

Сибирский региональный научно-исследовательский гидрометеорологический институт.

Научно- практическая школа-семинар для молодых ученых и специалистов Росгидромета.

31 октября - 2 ноября 2012 года, г. Новосибирск

Токарев В.М. (зав.ОИНИТ <СибНИГМИ>, г. Новосибирск)

Работа с метеорологическими данными: базы данных, обработка, графики, карты.

1. Базы данных.

Предпочтительные кроссплатформенные СУБД:

SQLite - "легкая", открытая;

MySQL - самая популярная в сети Интернет, свободно-распространяемая.

PostgreSQL - свободно-распространяемая, мощная.

Для чего?

- SQL
- универсальность
- переносимость
- забыть про "самоформатные личные файлы"

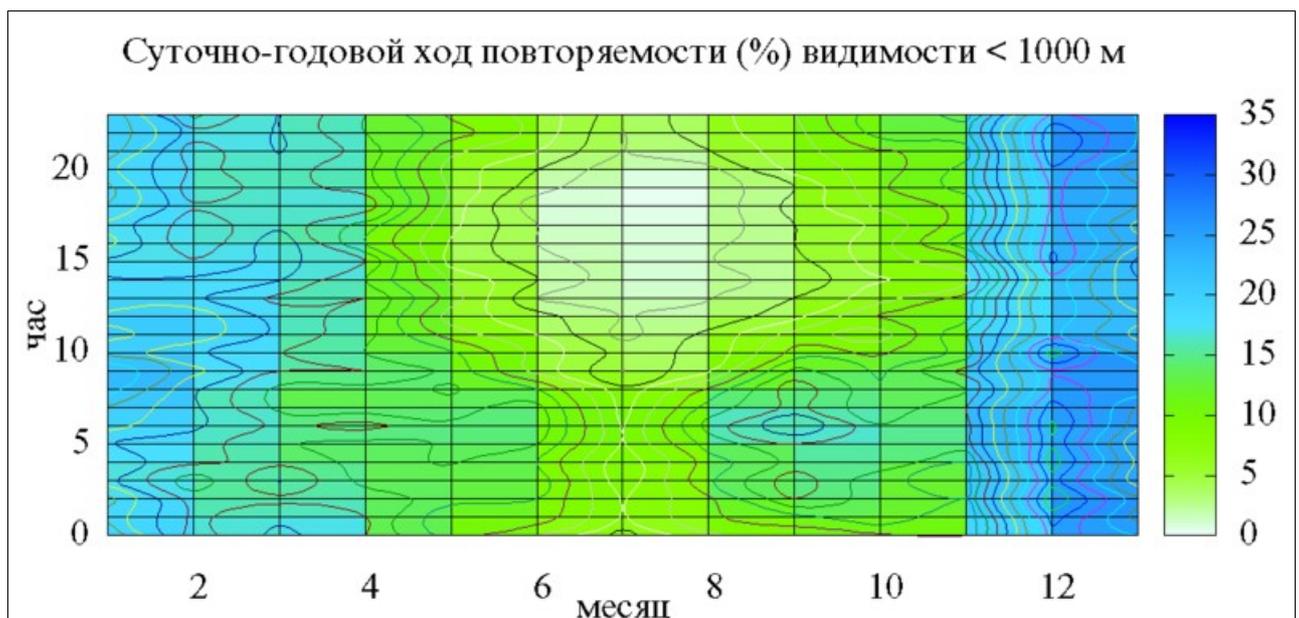
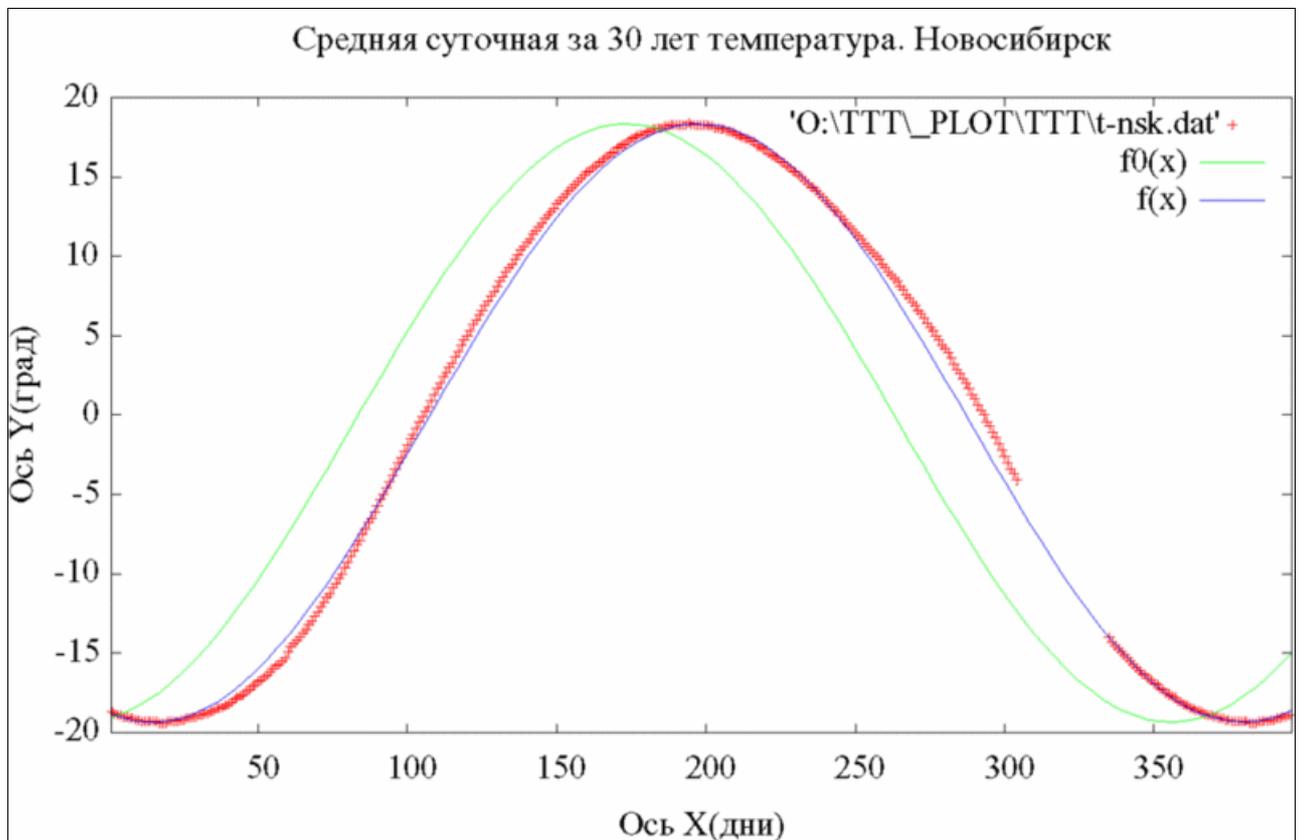
Математика, алгоритмы, программы:

MatLab - "матричная лаборатория"- интерактивная система обработки данных

Linpack - тщательно отлаженный пакет программ по вычислительной математике для программистов.

