

МЕТОДИЧЕСКИЕ АСПЕКТЫ РЫБОХОЗЯЙСТВЕННЫХ ИССЛЕДОВАНИЙ

УДК 639.2.053.7+573.22.087.1.001.57

**ПРИМЕНЕНИЕ ФИЛЬТРА КАЛМАНА В КОГОРТНОЙ МОДЕЛИ  
ДЛЯ КОРРЕКТИРОВКИ ОЦЕНОК ЗАПАСА  
ПРИ НАЛИЧИИ НЕУЧТЕННОГО ВЫЛОВА**

© 2016 г. А. А. Михеев

*Сахалинский научно-исследовательский институт рыбного хозяйства  
и океанографии, Южно-Сахалинск, 693023  
E-mail: a.mikheev@sakhniro.ru*

Поступила в редакцию 22.09.2015 г.

В статье рассмотрены вопросы анализа и корректировки оценок запаса в когортной модели. Установлено, что при наличии систематических погрешностей в данных о годовом вылове стандартная когортная модель оценивает запас со смещением. Показано, что фильтр Калмана помогает устранить это смещение. Обсуждается, как использовать фильтр Калмана для корректировок оценок запаса и общего допустимого улова при поступлении новых данных. Предложен способ стандартизации индекса запаса с применением указанного фильтра. Попутно рассмотрены некоторые вопросы тестирования моделей.

*Ключевые слова:* когортная модель, фильтр Калмана, оценка запаса.

## ВВЕДЕНИЕ

Рациональная эксплуатация морских биоресурсов невозможна без их количественной оценки и прогнозирования. Истинность данного тезиса является, на наш взгляд, очевидной и, пожалуй, общепризнанной. Меньшего согласия можно ожидать по поводу рассуждений о том, что применение математических моделей делает процедуру обоснования объемов допустимого изъятия по-настоящему полноценной и свидетельствует о высокой технологичности этого процесса. Вместе с тем, оставляя дискуссию по этому вопросу за рамками статьи, укажем только, что полагаем последнее утверждение справедливым.

Так называемые математические модели промысловой биологии, теории рыболовства и оценки запаса (по поводу терминологии см., например: Засосов, 1976; Бабаян, 2000), несмотря на почти вековую историю их применения, как и прежде, все еще нуждаются в совершенствовании. В последние пару-тройку десятилетий пришло осознание

ограниченности наших возможностей в получении идеальной информации о запасе и промысле, осознание того, что с этим придется считаться и учитывать при моделировании. Как следствие, возникли стимулы к развитию моделей, способных работать с зашумленными, неполными, смещенными данными и вместе с тем давать приемлемый результат.

Среди множества проблем, в том числе и концептуальных, присущих моделированию на основе искаженной информации, выделяется та, которая обусловлена так называемыми «ошибками в переменных» (Schnute, 1994). С позиций данной концепции, все расхождения между моделью и наблюдениями приписываются модели. На практике это означает, что результаты моделирования подгоняются как можно ближе к наблюдениям. Такой подход является следствием давней естественнонаучной парадигмы, в рамках которой факты всегда остаются мерилом истинности (Уайтхед, 1990). Однако следует признать, что современная наука, изучающая системы и объекты гораз-

до более разнообразные и сложные, чем те, которые изучала, например, классическая физика, все чаще сталкивается с ситуацией, когда невозможно точно и полно измерить представляющие интерес характеристики таких систем и объектов.

Эксплуатируемый запас (промысловая часть популяции или группы популяций), находящийся в изменчивом экосистемном окружении, также представляет собой весьма сложную систему, плохо поддающуюся контролю и точным измерениям. Мало того, что прямой тотальный учет популяции гидробионтов практически невозможен, но даже и уловы на усилие — из научных съемок и промышленные — далеко не всегда могут служить индексами запаса (Кадильников, 2001). Ошибки в определении возраста и размерная селективность орудий лова, хотя и усугубляют ситуацию, в принципе могут как-то контролироваться. В меньшей степени могут быть учтены ошибки, возникающие при пересчете учетных уловов в плотности запаса, а также ошибки экстраполяции на необловленные участки и интерполяции между станциями. Самый же неприятный момент состоит в том, что данные по годовому вылову, казалось бы, легко контролируемые, помимо случайных погрешностей часто еще и систематически занижаются. Например, в связи с выбросами или браконьерством. Выявить возникающие при этом смещения оценок запаса традиционными методами проблематично.

Большая часть прежде сказанного относится и к так называемым когортным моделям. Таким, например, как классический когортный анализ Поупа (Pope, 1972) или современные: ADAPT (Gavaris, 1988), Stock Synthesis (Methot, 1989), ISVPA (Васильев, 2001) и др. И это при том что модели указанного типа являются на сегодняшний день самыми продвинутыми в арсенале методов оценки запасов гидробионтов (Васильев, 2001).

Чтобы справиться с обозначенной выше проблемой, нужно отойти от концепции «ошибок в переменных» и использовать

фильтр Калмана (ФК) (Schnute, 1994). Как и любой другой, указанный фильтр предназначен для «выделения полезного сигнала» из зашумленного потока информации о состоянии непосредственно не наблюдаемой системы. (Напомним, что в контексте данной статьи такой ненаблюдаемой системой является запас, а наблюдения состоят из уловов, полученных в научных рейсах и на промысле.) Главное отличие ФК от прочих фильтров состоит в его рекуррентной основе, что дает ему огромные вычислительные преимущества. Наиболее известным на сегодня применением ФК является система ориентации головного и спускаемого модулей в исторической миссии «Аполлона XI» к Луне в 1969 г., разработанная NASA при личном участии Р. Калмана.

Оценка состояния системы в ФК определяется как взвешенная сумма наблюдений и основанного на модели прогноза (Kalman, 1960). Таким образом, общая неопределенность раскладывается на ошибки в наблюдениях и ошибки в переменных состояния. Доверительные интервалы оценок переменных состояния соответственно сужаются и довольно существенно.

Данная статья посвящена анализу ошибок в оценках обилия гидробионтов, возникающих в когортной модели при использовании искаженных данных о вылове. Цель работы — показать, как можно с помощью ФК скорректировать смещенные оценки запаса, используя данные по индексам запаса из различных независимых источников, включая промысловую статистику и научные съемки.

Одной из важных технических задач, требующих решения на подготовительном этапе исследования, стало преобразование уравнений ФК к виду, позволяющему обрабатывать пропущенные значения в наблюдениях. Постановка этой задачи обусловлена тем, что сбор информации об индексах запаса в научных съемках и на промысле в отдельные годы и периоды лет может не выполняться. Однако стандартное требование к данным при работе с ФК состоит в том,

чтобы они поступали на вход фильтра регуляционно (Harvey, 1989). Для одного источника информации подобная задача решается достаточно просто (Михеев и др., 2012). Появление в модели нескольких источников делает эту задачу менее тривиальной.

Еще одна непростая задача, вставшая в ходе исследования, заключалась в том, как соотносить друг с другом значимость каждого источника информации. Проблема взвешивания информационных вкладов от различных источников в так называемую функцию потерь (ФП), или целевую функцию, как ее определяют в теории управления, рассматривается в литературе достаточно давно (Cavaris, 1988), но полностью не решена до сих пор.

## МАТЕРИАЛ И МЕТОДИКА

Основная задача исследования состояла в том, чтобы проанализировать устойчивость модельных оценок к входным данным, содержащим систематические смещения и шум с аномальными значениями (аутлайерами). Отсюда выстраивалась методика исследования: первоначально имитировалась «истинная» динамика запаса и искаженная динамика вылова. Кроме того, производился ряд «истинных» наблюдений, который зашумлялся. Затем сравнивалось то, как когортная модель с ФК (модель получила название «когортный анализ с фильтром Калмана (КАФКА)») и без него оценивает «истинный» запас по искаженному вылову и зашумленным наблюдениям. Для имитации динамики запаса, уловов и наблюдений специально разработали операционную модель (приложение 2).

Сгенерированная матрица уловов охватывала период в 20 лет и 10 возрастных классов. Возраст пополнения был принят в 6 лет. В «истинные» уловы были внесены искажения двух типов: либо это был чистый шум различной интенсивности, как правило, логарифмический гауссовский, либо в указанный шум вносили аутлайеры или систематическое отрицательное смещение, которое по замыс-

лу исследования имитировало неучтенное изъятие или выбросы.

Наблюдения были сгенерированы в виде индексов запаса, полученных из различных информационных источников (приложение 2). Было предусмотрено три набора данных: 1) количество икры, обычно определяемое по сетным уловам ихтиопланктонных съемок; 2) промышленные уловы на единицу промыслового усилия (CPUE); 3) уловы на траление из учетных съемок. Данная конкретизация источников информации является непринципиальной. Принципиально здесь то, что каждый из указанных источников характеризовался своим переходным коэффициентом или функцией (например, коэффициентом улавливаемости для CPUE) и — в общем случае различными — погрешностями измерений. Количество источников при этом может быть произвольным. В наблюдаемые индексы запаса так же, как и в годовые уловы, включались логарифмические гауссовские шумы. Кроме того, в отдельные годы и периоды лет индексы запаса по какой-либо комбинации источников опускались, что является, как уже отмечалось выше, распространенным случаем на практике. Для сравнения с «истинными» значениями запаса его индексы масштабировались через соответствующие коэффициенты (например, коэффициенты улавливаемости).

