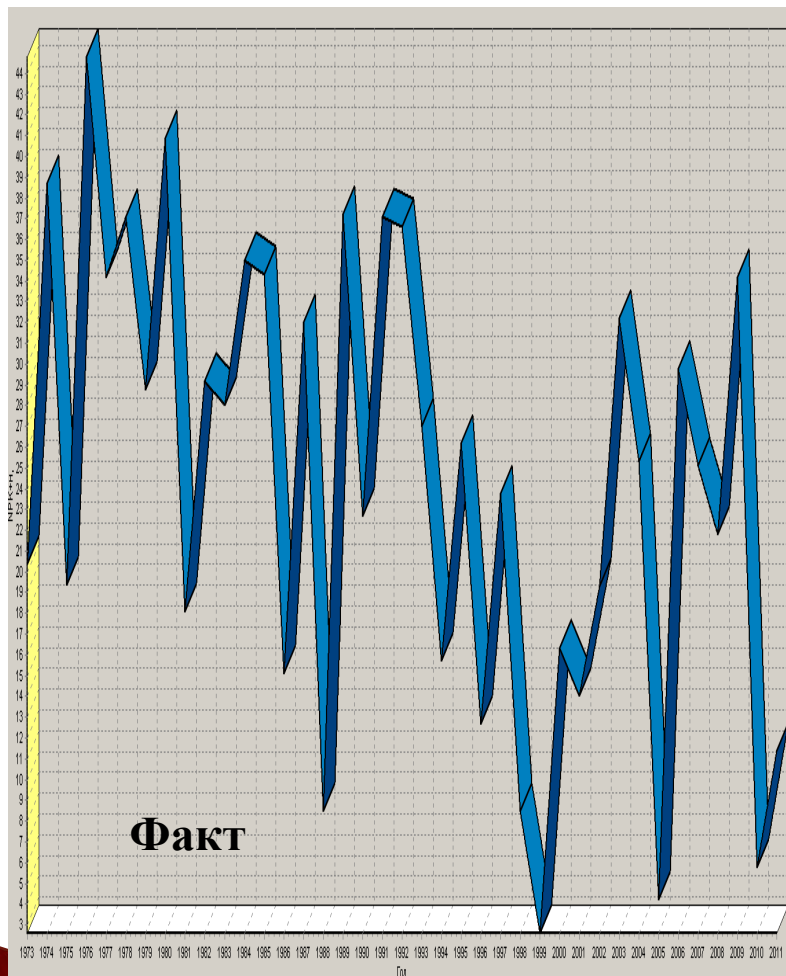
— O1 Апрель — O1 Май — O1 Июнь — O1 Июль

— O2 Апрель — O2 Май — O2 Июнь — O2 Июль

Осадки оказывают большее влияние на рост и развитие растений по сравнению с температурой, особенно осадки, выпавшие **в июне** – средняя корреляционная зависимость (коэффициент корреляции – **от 0,4 до 0,7**).