Графика:

GNUPlot- An Interactive Plotting Program



МНК: - почему МНК?

- переопределенные системы, математически некорректные, миним. норма
- зависимые переменные, неполный ранг, SVD

Дж.Голуб, Ч. Ван Лоун. Матричные вычисления. Пер. с англ. М:Мир, 1999г

Д. Каханер, К. Моулера, С. Нэш - Численные методы и программное обеспечение. Пер. с англ. М:Мир, 1998г

Е.Е.Тыртышников. Матричный анализ и линейная алгебра. М.: 2005г

Е.Е.Тыртышников. Методы численного анализа. М.: 2006г

2. Методология:

2.1 выборка должна быть однородной

- неоднородность подозревается: суточный, годовой ход, смешение масштабов(синоптический, мезо), пересечение вертикальных слоев (приземный, пограничный, свободная атмосфера), фазовые переходы(вода-лед, снег) и т.п.

- неоднородность проверяется: графический анализ, дисперсионный анализ,

- неоднородность нивелируется разбиением выборки, введением определяющей неоднородность переменной(многомерные выборки)

- цель: устойчивость МОДЕЛЕЙ должна преобладать над точностью!

2.2 бороться с неустойчивыми решениями(матрица неполного ранга, близкие к нулю сингулярные числа) не алгоритмическими ухищрениями (регуляризация), а корректировкой постановки задачи или упрощением модели данных.

2.3 МГУА

Метод группового учета аргументов, МГУА (Group Method of Data Handling, GMDH) - метод порождения и выбора регрессионных моделей оптимальной сложности.

А.Г. Ивахненко. Индуктивный метод самоорганизации моделей сложных систем. Киев: Наукова думка. 1981.

Теорема Гёделя о неполноте. Алгебра логики. Замкнутость, причинность.

- принцип внешнего дополнения
- принцип свободы выбора

Критерий качества модели - внешний, при оценке используются разные данные.

Метеорологи использовали давно!, но только для "независимой" оценки.

Внутренний критерий остается для параметров 1-й модели - среднеквадратическая ошибка. Модели - из семейства полиномов, наращивание степени, могут быть и другие семейства функций. Параметры : МНК-решение.

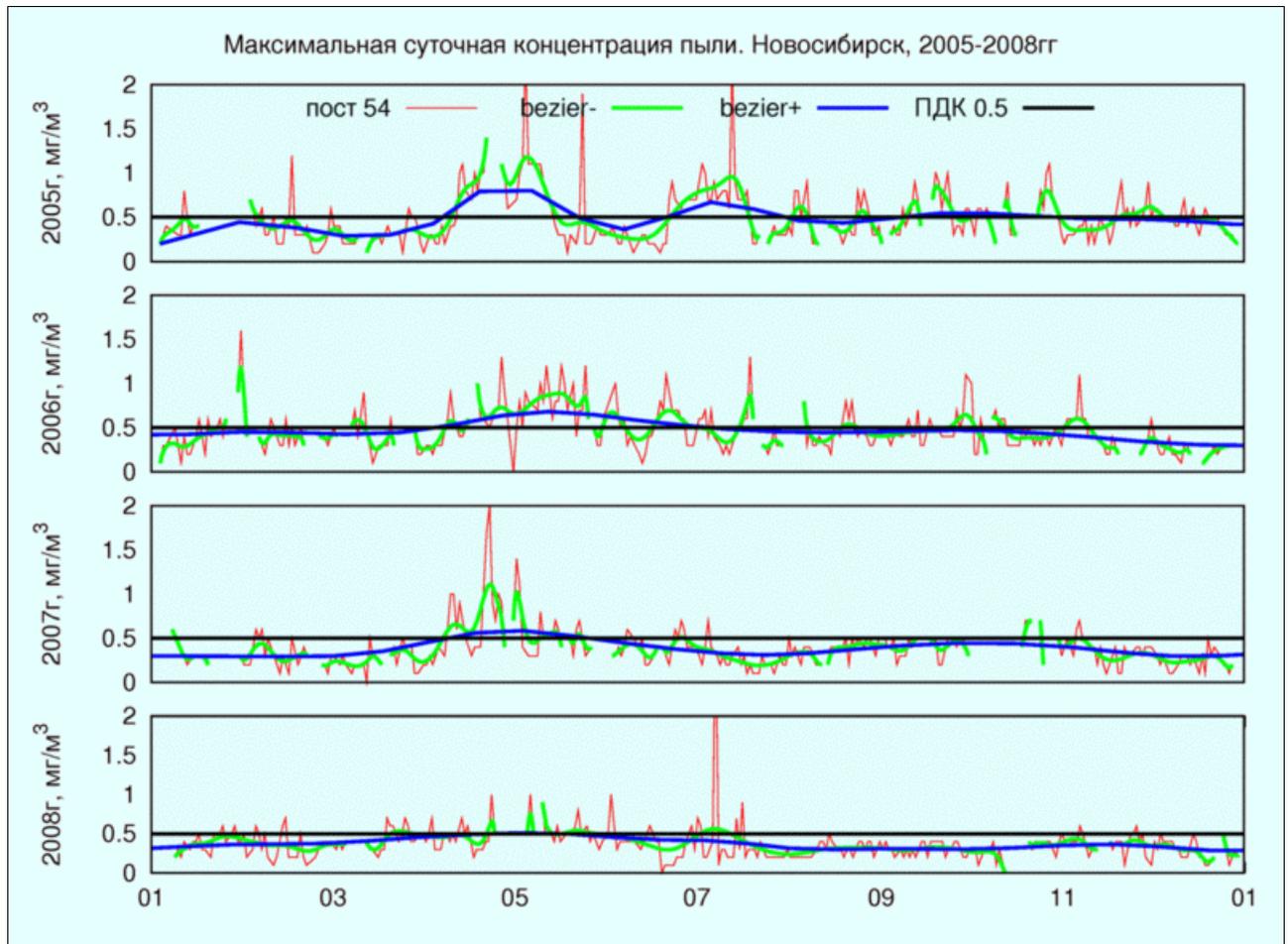
Внешние критерии МГУА(отбор моделей, свобода выбора):