Уравнения используемой в данной статье когортной модели аналогичны тем, что приведены в монографии Васильева (2001) с некоторыми изменениями (приложение 3). В частности, одно из отличий модели КАФКА состояло в том, что перед применением ФК матрица запаса свертывалась суммированием по возрастным классам (приложение 3, формула (ПЗ.4)). После работы фильтра запас для каждого года снова раскладывали на возрастные классы и таким образом получали скорректированную матрицу запаса с добавлением двух столбцов для прогнозируемых лет (приложение 3, формулы (ПЗ.11) — (ПЗ.16)).

Качество оценок определяли по взвешенной ФП. Были построены две

ФП — среднеквадратичное и медианное логарифмические отклонения модельных и наблюдаемых индексов запаса по годам и источникам данных (приложение 3). Медианную ФП использовали с тем, чтобы сделать результаты моделирования устойчивыми к влиянию аутлайеров в данных, как рекомендуется некоторыми исследователями (Васильев, 2001). Веса, устанавливающие информативную значимость каждого из источников данных в ФП, были выведены из уравнений ФК (приложение 4).

Параметрическую настройку модели выполняли, минимизируя ФП с помощью метода эволюционного стохастического поиска, известного как генетический алгоритм (ГА) (Рутковская и др., 2004; Михеев и др., 2006).

Уравнения ФК были модифицированы таким образом, чтобы присвоение пропущенным данным нулевых значений не нарушало корректность работы фильтра и не искажало результаты, как если бы эти нулевые значения принимались за нулевые индексы запаса (приложение 5). Вывод уравнений ФК в достаточно подробном виде опубликован ранее (Михеев, 2014), а сами уравнения приведены в приложении 4.

Отсутствующие в тексте пояснения к используемым символам следует искать в приложении 1. Поскольку в исследовании рассматривался только промысловый запас, будем далее — там, где это не вызовет недоумений, — называть его просто запасом.

## РЕЗУЛЬТАТЫ

На подготовительном этапе были получены необходимые условия для корректной работы ФК с пропущенными значениями из различных источников информации. В приложении 5 показано, что если информация из  $i$ -го источника для момента времени  $t$  отсутствует, то его влияние на результат для этого момента времени устраняется с помощью равенств  $I_{i,t} = 0$  и  $H_{i,t} = 0$  (приложение 4, формула (П4.4)). Уравнения ФК при этом модификации не требуют, а сочетание индексов  $i$  и  $t$  может быть произвольным.

Это означает, что пропущенные значения могут относиться как к отдельным годам, так и к периодам лет, и для различных источников данных могут не совпадать.

Еще один технический результат состоял в определении весов, устанавливающих информативную значимость каждого из источников данных. Для каждого  $i$ -го источника информации и каждого года  $t$  веса значимости  $w'_{i,t}$  могут быть определены через так называемые коэффициенты усиления  $K_{i,t}$  и переходные коэффициенты  $H_{i,t}$ , вычисляемые в ФК (приложение 4, формулы (П4.5), (П4.11)):

$$w'_{i,t} = \frac{K_{i,t} H_{i,t}}{\sum_i K_{i,t} H_{i,t}}, \quad i = 1, 2, 3.$$

Веса  $w'_{i,t}$  позволяют корректно строить ФП для параметрической оптимизации моделей вообще и для когортной модели в частности (приложение 3, формулы (П3.8), (П3.9)).

Кроме того, с помощью этих весов можно вычислить взвешенный индекс запаса  $y_{w_i}$  для года  $t$  (приложение 4, формула (П4.12)):

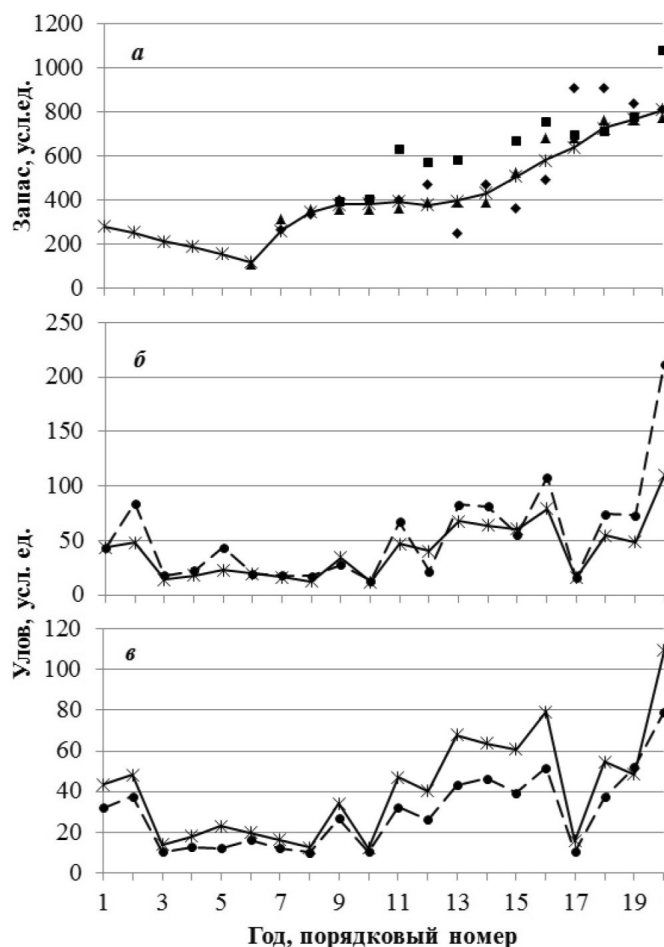
$$y_{w_i} = \sum_i w'_{i,t} y_{i,t},$$

где  $y_{w_i}$  — масштабированные индексы:

$$y_{i,t} = \frac{I_{i,t}}{H_{i,t}} \quad (\text{приложение 4}).$$

На основе операционной модели были сгенерированы ряды «истинного» запаса и зашумленные наблюдения (приложение 2). Перечисленные ряды представлены на рис. 1, а.

«По легенде», первые 5 лет промысла наблюдения за запасом отсутствовали полностью. В последующий период наблюдения за некоторыми индексами запаса отсутствовали в отдельные годы (рис. 1, а). Сгенерированные с помощью операционной модели «истинные» и искаженные ряды по годовому вылову за тот же период времени без систематического смещения и с оным показаны на рис. 1, б, в. Вариант со смещением был предназначен для имитации неучтенного вылова (рис. 1, в; приложение 2).



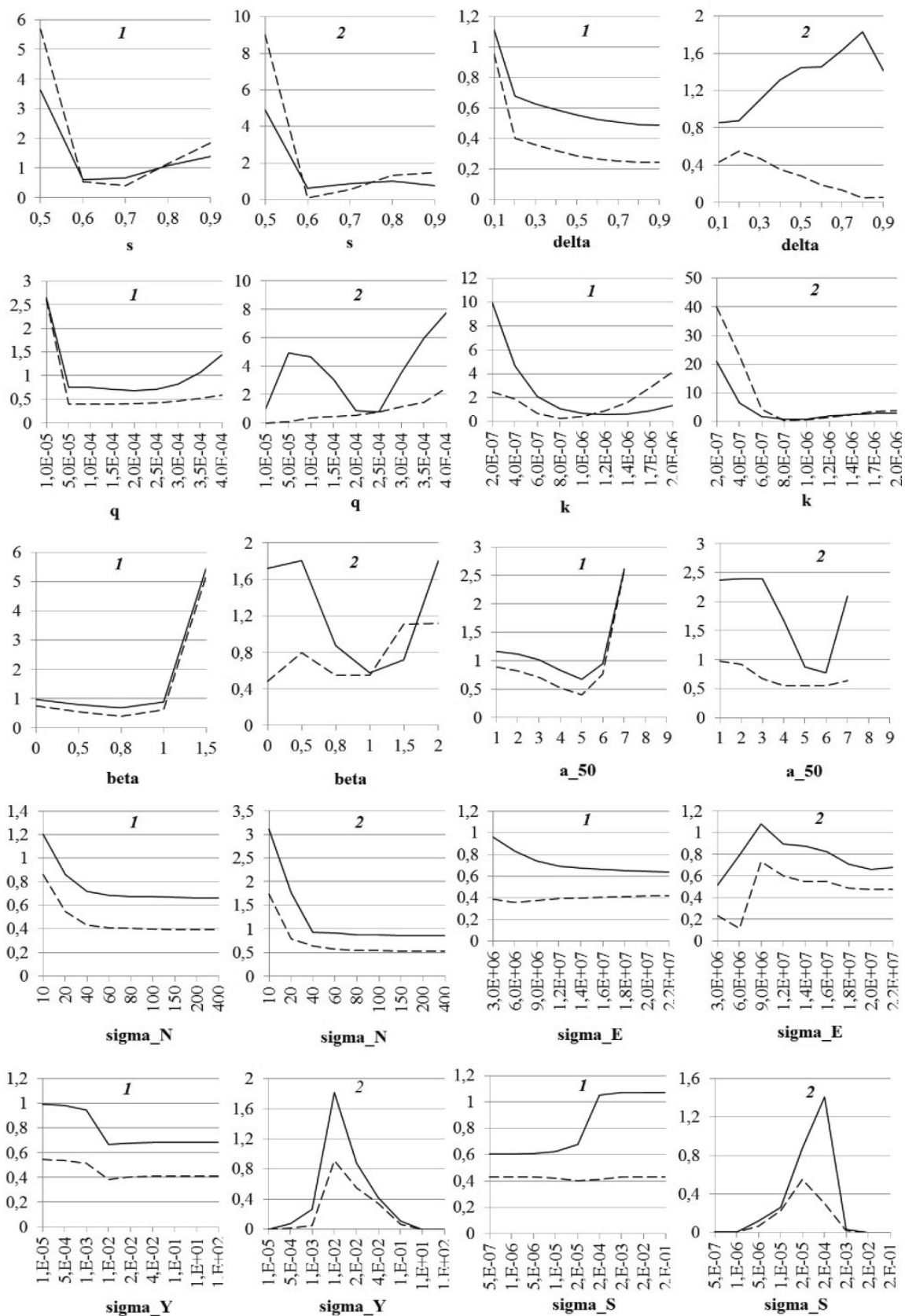
**Рис. 1.** Ряды исходных данных: *a* — масштабированные индексы запаса по трем источникам информации на фоне «истинных» значений: (—Ж—) — «истинный»; (■) — ихтиопланктонный; (◆) — СРУЕ; (▲) — учетный траловый; *b, в* — уловы «истинные» (—Ж—) и искаженные (---●---) стохастическим шумом с аутлайером и с систематическим отрицательным смещением соответственно.