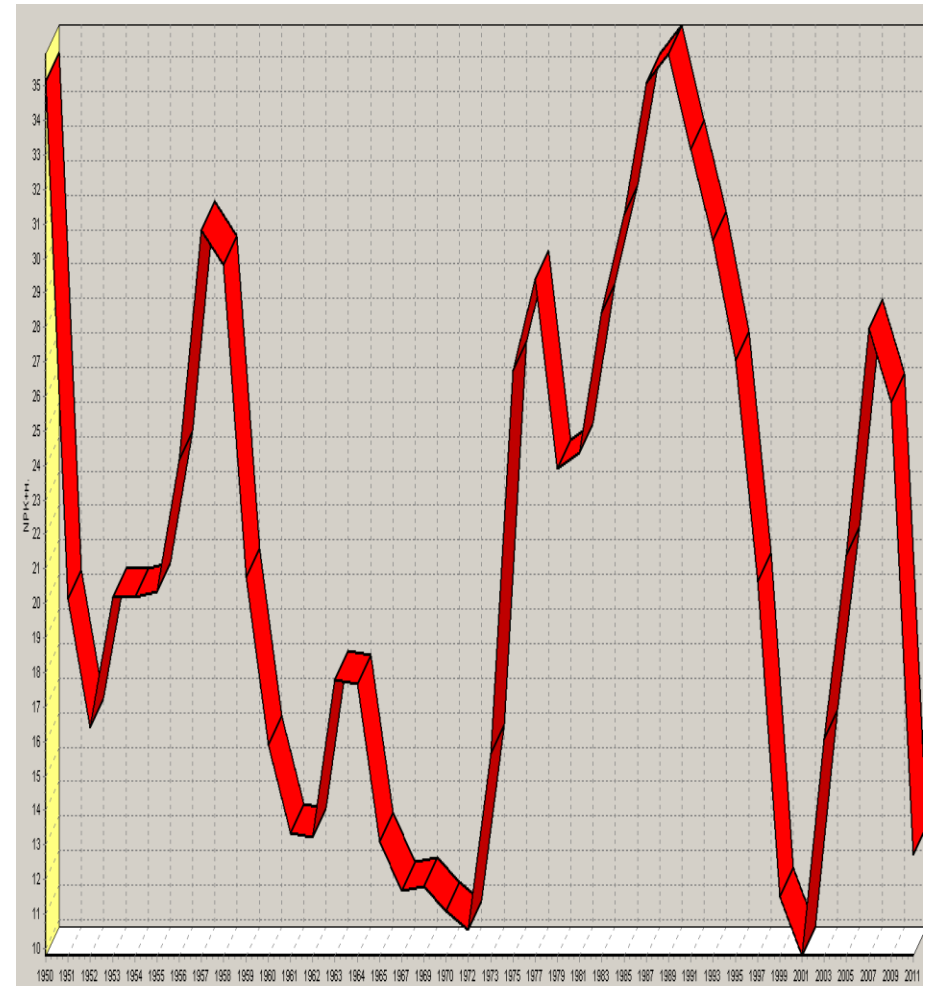
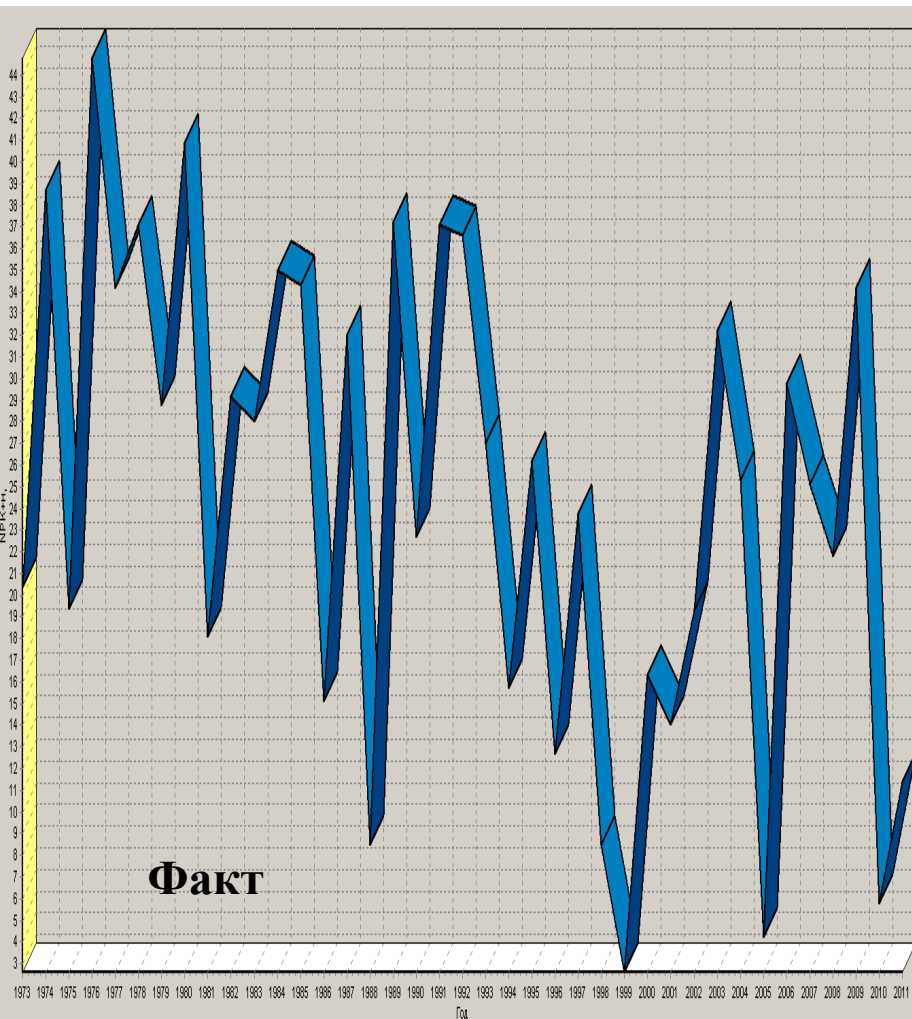
Вероятно, для определения определенных зависимостей необходимы детальные измерения агрометеорологических показателей **по основным фазам развития культур**.

