

ПРОГНОЗИРОВАНИЕ ПАРАМЕТРОВ ВРАЩЕНИЯ ЗЕМЛИ С ПОМОЩЬЮ СИНГУЛЯРНОГО СПЕКТРАЛЬНОГО АНАЛИЗА

Горшков В.Л., Миллер Н.О.

Дано краткое описание метода прогнозирования на основе сингулярного спектрального анализа в его реализации «Гусеница»-SSA. На материале параметров вращения Земли (ПВЗ) международной службы вращения Земли (МСВЗ) проведено исследование прогнозирования движения полюса и неравномерности вращения Земли этим методом. Исследованы различные возможности метода с целью выбора оптимальных параметров прогноза. Приводится сравнение ошибок прогнозирования данным методом с применяемыми в МСВЗ.

ВВЕДЕНИЕ

В МСВЗ с октября 2005 года по март 2008 года проводилась кампания по оценке прогнозирования ПВЗ различными методами (http://www.cbk.waw.pl/EOP_PCC). В кампании принимали участие 11 представителей разных организаций с различными подходами к прогнозированию ПВЗ, начиная от обычного Фурье анализа с последующей аппроксимацией методом наименьших квадратов (МНК) наиболее значимых гармоник и кончая использованием нейронных сетей и моделей на основе нечёткой логики. Методы применялись для краткосрочных (до 10 дней), среднесрочных (до 30 дней) и долгосрочных прогнозов до 500 дней. Долгосрочный прогноз вычислялся раз в месяц, а краткосрочные – раз в неделю.

Во многих прогнозах помимо рядов ПВЗ дополнительно использовались геофизические данные, а именно, ряды атмосферного углового момента, являющиеся одним из возбуждающих факторов динамики движения полюса и вращения Земли. Именно использование этих данных в сочетании с фильтрацией Калмана обеспечило лучший прогноз неравномерности вращения Земли. Параметры движения полюса наиболее качественно предсказываются с помощью МНК экстраполяции полигармонической модели и авторегрессионного прогнозирования.

Качество прогноза оценивалось по среднему абсолютному уклонению (MAE) прогнозируемой величины от впоследствии наблюдаемой для k -ого дня прогноза ε_{ki} , где $i = 1, 2, \dots, N$. Здесь N – количество прогнозов, варьировавшееся в данной кампании от 7 до 100.

$$MAE_k = \frac{1}{N} \sum_{i=1}^N |\varepsilon_{k,i}| \quad (1)$$

Среди применявшихся методов прогнозирования метод ССА не использовался для этих целей. В данной работе представлены прогностические возможности метода ССА и проведено их исследование с использованием данных МСВЗ и результатов вышеупомянутой международной кампании.

ОПИСАНИЕ МЕТОДА ПРОГНОЗИРОВАНИЯ

Метод сингулярного спектрального анализа (ССА) предназначен для исследования структуры временных рядов и совмещает в себе достоинства многих других методов, в частности, анализа Фурье и регрессионного анализа. Результатом применения метода является получение структуры ряда, т.е. разложение его на компоненты, которые могут быть интерпретированы как медленные тренды, сезонные и другие периодические или колебательные составляющие, а также шумовые компоненты. В случае прогнозирования выполняется продолжение структуры исходного ряда, выделенной при его анализе.

Метод «Гусеница»-SSA - в авторском названии модификации данного метода и его программной реализации (<http://www.gistatgroup.com/gus/>) неоднократно использовался для целей анализа ПВЗ в работах (Воротков и др., 2002; Горшков и др., 2000), и впервые был использован для целей прогноза ПВЗ в работе (Горшков, 2004).

Прогнозирование с помощью ССА включает в себя несколько различных этапов. Причём для каждого этапа, существуют различные возможности их реализации. Эти возможности можно рассматривать, как некоторые управляющие параметры. Исследование возможностей для каждого конкретного ряда позволяет подобрать оптимальные параметры для решаемой задачи прогнозирования. После выполнения исследования можно перейти в режим автоматического прогнозирования ряда. Подробное описание метода для целей прогнозирования можно найти в работах (Данилов, Жиглявский, 1997; Степанов, Голяндина, 2005). Описание того, как выбирать параметры при автоматическом выделении трендовых и периодических составляющих временного ряда, подробно описано в работе (Александров, Голяндина, 2005).

Можно предложить и другие подходы, которые просто использовать для исследования прогностических возможностей метода. Например, для этой цели можно использовать два критерия - либо выбирать одно и то же число компонент (K) для всех прогнозов, либо выбирать число компонент таким образом, чтобы суммарный вклад в общий процесс был не меньше определённого процента (например - 99.99%). Недостаток первого заключается в том, что на разных интервалах можно получать разное разложение по числу значимых компонент и при таком способе можно ухудшить качество аппроксимации и как следствие - качество прогноза. Второй критерий в этом смысле является более надёжным, но требует дополнительного анализа.

