

Благовещенский Ю.Н.

**СТАТИСТИЧЕСКИЕ ПРОБЛЕМЫ
ИССЛЕДОВАНИЯ ДАННЫХ ПО
COVID-19 В РАЗНЫХ СТРАНАХ**

Адрес электронной почты:

vkbun@yandex.ru

Комментарии к слайдам: [comment-stat-problems-of-COVID-19.pdf](#)

Главные задачи статистического анализа данных о распространении COVID-19 в разных странах:

- 1. Поиск индикаторов эффективности актов противодействия распространению эпидемии.**
- 2. Исследование динамики в разные периоды развития пандемии (экспоненциальный рост, окрестность максимума, период скатывания и особенности «мерцающего» затухания).**

Источники статистических данных

Открыв ссылку <https://indem.ru/russian.asp> вы откроете основную страницу сайта Фонда ИНДЕМ и увидите рисунок:



Щелкнув мышкой, вы перейдете на посвященную COVID-19 страницу сайта. В нижней части страницы находится раздел «**Источники данных об эпидемии COVID-19**».

Это официальная статистика ВОЗ, данные Роспотребнадзора, сайт **Johns Hopkins University**, портал открытых данных ЕС и ещё 5 источников, которые интересны каждый по своему.

Измеряемые характеристики

Для любого объекта наблюдения, будь это мир, страна, город или другая территория, собираются данные за одни сутки о числе на дату t

(1) инфицированных, $N(t)$,

(2) умерших, $D(t)$,

(3) выздоровевших, $R(t)$.

Кроме этого собираются данные о числе тестов на COVID-19, о числе коек и другие показатели, характеризующих состояние медицины.

В этом докладе анализ данных базируется только на первых трёх показателях.

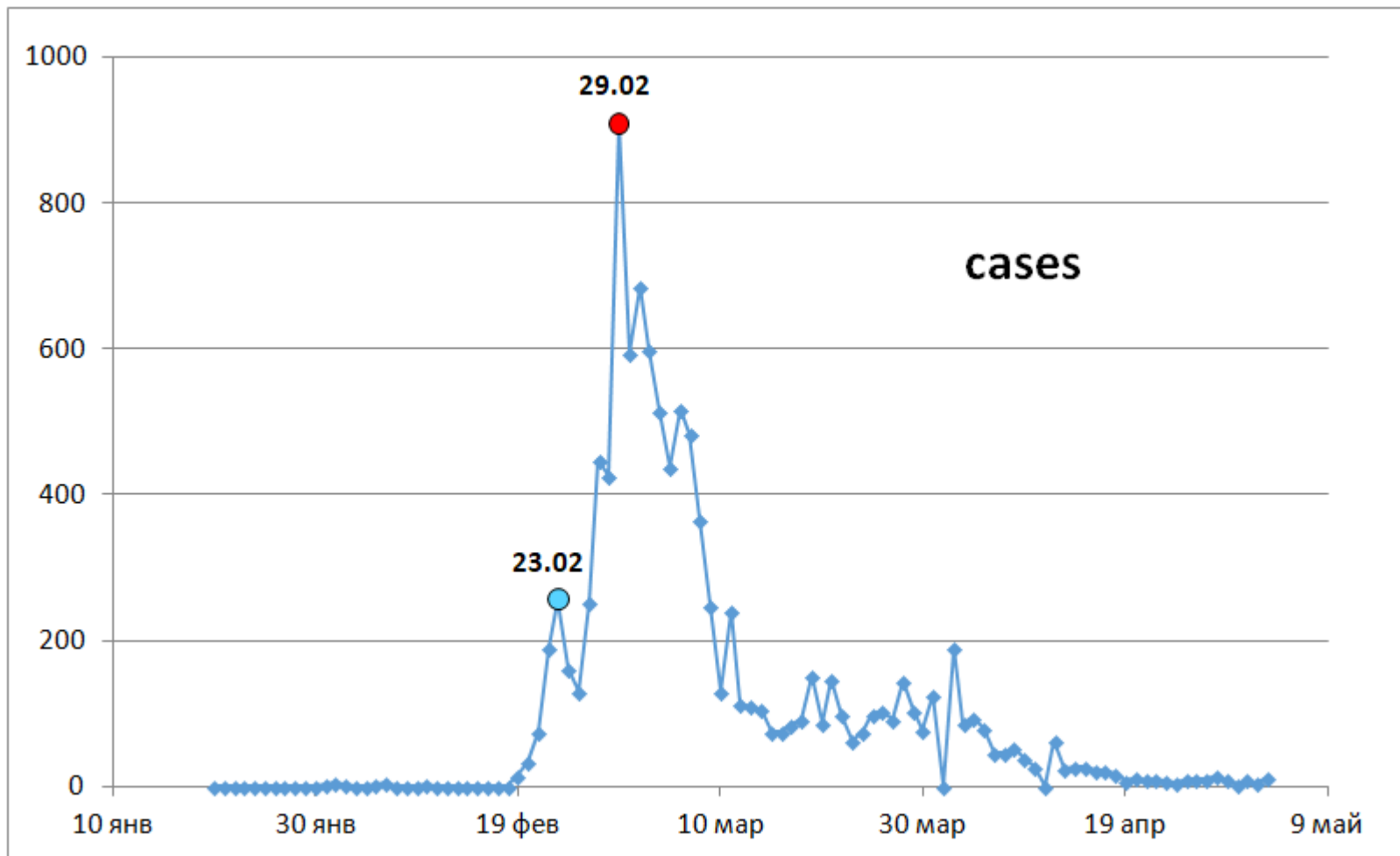
ПРЕДУПРЕЖДЕНИЕ

Во время пандемии за сутки может происходить тысячи, десятки тысяч заражений.

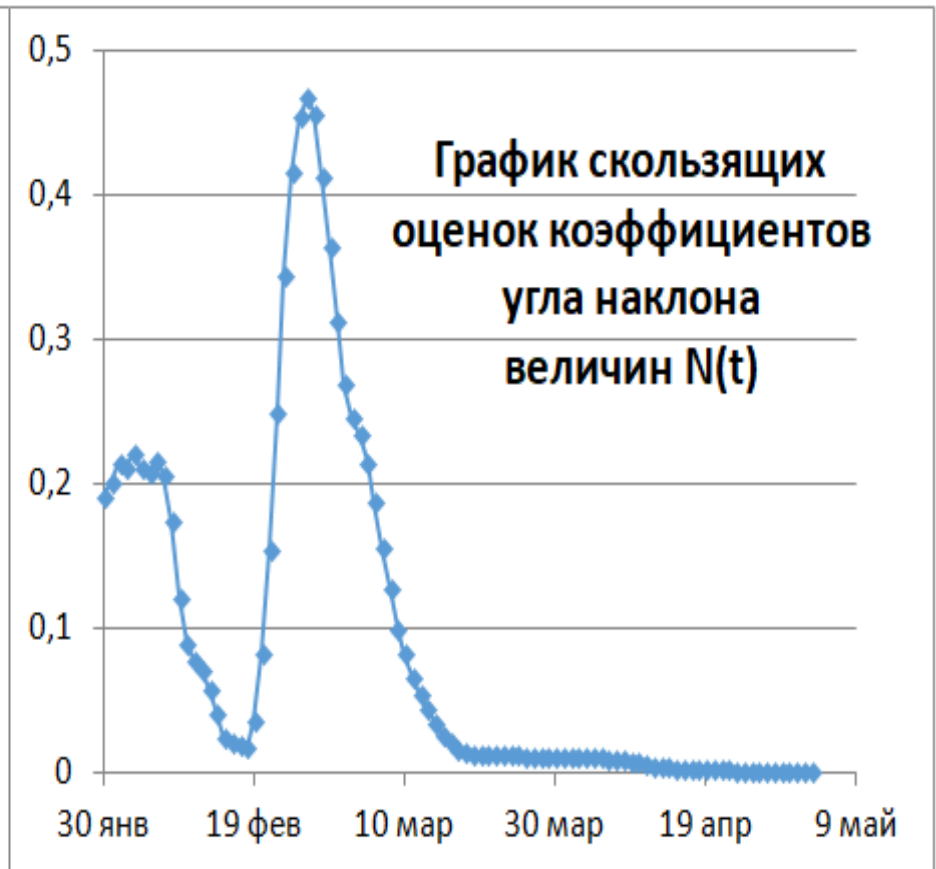
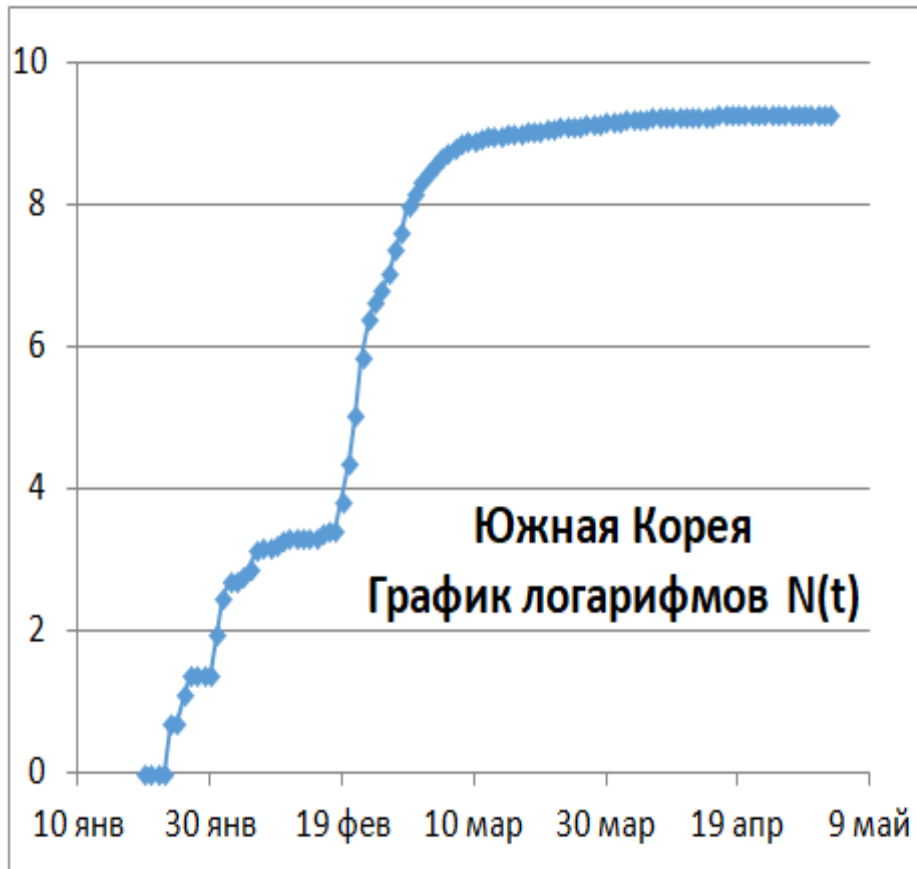
И в разных источниках данные могут существенно различаться из-за того, что данные относились к разному времени суток (по Москве), а иногда из-за трудно объяснимых ошибок (этим грешат и данные от университета Джонса Хопкинса).