Трансформация данных (аппроксимация, малая степень подавления аномальных значений, малая степень вычитания шума)



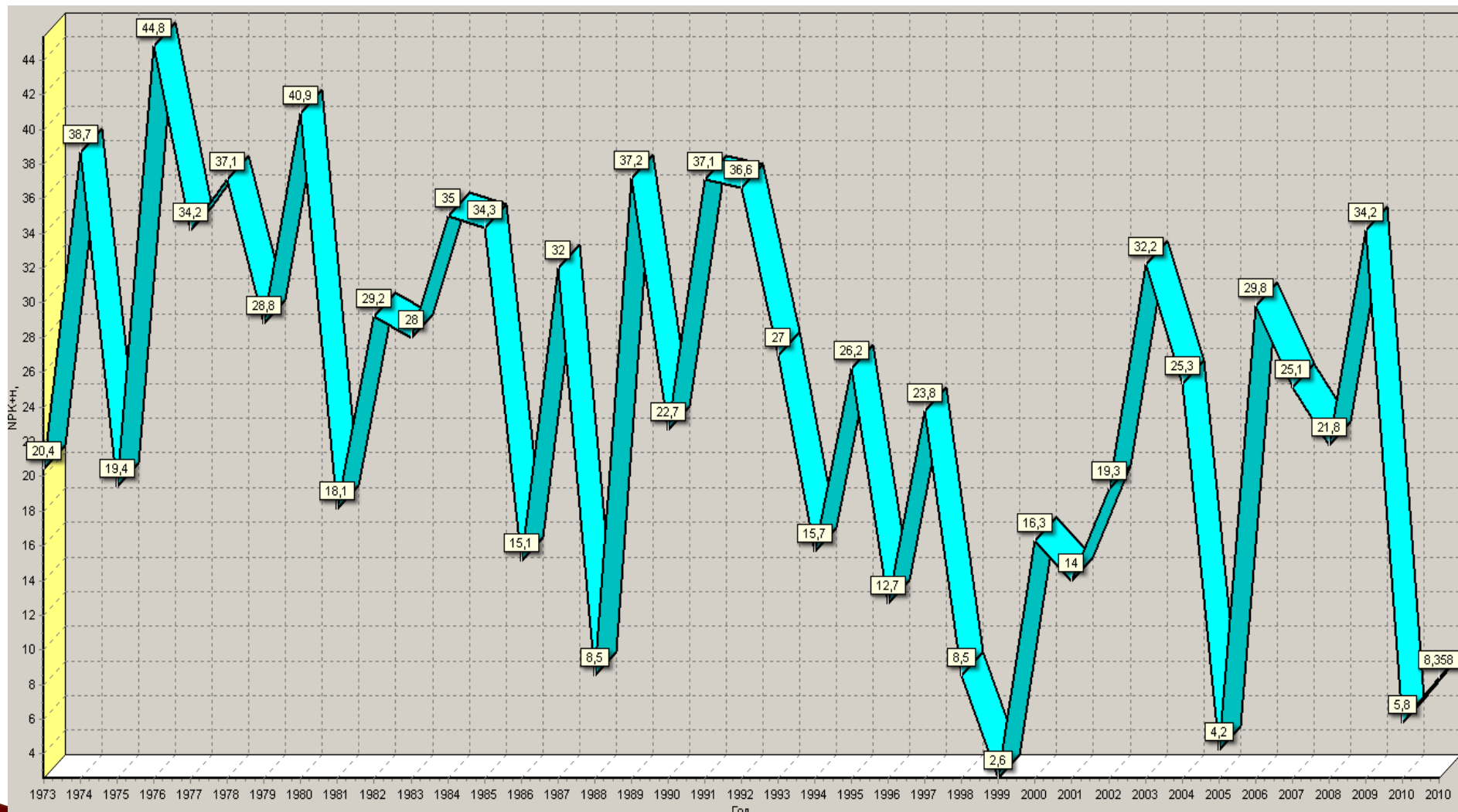
Циклы: Клемана Жюгляра (7 - 11 лет), Саймона Кузнеца (15 - 25 лет)