Перечислим этапы работы метода и укажем на различные возможности варьирования параметров на каждом этапе. На первом задаётся длина окна M для выполнения преобразования ряда в матрицу. Длина окна M должна быть выбрана достаточной, чтобы разделить шумовой и информативной компоненты и в то же время не должна быть очень большой. Надо отметить, что при прогнозировании нет необходимости выполнять полное разложение ряда, главное отделить полезный сигнал от шума. Чем больше M , тем больше размер характеристического полинома в рекуррентной формуле при прогнозировании, что может привести к появлению побочных корней полинома.

Затем выполняется разложение матрицы на собственные вектора и главные компоненты, и выбирается число компонент, по которым восстанавливается ряд и строится характеристический полином для аппроксимации ряда. Для ССА прогноза по виду и значимости главных компонент разложения можно оценить их количество необходимое для прогноза. После этого по выбранному набору главных компонент с помощью диагонального усреднения получается восстановленный ряд и затем выполняется аппроксимация ряда. Характеристический полином для прогнозирования в точности равен ба-

зису разложения. При этом нет необходимости в том, чтобы заранее знать параметрический вид тренда и периодических компонент.

Этап аппроксимации можно рассматривать как предварительный к этапу прогноза ряда. Как аппроксимация, так и прогноз заключаются в задании начальных точек и затем их продолжении. Используя аппроксимацию, можно понять, насколько ряд соответствует модели, заложенной в основу прогноза. Оба метода аппроксимации (и прогноза) используют рекуррентную формулу, сходную с формулой широко используемого в ПВЗ авторегрессионного прогнозирования, но имеющую точную размерность $M - 1$. Для прогнозирования и аппроксимации имеется два способа реализации - геометрический (векторный, V-метод) и аналитический (рекуррентный, R-метод). Для геометрического способа базовый вектор совпадает с восстановленным вектором, а для аналитического способа он может быть восстановленным (R_r), исходным (R_i) или аппроксимирующим (R_a). В дальнейшем на рисунках и в таблицах используются именно эти обозначения.

Геометрический (векторный) метод преобразует собственные вектора таким образом, чтобы первые $M - 1$ координат нового вектора были максимально близки к $M - 1$ последним координатам исходного вектора. Если выполнить несколько шагов такой процедуры, начав с базового восстановленного вектора, то в результате получится матрица, из которой с помощью процедуры диагонального усреднения получается ряд аппроксимации (или прогноза). В зависимости от того, делается ли диагональное усреднение до или после продолжения ряда, различают рекуррентный и векторный прогноз.

Аналитический (рекуррентный) метод аппроксимации использует рекуррентную формулу более явно. Начав с заданного вектора развертки восстановленного ряда, каждое следующее значение получается с помощью рекуррентной формулы, примененной к $M-1$ предшествующим точкам. Поскольку в рекуррентном прогнозе используются в качестве начальных данных последние точки ряда, это может вызвать увеличение ошибок прогноза. В этом смысле векторный прогноз более консервативен, но менее адаптивен к меняющемуся концу ряда.

Близость этих прогнозов может свидетельствовать об устойчивости полученного прогноза.

ПРОГНОЗ ПВЗ

Для прогнозирования движения полюса Земли (координат полюса X_p, Y_p) использовались ряд ПВЗ и их годичный прогноз USNO (ряд *finals*), вычисляемые срочной службой МСВЗ (<http://maia.usno.navy.mil/ser7/finals2000A.all>), и ряд комбинированных значений ПВЗ *C04* (http://hpiers.obspm.fr/iers/eop/eopc04_05/eopc04_IAU2000.62-now).

Перед тем как выполнять тестирование, было проведено небольшое исследование метода с целью выбора оптимальных параметров.

В результате сопоставления прогнозов для разных опорных интервалов от 3 до 24 лет (ряд *C04*) был сделан вывод, что для прогнозов разной длины нужно выбирать разные опорные интервалы. Оценка среднеквадратического отклонения (*rms*) была выполнена по 5 вариантам прогноза с длиной окна $M=250$ и числом восстановленных компонент $K=24$. В четырёх случаях вычисления проводились, начиная с начала ряда, со сдвигом в 10 дней и один вариант прогноза - из середины ряда. В итоге был выбран интервал в 12 лет, который является средним на интервале прогноза до 200 дней и лучшим на более длительных интервалах прогнозирования.

