

## Выбор метода для определения параметров модели Басса прогнозирования развития возобновляемой энергетики на примере ветрогенерации на языке программирования Python с использованием библиотеки scipy

Никифоров М.М., Куделин А.Г. (nikiforov1601@gmail.com)

*Ухтинский государственный технический университет, г. Ухта, Россия*

При выборе модели моделирования и прогнозирования на основе модели Басса, существует не мало методов для осуществления её работы. Каждый из методов является оптимальным в зависимости от данных, с которыми работает модель. Необходимо исследовать методы, которые можно применить к выбранной модели и выбрать тот метод, который даёт самое оптимальное решение.

В качестве модели Басса, было использовано следующее уравнение [1]:

$$S(t) = p * m + (q - p) * Y(t - 1) - \frac{q}{m} Y^2(t - 1)$$

Где  $S(t)$  – продажи в период  $t$ ;  $Y(t - 1) = \sum_{x=1}^{t-1} S(x)$  – совокупные продажи за период  $[0 \dots t - 1]$ ;  $p$  - коэффициент инновации,  $q$  - коэффициент имитации,  $m$  - общее количество всех покупок.

Модель Басса описывает продажи за период, которые являются производной от ветрогенерации за прошлый период. Определив продажи за период, мы высчитываем текущую ветрогенерацию. Таким образом, определив параметры  $p$ ,  $q$ ,  $m$ , так, что смоделированные данные будут максимально равны фактическим, мы в дальнейшем можем давать прогноз на необходимые промежутки времени.

Для дальнейшего исследования определим данные, над которыми будем выполнять все операции. Будем использовать данные по ветрогенерации за период с 1995 по 2020 гг в следующих регионах: Суммарные данные по миру, Европа, Северная Америка, Центральная и Южная Америка, Африка, Азиатско-Тихоокеанский регион, Средний Восток.

Реализуем поиск оптимальных параметров  $p$ ,  $q$ ,  $m$  с помощью MS Excel (данные-Поиск решения) на одном наборе данных (Суммарные данные по миру):

СУММ : $=SA4*SC4+(SB4-SA4)*I6-(SB4/SC4)*(I6^2)$														
	A	B	C	D	E	F	G	H	I	J	K	L	M	
1							Year		1995	1996	1997	1998	1999	
2	Мировые данные						Generation (TW*h)		8,3	9,2	12,0	15,9	21,2	
3	p	q	m					"Sales"		0	0,9	2,813216	3,903444	5,294914
4	0,000572651	0,249510681	2407,359254					Prognose "Sales"		0	$=SA4*SC4+(SB4-SA4)*I6-(SB4/SC4)*(I6^2)$	4,274531	5,326373	6,631742
5							Square Mista	4722,499296		0	6,177868128	15,5776	28,8345	44,97855
6							Prognose Cumulative		8,261923	11,69013242	15,96466	21,29104	27,92278	
СУММ : $=I6+J4$														
	A	B	C	D	E	F	G	H	I	J	K	L	M	
1							Year		1995	1996	1997	1998	1999	
2	Мировые данные						Generation (TW*h)		8,3	9,2	12,0	15,9	21,2	
3	p	q	m					"Sales"		0	0,9	2,813216	3,903444	5,294914
4	0,000572651	0,249510681	2407,359254					Prognose "Sales"		0	3,428208977	4,274531	5,326373	6,631742
5							Square Mista	4722,499296		0	6,177868128	15,5776	28,8345	44,97855
6							Prognose Cumulative		8,261923	$=I6+J4$	15,96466	21,29104	27,92278	
СУММ : $=(J6-J2)^2$														
	A	B	C	D	E	F	G	H	I	J	K	L	M	
1							Year		1995	1996	1997	1998	1999	
2	Мировые данные						Generation (TW*h)		8,3	9,2	12,0	15,9	21,2	
3	p	q	m					"Sales"		0	0,9	2,813216	3,903444	5,294914
4	0,000572651	0,249510681	2407,359254					Prognose "Sales"		0	3,428208977	4,274531	5,326373	6,631742
5							Square Mista	4722,499296		0	$=(J6-J2)^2$	15,5776	28,8345	44,97855
6							Prognose Cumulative		8,261923	11,69013242	15,96466	21,29104	27,92278	