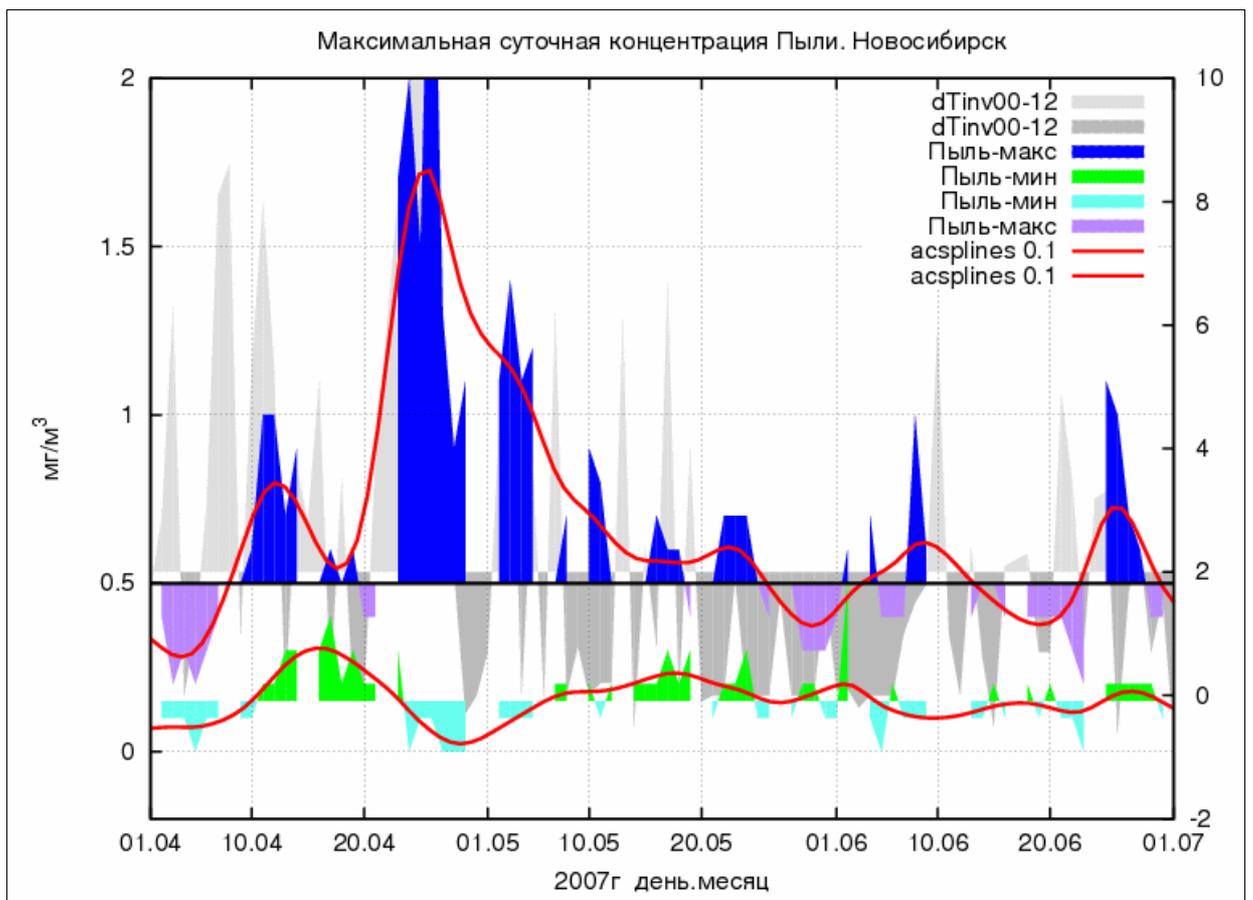
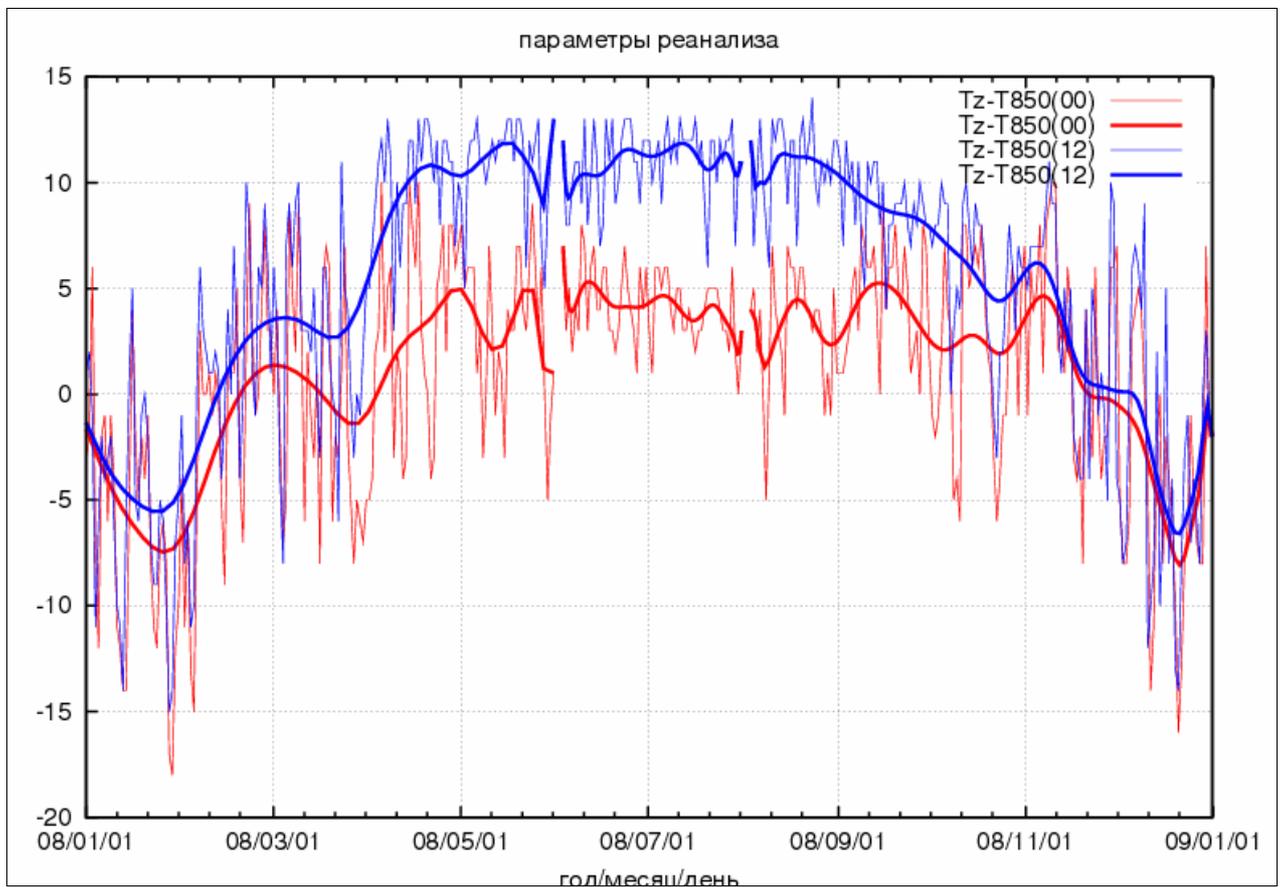
- критерий регулярности
- критерий минимума смещения
- критерий баланса и др.