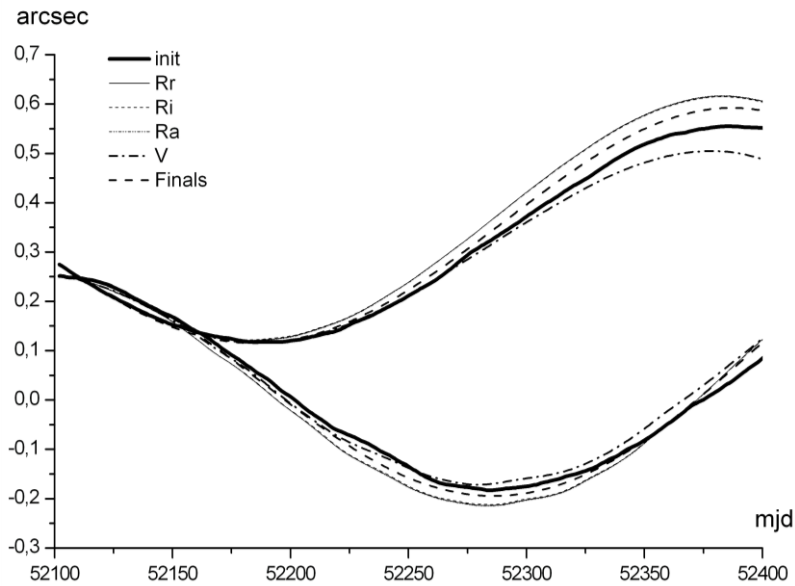


Рис.1. Пример прогноза координат полюса USNO, полученного разными методами и сравнение его с прогнозом МСВЗ (нижние кривые - X_p , в верхние - Y_p , прогноз МСВЗ - Finals).

На рис. 1. приводится пример прогноза ряда Finals, полученного разными модификациями метода ССА. База прогноза составляла 9 лет, длина окна $M=2$ года (730 точек). Этот рисунок наглядно показывает разницу между прогнозом, полученным векторным методом и прогнозами, полученными аналитическими методами. Кроме того, он даёт возможность качественно сопоставить полученные прогнозы с исходными рядами и с прогнозами срочной службы МСВЗ.

Для тестирования метода был выбран далеко не самый простой интервал для прогнозирования с существенной особенностью в поведении X_p - координаты полюса (рис. 2). Представленный прогноз был получен с использованием восстановленного ряда в качестве базового вектора (тёмная линия). Из рисунка видно хорошее совпадение исходного и восстановленного ряда до начала прогноза и в первых точках самого прогноза. Этот прогноз выполнен для координаты X_p и является одним из прогнозов (первым) вошедших в конечный результат тестирования метода.

Тестирование метода производилась по 15 прогнозам координат полюса с шагом в 30 дней, длина опорного интервала -12 лет (4380 точек) и длина “гусеницы” (окна) M - 438 точек. На этапе восстановления выбиралось постоянное число компонент для X_p равное 24 и для Y_p - 22. Число компонент было выбрано таким образом, чтобы заведомо произошло отделение полезной информации от шума. Т.е. для некоторых разложений число компонент могло бы быть увеличено, но не уменьшено. Для аппроксимации был выбран рекуррентный модифицированный метод. Максимальная разница между исходным и аппроксимирующим рядами составила менее 0.6 mas.

На рис.3 и в табл.1 представлены оценки тестирования прогноза координат полюса разными модификациями метода ССА. В таблице 1 даны среднеквадратичные отклонения между предсказанными координатами полюса и рядом $C04$. На рис.3. приводится среднее абсолютное уклонение прогнозируемой величины от наблюдаемой (MAE), вычисленное по формуле (1).