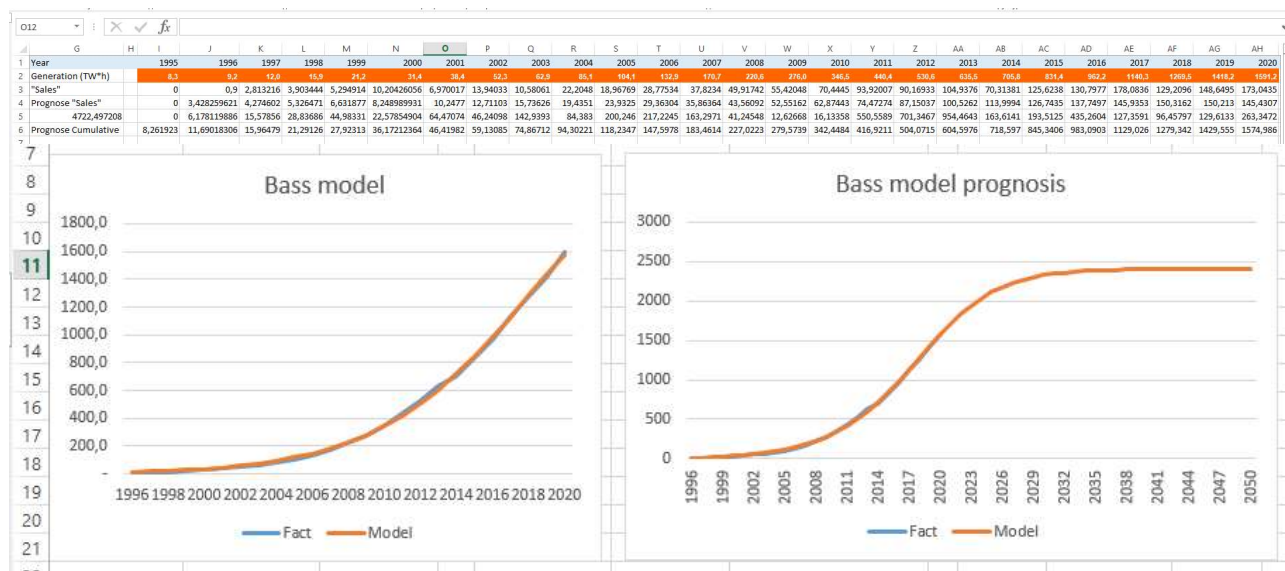


Рисунок 1

Из полученных данных видно, что данный метод наиболее максимально подбирает параметры  $p$ ,  $q$ ,  $m$ . Но, что затрудняет его применение при дальнейших исследованиях? Основные проблемы:

1. При необходимости исследовать большое количество данных, уходит очень много времени на подстановку данных;
2. Первоначальные параметры  $p$ ,  $q$ ,  $m$  необходимо подбирать вручную, что не всегда дает нужный результат;

Данный способ необходимо автоматизировать, избежав проблем метода с помощью MS Excel.

Для автоматизации был выбран язык программирования – python, являющийся одним из основных используемых в научных вычислениях и имеющий множество библиотек, специализированных для работы с математическими вычислениями. Использование данного языка программирования решит нашу первую проблему предыдущего способа.

Для решения поставленной задачи, были выбраны следующие библиотеки:

- numpy (Numerical Python), для работы с массивами, матрицами;
- pandas, для работы с DataFrame;
- scipy, предназначенная для выполнения научных и инженерных расчётов.

Для решения второй проблемы, будем использовать библиотеку scipy и её модуль optimize, содержащий функцию curve\_fit – использующая нелинейный метод наименьших квадратов, чтобы подогнать функцию  $f$  к данным.

Функция minimize (минимизация скалярной функции одной или нескольких переменных) модуля optimize отлично подойдет для оптимального подбора параметров  $p$ ,  $q$ ,  $m$ , так, что полученные данные будут максимально приближены к фактическим. Обратившись к описанию функции minimize видим, что если наши параметры имеют ограничения (в нашем случае параметры  $p$ ,  $q$ ,  $m$  положительные целые числа), то используются определенные методы «Bounds on variables for Nelder-Mead, L-BFGS-B, TNC, SLSQP, Powell, and trust-constr methods»[2], соответственно их все надо проверить и определить самый оптимальный, но сначала посмотрим описание данных методов:

Nelder-Mead – также известный как метод деформируемого многогранника и симплекс-метод, — метод безусловной оптимизации функции от нескольких переменных, не использующий производной (точнее — градиентов) функции, а поэтому легко применим к негладким и/или зашумлённым функциям.[3]

Powell – представляет собой алгоритм, предложенный Майклом Дж. Д. Пауэллом для нахождения локального минимума функции. Функция не обязательно должна быть дифференцируемой.[4]

L-BFGS-B – метод реализованный с уменьшенным потреблением памяти за счет частичной загрузки векторов из матрицы Гессе.[5]

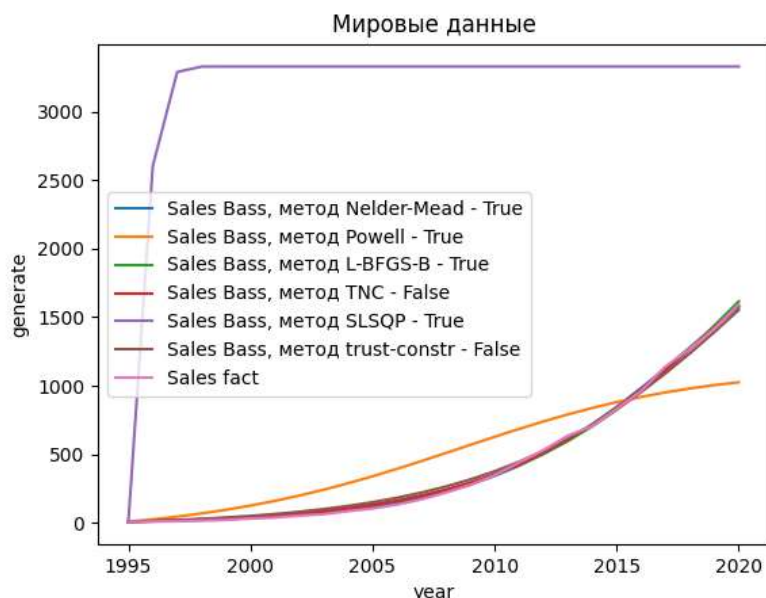
TNC – Truncated Newton Constrained, ограниченное число итераций, хорош для нелинейных функций с большим числом независимых переменных.[5]

SLSQP – последовательное квадратичное программирование с ограничениями, ньютоновский метод решения системы Лагранжа.[5]

trust-constr – поиск локального минимума в доверительной области.[5]

Для дальнейшего исследования создадим функцию Bass, принимающую предыдущие значения ветрогенерации и параметры  $p$ ,  $q$ ,  $m$ , рассчитывает продажи. Функцию `def squareMistake`, принимающую первоначальные параметры  $p$ ,  $q$ ,  $m$  и кортеж продаж, рассчитывает сумму квадратов разностей значений прогнозируемой и реальной генерации. Наилучший метод оптимизации будем определять по значению RSS, разности суммы квадратов между реальными значениями и значениями, полученными при использовании каждого метода.

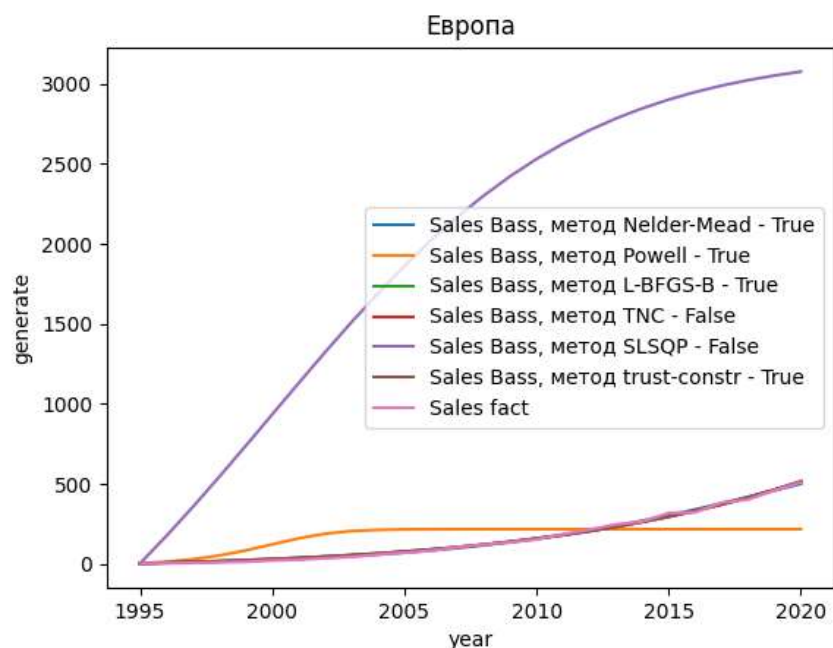
Для каждого набора данных выведем значения, методов (с помощью MS Excel, `curve_fit` первоначальными значениями, Nelder-Mead, L-BFGS-B, TNC, SLSQP, Powell, trust-constr), минимизации (True – минимизация прошла успешно, false - нет), параметров  $p$ ,  $q$ ,  $m$  и значение RSS. Сначала выводим график полученных данных, ниже его значения.