Использование в СибНИГМИ: М.Я.Здерева, прогностические модели, интерпретация, уточнение расчетных полей.

2.4 Многомерные выборки:

Первый этап исследования - **разведочный анализ данных, включающий графическое представление** многомерных временных рядов для предварительной оценки стационарности, эргодичности, непрерывности и полноты данных





и численные оценки условных вероятностных оценок из теории дисперсионной идентификации.

Дисперсионная идентификация/Под ред. Н.С.Райбмана.- М.: Наука,1981-336с.

Основой условных статистических оценок служит разложение дисперсии:

$$DY = DM(Y|x) + MD(Y|x),$$

где: $DM(Y|x)$ - дисперсия Y за счет влияния X ,

$MD(Y|x)$ - неопределенная(остаточная) дисперсия Y ,

$$D(Y|x) = M[Y - M(Y|x)]^2$$

$DM(Y|x)/DY$ - дисперсионное(корреляционное) отношение, мера(нелинейной) связи.

	по городу				N	посты		тренд		сезонность	
	макс	мин	сред	дисп		$DM(Y/X)$	$DM(Y/X)/DY$	$DM(Y/X)/DY$		$DM(Y/X)/DY$	
	M_Y	M_Y	M_Y	DY				выброс	фон	выброс	фон
								%	%	%	%
пыль	0.55405	0.12934	0.30519	0.04998	10	0.00293	5.86	2.71	2.24	6.95	13.96
SO2	0.01860	0.00136	0.00782	0.00026	6	0.00001	6.18	8.42	5.07	22.53	13.32
CO	4.78524	1.28747	2.58200	4.07565	10	0.27580	6.76	11.38	0.25	2.46	19.45
NO2	0.16659	0.03972	0.09193	0.00533	10	0.00030	5.66	12.62	53.07	2.62	0.61
NO	0.08238	0.02959	0.05714	0.00274	7	0.00066	24.32	37.35	51.67	1.80	0.85
H2S	0.00046	0.00046	0.00046	0.00000	1	0.00000	0.00	5.05	5.05	0.46	0.46
Фенол	0.00740	0.00124	0.00415	0.00004	3	0.00000	0.47	11.80	10.21	0.42	0.37
Сажа	0.09714	0.00164	0.03330	0.00527	10	0.00038	7.38	0.67	4.26	7.49	9.87
HF	0.00763	0.00763	0.00760	0.00008	1	0.00000	0.00	1.49	1.49	3.21	3.21
Аммиак	0.13945	0.02303	0.07739	0.01305	3	0.00067	5.17	5.27	1.27	2.20	1.90
Форм	0.02357	0.00734	0.01525	0.00018	6	0.00001	10.65	21.87	52.55	0.47	0.86

2.5 Распознавание образов.

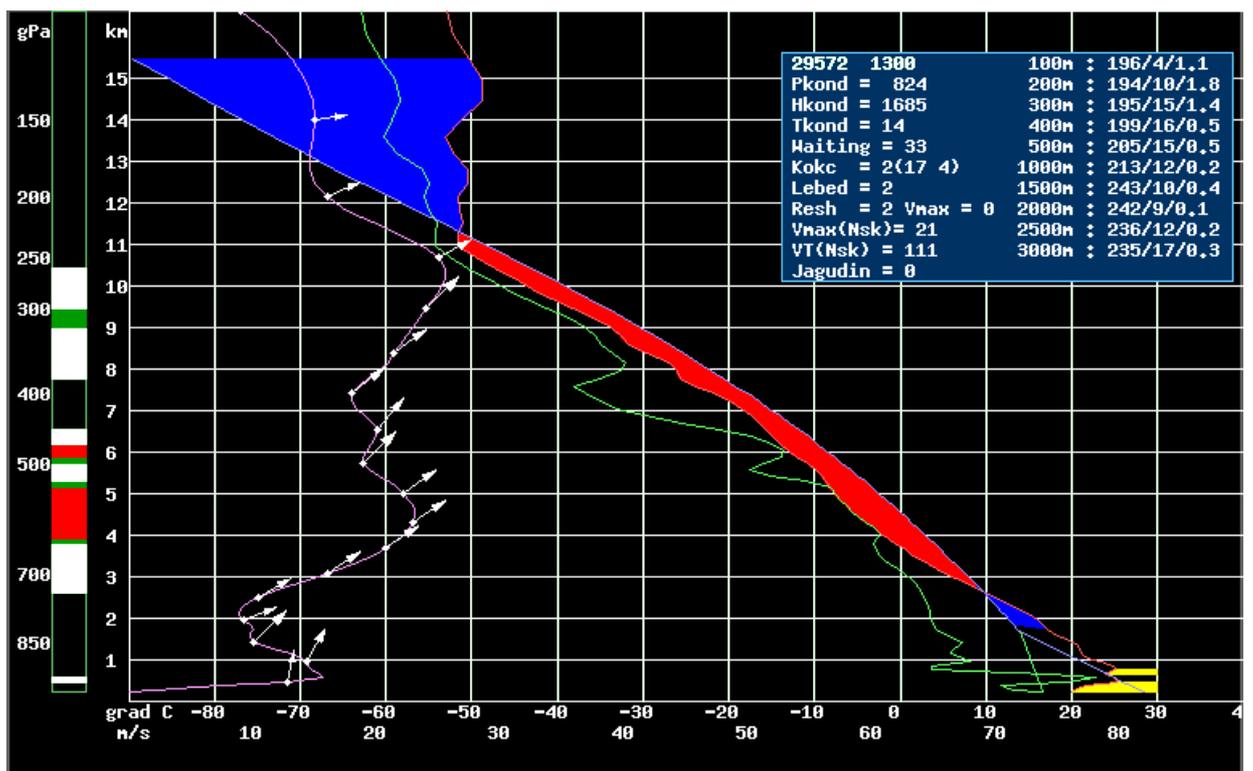
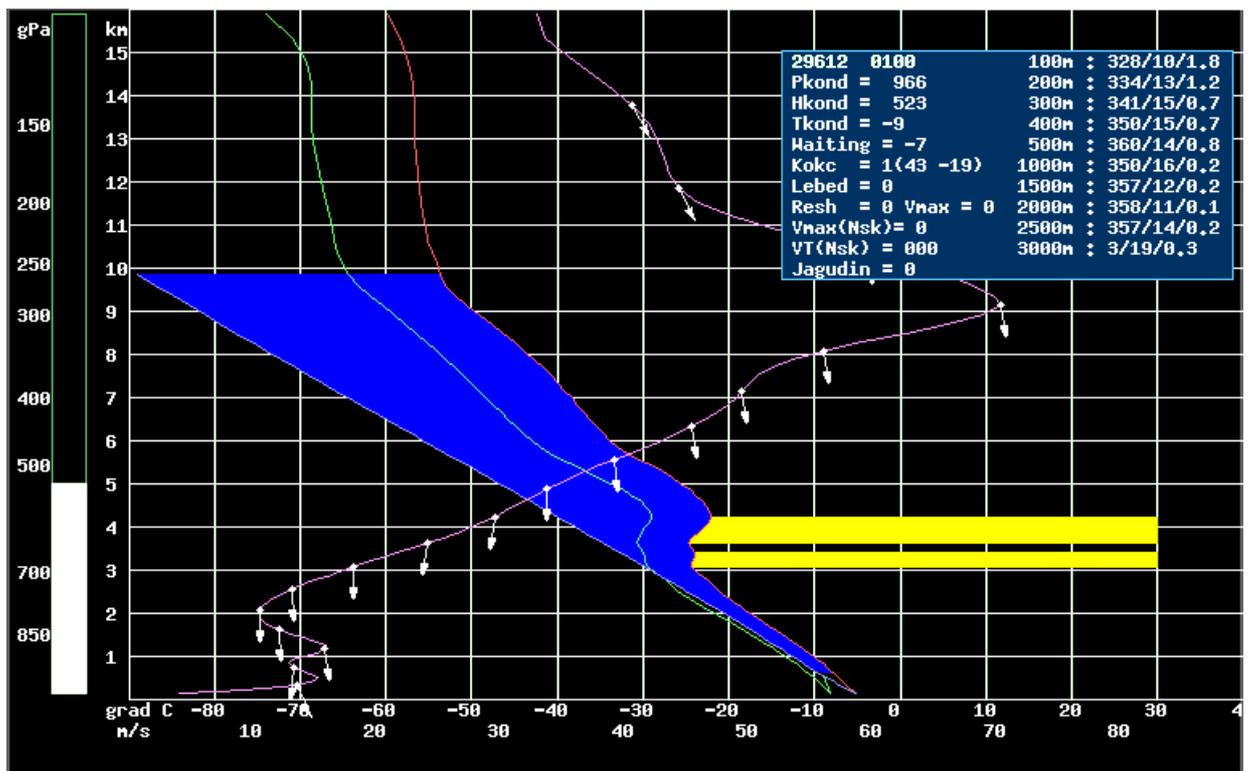
Дж.Ту, Р.Гонсалес. Принципы распознавания образов. Пер. с англ. М:Мир, 1978г