Трансформация данных (аппроксимация, вейвлет преобразование: глубина разложения – 2, порядок вейвлета – 6)



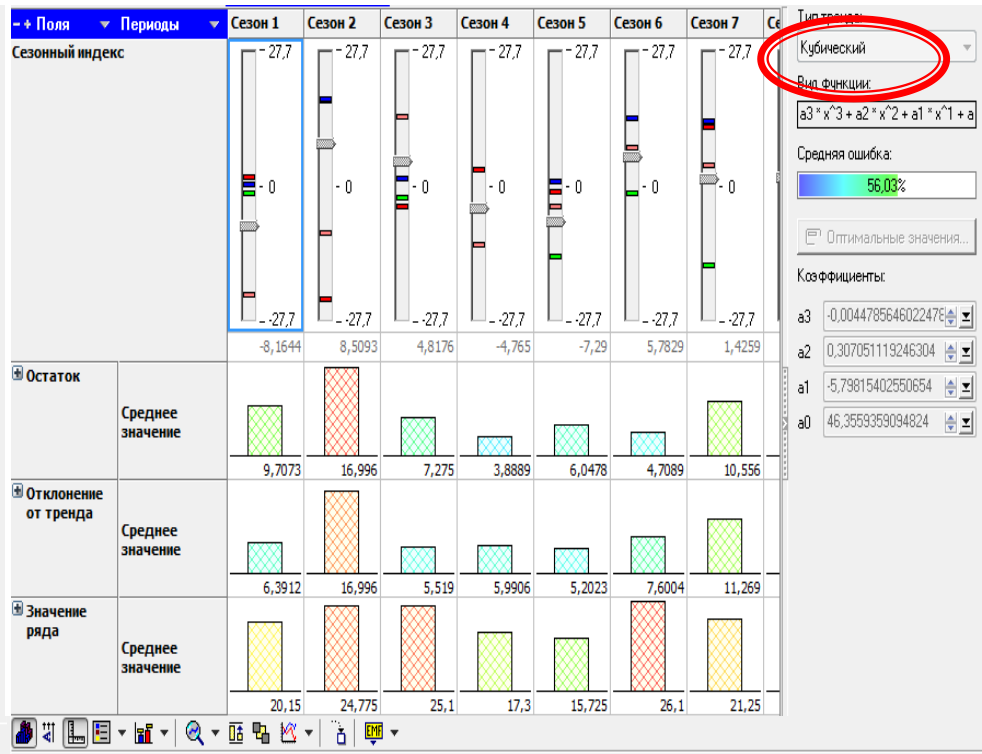
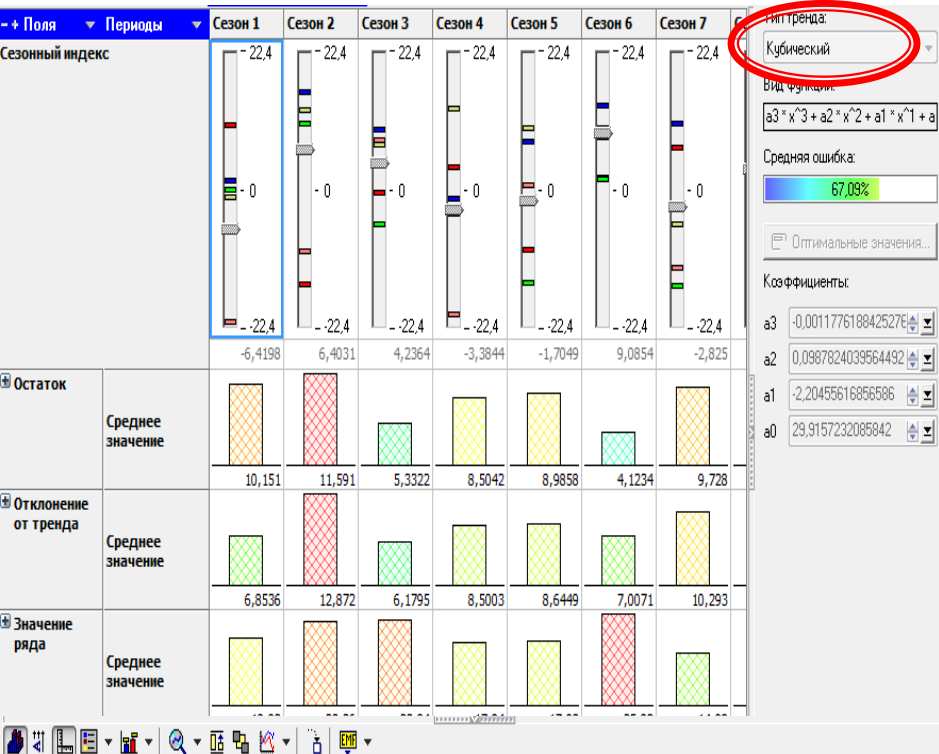
Циклы: Клемана Жюгляра (7 – 11 лет), Саймона Кузнеца (15 – 25 лет)