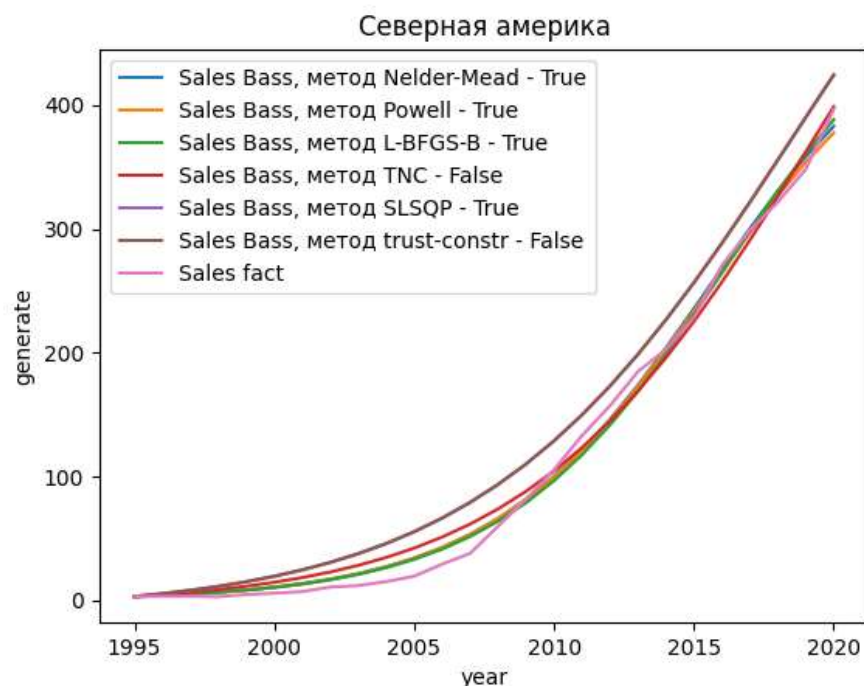
Генерация	Метод	Минимизирована	P	Q	M	RSS
Мировые данные	Вручную	---	0.0005726510	0.2495106815	2407.3592536598	4722.4992960118
Мировые данные	curve_fit	---	0.0008791697	0.1925282635	3328.9194507414	456497.3094664136
Мировые данные	Nelder-Mead	True	0.0005724562	0.2495302410	2407.0349034294	4722.4961073601
Мировые данные	Powell	True	0.0141298677	0.1824165251	1121.2597409441	1337418.1068945487
Мировые данные	L-BFGS-B	True	0.0008474944	0.2143716594	3328.9194511150	9997.4786109689
Мировые данные	TNC	False	0.0008379150	0.2285403501	2737.8717016444	6490.7910221679
Мировые данные	SLSQP	True	0.7813909630	0.2089801208	3328.9194510486	209331090.9082589447
Мировые данные	trust-constr	False	0.0012468316	0.1959540975	3328.9194510386	23158.4081252461