Такое рассогласование надо учитывать, когда работаешь с разными источниками и странами.

Индикаторы борьбы с COVID-19

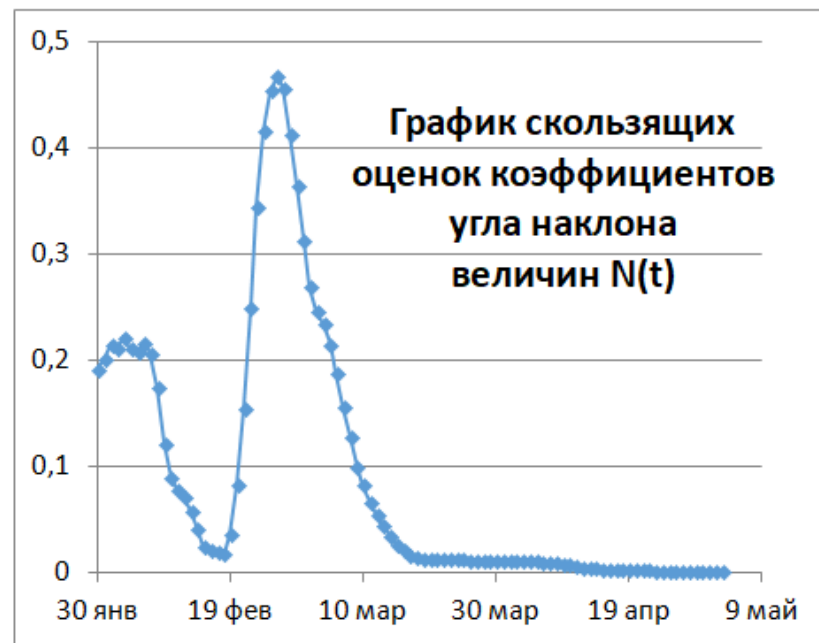
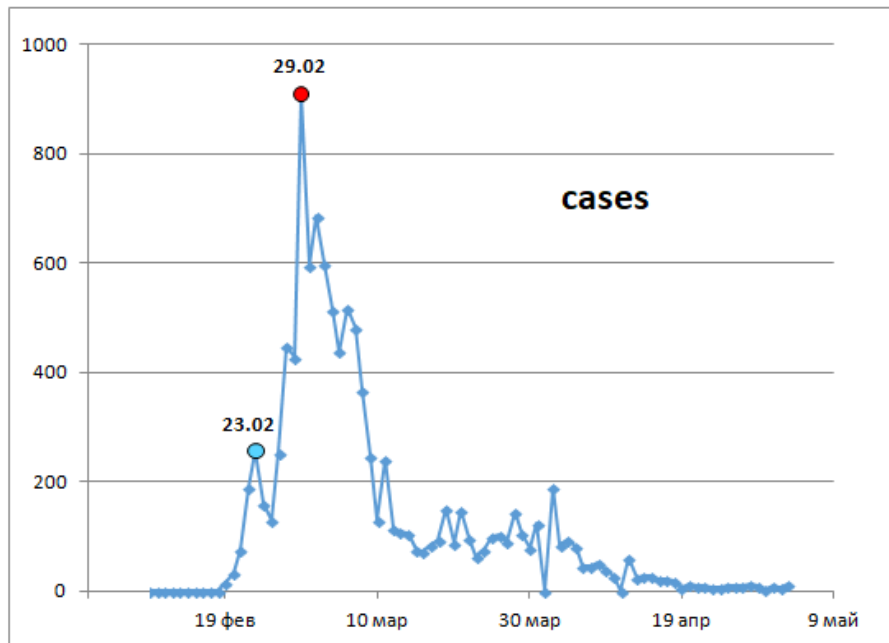


Это посуточное число выявляемых зараженных COVID-19 в Ю. Корее (20.01-03.05).



**Натуральные логарифмы
Накопленного числа
инфицированных**

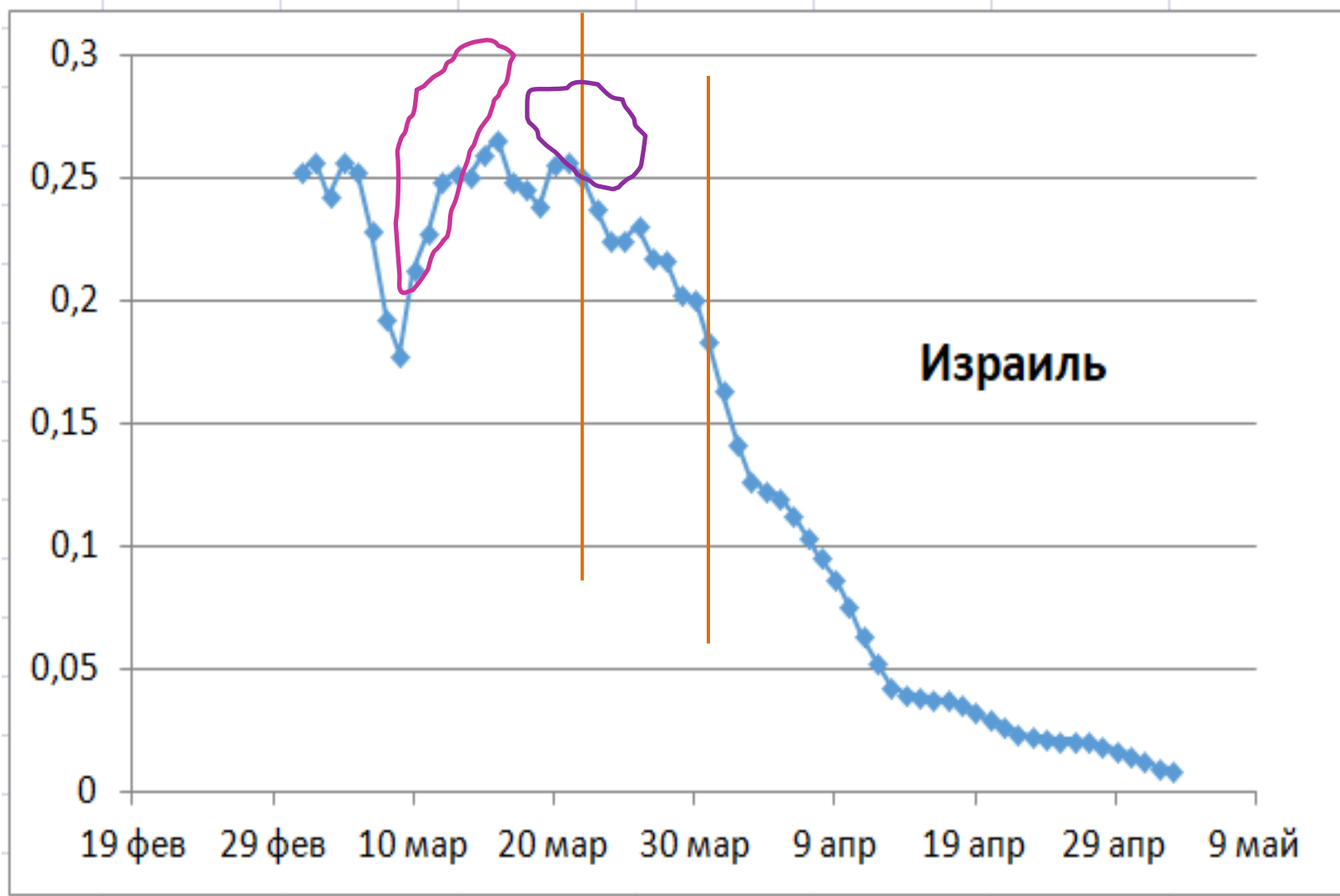
**Наклон касательной к
левой кривой (оценка
по 10 датам до неё)**