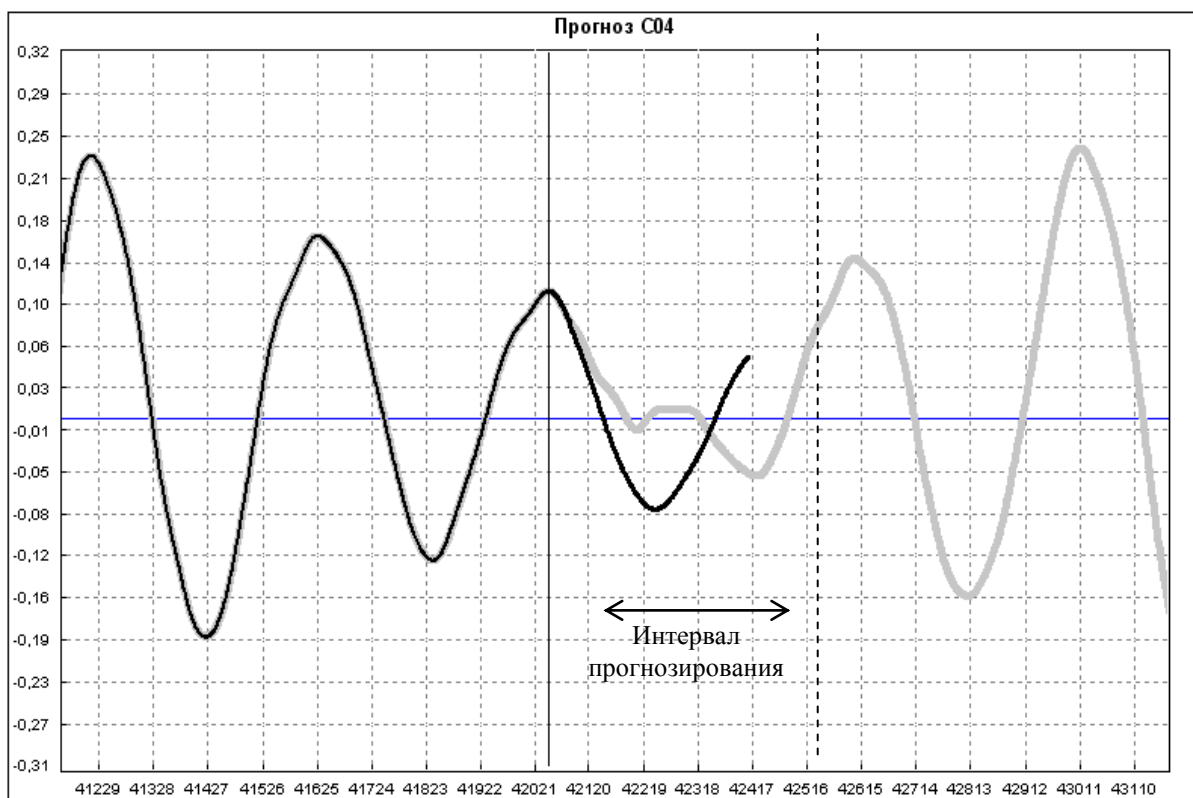


Рис.2. Интервал, выбранный для оценки метода. Исходный ряд координаты X_p - серая линия, базовый вектор – черная линия, начало прогноза отмечено непрерывной вертикальной линией.

Таблица 1. Среднеквадратичная (mas) ошибка прогноза координат полюса.

ПВЗ	Метод	Длина прогноза, сутки									
		10	20	30	40	60	90	120	150	180	365
X_p	V	1.31	4.24	8.91	13.43	17.18	15.00	26.60	36.06	33.29	45.07
	Rr	1.89	4.24	7.61	11.42	13.51	14.82	18.30	24.27	20.93	37.02
	Ri	1.38	3.51	7.3	11.72	15.52	18.12	22.13	23.85	20.65	37.70
	Ra	2.40	5.90	10.72	16.45	21.14	23.43	24.63	26.40	22.60	39.38
Y_p	V	1.69	4.42	7.93	10.60	13.56	15.78	22.58	26.93	27.82	32.57
	Rr	1.87	4.44	6.40	8.47	13.06	17.26	22.44	25.84	27.61	33.48
	Ri	1.42	3.53	5.95	9.29	16.47	21.70	24.71	27.78	28.67	33.97
	Ra	2.85	6.14	8.88	10.76	17.27	21.35	27.96	30.67	30.14	35.31

Из приведённых результатов видно, что для прогнозов разной длины разные методы являются лучшими. На интервале до 30 дней лучшим является аналитический прогноз, с исходным базовым вектором, а также неплохие результаты получаются при использовании векторного метода. Причём все прогностические возможности векторного метода для прогноза координат полюса в данной работе не исследовались.