Перед тем как приступить к моделированию динамики запаса, исследовали поведение среднеквадратичной и медианной ФП (приложение 3, формулы (ПЗ.8) и (ПЗ.9) соответственно). Результаты этого исследования показаны на рис. 2.

Следует отметить, что построенные профили ФП не обязательно проходят через ее глобальный минимум. Тем не менее, видно, что характер изменения этого функционала по большинству параметров модели делает поиск глобального минимума ФП очень затруднительным, а порой и невыполнимым в связи с отсутствием четкого сигнала. При этом медианная ФП в ряде случаев

дает парадоксальный результат, демонстрируя максимум там, где можно было бы ожидать минимума. Это касается ее поведения относительно таких параметров, как  $\Delta$ ,  $q$  и всех  $\sigma$ , связанных с наблюдениями, —  $\sigma_Y$ ,  $\sigma_E$  и  $\sigma_S$  (рис. 2). Почти в половине случаев ее оценочные профили были дальше от «истинных» по сравнению со среднеквадратичной ФП.

Таким образом, анализ профилей ФП показывает, что использование медианной ФП в лучшем случае не дает никаких видимых преимуществ в параметрической оптимизации модели КАФКА, а в худшем приводит к неверным результатам. В связи



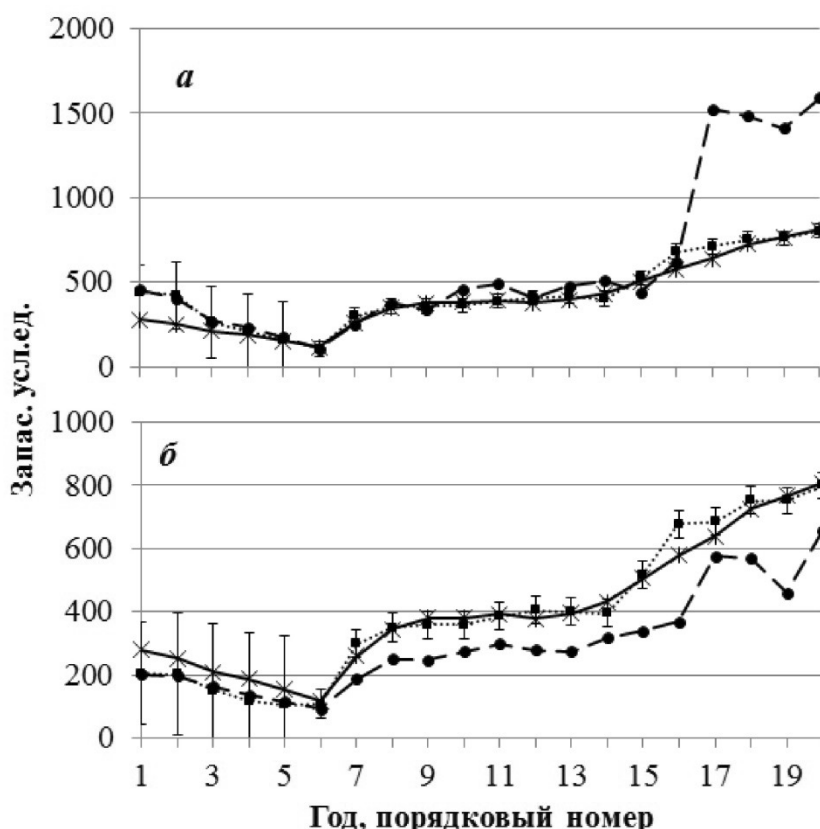
**Рис. 2.** Профили среднев квадратичной (1) и медианной (2) функций потерь для каждого параметра модели КАФКА относительно наблюдений (—) и «истинного» запаса (---) соответственно.

с этим было принято решение: исследовать влияние случайных и систематических смещений в данных по уловам на оценку запаса на основе только среднеквадратичной ФП.

Чтобы при моделировании динамики запаса не отвлекаться на оптимизацию по той части параметров, которые связаны с когортным анализом, из операционной модели были взяты их «истинные» значения (приложение 2). Поскольку фильтр (приложения 1, 3) в генерировании исходных данных не использовался, «истинных» значений параметров он не имеет. Поэтому их требовалось оценить, что и было сделано с помощью ГА. Полученные оценки параметров ФК по результатам оптимизации с применением ГА указаны ниже наряду с соответствующим значением среднеквадратичной ФП ( $L$ ), усл. ед.

- $\rho_0 - 10^2$
- $\sigma_N - 80$
- $\sigma_E - 1,4 \times 10^7$
- $\sigma_Y - 0,02$
- $\sigma_S - 2 \times 10^{-5}$
- $L - 0,785$

На рис. 3 приведены смоделированные ряды запаса на фоне «истинных» значений. В качестве исходных, как уже упоминалось, были использованы данные, показанные на рис. 1. Рассматривалось два варианта: с аутлайером в шуме на конце ряда и с отрицательным систематическим смещением, добавленным в шум умеренной



**Рис. 3.** Динамика оценок запаса, полученных с фильтром Калмана (.....■.....) и без него (---●---), на фоне «истинных» значений (—ж—) в случае добавления стохастического шума: *a* — с аутлайером; *b* — с систематическим отрицательным смещением; показаны 95%-ные доверительные интервалы.

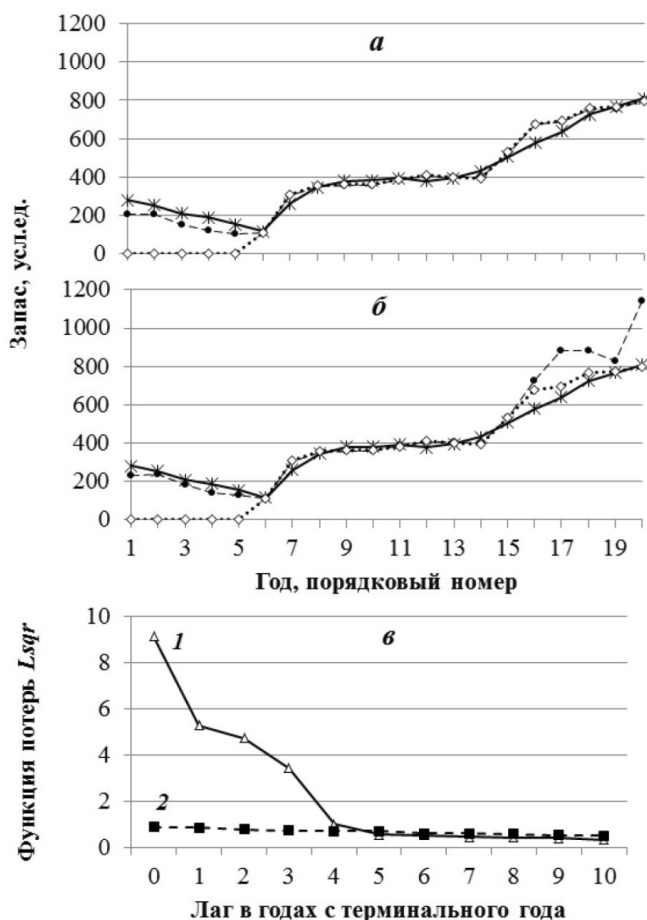
интенсивности. Моделирование выполняли с помощью только когортного анализа и когортного анализа с добавлением ФК.