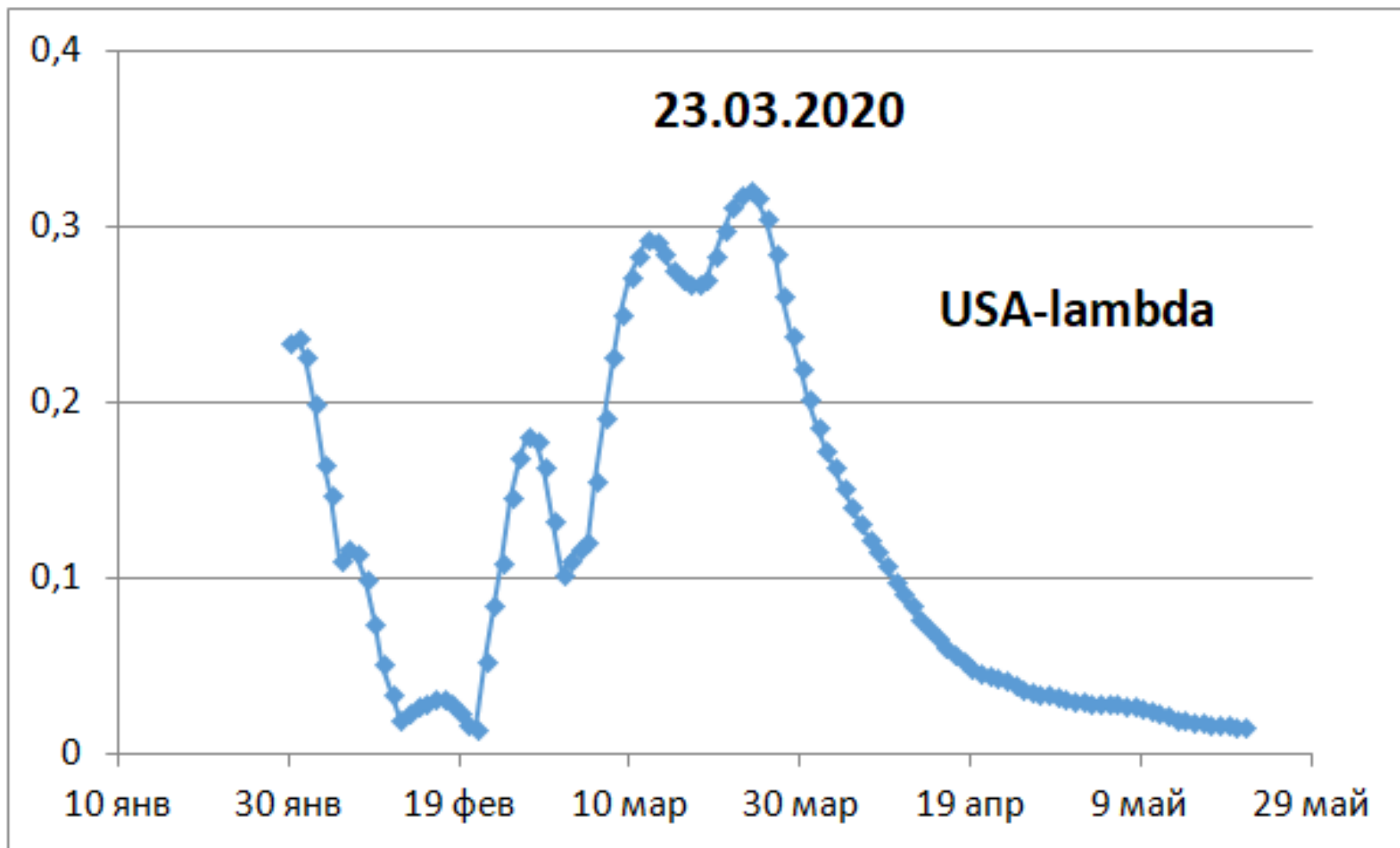


Эти два графика чуть-чуть похожи, но для правого

- (1) вычисляемые величины безразмерны,**
- (2) инвариантны к выбору масштабов, что дает право сравнивать разные страны между собой,**
- (3) концентрируют в себе влияние всех событий, которые происходят в предшествующие 10 дней по отношению к наблюдаемому моменту времени.**

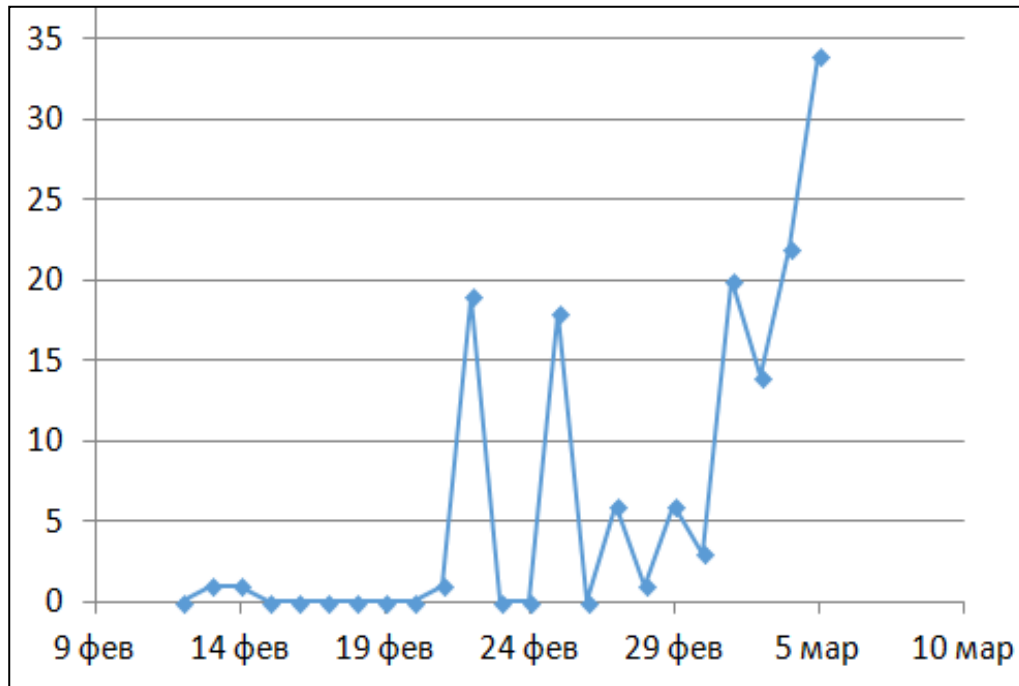
Работоспособность метода скользящих окон





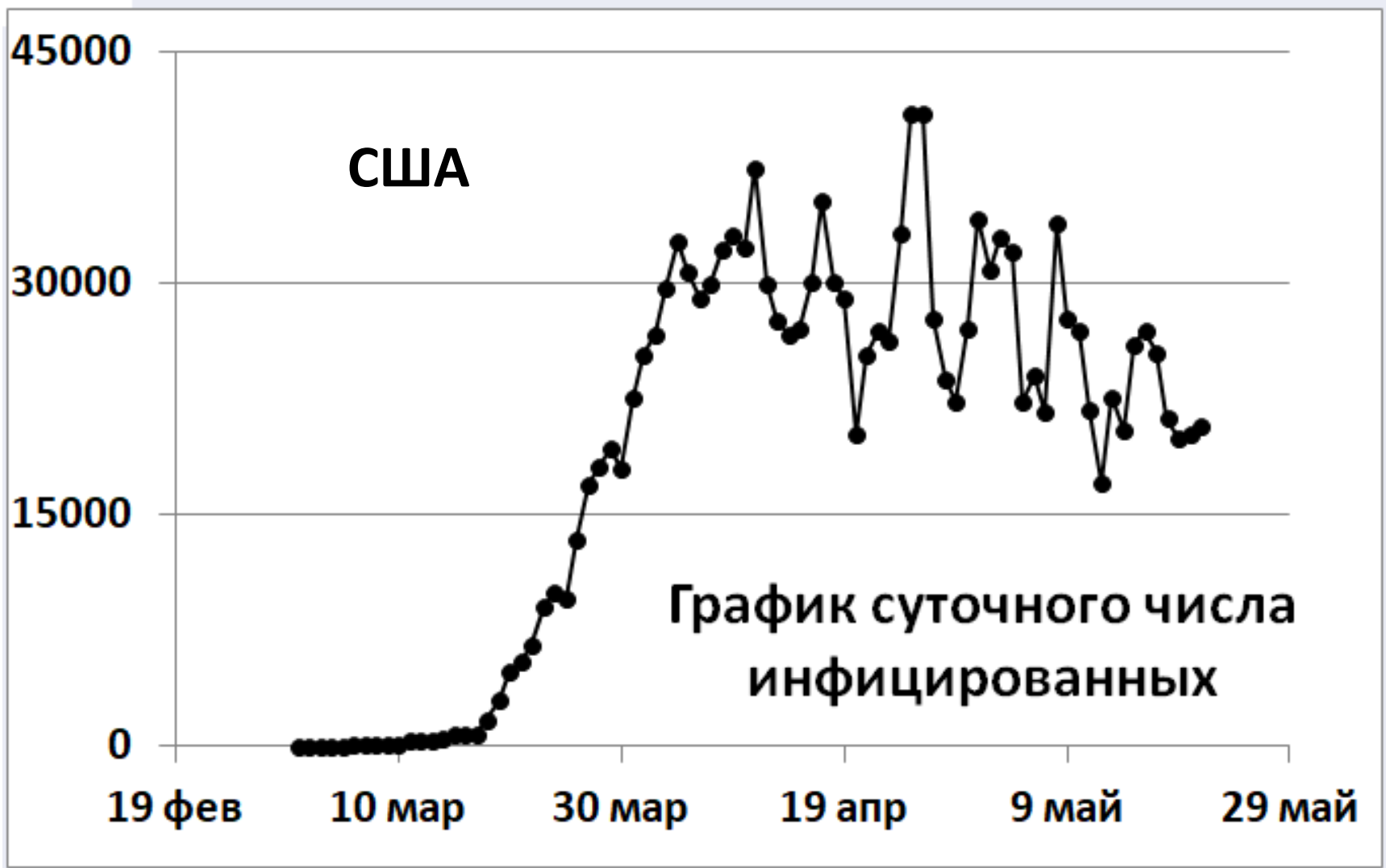
Даты вершин: 27 февраля, 12 и 23 марта

Тут помогает схожесть графиков России и США с точки зрения холмистости и федерального устройства.



5 марта
Накоплено
159
зараженных

Февраль	21	22	23	24	25	26
Было	15	16	35	35	35	53
Добавок	1	19	0	0	18	0



В отличие от графика на слайде 10 здесь нет точек индикации, метящих действия власти

Этап пандемии

Пусть $t = t_0, t_0 + 1, t_0 + 2, \dots$ – даты,
 $n(t)$ – число зараженных за сутки,

$$N(t) = \sum_{s=t_0}^t n(s),$$

t_0 – дата появления 1-го зараженного.