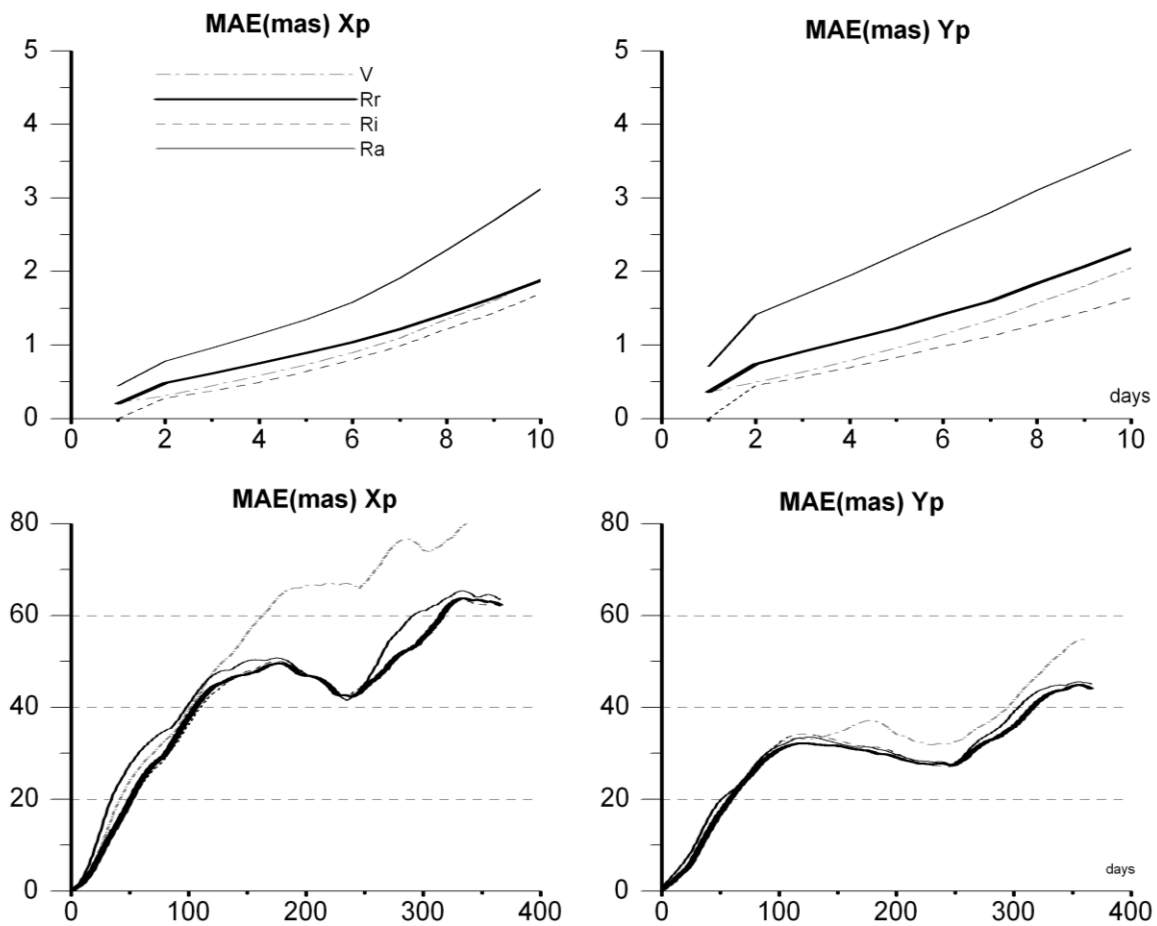


Рис.3. Оценка прогноза координат полюса, полученного методом ССА.

На всем интервале в 365 дней прогноз координаты полюса Y_p получается лучше, чем прогноз X_p и наблюдается хорошее согласование между всеми построенными прогнозами. На интервале до 120 дней векторный и аналитический прогнозы практически совпадают. Это может говорить об устойчивости данного прогноза на этом временном интервале. Таким образом подтверждается вывод о том, что для более длительных прогнозов нужно выбирать другие управляющие параметры.

Для сравнения на рис. 4 представлены ошибки полученных прогнозов по данным МСВЗ кампании. Надо отметить, что прогноз от 2-х до 10 дней для координаты X_p лучше всех представленных на рис.5. Из анализа рисунков следует, что прогноз, полученный с помощью ССА, не хуже прогнозов срочной службы МСВЗ.