Из результатов моделирования можно видеть, что качество подгонки к «истинным» значениям между упомянутыми моделями существенно различается (рис. 3).

Наличие аутлайера на конце ряда данных о годовом вылове приводит к сильному завышению оценок запаса в последние четыре условных года в случае моделирования без ФК. При использовании отрицательно смещенных данных о годовом вылове запас получается заниженным, если его оценивание также производится средствами только когортного анализа (рис. 3, б). ФК со своей стороны справляется с «плохими» данными

и выдает несмещенные оценки запаса. Кроме того, следует отметить, что отсутствие наблюдений в первые пять условных лет (рис. 1) в разы увеличивает доверительный интервал оценок запаса (рис. 3).

В рамках исследования был изучен традиционный для когортных моделей вопрос об устойчивости оценок к неопределенности в терминальных коэффициентах эксплуатации. На рис. 4 представлена динамика оценок запаса при оптимальных по ФП терминальных коэффициентах эксплуатации и аналогичных коэффициентах, уменьшенных более чем вдвое. При моделировании в данные по годовому вылову был добавлен стохастический шум и систематическое отрицательное смещение.



**Рис. 4.** Динамика скорректированных оценок запаса (—●—) и индекса  $y_w$  (····◇····) на фоне «истинных» значений (—ж—): а — при оптимальном значении  $a_{50} = 5$  лет ( $L = 0,8$ ) (2); б — при  $a_{50} = 9$  лет ( $L = 9,2$ ) (1); в — изменение функции потерь при последовательном сокращении ряда.



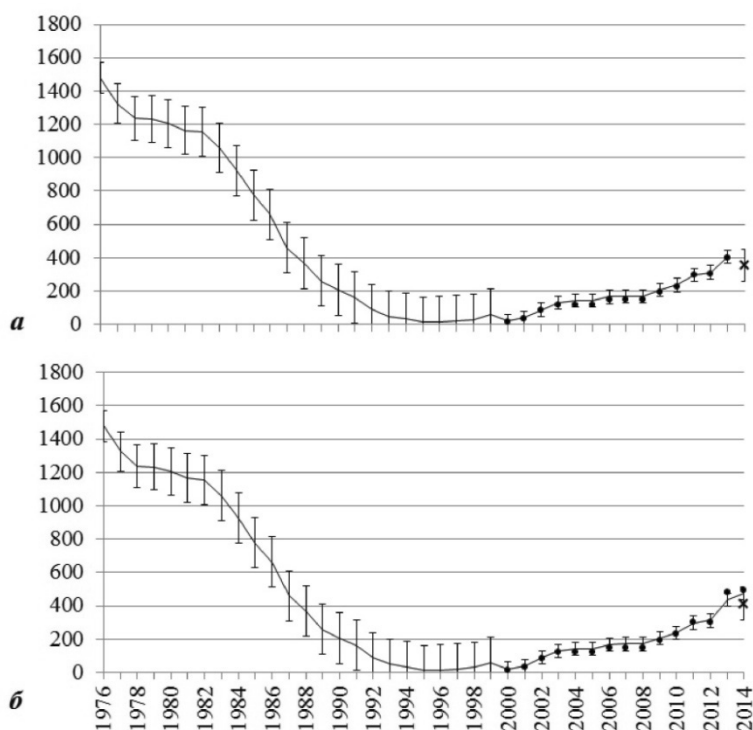
Напомним, что величина терминальных коэффициентов эксплуатации определена через два параметра, один из которых  $a_{50}$  (приложение 3). Указанный параметр задает возрастной класс, из которого изымается промыслом половина особей.

Было установлено, что при увеличении  $a_{50}$  (т.е. при уменьшении коэффициентов эксплуатации для большего числа ранних возрастных классов) оценки запаса завышаются, но только по тем последним четырем годам, для которых запас включал несколько поколений, только начинающих облавливаться (напомним, что в каждом поколении моделировалось 10 возрастных классов) (рис. 4, б). Взвешенный масштабированный индекс запаса  $y_w$  при этом проявлял устойчивость к изменениям в терминальных коэффициентах эксплуатации и варьировался незначительно (рис. 4, а, б). Профили ФП также подтверждают, что модель была устойчива

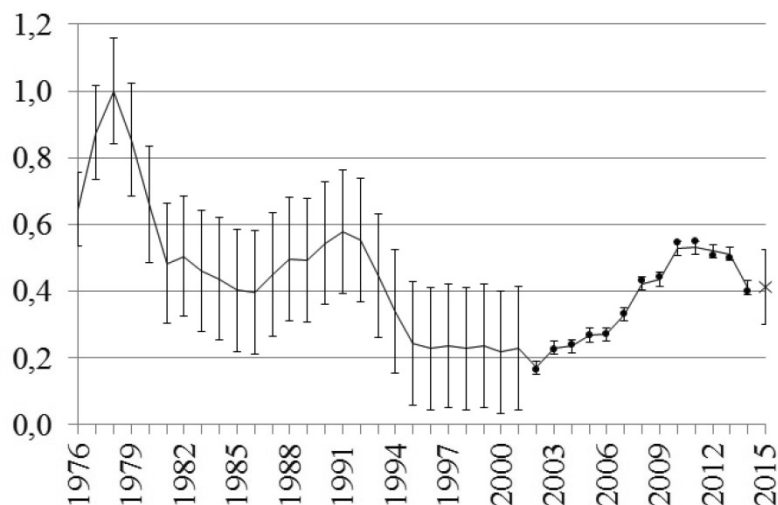
к изменениям в терминальных коэффициентах эксплуатации при исключении из рассмотрения четырех последних лет промысла (рис. 4, в).

Чтобы продемонстрировать, как модель КАФКА может быть применима к корректировке оценок запаса, представим одно ее реальное приложение. Оно касается корректировки прогноза запаса и ОДУ минтая подзоны Восточно-Сахалинской в 2014 г. Результаты моделирования для указанного случая показаны на рис. 5.

Первоначальная прогнозная оценка указанного запаса на 2014 г., по данным ихтиопланктонной съемки на НИС «Дмитрий Песков» (СахНИРО), выполненной в 2012 г., составила 354,4 тыс. т (рис. 5, а), а соответствующий ОДУ был определен согласно правилам регулирования промысла (ПРП) на уровне 18% от запаса, или 64 тыс. т в абсолютном выражении. Ана-



**Рис. 5.** Динамика промыслового запаса минтая подзоны Восточно-Сахалинской, тыс. т: скорректированная оценка (—); взвешенные масштабированные индексы запаса (●); прогноз на 2014 г. (×): а — по съемкам СахНИРО 2012—2013 гг.; б — по съемкам ТИНРО-центра 2013—2014 гг.; показаны 95%-ные доверительные интервалы.



**Рис. 6.** Динамика промыслового запаса минтая зоны Восточно-Камчатской, доли от максимума скорректированной оценки: скорректированная оценка (—), взвешенные масштабированные индексы запаса (●), прогноз на 2015 г. (×); показаны 95%-ные доверительные интервалы.

логичная съемка, произведенная в 2013 г., не потребовала корректировки ОДУ. Однако, по данным промысловой статистики (ИС «Рыболовство»), в первой половине 2014 г. было выловлено почти 60 тыс. т, т.е. практически весь ОДУ. При этом улов на судо-сутки составлял около 75 т, что в историческом плане соответствовало высоким значениям запаса. Новые данные по СРУЕ и материалы ихтиопланктонных съемок, проведенных в 2013 и 2014 гг. на НИС «Профессор Кагановский», любезно предоставленные коллегами из ТИНРО-центра, позволили ФК скорректировать оценку запаса до 475,5 тыс. т в 2014 г. (рис. 5, б). Тем самым ОДУ на 2014 г., согласно ПРП, был определен в 24% от запаса (115 тыс. т) и до конца года был освоен не менее чем на 97%.

Результаты еще одного приложения модели КАФКА, впрочем, как и предыдущего, иллюстрируют преимущества ФК перед концепцией «ошибок в переменных». На рис. 6 показаны результаты моделирования динамики запаса минтая зоны Восточно-Камчатской.

Как и в случае с минтаем подзоны Восточно-Сахалинской, интервальная оцен-

ка для периода лет с отсутствующими наблюдениями многократно превышает аналогичную оценку для периода лет с наблюдениями (рис. 5, 6). Следует отметить, что интервальная оценка для периода лет с отсутствующими наблюдениями превышает и прогнозируемый интервал.

## ОБСУЖДЕНИЕ

Нетрудно задаться вопросом: зачем нужна еще одна когортная модель, когда проверенных на сегодня существует десяток-другой? Истоки ответа на данный вопрос лежат, на наш взгляд, в истории развития когортных моделей. Достаточно подробный исторический экскурс по этой теме имеется в литературе (см., например: Васильев, 2001; Михеев, 2003), и поэтому выносится за рамки статьи. Вкратце же ответ на поставленный вопрос можно сформулировать следующим образом. Во-первых, когортные модели совершенствуются постоянно, впитывая в себя новые достижения в области анализа данных. Модель КАФКА в этом смысле не является исключением. Она соединила в себе две новации, не свойственные существующим моделям данного класса: оптимизатор на базе генетиче-

ского алгоритма и ФК. Во-вторых, хотелось получить когортную модель, не обремененную громоздкими настройками, и достичь в ней максимума простоты без потери вычислительной и оценочной эффективности. Насколько данный замысел удался — судить читателю. В любом случае ФК обеспечивает, насколько это возможно, указанную эффективность.