А.Н. Манохин Алгоритм DW для распознавания образов: Пакет прикладных программ ОТЭКС - Новосибирск: изд-во Новосибирского государственного университета, 1981

Логические деревья распознавания случаев концентрации примесей выше ПДК (2 и больше постов).
Обучение. Новосибирск. Выборка 2005-2008гг. Вариант без признака серии.

Пыль 141/1160=0.12					
H500(12)					
<=549(0.43) 27/497=0.05			>549(0.57) 114/663=0.17		
r(00)			VV(00)		
>72(0.18) 5/214=0.02	<=72(0.24) 22/283=0.08		>5(0.31) 41/354=0.12		<=5(0.27) 73/309=0.24
H500(12)		r(00)		Tz-T850	
<=533(0.07) 0/81=0.00		≥533(0.17) 22/202=0.11		>64(0.13) 8/148=0.05	<=64(0.18) 33/206=0.16
				>2(0.17) 33/198=0.17	<=2(0.10) 40/111=0.36
106	35	141	Предупрежденность: 75%		
409	610	1019			
515	645	1160	Общая оправдываемость: 62%		

2.6 Аппроксимация, интерполяция.



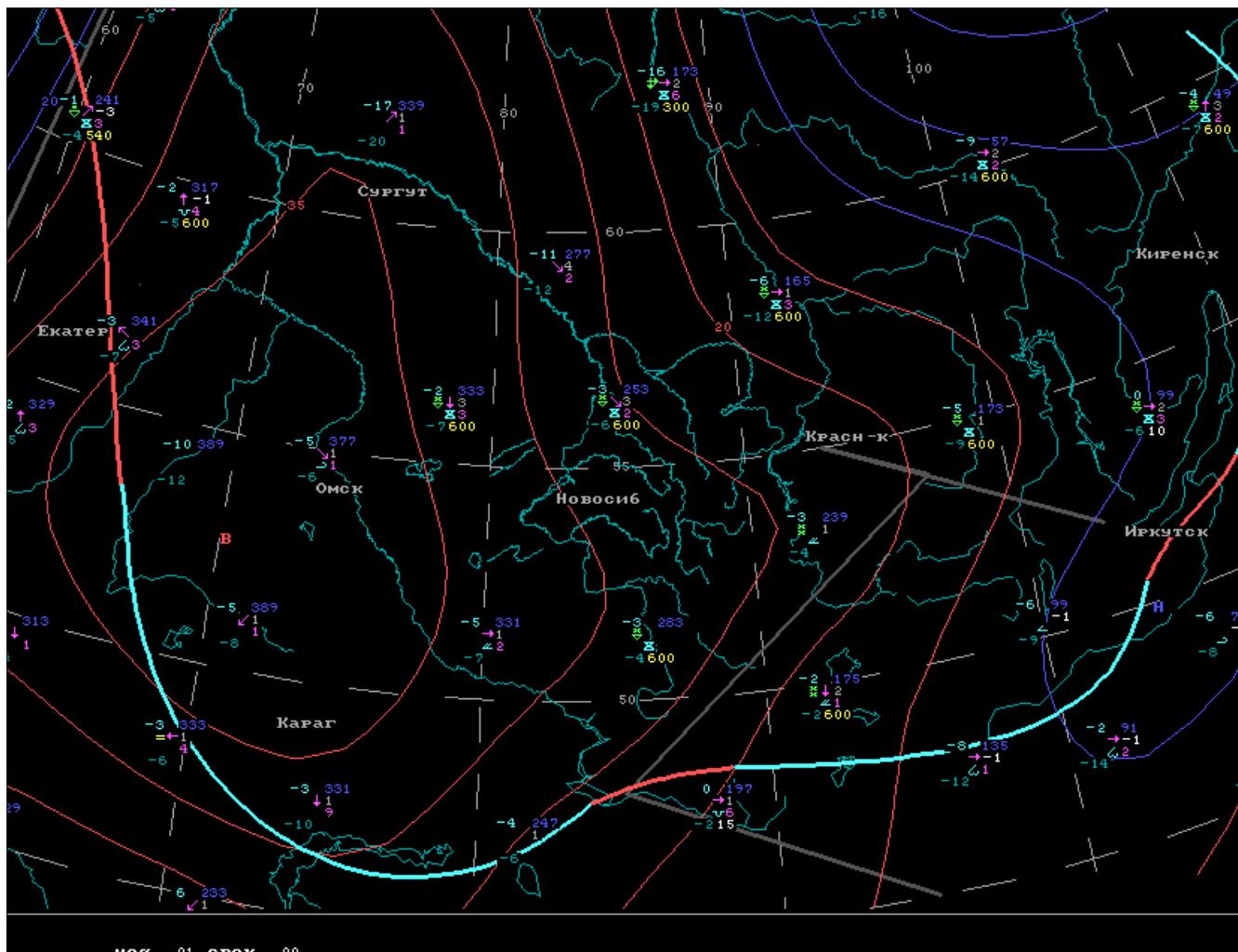
2.7 Карты: спец.АРМы, ГИСы, Яндекс-Гугл-"стриты", GPS..