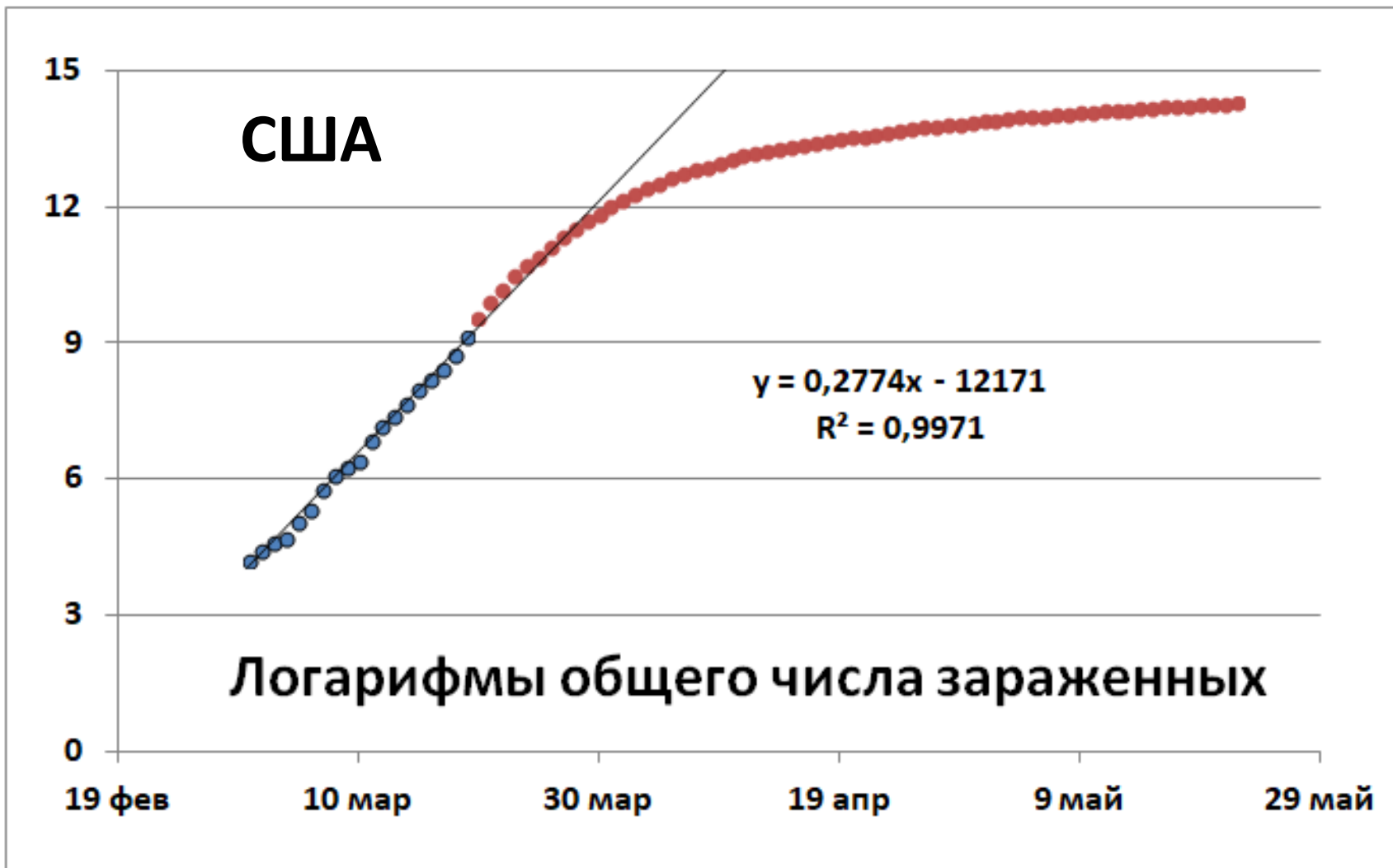
Критерий пандемии

Это – приближенное равенство:

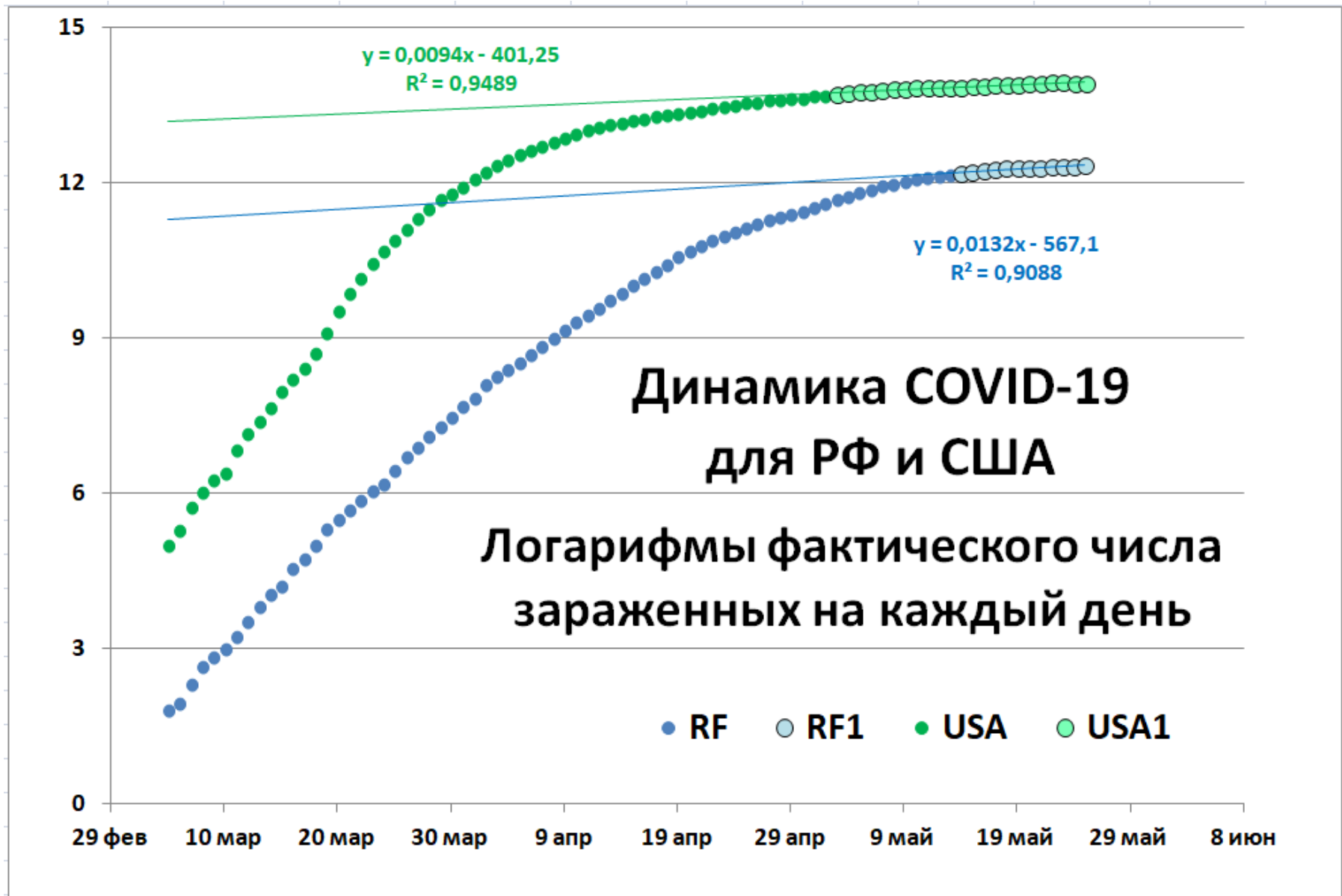
$$n(t + 1) \approx g \cdot N(t),$$

из которого уже следует, что

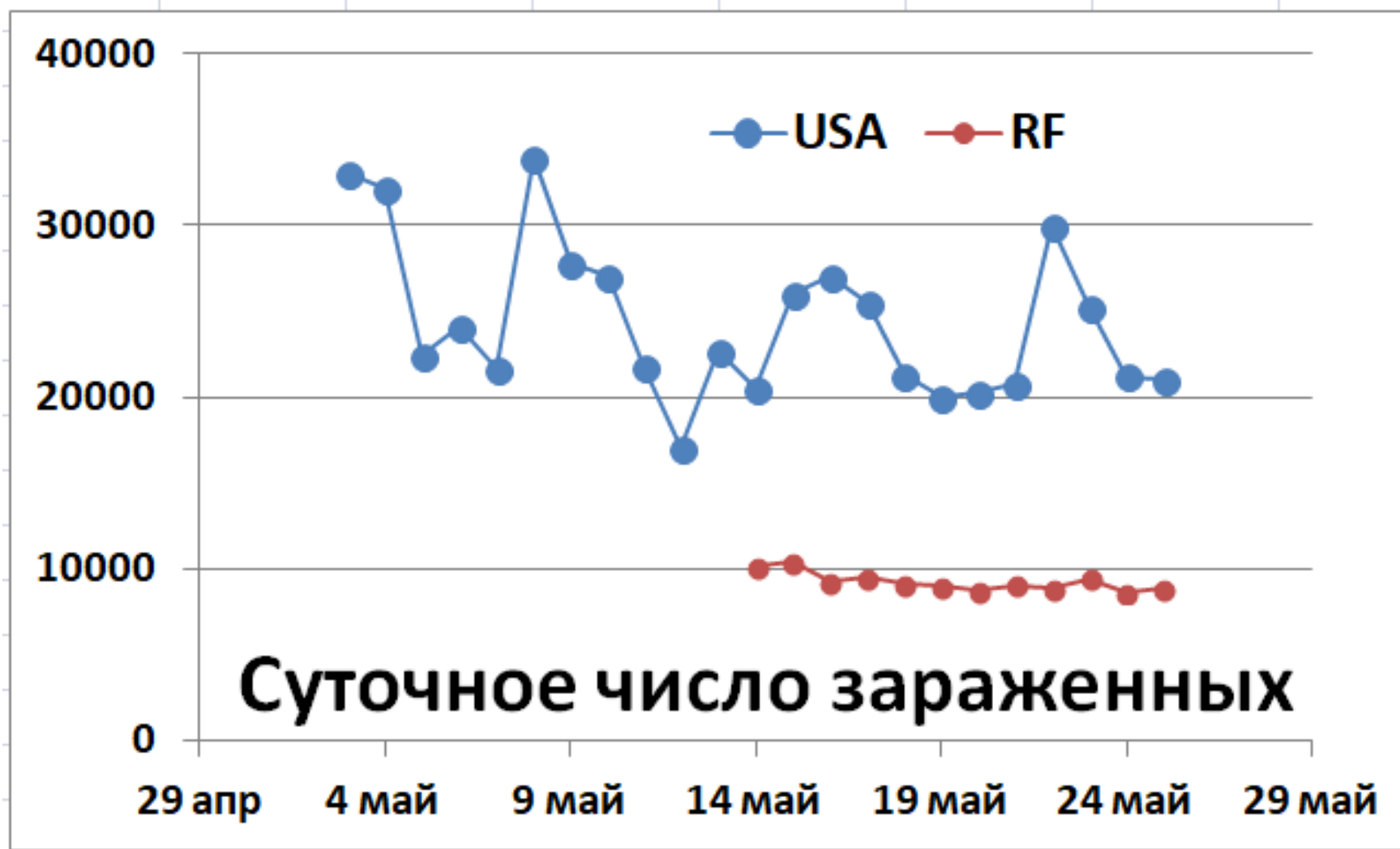
$$N'(t) \approx \frac{\Delta N(t)}{\Delta t} = N(t + 1) - N(t) \approx g \cdot N(t).$$