Если как альтернативу рассматривать когортные модели процессов, то при значительно большей реалистичности в сравнении со стандартными версиями их нелегко применять. Причина — в большом числе уравнений и параметров в таких моделях (Михеев, 2003). Отсюда проистекают сложности решения оптимизационной задачи, имеющей в данном случае очень высокую размерность. Отсюда же неоправданно колоссальные затраты времени на вычисления и, как следствие, медленное накопление различных сценариев, возникающих в процессе проверки тех или иных предположений.

Переходя собственно к обсуждению полученных результатов, начнем с того, что был однозначно прояснен ряд свойств уравнений когортной модели в целом и модели КАФКА в частности. Первоначально в модели КАФКА предусматривалось наряду с обычным еще и сепарабельное представление коэффициентов эксплуатации. Согласно рекомендациям, расчет сепарабельных коэффициентов эксплуатации ведется следующим образом (Васильев, 2001). Фактор года вычисляется по формуле:

$$f_t = \sum_{a=1}^A \varphi_{at}$$

Отсюда можно найти фактор возраста (по сути, коэффициент селективности) по формуле:

$$s_a = \sum_{t=1}^T \varphi_{at} / \sum_{t=1}^T f_t, a = \overline{1, A},$$

если выполняется условие нормировки  $\sum_{a=1}^A s_a = 1$  (Васильев, 2001. С. 29. Формула (2.5)). Это, на наш взгляд, довольно жесткое и неясное требование. Окончательное сепарабельное представление получается перемножением факторов года и возраста:

$$\varphi_{sep_{at}} = s_a f_t.$$

Как оказалось, при таком сепарабельном представлении и так называемом обратном решении для ряда поколений в отдельных возрастных классах, как правило старших, возникают некорректные оценки коэффициентов эксплуатации, превышающие единицу. Результаты подобных расчетов показаны на рис. 7.

Вместе с тем сама по себе гипотеза о сепарабельности представляется весьма разумной не только потому, что позволяет понизить размерность оптимизационной задачи, но и с содержательной точки зрения: паттерн селективности для однородных промысловых усилий, скорее всего, должен быть сходным. И все же из уравнений КАФКА сепарабельное представление коэффициентов эксплуатации было исключено.

Исследование робастных свойств медианной ФП показало, по крайней мере, отсутствие ее видимых преимуществ перед среднеквадратичной ФП. В ряде случаев медианная ФП давала парадоксальный результат, демонстрируя максимум там, где можно было бы ожидать минимум (рис. 2). В то же самое время у среднеквадратичной ФП оценочные профили были более близки к истинным.



**Рис. 7.** Профили матрицы сепарабельных коэффициентов эксплуатации  $\varphi_{sep_{at}}$ , рассчитанных по стандартным уравнениям когортного анализа для данных, сгенерированных с помощью операционной модели; годы: (—●—) — 1-й; (---□---) — 2-й; (····Δ····) — 20-й.

Попытки достичь при использовании ФК устойчивости оценок к аутлайерам в данных наблюдений (не в уловах, а в индексах запаса) за счет повышения их робастности, по-видимому, избыточны. Действительно, отклонения в наблюдениях за индексом запаса от «истинных» значений, включая аутлайеры (рис. 1, а, например, в ихтиопланктонном индексе за период лет с номерами 11–15 или в СРУЕ за годы 13, 15, 17 и 18), статистически эффективно сглаживаются фильтром (рис. 3). Зачем тогда применять к наблюдениям еще и различные усечения или проецирования «хвостов» распределений? Наверное, эти процедуры будут уместны в тех когортных моделях, в которых не используются ФК либо иной статистически эффективный метод оценивания и сглаживания данных. Так или иначе, но в окончательную версию модели КАФКА медианная ФП включена не была. Приложения данной модели к оцениванию и прогнозированию реальных запасов основаны на использовании среднеквадратичной ФП (приложение 3, формула (П3.8)).

Еще одним значительным отличием модели КАФКА от прочих когортных моделей является свертывание матриц запаса перед применением ФК и их последующее развертывание после коррекции оценок (приложение 3, формула (П3.4)). Это позволило использовать скалярные уравнения наблюдений. Указанная процедура была во многом необходима, поскольку индексы запаса, как правило, определяются без возрастной структуры. Для СРУЕ возрастной состав в любом случае был бы идентичен возрастному составу уловов. Таким образом, в уравнения наблюдений все равно пришлось бы подставлять запас без разложения его на возрастные классы. В данном случае свертывание матрицы запаса в уравнении процесса (приложение 3, формула (П3.4)) с последующей корректировкой оценок и развертыванием практически эквивалентно уравнению процесса в форме пространства состояний (Harvey, 1989), где каждое состояние относилось бы к отдельному воз-

растному классу, а расчет велся бы отдельно по поколениям. С другой стороны, исключение матричных вычислений за счет свертывания матрицы запаса и применения уравнения процесса в скалярном виде значительно ускоряет работу фильтра.

Тестирование модели КАФКА показало, что у нее в целом неплохая точность. Несмотря на разнообразные искажения, внесенные в исходные данные, указанная модель оценивала запас близко к его «истинному» значению (рис. 3). Самый же, на наш взгляд, важный из полученных результатов состоит в том, что при наличии несмещенных наблюдений оценки запаса также являются несмещенными, несмотря на систематическое занижение данных о годовом вылове. Заметим, что само по себе использование несмещенных данных по индексам запаса — без применения ФК — полностью не решает проблему, так как без данного фильтра проблематично оценить погрешность измерений.

Если имеется несколько независимых источников поступления информации о среднегодовых индексах запаса или есть всего один источник, и он дает несмещенные наблюдения, то шансы получить несмещенную оценку практически стопроцентны. Трудно ожидать, что указанные индексы, выраженные, например, через средние уловы научных съемок, могут быть смещены. Действительно, при регулярной сетке станций вероятность попадания на пятна с определенным уровнем биомассы не может не меняться по годам. Как следствие, более правдоподобными в рассматриваемых индексах видятся случайные ошибки. С еще большим основанием данный тезис применим к СРУЕ. Довольно проблематично, на наш взгляд, чтобы СРУЕ, например, занижались или завышались из года в год в течение длительного срока. Таким образом, достаточно правдоподобной гипотезой видится утверждение о случайных погрешностях в измерениях индекса запаса, т.е. о несмещенности этих измерений.

Особую роль в этой связи приобретают так называемые стандартизированные индексы, или стандартизированные СРУЕ.

Здесь уместно еще раз отметить, что взвешенный индекс запаса  $y_{w_i}$ , вычисляемый с помощью ФК по формуле (П4.12) (приложение 4), представляет собой статистически эффективно сглаженную оценку. В таком случае если под каждым источником информации рассматривать отдельное промысловое судно, то указанный или подобный ему взвешенный индекс, на наш взгляд, может рассматриваться в качестве стандартизированного индекса запаса. Вместе с тем его вычисление происходит в процессе моделирования и средствами самой модели, а не на его предварительном этапе независимым методом (например, с помощью обобщенных линейных моделей GLM), как того требуют когортные модели без ФК. И это, на наш взгляд, преимущество. Представляется интересным сравнить индексы, полученные на одних и тех же данных, — взвешенный с помощью ФК и стандартизированный на основе GLM. Но это задача на будущее.

Здесь будет также уместным сказать пару слов о том, как в данном исследовании была решена проблема взвешивания вкладов в ФП от нескольких источников наблюдений. Напомним, что в ФК каждый источник наблюдений взвешивается, а веса определяются через коэффициенты усиления Калмана и соответствующие переходные функции  $H_{i,t}$  и нормируются на единицу (приложение 4, формулы (П4.11), (П4.12)):

$$w'_{i,t} = \frac{K_{i,t}H_{y,t}}{\sum_i K_{i,t}H_{i,t}}, i = 1, 2, 3, \text{ (П4.11)}$$

$$y_{w_i} = \sum_i w'_{i,t} y_{i,t}. \text{ (П4.12)}$$

Вместе с тем формулу (П4.1) можно интерпретировать таким образом, что с вероятностью  $w_i$  скорректированная оценка запаса  $\hat{N}_i$  примет прогнозное значение  $N_i^+$  (полученное на этапе выполнения когортного анализа), а с вероятностью  $(1-w_i)$  будет равна наблюдению  $y_i$ .

$$\hat{N}_i = w_i N_i^+ + (1-w_i) y_i. \text{ (П4.1)}$$

Но, как следует из формулы (П4.6) и условий независимости источников наблюдений, эта последняя вероятность раскладывается на компоненты  $K_{i,t}H_{i,t}$ . Иначе говоря, сумма  $\sum_i K_{i,t}H_{i,t}$  — это полная вероятность того, что переменная  $\hat{N}_i$  будет равна наблюдению  $y_i$ .

$$\sum_i K_{i,t}H_{i,t} + w_i \equiv 1. \text{ (П4.6)}$$

Отсюда, на наш взгляд, вытекает, что нормированные компоненты указанной вероятности вполне могут играть роль весов для каждого из источников наблюдений в ФП (приложение 3, формула (П3.8)). Таким образом, веса значимости  $w'_{i,t}$  позволяют достаточно корректно строить многокомпонентную ФП для параметрической оптимизации моделей с несколькими источниками входных данных и, в частности, для когортной модели.