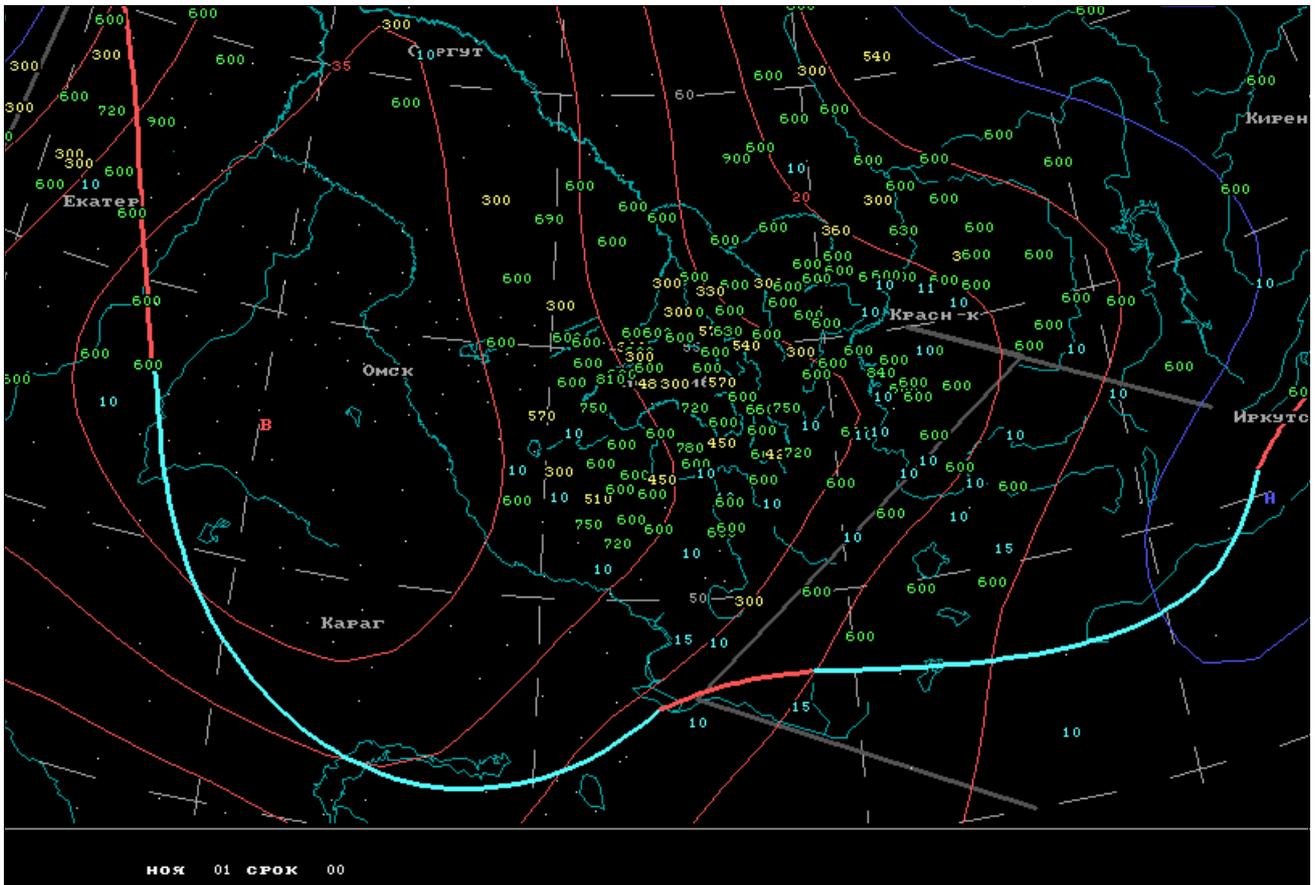
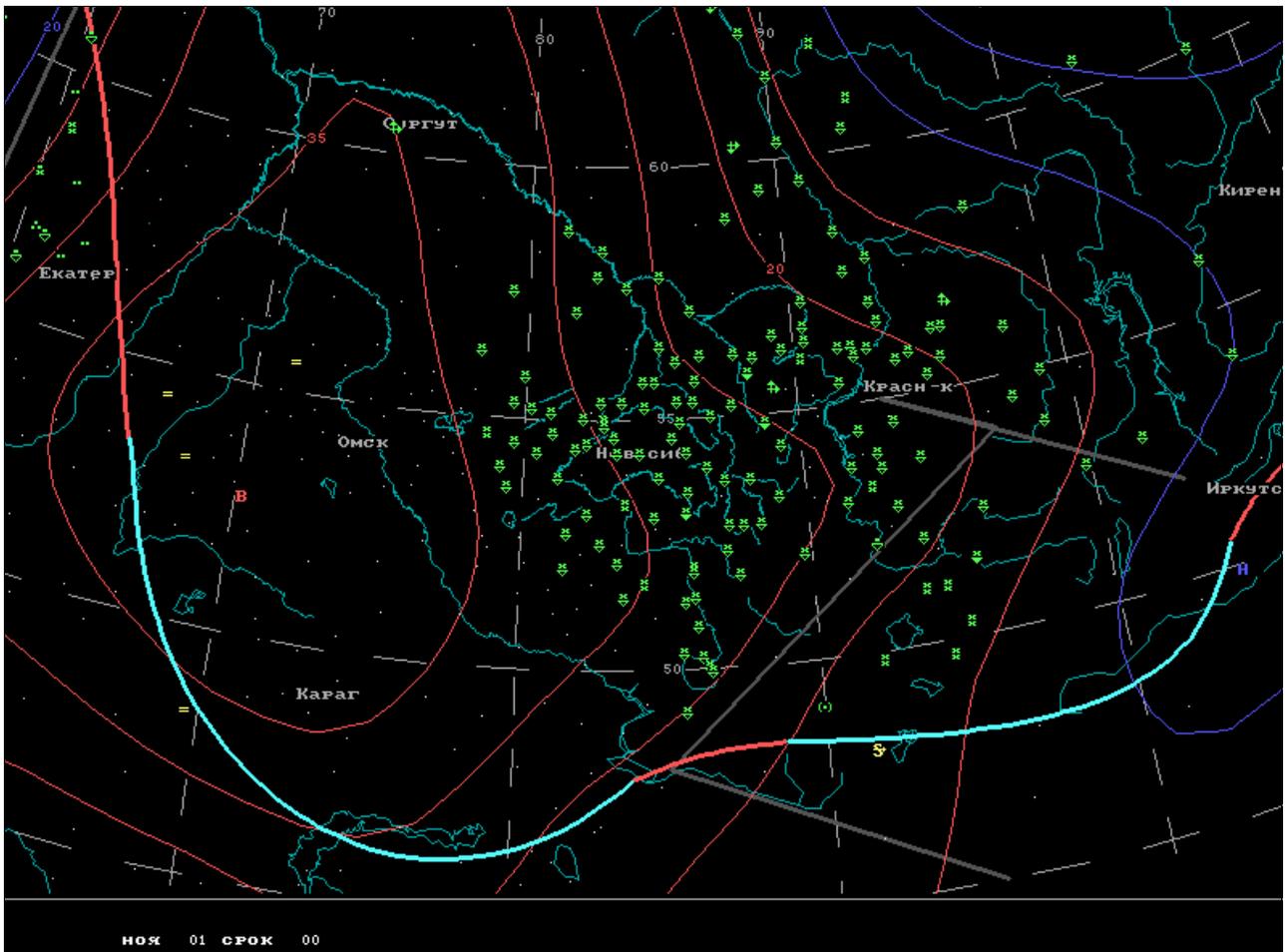
The Generic Mapping Tools (**GMT**) is free software;
 Copyright (c)1991 - 2011 by Paul Wessel and Walter H. F. Smith