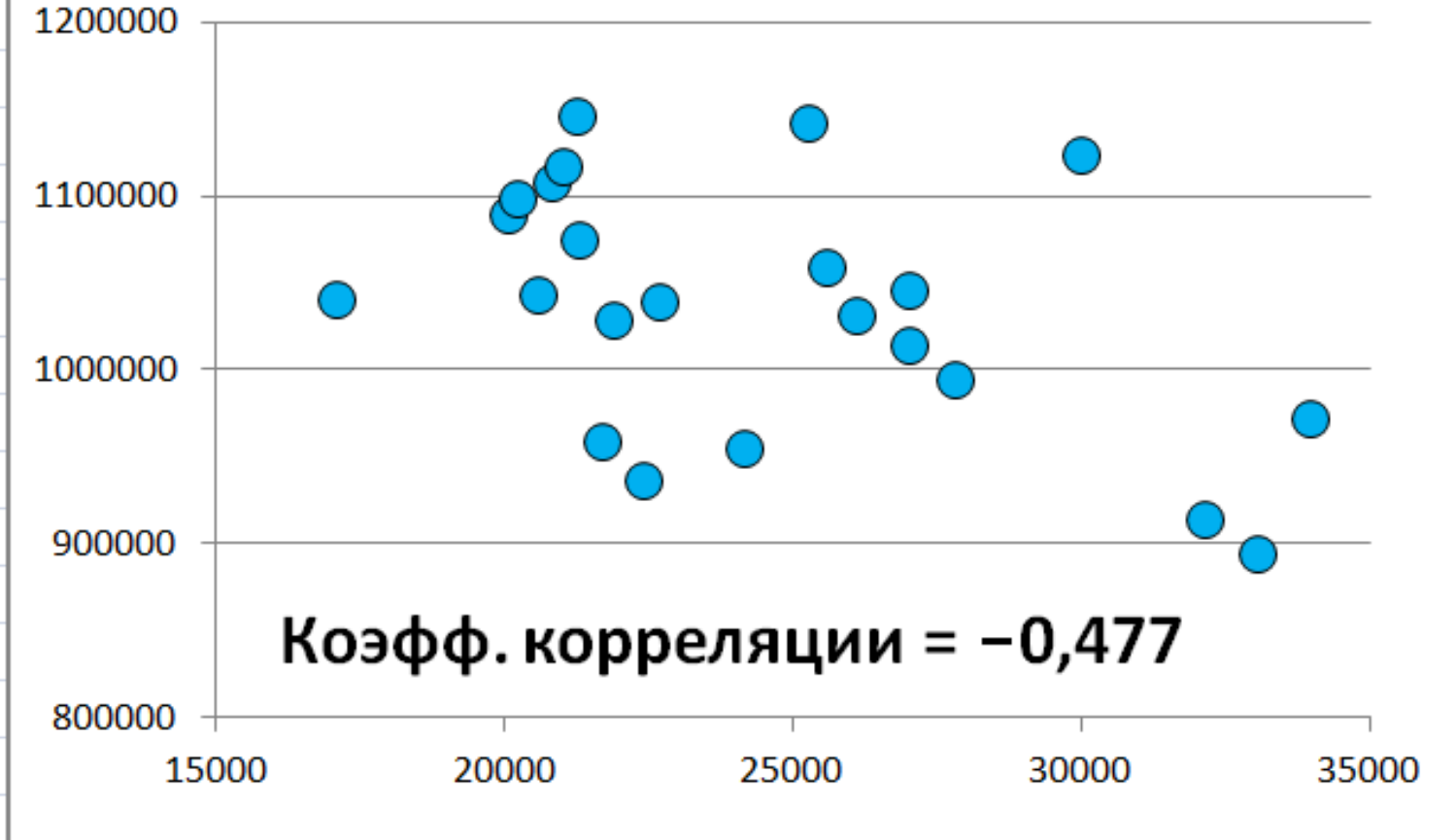
**Участок экспоненты (пандемия в США),
 $g = 0,2774$**



Анализ вблизи плато

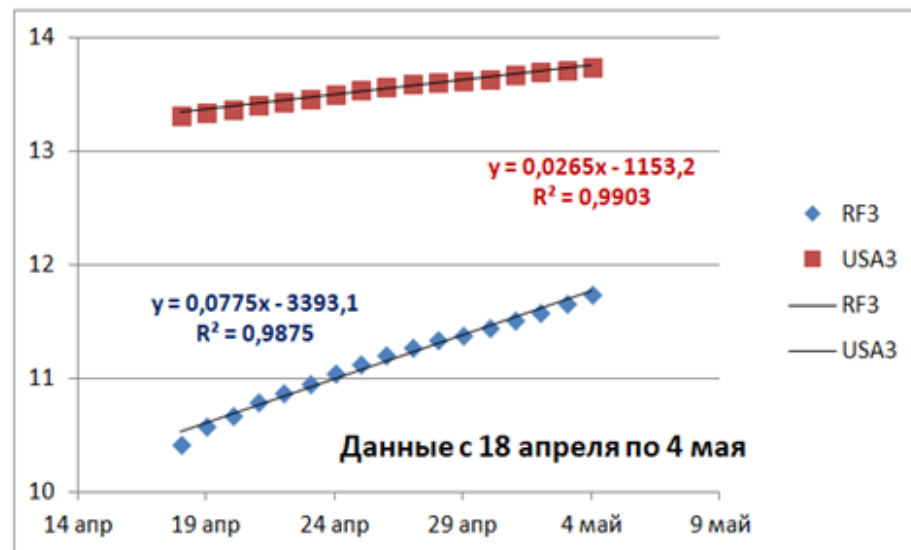
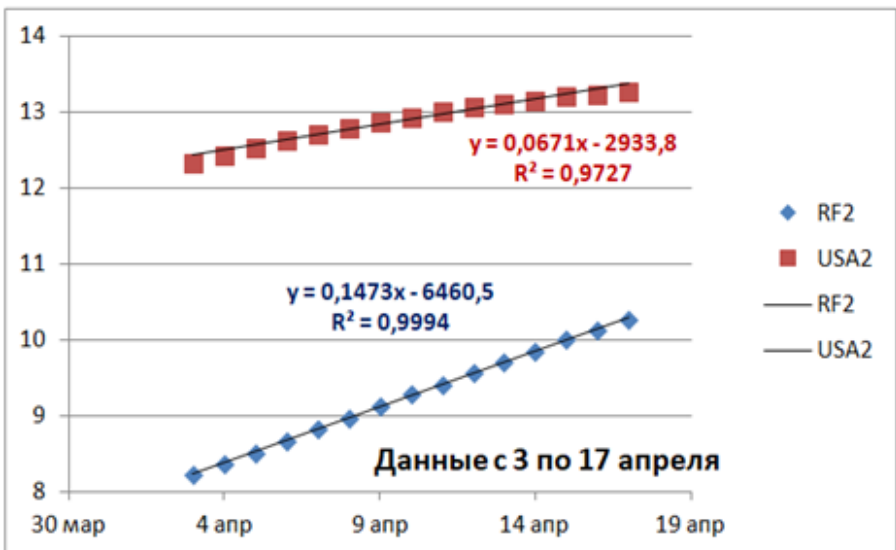
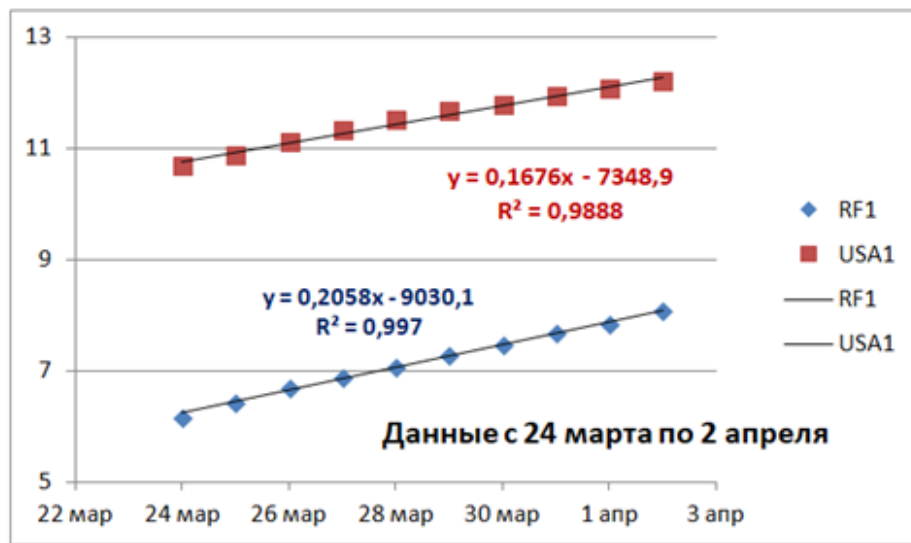
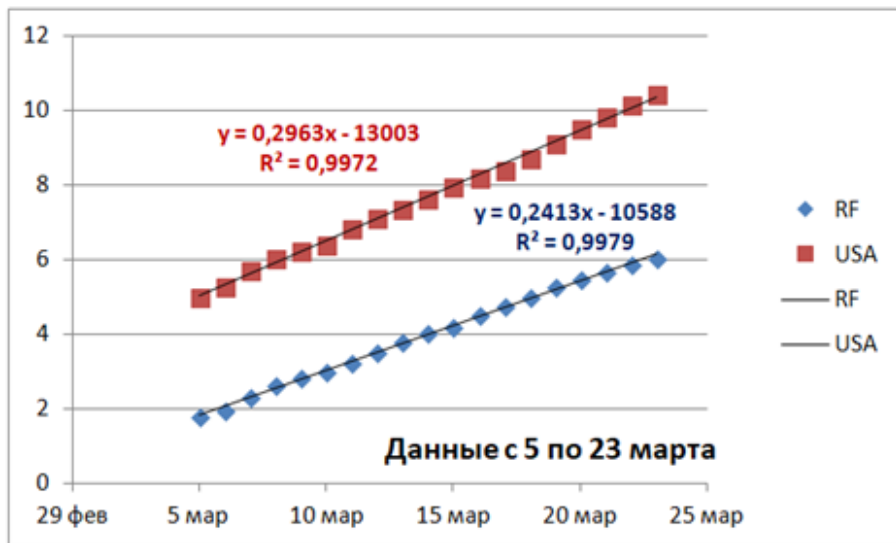