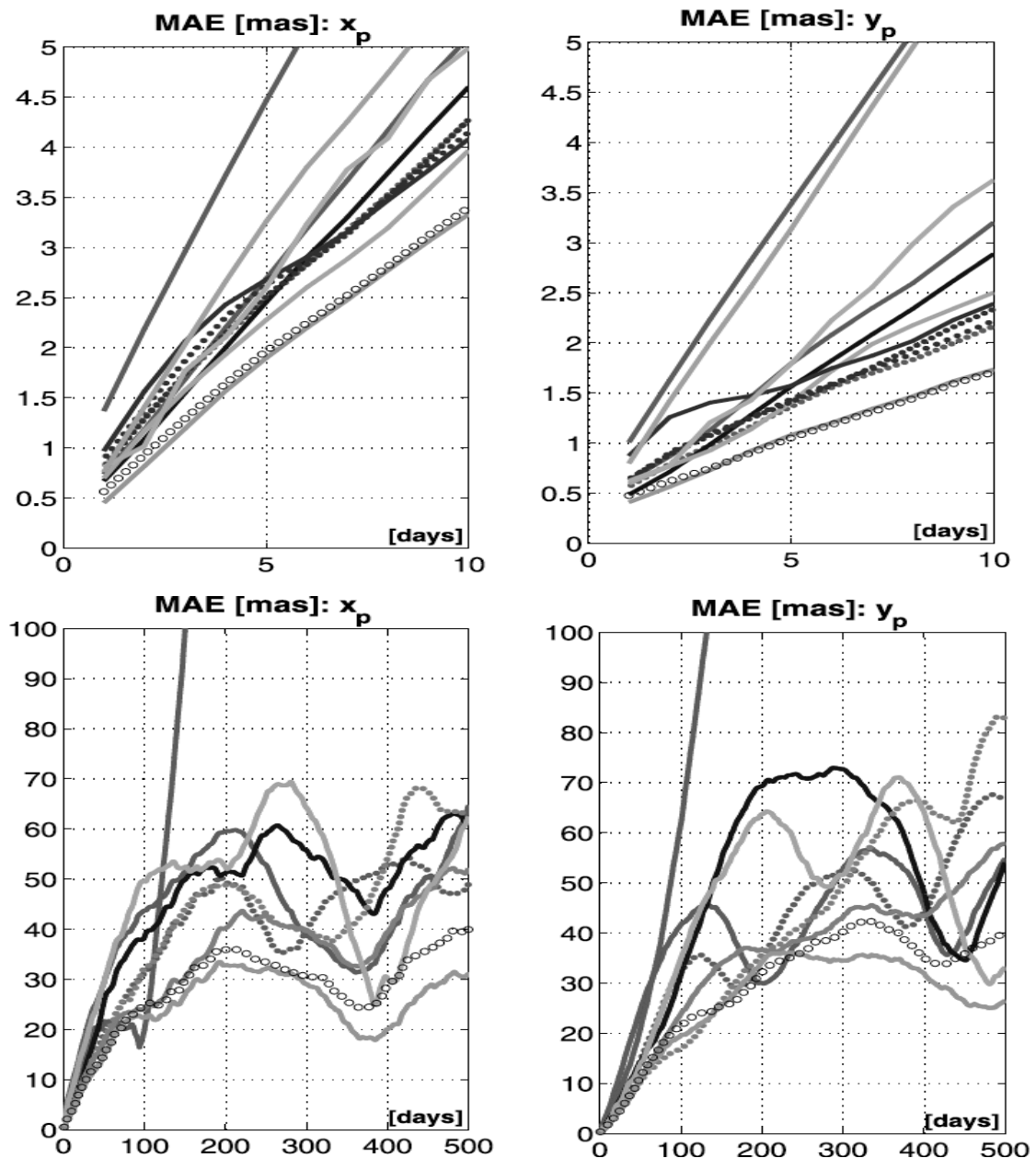


Рис.5. Прогнозы координат полюса срочной службы МСВЗ.

Прогнозирование неравномерности вращения Земли отличается тем, что ряд $UTI - UTC$ имеет очень мощную нерегулярную низкочастотную составляющую (тренд и декадные вариации во вращении Земли), на порядки превосходящую прочие квазипериодические вариации во вращении Земли. Поэтому для прогноза $UTI - UTC$ приходится раздельно исследовать низкочастотные и высокочастотные составляющие ряда, именно ввиду экономии вычислительного ресурса и однородности получаемых результатов. Как показал численный эксперимент, исключение линейного тренда иногда оказывается достаточным для совместной оценки прогноза по обеим составляющим в $UTI - TAI$.

Для анализа и прогноза неравномерности вращения Земли естественно использовать данные, во-первых, приведенные к непрерывной (без эпизодических секундных скачков) атомной шкале времени, то есть использовать $UTI - TAI$, а не $UTI -$

UTC. Во-вторых, необходимо исключить из них известные зональные приливные вариации во вращении Земли в интервале периодов от 5.6 дней до 18.6 лет, то есть использовать ряд *UTIS – TAI* (<http://hpiers.obspm.fr/eop-pc/models/UT1/>).

Использовался ряд неравномерности вращения Земли *Finals* с 1986 года до начала 2008 года, на основе которого были проведены ежемесячные прогнозы на год вперед, начиная с февраля 2006 года до августа 2007 года. Выбор 1986 года в качестве начальной точки ряда для построения прогноза определяется резким уменьшением стохастической составляющей ввиду отказа МСВЗ от оптических определений ПВЗ именно с этого года. Этот момент существенен для прогнозирования, т.к. ряд должен быть стохастически однородным.

Для прогноза трендовой составляющей применялся векторный метод с максимальным лагом (M) равным половине реализации ряда. Как показал численный эксперимент возможно уменьшение M до четверти длины ряда практически без изменения первых трех используемых трендовых составляющих разложения. Это сильно (в 4 раза) сокращает вычислительное время, весьма значительное ввиду длины ряда (22 года ежедневных данных). Сократить вычислительное время можно используя прореживание ряда, однако в этом случае необходима последующая интерполяция при получении суммарного прогноза по обеим составляющим.

Однако векторный прогноз оказался нечувствительным к намечающейся к концу ряда изменению его тенденции (к смене производной, характерной для 2007 года). Поэтому стратегия прогнозирования, к сожалению, включает некоторую трудно алгоритмируемую изменчивость. В этом случае надо либо переходить на один и вариантов рекуррентного прогноза, либо увеличивать исходный ряд, включая данные до 1986 года, и/или увеличивать лаг M .

Для прогноза высокочастотной (оперативной) составляющей применялся модифицированный рекуррентный метод с лагом M не менее 4 лет (1460 точек). В большинстве случаев использование в качестве базового вектора исходного (R_i в применяемых здесь обозначениях) давало лучшие прогнозы, особенно существенные в долгосрочном прогнозировании. Выбор числа компонент для восстановления ряда осуществлялся по их суммарному вкладу не менее 99.95%, что составляет обычно около 100 компонент.