Прогнозирование

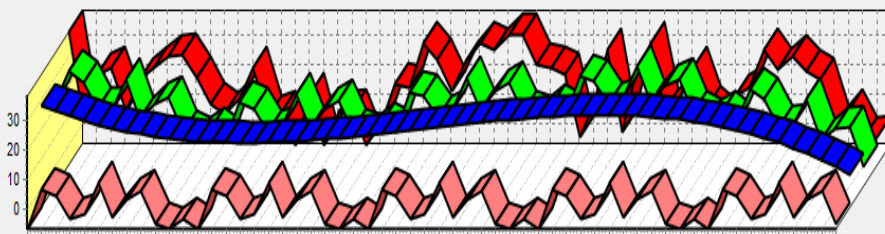


**Фактическое значение урожайности в 2011 г. – 11,4 ц/га,
прогнозное – 8,4 ц/га.**

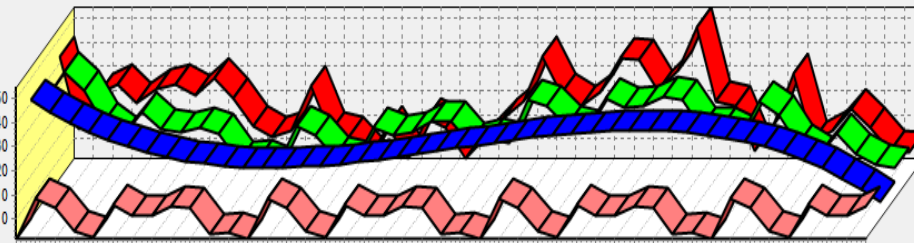
Декомпозиция



Кубический тренд: $-0,0011776 * x^3 + 0,098782 * x^2 - 2,2046 * x^1 + 29,916$



Кубический тренд: $-0,0044786 * x^3 + 0,30705 * x^2 - 5,7982 * x^1 + 46,356$



Фон 1 – без извести

Фон 2 – с известью

Метод главных компонент

Фон 1 (без извести)

	Переменные	Окончательные факторы (Варимакс метод)						
		Фактор 1	Фактор 2	Фактор 3	Фактор 4	Фактор 5	Фактор 6	Фактор 7
<input checked="" type="checkbox"/> O	Навоз, т/га	0,9304						
<input checked="" type="checkbox"/> N	N, кг д.в. на 1 га	0,9844						
<input checked="" type="checkbox"/> P	P205	0,9204						
<input checked="" type="checkbox"/> K	K20	0,9844						
<input checked="" type="checkbox"/> NP	o Апрель					-0,9705		
<input checked="" type="checkbox"/> NK	o Май			-0,9752				
<input checked="" type="checkbox"/> PK	o Июнь		-0,8865					
<input checked="" type="checkbox"/> NPK	o Июль				-0,9167			
<input checked="" type="checkbox"/> NPK+н.	t Апрель						0,9435	
	t Май							0,8971
<input checked="" type="checkbox"/> Навоз, т/га	t Июнь		0,8111					
<input checked="" type="checkbox"/> N, кг д.в. на 1 га	t Июль		0,3476		0,6595			

Фон 2 (с известью)

	Переменные	Окончательные факторы (Варимакс метод)						
		Фактор 1	Фактор 2	Фактор 3	Фактор 4	Фактор 5	Фактор 6	Фактор 7
<input checked="" type="checkbox"/> O	Навоз, т/га	0,9230						
<input checked="" type="checkbox"/> N	N, кг д.в. на 1 га	0,9856						
<input checked="" type="checkbox"/> P	P205	0,9088						
<input checked="" type="checkbox"/> K	K20	0,9856						
<input checked="" type="checkbox"/> NP	o Апрель					0,9731		
<input checked="" type="checkbox"/> NK	o Май				-0,9707			
<input checked="" type="checkbox"/> PK	o Июнь		-0,9169					
<input checked="" type="checkbox"/> NPK	o Июль			-0,8789				
<input checked="" type="checkbox"/> NPK+н.	t Апрель						0,9447	
<input checked="" type="checkbox"/> Навоз, т/га	t Май							0,8738
<input checked="" type="checkbox"/> N, кг д.в. на 1 га	t Июнь		0,7452					-0,3922
<input checked="" type="checkbox"/> P	t Июль		0,3878	0,7339				