**Диаграмма рассеяния точек:
(добавок; всего инфицированных)**



США: 3-25 мая: точки $(n(t+1); N(t))$

Прогнозы максимума заражений

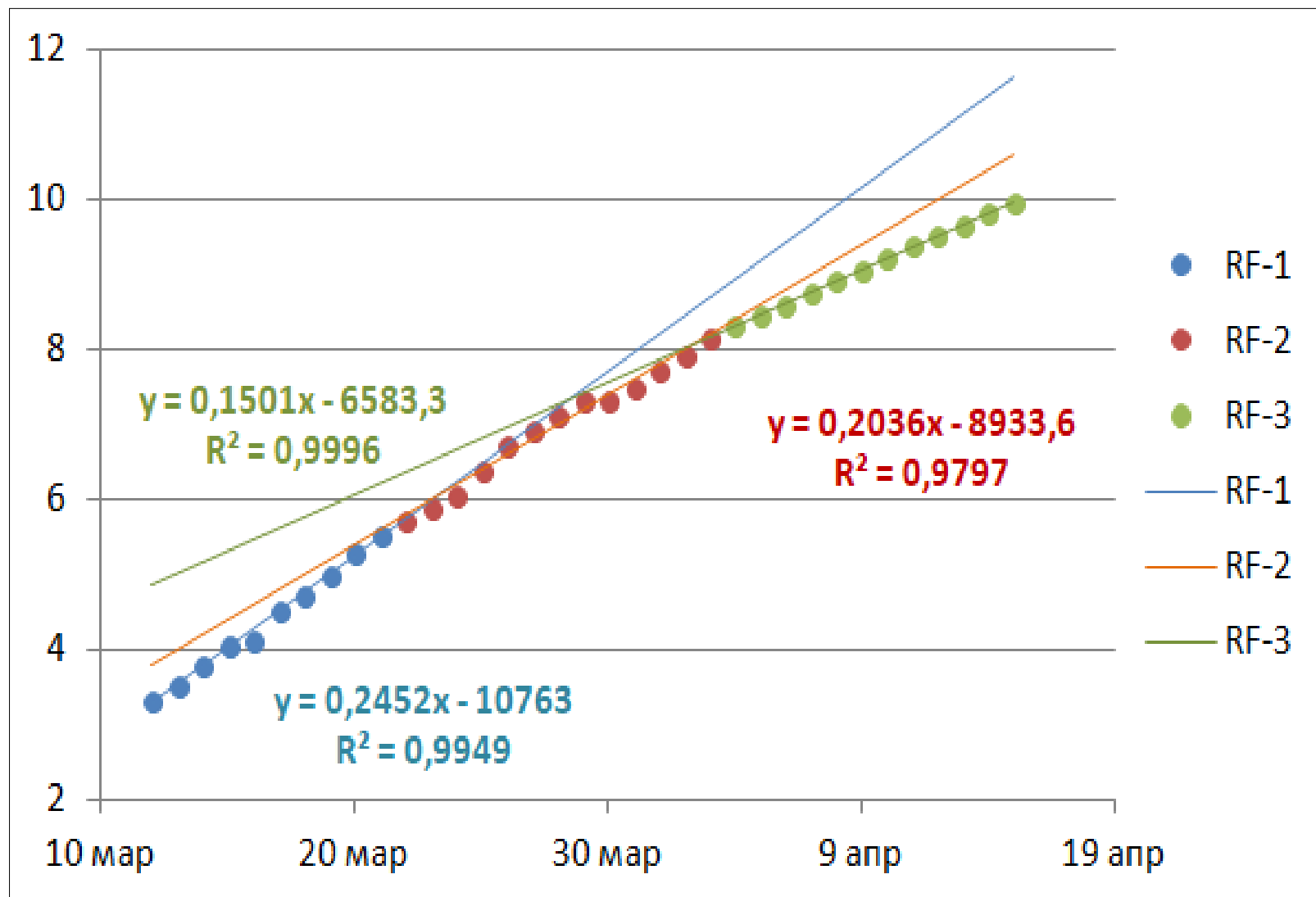


Результаты аппроксимации

Интервал дат	Коэфф. g		Коэфф. R ²	
	США	РФ	США	РФ
5-23 марта	0,296	0,241	0,997	0,998
24.03 – 02.04	0,168	0,206	0,989	0,997
3-17 апреля	0,067	0,147	0,973	0,999
18.04 – 04.05	0,027	0,076	0,99	0,987

Следствия. (1) В США и РФ пандемия имеет общую природу развития; (2) В США ситуация тяжелей, но действия властей эффективней.

Россия: Прогноз от 16 апреля



Методика прогноза

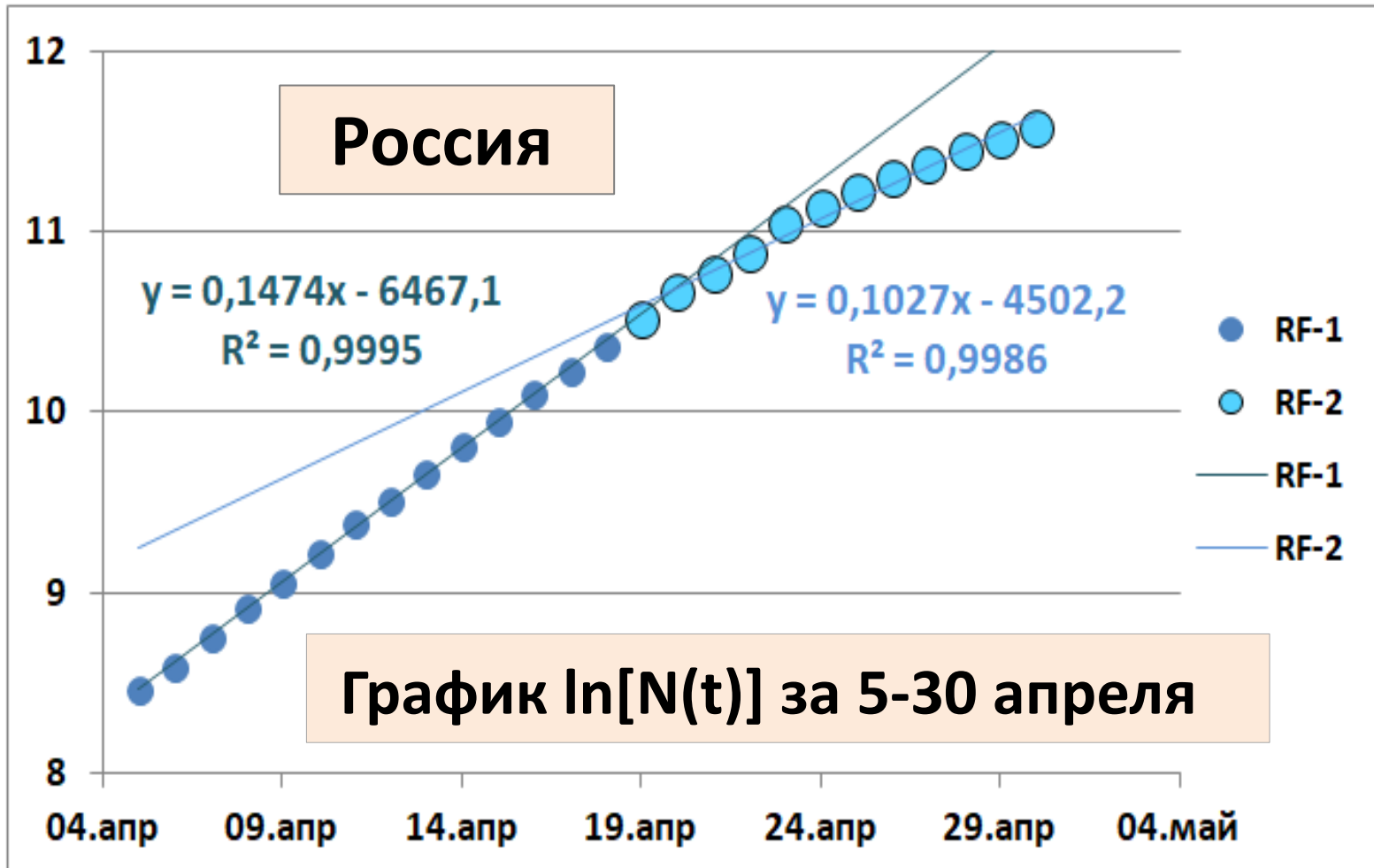
1. Отрезок 12-21 марта (середина принята за 0);
2. 22 марта – 3 апреля (середина равна 11,5);
3. Отрезок 4-15 апреля (середина равна 24).

Коэффициенты углов наклона по датам:

Отрезок	1	2	3
Абсцисса	0	11,5	24
Ордината	0,2452	0,2036	0,1501

... и смотрим, при каком «икс» падение «игрека» достигнет нуля. Это примерно 9 мая.

Другой подход к прогнозу



$$y = (0,1442x + 10,512) - 0,0044x^2$$

Результаты прогнозов

Прогноз максимума	Используемые данные		
	12.03- 15.04	19.04- 30.04	23.04- 30.04
Какой день	9 мая	6 мая	13 мая
Сколько	160 тыс.	120 тыс.	550 тыс.