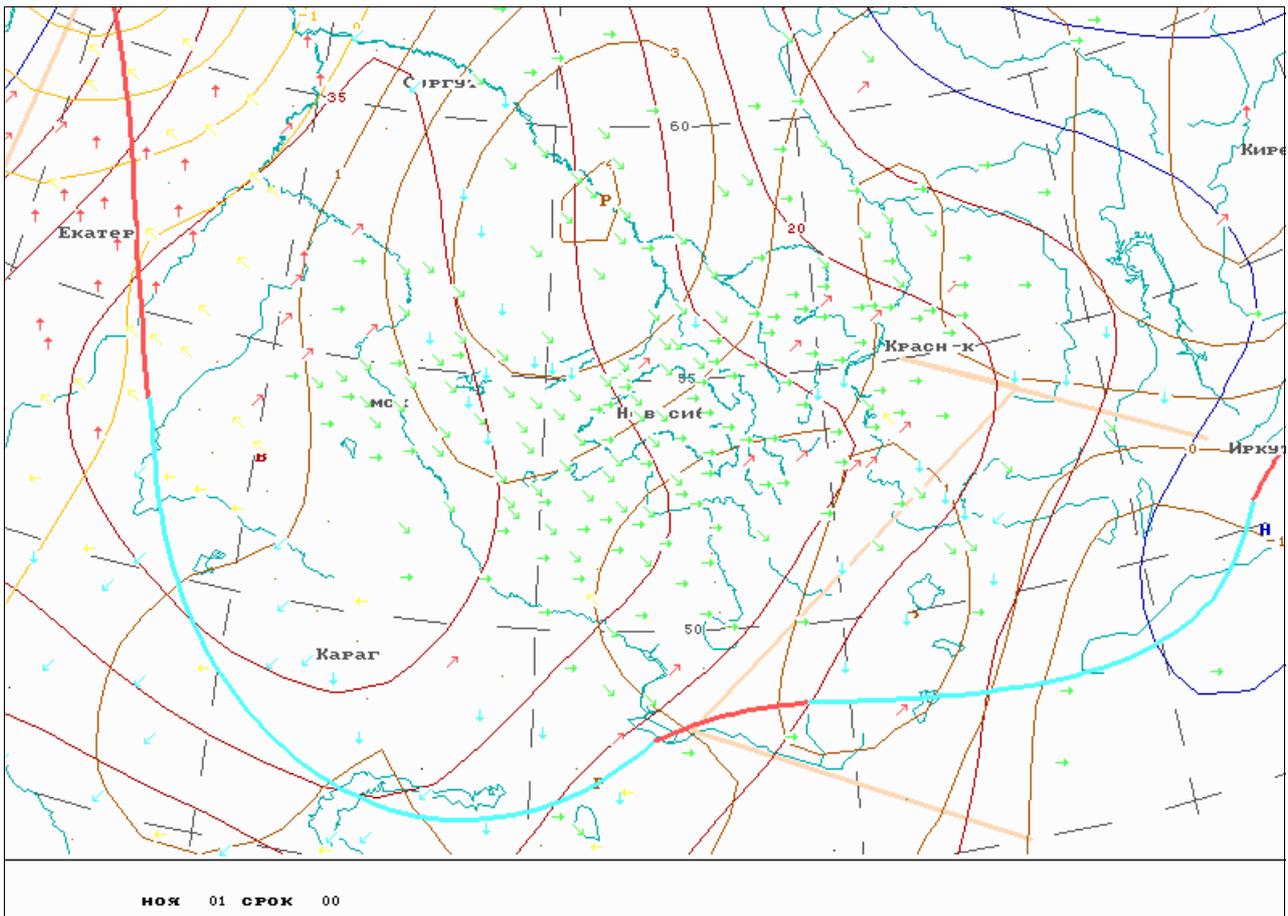
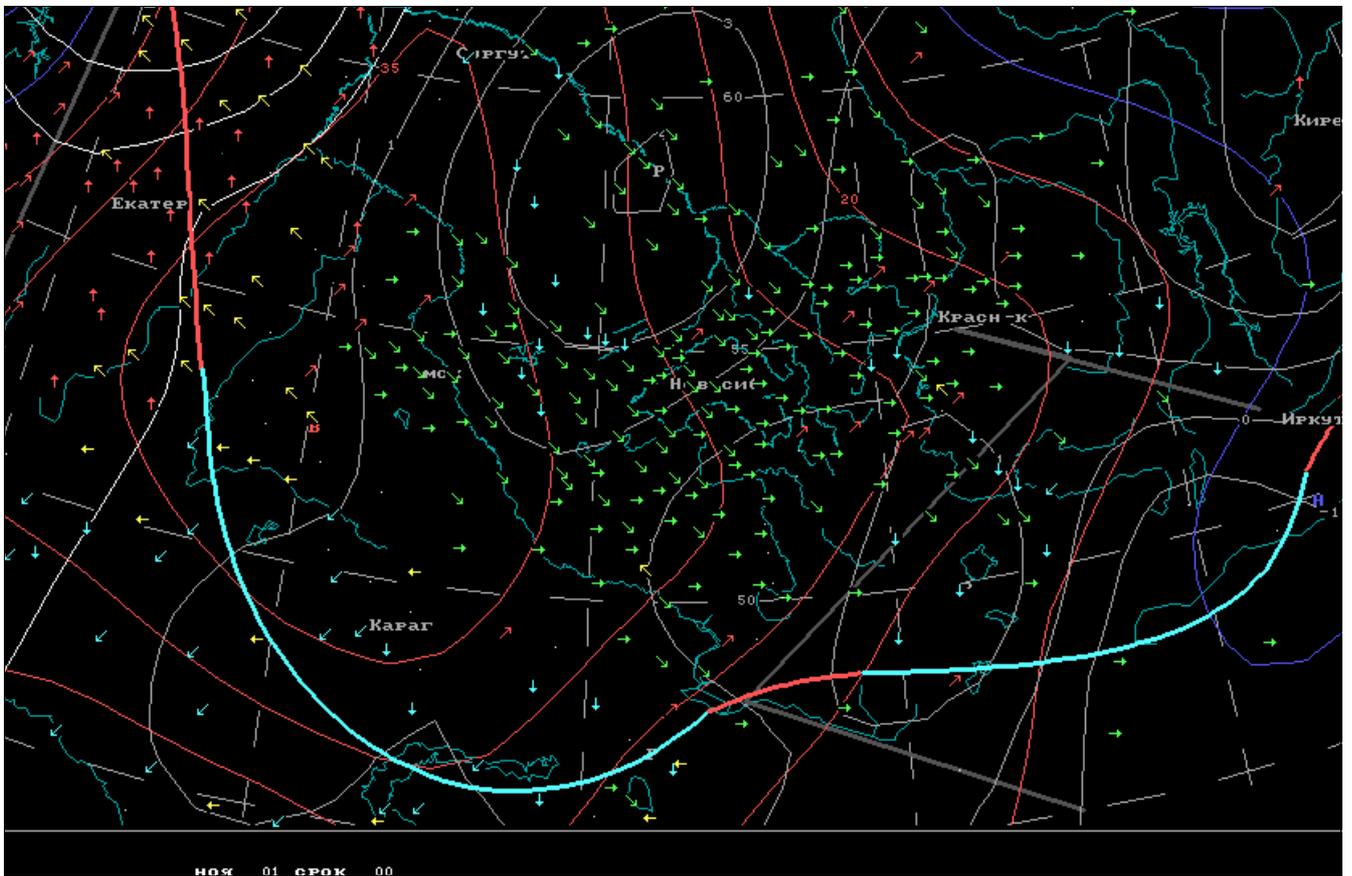
Слайны на функции Грина.