Рисунок 2





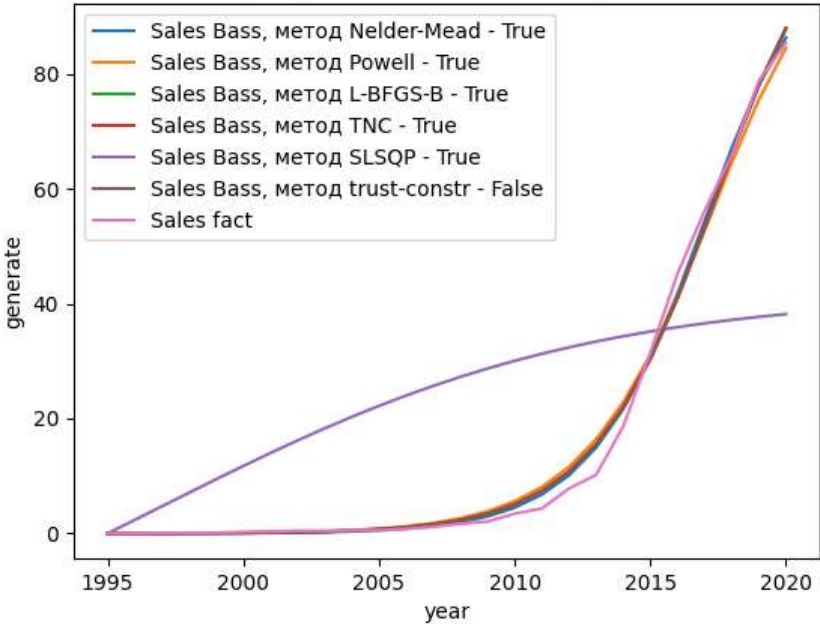
Европа	Вручную	---	0.0026905792	0.1656779601	968.8619048751	1294.4645554770
Европа	curve_fit	---	0.0010605382	0.1107641829	3205.5577891014	54900.1605036525
Европа	Nelder-Mead	True	0.0026908364	0.1656931974	968.6466570607	1294.4643838107
Европа	Powell	True	0.0367084999	0.6361559325	217.7948606388	412085.3361803551
Европа	L-BFGS-B	True	0.0011677141	0.1227965322	3205.5577891287	2111.0148828319
Европа	TNC	False	0.0012421014	0.1234550609	3006.3657906855	2077.2734180703
Европа	SLSQP	False	0.0536499403	0.1116237588	3205.5577891203	102127624.5173335224
Европа	trust-constr	True	0.0011678490	0.1227893290	3205.5474182261	2111.0128094999



Северная америка	Вручную	---	0.0000000000	0.2714272786	479.2328093480	1880.4015926827
Северная америка	curve_fit	---	0.0023187286	0.1841640722	730.8623375931	16181.4101913008
Северная америка	Nelder-Mead	True	0.0000000000	0.2714274563	479.2329731698	1880.4015909381
Северная америка	Powell	True	0.0000000000	0.2713342911	470.7796309894	1970.2865897525
Северная америка	L-BFGS-B	True	0.0000000000	0.2672133164	503.8082336088	2062.4027198326
Северная америка	TNC	False	0.0012493774	0.2031586779	707.9320610910	3908.5496129588
Северная америка	SLSQP	True	0.0023187286	0.1841640722	730.8623375931	16181.4101913008
Северная америка	trust-constr	False	0.0023187286	0.1841640722	730.8623375931	16181.4101913008