Отличительная особенность моделирования на основе ФК ярко проявляется в ситуации, когда временной ряд наблюдений включает пробелы, как, например, ряды на рис. 1, сгенерированные с помощью операционной модели. Отсутствие наблюдений за индексами запаса в первые пять условных лет (рис. 1) обуславливает широкий доверительный интервал оценок запаса, который в разы уменьшается, как только появляются наблюдения (рис. 3). Сходным образом ведут себя доверительные интервалы и в случаях приложений КАФКА к реальным данным (рис. 5, 6).

Перечисленные примеры наглядно иллюстрируют преимущества от реализации ФК в сравнении с концепцией «ошибок в переменных». Способность ФК вычислять ошибки измерения существенно снижает неопределенность в оценках при наличии наблюдений. В случае, когда наблюдения отсутствуют, на скорректированные оценки ложится вся неопределенность. Там, где наблюдения имеются, общая неопределенность разделена между ошибками наблюдений и ошибками модели. Как следствие, в скорректированных оценках ФК неопределенность сокращается, поскольку ошибки

измерений оцениваются фильтром и из указанных оценок исключаются. Неопределенность в оценках для периода лет без наблюдений всегда будет выше и неопределенности прогноза, порой существенно. Это объяснимо тем, что в прогноз не включена ошибка еще не сделанных наблюдений.

Далее рассмотрим вопрос ежегодных корректировок ОДУ. Согласно нормативным документам, допустимы ежегодные корректировки прогнозируемых ОДУ при условии поступления новых данных о состоянии запаса и/или промысла. Поскольку ПРП при этом остаются неизменными, то, как правило, корректировка ОДУ эквивалентна корректировке оценки запаса. Алгоритм коррекции прогнозных оценок в ФК также построен на новых наблюдениях (приложение 4, формула (П4.9)). Пример корректировки прогноза реального запаса с помощью ФК показан на рис. 5 и там же вкратце обсуждается. Заметим, что при выполнении коррекции можно учитывать не только каждое новое наблюдение за индексом запаса, но даже некоторое множество новых наблюдений, полученных из нескольких источников (промысловых или научно-исследовательских судов).

Выполненный в данной статье анализ высвечивает, на наш взгляд, ряд базовых элементов в принципиальной схеме тестирования любых моделей промысловой биологии. Краеугольным камнем указанной схемы должны стать некие «истинные» данные. Почему? Дело в том, что первоочередной целью названных моделей является оценка запаса, а он непосредственно не наблюдаем. На сегодняшний день это факт, с которым приходится считаться. Как бы хорошо модель ни воспроизводила наблюдения, о ее точности все равно ничего нельзя будет сказать с полной уверенностью. Причина в том, что наблюдения связаны с запасом еще одной моделью (второго уровня, так сказать), выраженной уравнением наблюдения, которое содержит свою ошибку (приложение 4, формула (П4.4)). Следовательно, на начальном этапе тестирования нужно сгенерировать «истинные» данные о запасе и вылове

и получить соответствующие наблюдения. Затем внести искажения в уловы и наблюдения и по ним с помощью модели попытаться восстановить «истинные» данные о запасе.

Отсюда неизбежно вытекают вопросы, касающиеся разработки операционной модели, т.е. модели, с помощью которой генерируются исходные данные. Например: какие уравнения должны быть положены в основу производства «истинных данных»? Как должны зашумляться и смещаться генерируемые «истинные» данные и параметры? Какие из параметров должны быть фиксированы, а какие могут варьироваться произвольно? Насколько должны сочетаться простота и реалистичность модели? Очевидно, что первый и четвертый вопросы тесно связаны. Скорее всего, операционная модель должна быть достаточно реалистичной и, как следствие, сложной. Ее уравнения должны описывать все основные процессы воспроизводства, роста и убыли. Иначе говоря, это должна быть модель процессов (Михеев, 2003). Поскольку смертность от причин, не связанных с промыслом, непрерывно и синхронно связана с промысловой смертностью, то наиболее подходящими уравнениями для убыли являются известные уравнения Ф.И. Баранова. В уравнения воспроизводства лучше не закладывать никаких предположений о зависимости типа запас—пополнение, а пополнение описывать напрямую — через соотношение полов, плодовитость и смертность на ранних стадиях жизненного цикла. Уравнения роста достаточно хорошо изучены и могут быть напрямую включены в операционную модель. Но для целей тестирования, на наш взгляд, достаточно обойтись и средними навесками по возрастным классам и полам или вообще по всей промысловой части популяции. Параметры, относительно которых с достаточной точностью известны реальные значения, лучше фиксировать, а остальные варьировать в заданных пределах.

Необходимо составить план тестирования, к которому следует приложить полное подробное описание модели и инструкцию по использованию. Для тестирования

каждой конкретной модели целесообразно создавать в рамках Рабочей группы по методологии комиссию, которая совместно с разработчиками модели принимает по акту соответствующую операционную модель. Два представителя этой комиссии с помощью указанной операционной модели генерируют «истинные» данные и зашумляют их. Запечатанный пакет (компьютерный архив — зашифрованный или с паролем) с «истинными» и зашумленными данными по акту передается председателю комиссии на хранение. Зашумленные данные передаются разработчикам, которые применяют к ним свою модель. Результаты сравнения «истинных» данных и их смоделированных оценок рассматриваются на комиссии с участием разработчиков, а затем принимается коллегиальное решение о результате тестирования и рекомендации по применению модели.

И только пройдя первую ступень отбора, тестируемая модель может быть допущена к проверке с привлечением реальных данных. Требования на этом уровне тестирования уже должны быть иными. Главным испытанием для модели здесь должна стать проверка на адекватность и реалистичность. В этом случае на смену строгому анализу, возможно, должны прийти заключения экспертов и пользователей. Окончательное решение о прохождении теста необходимо принимать коллегиально, всеми членами рабочей группы или ее профильной комиссии.

## ЗАКЛЮЧЕНИЕ

Подводя итог проделанным исследованиям, сформулируем их основные результаты.

Анализ свойств уравнений модели показал, что они обеспечивают корректную процедуру обратного расчета для коэффициентов изъятия в когорте.

Выполненные тесты позволили сделать вывод о том, что применение ФК позволяет получать устойчивые и несмещенные оценки запаса даже в случае систематических искажений данных о годовом вылове

и при наличии в них аутлайеров, а также при зашумлении наблюдений.

С помощью разработанной модели была реконструирована динамика запасов минтая подзоны Восточно-Сахалинской в период 1976—2014 гг. и зоны Восточно-Камчатской в период 1975—2014 гг., а также сделан прогноз. Тем самым были получены результаты, которые могут быть рассмотрены в сравнении с ранее найденными оценками для указанных запасов. Такое сравнение будет полезным для уточнения будущих прогнозов.

Напрашивается вывод о том, что использование ФК в моделях промысловой биологии и, в частности, в когортной модели делает процедуру корректировки ОДУ теоретически обоснованной, т.е. до определенной степени объективной.

Результаты тестирования модели КАФКА позволяют надеяться на то, что она сможет пополнить арсенал методов оценки запасов морских рыб, в особенности охотоморского минтая. Эти надежды приобретают важность в связи с тем, что сегодня научное обоснование стратегий управления промыслом становится неотъемлемой частью международных соглашений в сфере эксплуатации трансграничных и далеко мигрирующих морских биоресурсов.

## БЛАГОДАРНОСТИ

Работа была выполнена во многом благодаря финансовой поддержке НКО «АДМ» при содействии ее президента Германа Станиславовича Зверева, за что автор выражает ему и его коллегам свою признательность.

## СПИСОК ЛИТЕРАТУРЫ

Бабаян В. К. Предосторожный подход к оценке общего допустимого улова (ОДУ). Анализ и рекомендации по применению. М.: ВНИРО, 2000. 192 с.

Васильев Д. А. Когортные модели и анализ промысловых биоресурсов при дефиците информационного обеспечения. М.: ВНИРО, 2001. 111 с.