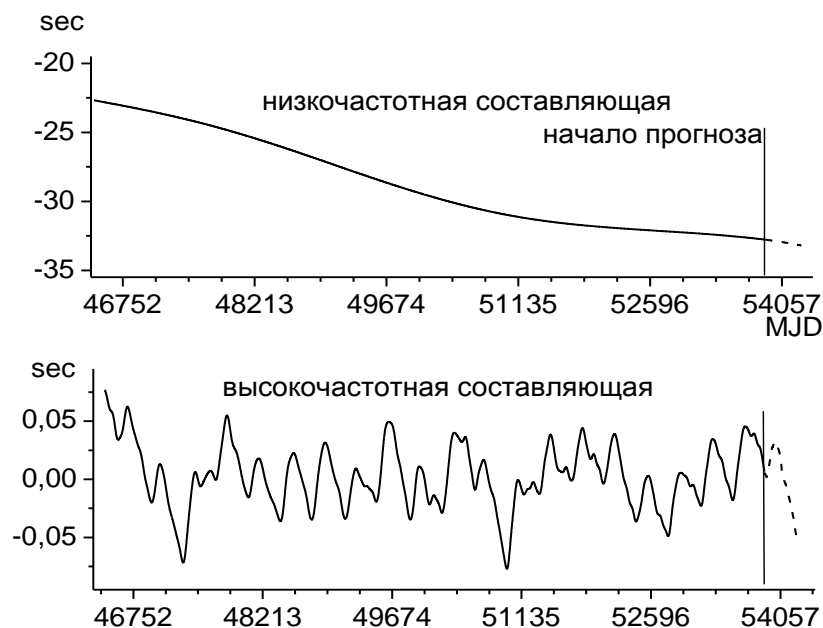


Рис.6 Годичный прогноз с мая 2006. Каждое деление шкалы MJD соответствует году.

На рис. 6 представлены пример аппроксимации и прогноза обеих составляющих в неравномерности вращения Земли. Видно, что годичный компонент является доминирующим в высокочастотной составляющей ряда. Это, вероятно, и определило успех метода прогнозирования в кампании МСВЗ, основанного на использовании данных эффективного атмосферного углового момента, т.к. он является основным возбуждающим фактором годичной составляющей в неравномерности вращения Земли. Этот факт дает основание для дальнейшего развития данного (ССА) метода в плане использования его многомерной реализации с включением данных атмосферного углового момента.

На рис. 8 представлены ошибки полученных прогнозов по данным прогностической кампании МСВЗ и на рис. 9 по нашим прогнозам, полученным по 15 реализациям. Видно, что прогноз *UTIS – TAI* с помощью метода «Гусеница»-SSA уступает большинству прогнозов кампании в краткосрочном прогнозировании, сравним по качеству в среднесрочном прогнозировании (до 30 дней) и превосходит даже лучший долгосрочный прогноз в проведенной в течение трех лет кампании МСВЗ по изучению методов прогнозирования ПВЗ.

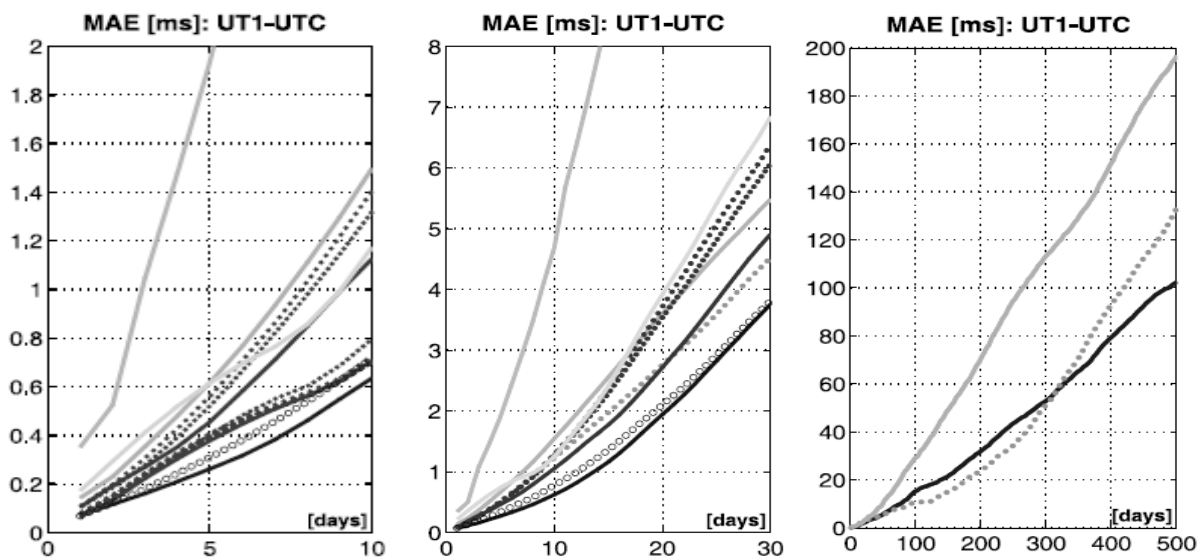


Рис. 8 *UT1 – UTC* прогнозы срочной службы МСВЗ.