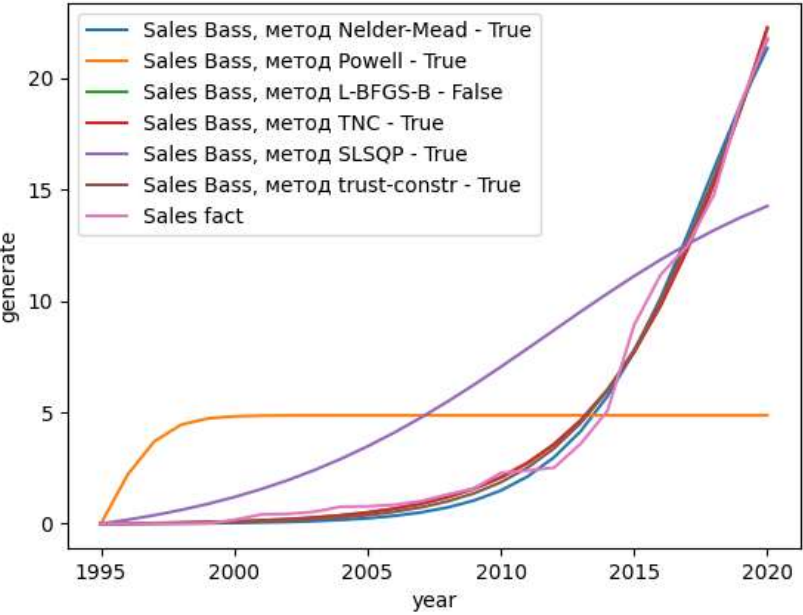
Рисунок 3

Центральная и Южная Америка



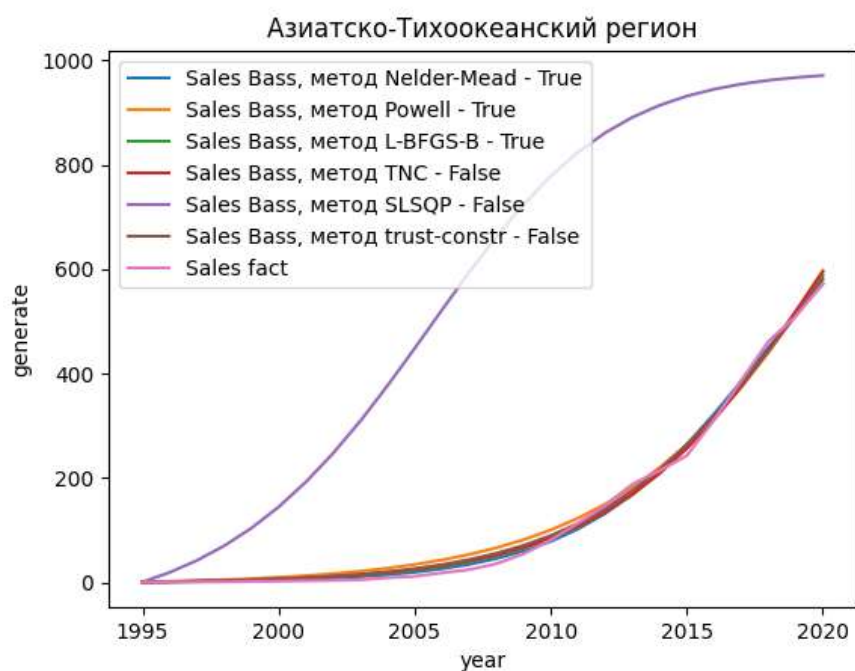
Центральная и Южная Америка	Вручную	---	0.0000000000	0.5343151234	93.3138680130	72.3611589272
Центральная и Южная Америка	curve_fit	---	0.0000000000	0.4885059692	105.2386309378	1600.8021971074
Центральная и Южная Америка	Nelder-Mead	True	0.0000000000	0.5310004335	97.2837564361	64.1956702870
Центральная и Южная Америка	Powell	True	0.0000486846	0.4684022585	101.3486700641	135.2483165314
Центральная и Южная Америка	L-BFGS-B	True	0.0000259445	0.4885059792	105.2386309378	96.5945892840
Центральная и Южная Америка	TNC	True	0.0000260535	0.4883812263	105.2371067047	96.6903170737
Центральная и Южная Америка	SLSQP	True	0.0562409115	0.0668376690	41.7828647864	12449.4392748953
Центральная и Южная Америка	trust-constr	False	0.0000260354	0.4883731807	105.2386317061	96.6989177510