- Засосов А. В.* Динамика численности промысловых рыб. М.: Пищ. пром-сть, 1976. 312 с.
- Кадильников Ю. В.* Вероятностно-статистическая теория рыболовных систем и технической доступности для них водных биологических ресурсов. Калининград: АтлантНИРО, 2001. 275 с.
- Михеев А. А.* Стохастическая когортная модель для беспозвоночных с прерывистым ростом // Тр. СахНИРО. 2003. Т. 5. С. 216–242.
- Михеев А. А.* Приложение модели открытой эксплуатируемой популяции к оценке локальных запасов // Тр. ВНИРО. 2014. Т. 151. С. 112–125.
- Михеев А. А., Букин С. Д., Первеева Е. Р. и др.* Оценка запасов беспозвоночных в Сахалино-Курильском районе на основе анализа временных рядов уловов с применением фильтра Калмана // Изв. ТИРО. 2012. Т. 168. С. 93–114.
- Михеев Ф. А., Дмитриевич Г. Д., Павлушин В. А., Марков М. В.* Свидетельство об официальной регистрации программ для ЭВМ № 2006612785. Реестр программ для ЭВМ // Офиц. бюл. ФС по ИСПТЭ. Сер. Программы для ЭВМ. Базы данных. Топология интегральных микросхем. 2006. RU ОБПБТ № 3 (56).
- Рутковская Д., Пилиньский М., Рутковский Л.* Нейронные сети, генетические алгоритмы и нечеткие системы: монография. М.: Горячая линия – Телеком, 2004. 452 с.
- Уйтхед А. Н.* Избранные работы по философии. М.: Прогресс, 1990. 716 с.
- Gavaris S.* An adaptive framework for the estimation of population size // Can. Atl. Fish. Sci. Adv. Comm. (CAFSAC). 1988. Res. Doc. 88/29. 12 p.
- Harvey A.* Forecasting, structural time series models and the Kalman filter: the monography. N. Y.: Cambridge Univ. Press, 1989. 554 p.
- Kalman R. E.* A new approach to linear filtering and prediction problems // J. Basic Eng. 1960. V. 82. P. 34–45.
- Methot R. D.* Synthesis estimates of historical abundance and mortality on northern anchovy // Mathematical analysis of fish stock dynamics: reviews, evaluations, and current applications. V. 6. Bethesda, Maryland: Am. Fish. Soc. Symp., 1989. P. 66–82.
- Peterman R. M., Pyper B. J., MacGregor B. W.* Use of the Kalman filter to reconstruct historical trends in productivity of Bristol Bay sockeye salmon (*Oncorhynchus nerka*) // Can. J. Fish. Aquat. Sci. 2003. V. 60. P. 809–824.
- Pope J. G.* An investigation of the accuracy of virtual population analysis using cohort analysis // Res. Bull. 1972. V. 9. P. 65–74.
- Schnute J. T.* A general framework for developing sequential fisheries models // Can. J. Fish. Aquat. Sci. 1994. V. 51. P. 1676–1688.

## KALMAN FILTER APPLICATION IN COHORT MODEL FOR ADJUSTING STOCK ASSESSMENTS IN THE CASE OF UNREPORTED CATCH

© 2016 y. A. A. Mikheyev

*Sakhalin Research Institute of Fisheries and Oceanography, Yuzhno-Sakhalinsk, 693023*

This article presents the analysis and adjustment of stock assessments in the cohort model. It was found that a standard cohort model could be used to evaluate stock with a bias in the case of systematic errors in the annual catch data. The Kalman filter promotes elimination of this bias. The use of Kalman filter for adjusting stock assessments and a total allowable catch when new data appear is discussed. A method of standardizing stock index using this filter is proposed. Along with the stated topic, some questions on testing the models are considered.

*Keywords:* cohort model, Kalman filter, stock assessment.



ПРИЛОЖЕНИЯ

**Приложение 1. Принятые обозначения**

**Переменные и параметры (константы) когортной модели**

*Индексация.*  $a = 1, 2, \dots, A$  — индекс возрастного класса в запасе;  $t = 1, 2, \dots, T$  — индекс года.

*Процессная переменная.*  $N_t$  — запас в год  $t$ :  $N_t = \sum_a n_{at}$ ,  $a = 1, 2, \dots, A$ , где  $n_{at}$  — численность когорты в запасе в возрастном классе  $a$  в год  $t$ .

*Управляющая переменная.*  $C_t$  — улов в год  $t$ :  $C_t = \sum_a c_{at}$ ,  $a = 1, 2, \dots, A$ , где  $c_{at}$  — численность возрастного класса в улове в год;  $\varphi_{at}$  — коэффициент изъятия (эксплуатации) (доля промысловой убыли за год).

*Задаваемые параметры когортной модели.*  $\Delta$  — период с начала года до начала промысла в долях года,  $\Delta \in (0, 1)$ ;  $\Delta_S$  — период с начала года до начала съемки в долях года,  $\Delta_S \in (0, 1)$ ;  $\lambda$  — индивидуальная абсолютная плодовитость;  $W$  — средняя масса особи в улове;  $a_R$  — возраст вступления в промысел;  $\{m_a\}_{a=1}^A$  — огиба половозрелости;  $\epsilon$  — символ принадлежности; запись  $\Delta \in (0, 1)$  означает, что значения параметра  $\Delta$  принадлежат интервалу от 0 до 1.

*Оптимизируемые параметры когортной модели.*  $s$  — годовая доля выживших рыб в промысловом запасе от причин, не связанных с промыслом (коэффициент естественной выживаемости),  $s \in (0, 1)$ ;  $q$  — коэффициент улавливаемости для промысловых орудий лова (доля запаса, изымаемая единицей промыслового усилия),  $q \in (0, 1)$ ;  $k$  — коэффициент улавливаемости для учетного трала,  $k \in (0, 1)$ ;  $a_{50}$  и  $\beta$  — параметры кривой селективного изъятия.

**Переменные и параметры (константы) ФК**

*Переменные наблюдения.*  $E_t$  — численность икры в год  $t$ ;  $Y_t$  — численность улова на единицу промыслового усилия (CPUE) в год  $t$ ;  $S_t$  — индекс численности по данным учетных тралений в год  $t$ .

*Задаваемые параметры ФК.*  $P_0$  — начальная дисперсия ошибки оценивания.

Если нет других соображений, то  $P_0$  может задаваться как дисперсия так называемого диффузионного прайера (Peterman et al., 2003).

*Оптимизируемые параметры ФК.*  $\sigma_N$  — интенсивность шума в динамике запаса;  $\sigma_E$  — интенсивность шума в измерениях запаса через обилие икры  $E_t$ ;  $\sigma_Y$  — интенсивность шума в измерениях запаса через улов на единицу промыслового усилия  $Y_t$ ;  $\sigma_S$  — интенсивность шума в измерениях запаса через улов на учетное траление  $S_t$ .

**Приложение 2. Описание операционной модели**

Данные зашумлялись с помощью генератора случайных чисел из логнормального распределения с нулевым средним. Истинные данные по уловам помимо зашумления либо смещались с множителем 0,7, имитируя наличие неучтенного изъятия, либо не смещались. Зашумлению подвергались и наблюдения: промысловые уловы на усилие, запасы выметанной икры и траловые индексы запаса. Также при генерации истинных данных зашумлялись гауссовым шумом параметры кривой изъятия  $a_{50}$  и  $\beta$ , что предполагало случайную изменчивость селективности по годам.

«Истинную» популяцию формировали из  $A = 15$  возрастных классов с начальной численностью  $N_1$  в 1 млрд экз. и возрастным составом в первый год эксплуатации, определяемым с помощью обратной логистической функции:

$$n_{a,1} = \frac{\pi_a}{\sum_a \pi_a} N_1; a = \overline{1, A}; \pi_a = 1 - \frac{1}{1 + \exp[0,8(7-a)]}$$

«Истинную» динамику эксплуатируемой популяции за период  $T = 20$  лет построили, используя в цикле по  $t = \overline{1, T-1}$  и  $a = \overline{1, A-1}$  следующие уравнения (приложение 1).

$$n_{a+1,t+1} = (1 - \varphi_{a,t}) s n_{a,t};$$

$$n_{1,t+1} = s'_R E_t;$$

$$N_t = \sum_{a=1}^A n_{a,t};$$

$$E_t = \lambda s \sum_{a=a_m}^A m_a n_{a,t};$$

$$m_a = \frac{1}{1 + \exp[1,1(7-a)]};$$

$$s_a = \begin{cases} 0, & \text{если } a < a_R, \\ \frac{1}{1 + \exp[\beta(a_{50} - a)]}, & \text{иначе.} \end{cases}$$

$$\varphi_{a,t} = \begin{cases} s_a \exp(-0.02) \in_t, & \text{если } < 1 \\ 1, & \text{иначе,} \end{cases}$$

$$k = 10^{-6}$$

$$\Delta_s = 0,5$$

$$\lambda = 5 \times 10^5$$

$$a_R = 6$$

$$a_m = 4$$

$$s'_R = 2 \times 10^{-5}.$$

### Приложение 3. Описание модели КАФКА (когортный анализ с фильтром Калмана)

где  $\in_t$  — случайная логнормально распределенная величина с нулевым средним и стандартным отклонением 0,15.