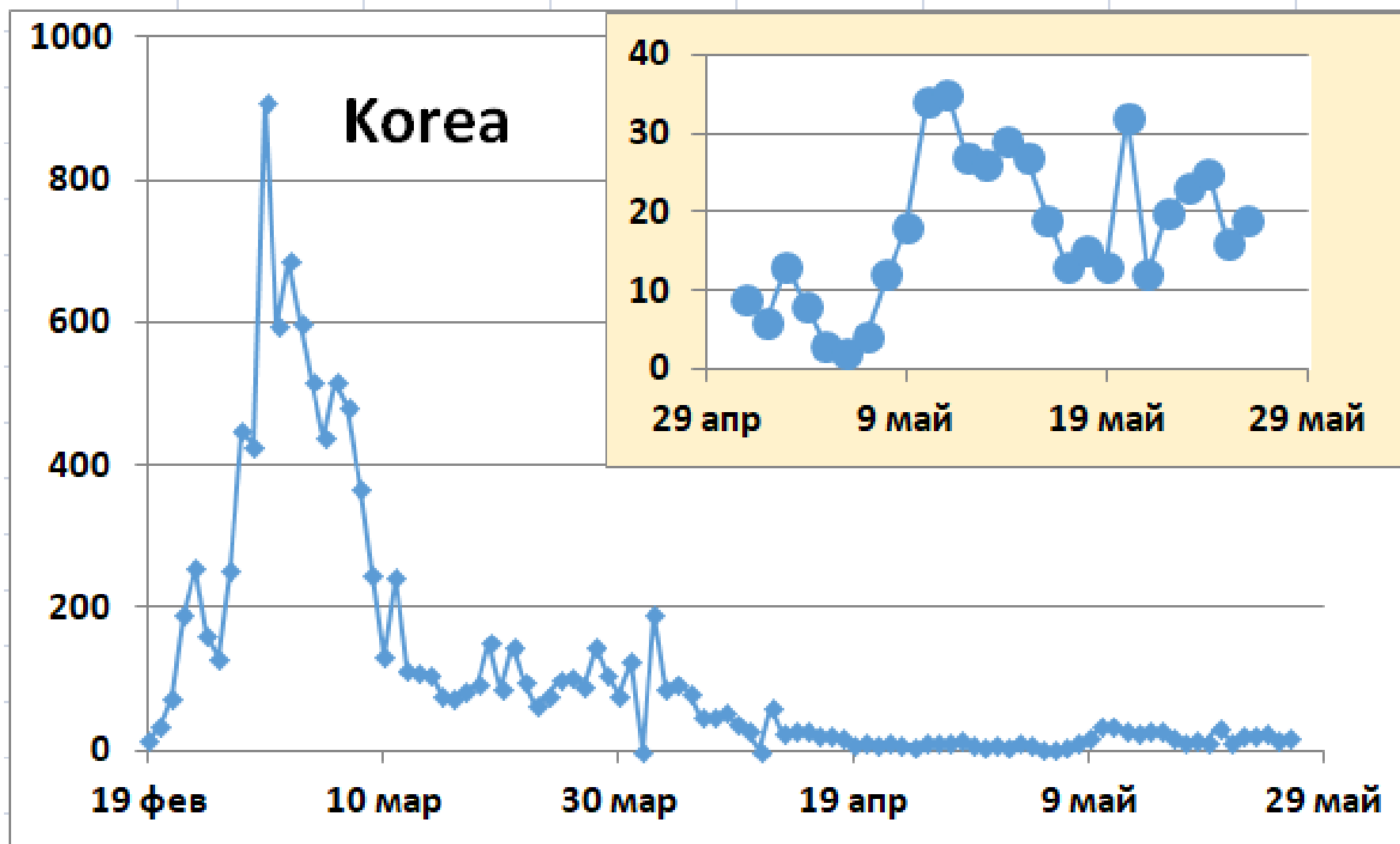
Выставлено 15 мая

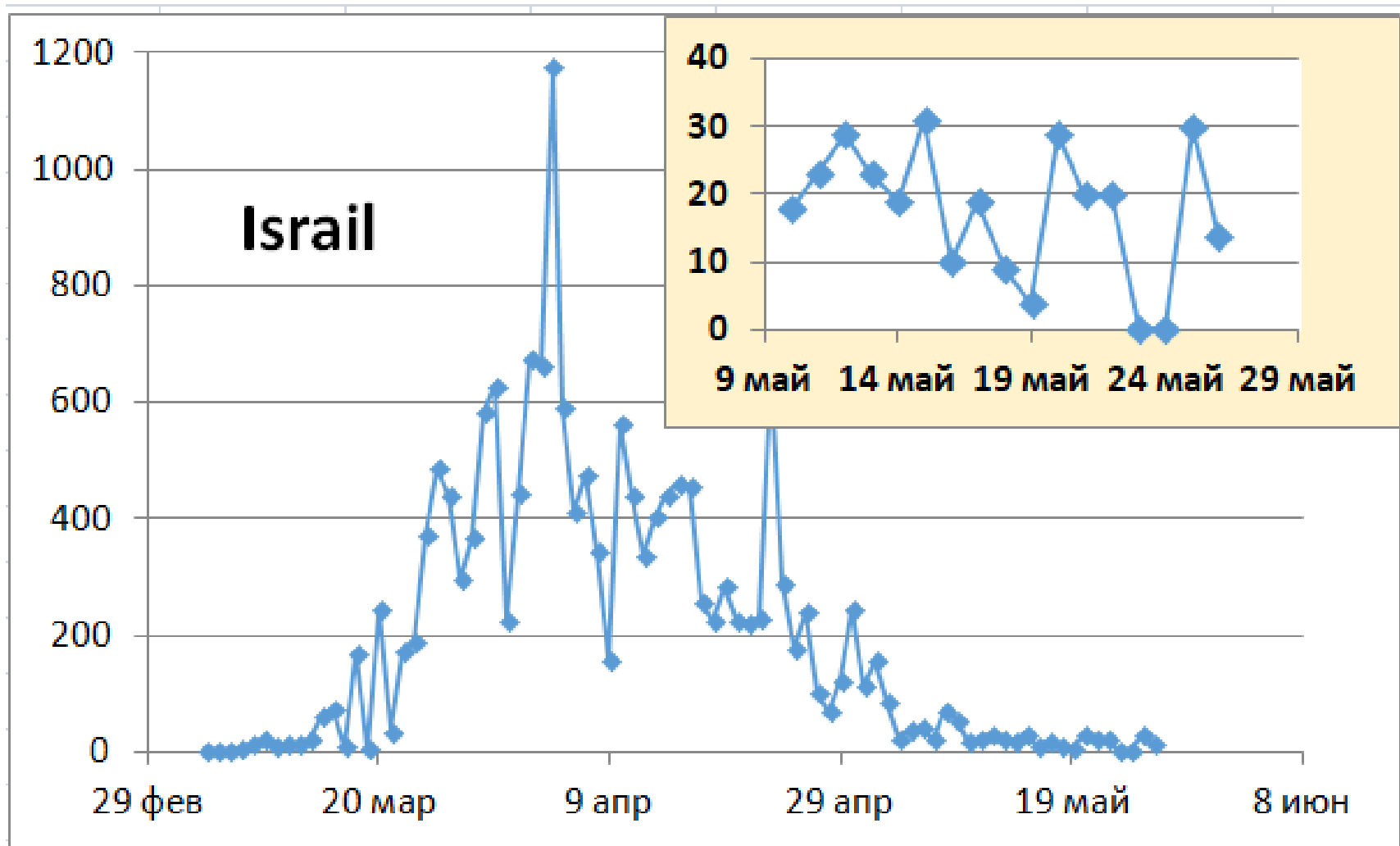
[https://indem.ru/coronavirus/
что-стало-прогноз-ross1.pdf](https://indem.ru/coronavirus/что-стало-прогноз-ross1.pdf)

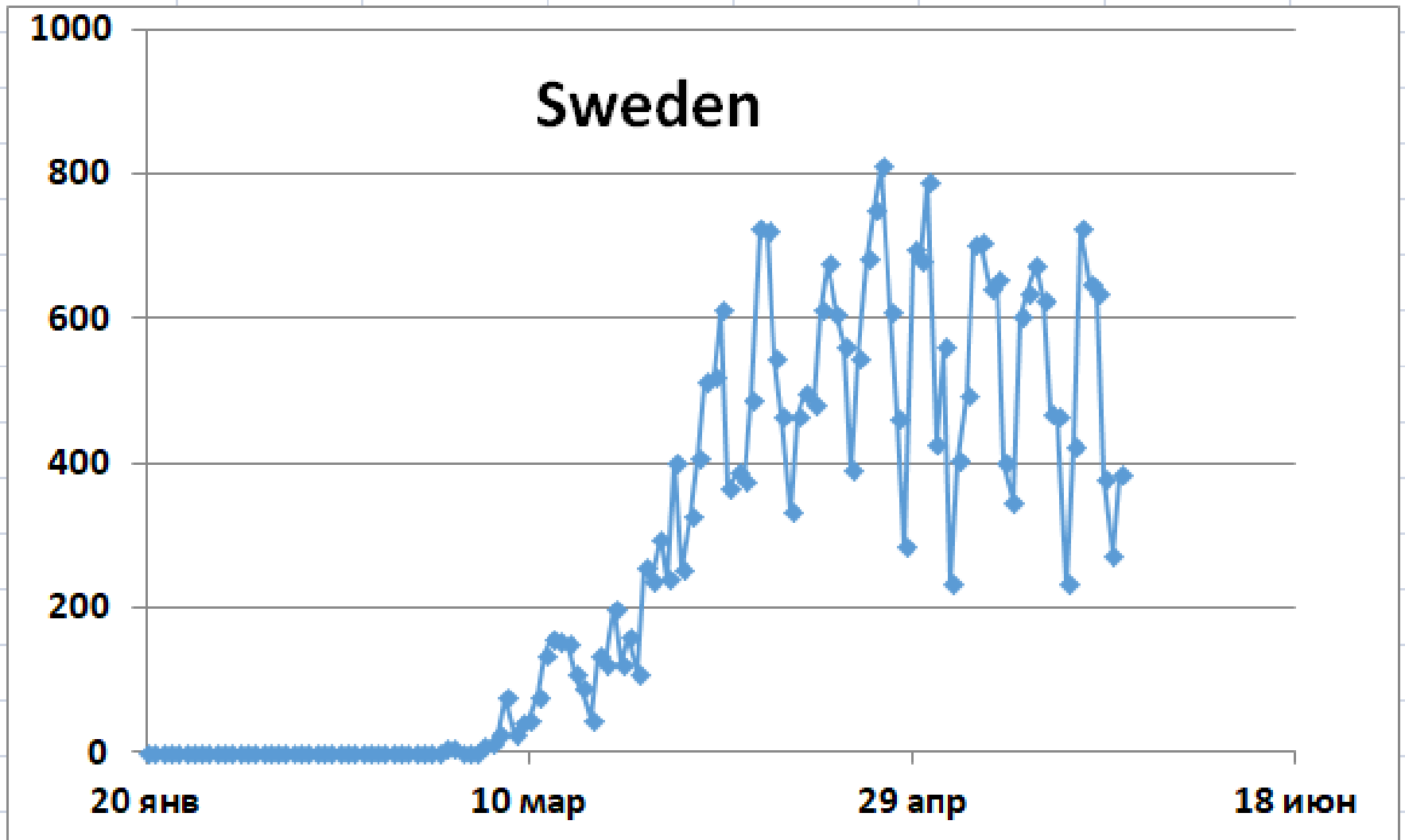


Область максимума: 4-13 мая; накопленное число инфицированных: от 134687 до 232243.