Африка

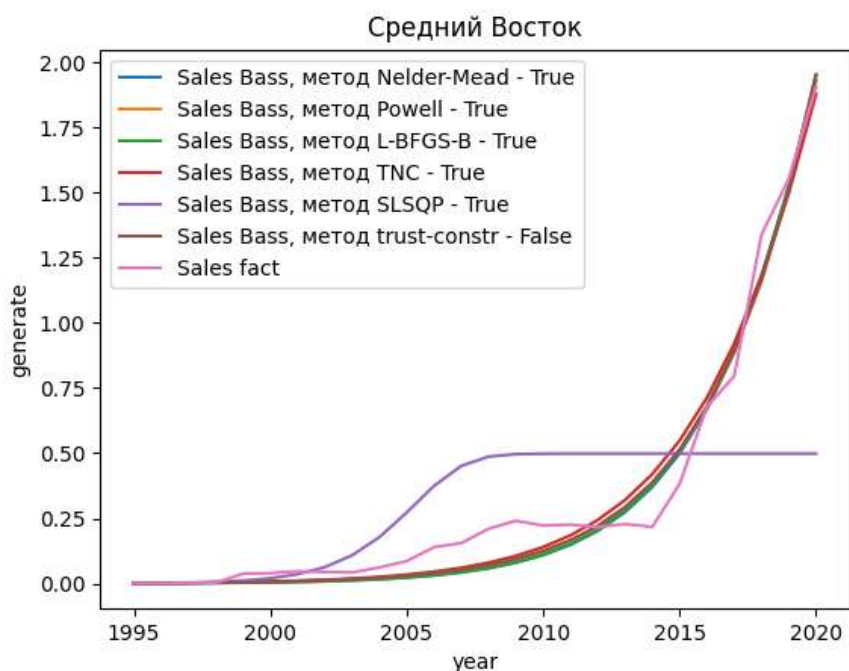


Африка	Вручную	---	0.0002253365	0.3183026783	43.1086940618	7.8926096734
Африка	curve_fit	---	0.0000000000	0.2843527410	45.6377725207	1113.3650994329
Африка	Nelder-Mead	True	0.0000000000	0.4434814504	26.9235036962	8.2059568444
Африка	Powell	True	0.4514442981	0.2454956990	4.8599471088	941.7208689545
Африка	L-BFGS-B	False	0.0001784864	0.3263036312	45.6378457411	7.5464721687
Африка	TNC	True	0.0001807924	0.3213845586	48.0899265771	7.7090582492
Африка	SLSQP	True	0.0100862914	0.1727331425	17.1798511336	328.9471456094
Африка	trust-constr	True	0.0001193341	0.3712871665	33.5372679355	6.9315689227

Рисунок 4



Азиатско-Тихоокеанский регион	Вручную	---	0.0000001109	0.3349636857	774.1021683152	1968.8661404116
Азиатско-Тихоокеанский регион	curve_fit	---	0.0005610691	0.2600558920	981.2284218616	6277.5667057539
Азиатско-Тихоокеанский регион	Nelder-Mead	True	0.0000110027	0.3329251741	779.3372969105	1957.5268378333
Азиатско-Тихоокеанский регион	Powell	True	0.0005556505	0.2300566171	1566.1718340505	6717.9927076976
Азиатско-Тихоокеанский регион	L-BFGS-B	True	0.0002423746	0.2922963955	981.2284249206	2676.4244228839
Азиатско-Тихоокеанский регион	TNC	False	0.0002933489	0.2789791575	1114.2398081834	3369.2489644990
Азиатско-Тихоокеанский регион	SLSQP	False	0.0185561791	0.2603569531	981.2284218864	6266514.0205576271
Азиатско-Тихоокеанский регион	trust-constr	False	0.0003699639	0.2797887386	981.2283911807	3120.0111272986



Средний Восток	Вручную	---	0.0000217807	0.3441681293	7.3126511883	0.1832602763
Средний Восток	curve_fit	---	0.0000000000	0.3059547203	5.4576438435	3.8406491790
Средний Восток	Nelder-Mead	True	0.0000000000	0.3682698062	6.4184066729	0.1831050530
Средний Восток	Powell	True	0.0000187370	0.3342713973	11.6685439043	0.1787976017
Средний Восток	L-BFGS-B	True	0.0000000000	0.3705188937	5.4585160218	0.1843092120
Средний Восток	TNC	True	0.0000521697	0.3249524017	6.8609583970	0.1938955908
Средний Восток	SLSQP	True	0.0000000000	0.8253401253	0.4984808910	4.6454169750
Средний Восток	trust-constr	False	0.0000196332	0.3310019896	12.4911484614	0.1785492859

Рисунок 5

Закключение.

Проведя исследование по нахождению метода для определения параметров модели Басса на языке программирования Python с использованием библиотеки `scipy`, можно сделать вывод, что нет универсального метода. Следовательно, необходимо использовать тот метод, который является самым оптимальным при подборе параметров.

Используемый код находится по ссылке  
[https://github.com/Misha1601/magistr/blob/main/bass\\_doclad\\_30032023.py](https://github.com/Misha1601/magistr/blob/main/bass_doclad_30032023.py)

#### **Библиографический список:**

1. F.M. Bass, Bass 1969 New Prod Growth Model.pdf, Manage. Sci. 15 (1969) 215–227.
2. <https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html>
3. [https://ru.wikipedia.org/wiki/Метод\\_Нелдера\\_—\\_Мида](https://ru.wikipedia.org/wiki/Метод_Нелдера_—_Мида)
4. [https://en.wikipedia.org/wiki/Powell%27s\\_method](https://en.wikipedia.org/wiki/Powell%27s_method)
5. <https://habr.com/ru/companies/ods/articles/448054/>