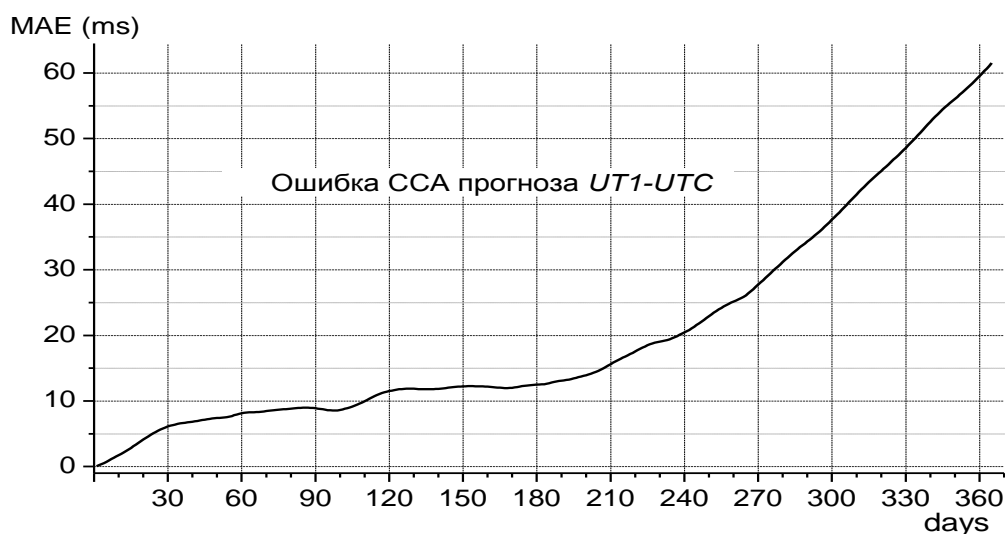


Рис. 9 Оценка прогноза *UT1 – UTC*, полученного методом ССА.

ВЫВОДЫ

Прогноз координат полюса, полученный с помощью ССА, не хуже прогнозов срочной службы МСВЗ. В частности, краткосрочный ССА прогноз для координаты X_p (до 10 дней) лучше всех представленных в проведенной в течение трех лет кампании МСВЗ по прогнозированию ПВЗ.

Прогноз неравномерности вращения Земли ($UTI - UTC$) уступает большинству прогнозов МСВЗ кампании в краткосрочном прогнозировании, сравним по качеству в среднесрочном прогнозировании (до 30 дней) и превосходит даже лучший долгосрочный прогноз (особенно до 250).

Вероятно полученный результат может быть улучшен, учитывая большое количество возможностей параметризации этого метода прогнозирования.

ЛИТЕРАТУРА:

- Александров Ф., Голяндина Н.Э.**, 2005, *Выбор параметров при автоматическом выделении трендовых и периодических составляющих временного ряда в рамках подхода «Гусеница»-SSA*. Труды IV Международной конференции "Идентификация систем и задачи управления" SICPRO'05. Москва, 2005, с. 1849-1864
- Воротков М.В., Горшков В.Л., Миллер Н.О., Прудникова Е.Я.**, 2002, *Исследование основных составляющих в движении полюса земли*, Изв. ГАО, № 216, с. 406-409.
- Горшков В.Л., Миллер Н.О., Персиянинова Н.Р., Прудникова Е.Я.**, 2000, *Исследование геодинамических рядов методом главных компонент*, Изв. ГАО, № 214, с. 173-180.
- Горшков В.Л.**, 2004, *О методах прогнозирования в геодинамике*, Изв. ГАО РАН, №217, с. 365-378.
- Данилов Д.Л., Жиглявский А.А. (ред.)**, 1997, *Главные компоненты временных рядов: метод «Гусеница»*, СПбГУ, с. 308.
- Степанов Д., Голяндина Н.Э.**, 2005, *Варианты метода «Гусеница»-SSA для прогноза многомерных временных рядов*. Труды IV Международной конференции "Идентификация систем и задачи управления" SICPRO'05. Москва, 2005, с. 1831-1848.

THE FORECASTING OF EARTH ORIENTATION PARAMETERS

BY SINGULAR SPECTRUM ANALYSIS

Gorshkov V.L., Miller N.O.

ABSTRACT

The short description of the prognostic possibilities of the singular spectrum analysis in the "Caterpillar"-SSA program realization is given. The Earth orientation parameters of the IERS were used for search of the optimal parametrization of this forecasting method. Some tens of the forecasts of the polar coordinates and the Earth's rotation data ($UTI-UTC$) were made for that during two year. The quality of these prediction data were compared with similar data of the Prediction Comparison Campaign of the IERS. The main result of this comparison consist in the large forecasting possibilities of SSA method. In many cases the SSA forecasts were better than the better ones of IERS Campaign.