Василенко В.А. Слайн-функции: теория, алгоритмы, программы.-

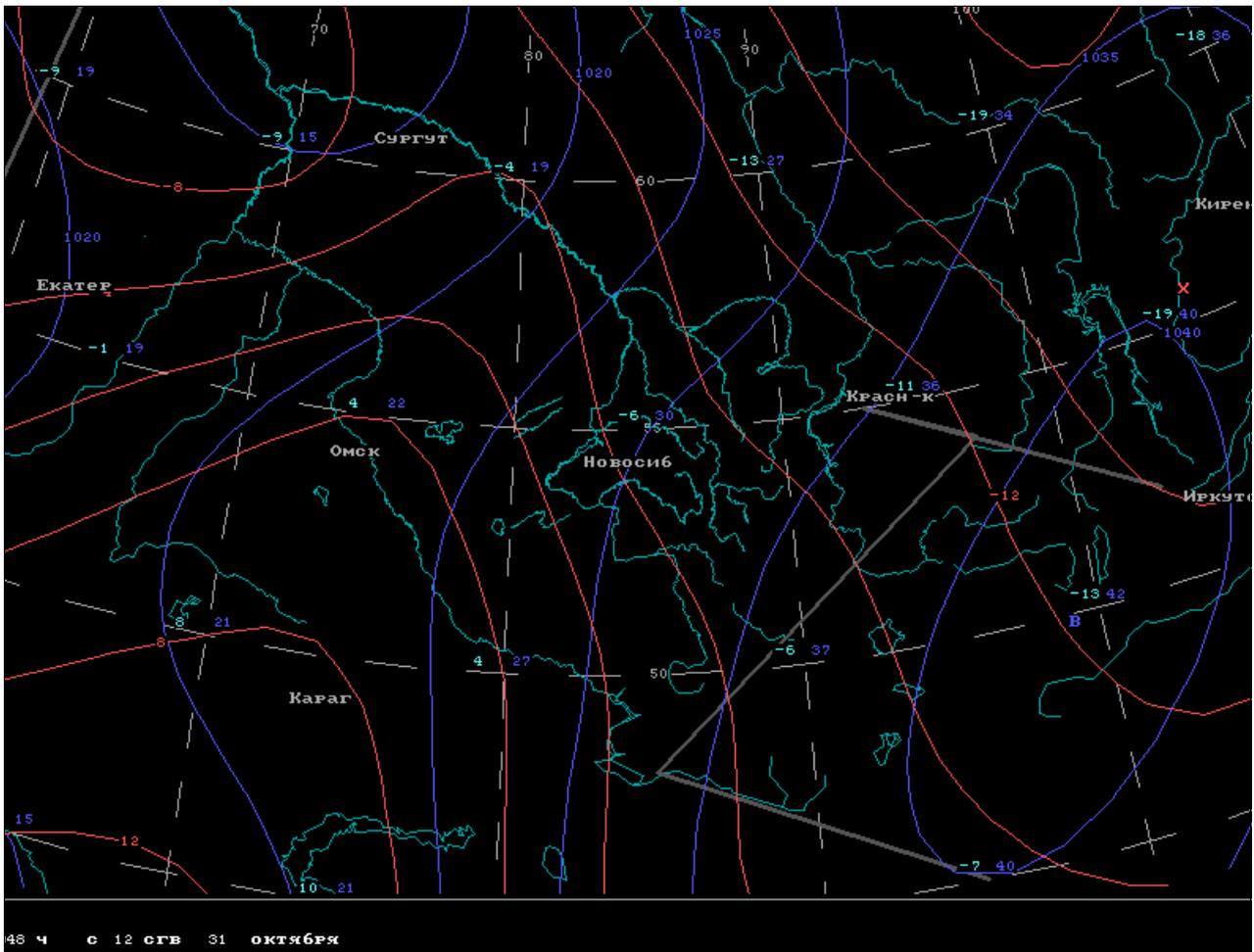
Профессиональные карты, проект "Контур -> Аеро ->":







проект "Аеро-TV":



Прогностическое поле на пятницу 02 ноября (с 12 сгв 31 октября на 048 ч)



Прогностическое поле на субботу 03 ноября (с 12 сгв 31 октября на 072 ч)



Прогностическое поле на воскресенье 04 ноября (с 12 сгв 31 октября на 096 ч)



Прогностическое поле на понедельник 05 ноября (с 12 сгв 31 октября на 096 ч)