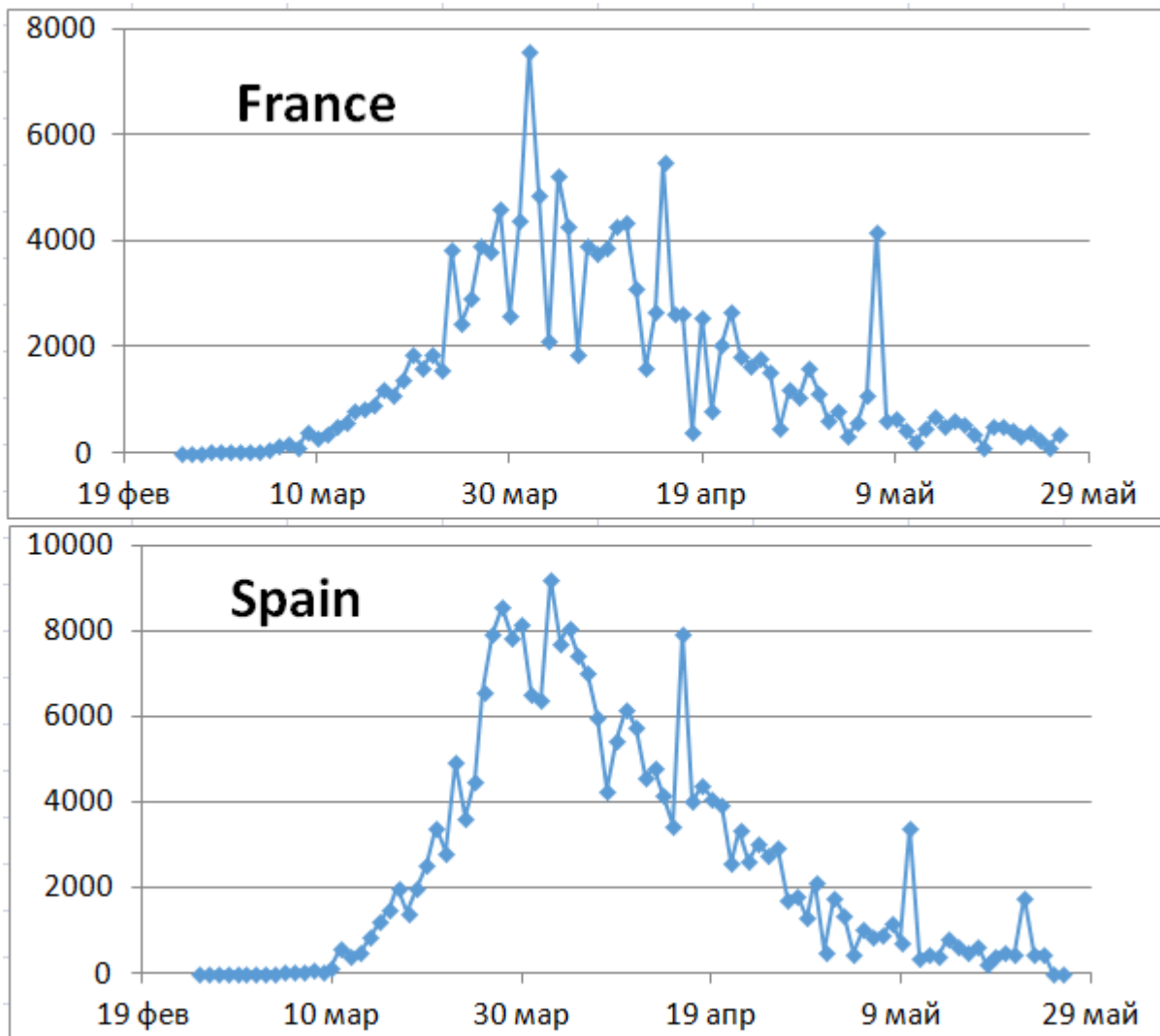
Этап угасания эпидемии



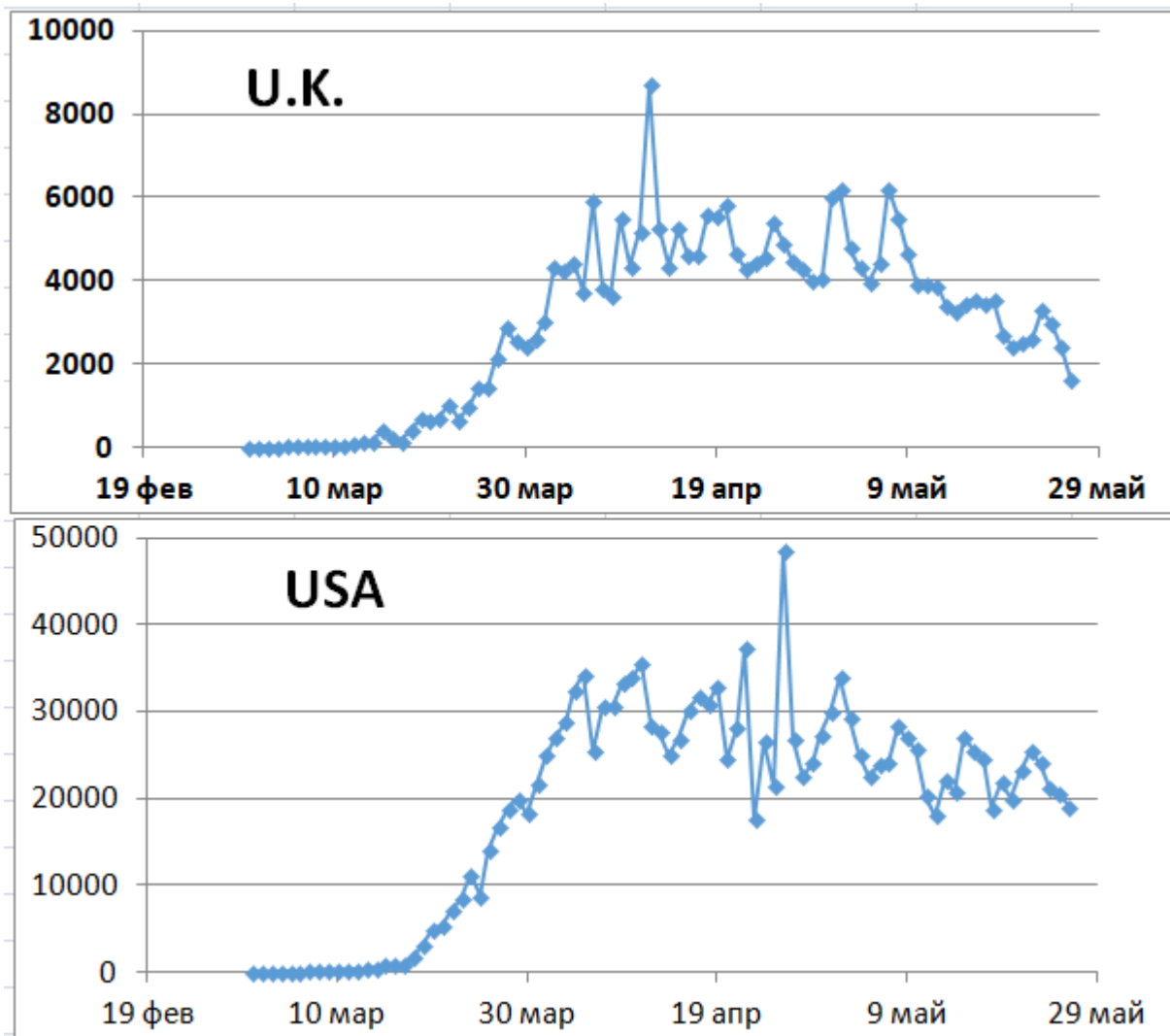




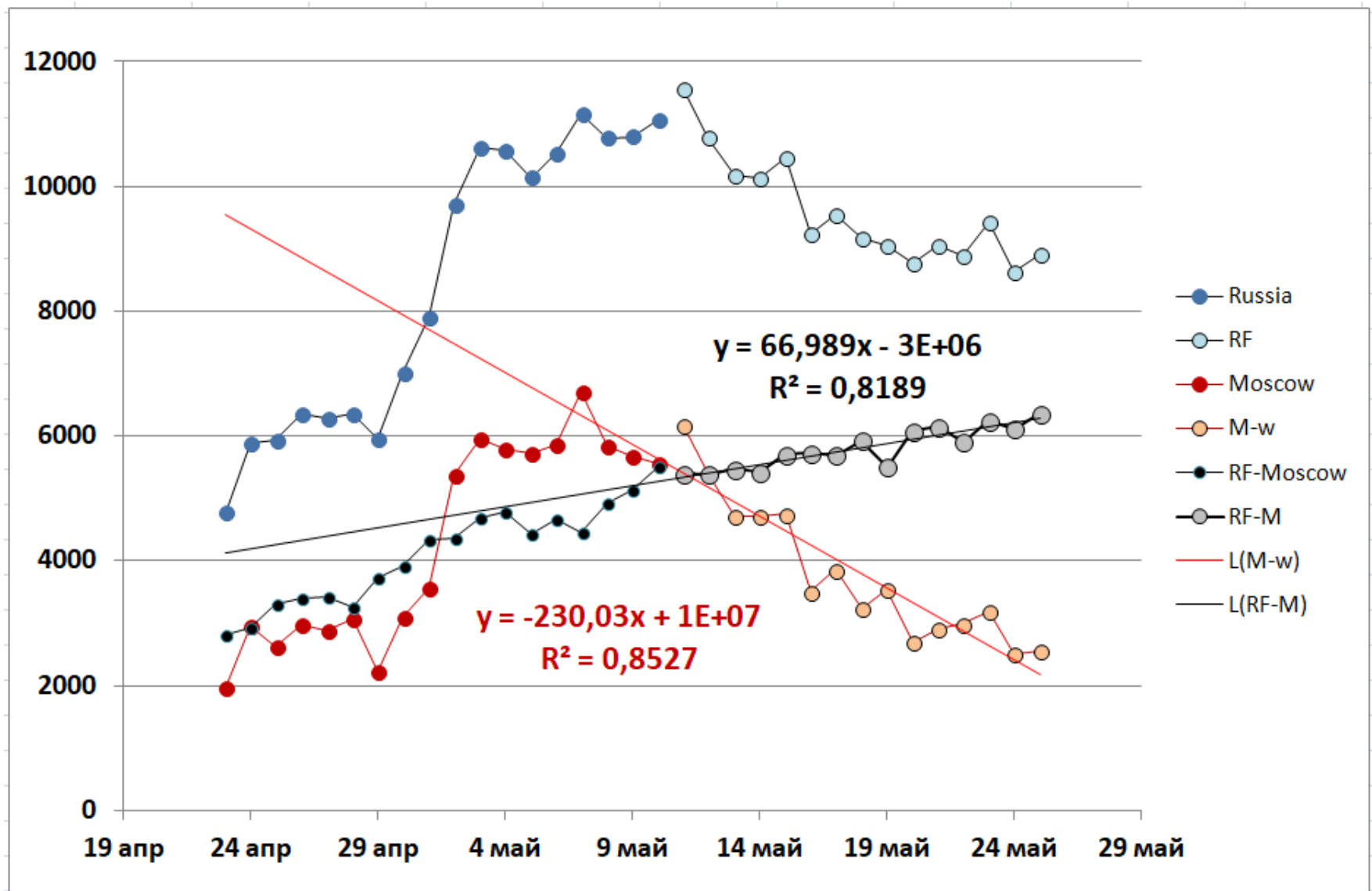
**Швеция – это особая форма течения COVID-19.
И этим всё сказано.**



Очень похожие по течению COVID-19 страны, но они только-только вышли в зону флуктуации.



По течению COVID-19 эти страны похожи, но они ещё, пройдя пик, не добрались до зоны флуктуации.



Это РФ, РФ без Москвы и Москва. Очевидно, что РФ и Москва пики свои проскочили, но до низа далеко.

ВЫВОДЫ

- 1.** Естественный ход распространения заражений обладает гладкостью, а вмешательство приводит: (1) к смене параметров процесса или его типа скачком, или (2) вызывает турбулентность, хаос. Т.е. индикаторы борьбы – динамика производных.
- 2.** Возврат пандемии в эпидемию – отсутствие корреляции между локальными и общими числами.
- 3.** На разных этапах значимость разных данных для исследований и принятия решений меняется.
- 4.** Любая грамотная модель обязана учитывать то, что на разных этапах пространство для пандемии оказывается разным, т.к. меняется поведение и людей, и властей. Модель должна меняться тоже.



спасибо за внимание!