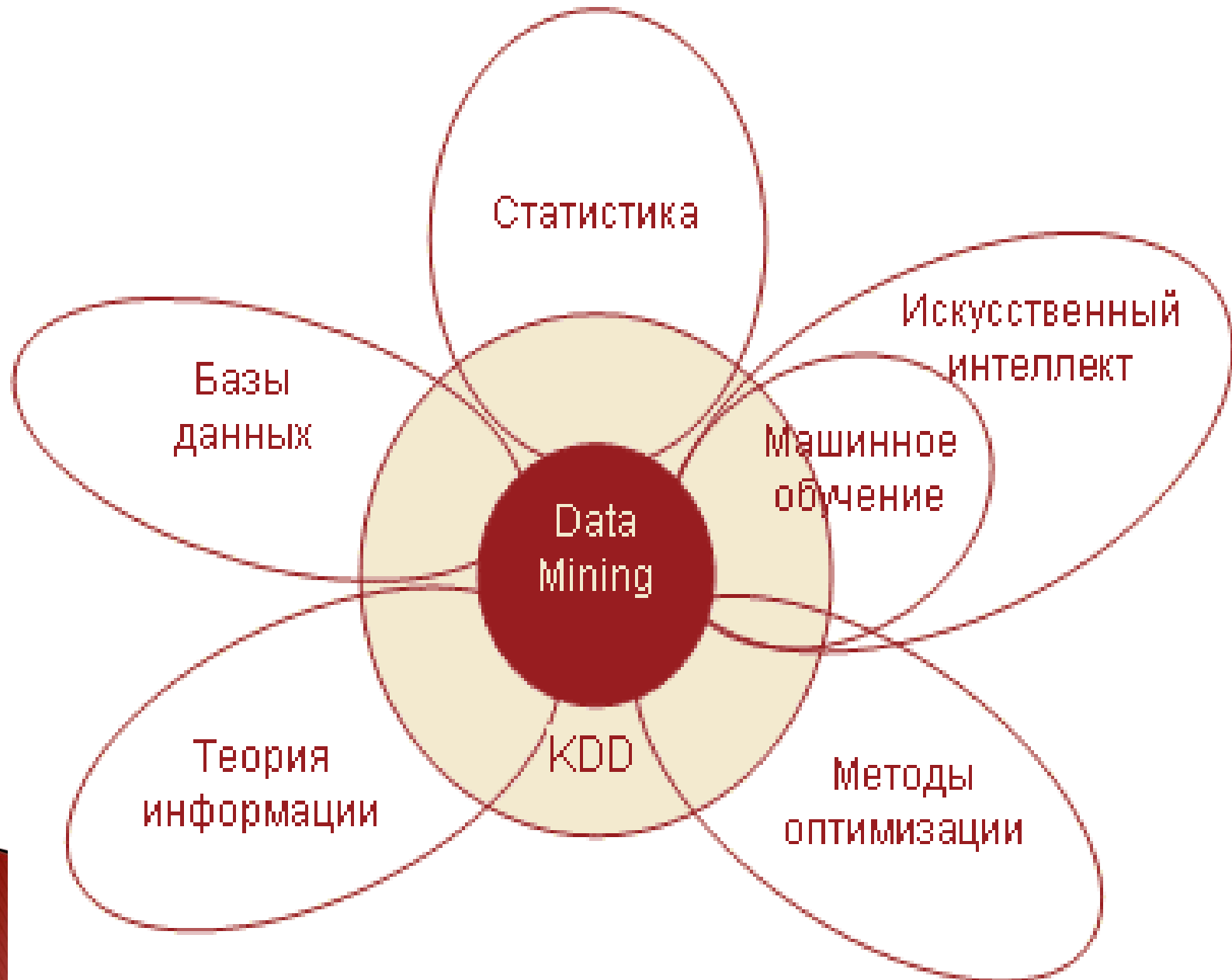
- O
- N
- P
- K
- NP
- NK
- PK
- NPK
- NPK+н.
- Навоз, т/га
- N, кг д.в. на 1 га
- P205
- K20
- o Апрель
- o Май
- o Июнь
- o Июль
- o Сумма за год
- t Апрель
- t Май
- t Июнь
- t Июль
- t Средняя за год
- Год

Заключение

Список используемых обработчиков:

- **метод главных компонент;**
- **методы кластеризация;**
- **линейная регрессия;**
- **корреляционный анализ;**
- **методы трансформации данных;**
- **прогнозирование и др.**

Мультидисциплинарный характер Data Mining



**Благодарим
за внимание!**