«Истинную» и искаженную матрицу уловов рассчитали по формулам, в которых верхним штрихом здесь и далее обозначены искаженные величины:  $c_{a,t} = \varphi_{a,t} s^{\Delta} n_{a,t}$ ;  $c'_{a,t} = \vartheta c_{a,t} \in'_t$ , где  $\in'_t$  — случайная логнормально распределенная величина с нулевым средним и стандартным отклонением 0,7; при смещении уловов множителю  $\vartheta$  придавали значение 0,7, а в остальных случаях — 1. «Истинную» и искаженную динамику уловов по годам определили по формулам:

$$C_t = \sum_{a=a_R}^A c_{a,t}; \quad C'_t = \sum_{a=a_R}^A c'_{a,t}.$$

«Истинные» и искаженные данные наблюдений за индексами запаса по годам определили следующим образом:  $Y_t = qs^{\Delta} N_t$ ;  $S_t = ks^{\Delta_s} N_t$ ;  $Y'_t = 0,99Y_t \in_{Y_t}$ ;  $S'_t = S_t \in_{S_t}$ ;  $E'_t = E_t \in_{E_t}$ , где  $\in_{Y_t}$ ,  $\in_{S_t}$ ,  $\in_{E_t}$  — случайные логнормально распределенные величины с нулевым средним и стандартными отклонениями 0,3, 0,2 и 0,2 соответственно.

Значения, присвоенные параметрам уравнений операционной модели:

$$s = 0,7$$

$$\Delta = 0,2$$

$$W = 1$$

$$\beta = 0,8$$

$$a_{50} = 11$$

$$q = 2 \times 10^{-4}$$

Отметим, что все шумы в модели КАФКА полагаются гауссовыми, т.е. нормально распределенными с нулевым средним.

В модели КАФКА использованы следующие динамические уравнения:

$$n_{a+1,t+1} = sn_{at} - s^{1-\Delta} c_{at}, \quad (\text{ПЗ.1})$$

$$c_{at} = \varphi_{at} s^{\Delta} n_{at}, \quad (\text{ПЗ.2})$$

где  $n_{at}$  — численность когорты в запасе в возрасте  $a$  в год  $t$ ;  $c_{at}$  — численность когорты в улове в возрасте  $a$  в год  $t$ ;  $\varphi_{at}$  — доля когорты в улове в возрасте  $a$  в год  $t$  (коэффициент изъятия).

Из уравнений (ПЗ.1) и (ПЗ.2) было выведено рекуррентное уравнение для коэффициентов изъятия в когорте:

$$\varphi_{at} = \frac{\varphi_{a+1,t+1}}{\alpha_{at} + \varphi_{a+1,t+1}}; \quad a = \overline{1, A-1}, t = \overline{1, T-1}, \quad (\text{ПЗ.3})$$

$$\text{где } \alpha_{at} = \frac{c_{a+1,t+1}}{s c_{at}}.$$

Терминальные значения для коэффициентов изъятия параметризовали с помощью  $s$ -образной функции:

$$\varphi_{aT} = \frac{1}{1 + \exp[\beta(a_{50} - a)]}; \quad a = \overline{1, A}.$$

Из уравнения (ПЗ.3) следует, что при обратном расчете величины  $\varphi_{at}$  всегда меньше единицы. Следовательно, если терминальные значения  $\varphi_{At}$  и  $\varphi_{aT}$  не превышают единицу, уравнения (ПЗ.1) и (ПЗ.2) дают корректную процедуру обратного расчета

для коэффициентов изъятия в когорте для всех  $a = \overline{1, A}$  и  $t = \overline{1, T}$ .

Уравнение (ПЗ.1) суммированием всех переменных по индексу  $a$  было сведено к следующему уравнению процесса:

$$N_{t+1} = sN_t - s^{1-\Delta}C_t + R_t + \varepsilon_{N_t}, \quad (\text{ПЗ.4})$$

где  $R_t = n_{1,t}$  — пополнение;  $\varepsilon_{N_t}$  — шум процесса с неизвестной дисперсией  $\sigma_N^2$ .

Кроме того, для реализации ФК было построено три уравнения наблюдений соответственно для трех источников данных:

$$E_t = H_{1,t}N_t + \varepsilon_{E_t}, \quad (\text{ПЗ.5})$$

$$Y_t = H_{2,t}N_t + \varepsilon_{Y_t}, \quad (\text{ПЗ.6})$$

$$S_t = H_{3,t}N_t + \varepsilon_{S_t}, \quad (\text{ПЗ.7})$$

где  $H_{i,t}$ ,  $i = 1, 2, 3$  — переходные коэффициенты наблюдений:

$$H_{1,t} = \lambda s \sum_{a=1}^A m_a p_{a,t};$$

$$H_{2,t} = qs^{\Delta};$$

$$H_{3,t} = ks^{\Delta_s},$$

$p_{at}$  — доли возрастных классов в запасе:

$$p_{at} = \frac{n_{at}}{N_t}; \varepsilon_{E_t}, \varepsilon_{Y_t}, \varepsilon_{S_t} — \text{шумы (погрешности)}$$

наблюдений с неизвестными дисперсиями  $\sigma_E^2$ ,  $\sigma_Y^2$ ,  $\sigma_S^2$  соответственно.

Формирование массивов  $H_{2,t}$  и  $H_{3,t}$  из констант необходимо для учета отсутствующих наблюдений в ФК.

Для каждого  $i$ -го источника информации в год  $t$  вычисляли веса значимости по формуле (П4.11):

$$w'_{i,t} = \frac{K_{i,t}H_{i,t}}{\sum_i K_{i,t}H_{i,t}}, \quad i = 1, 2, 3.$$

Модель оптимизировали по вектору параметров  $\theta$  с помощью одной из следующих функций потерь (ФП):

$$L_{\text{med}}(\theta | Data) = \text{med}_{i,t} \left[ w'_{i,t} \left( \ln \frac{y_{i,t}}{N_t^+} \right)^2 \right], \quad (\text{ПЗ.8})$$

$$L(\theta | Data) = \sum_t \sum_i w'_{i,t} \left( \ln \frac{y_{i,t}}{N_t^+} \right)^2; \quad (\text{ПЗ.9})$$

где  $\text{med}_{i,t}[\cdot]$  — оператор вычисления медианы по всем  $i = 1, 2, 3$  и  $t = \overline{1, T}$ .

В соответствии с существующими рекомендациями наряду со среднеквадратичной ФП (ПЗ.8) была применена медианная ФП (ПЗ.9) с тем, чтобы повысить статистическую устойчивость (робастность) оценок к аутлайерам в данных (Васильев, 2001).

Перед началом работы ФК для поколений, еще не вышедших в текущем году из промысла, корректировали оценки численности в возрасте пополнения, полученные с помощью когортного анализа. Для этого использовали линейные тренды в динамике пополнения для ряда последних поколений из числа полностью обловленных на текущий год. Необходимость в данной процедуре обусловлена следующим. Для поколений, недолго находящихся в промысле, стандартный расчет не позволяет в достаточной мере уменьшить возможные ошибки в терминальных коэффициентах эксплуатации. Как следствие, для указанных поколений возникают ошибки и в оценках пополнения. В уравнении процесса использовали уже скорректированные оценки пополнения  $R'_t$  (П4.7).

На завершающей стадии работы модели КАФКА возникает необходимость обновить исходную матрицу запаса (ПЗ.1) с тем, чтобы суммирование столбцов по возрастам давало в точности скорректированные  $\hat{N}_t$  и прогнозные оценки  $N_{T+1}^+$  и  $N_{T+2}^+$  (П4.9), а также (П4.7) соответственно. Кроме того, требуется определить возрастной состав уловов на предстоящие два года и обеспечить при этом выполнение условия равенства суммарной численности уловов в возрастных классах установленному ОДУ  $C_{T+1}$  и  $C_{T+2}$ :

$$C_{T+k} = \sum_a c_{a,T+k}, \quad k = 1, 2. \quad (\text{ПЗ.10})$$

В этой связи использовали следующий алгоритм. Задавали первую строку и столбец матрицы запаса:

$$n'_{1,t} = R'_t; t = 1, 2 \dots T+2; \quad (\text{ПЗ.11})$$

$$n'_{a,1} = n_{a,1}; a = 1, 2 \dots A. \quad (\text{ПЗ.12})$$

Затем в цикле по  $k = 1, 2$  вычисляли:

$n'_{a+1,t+1} = sn'_{a,t} - s^{1-\Delta}c_{a,t}$ , если результат неотрицательный, иначе:

$$n'_{a+1,t+1} = 0; t = 1, 2 \dots T+1; \quad (\text{ПЗ.13})$$

$$c_{a,T+k} = c_{T+k} \frac{c_{a,T} \frac{n'_{a,T+k}}{n_{a,T}}}{\sum_a c_{a,T} \frac{n'_{a,T+k}}{n_{a,T}}}, \quad (\text{ПЗ.14})$$

$$\hat{n}_{a,t} = \hat{N}_t \frac{n'_{a,t}}{\sum_a n'_{a,t}} \quad (\text{ПЗ.15})$$

$$n'_{a,T+k} = N_{T+k}^+ \frac{n'_{a,T+k}}{\sum_a n'_{a,T+k}}. \quad (\text{ПЗ.16})$$

Формулы (ПЗ.11) – (ПЗ.16) дают на выходе алгоритма требуемую скорректированную матрицу запаса  $\{\hat{n}_{a,t}\} + \{n'_{a,T+k}\}$ . Формула (ПЗ.14) подразумевает, что паттерны селективности и выживаемость до начала промысла в прогнозируемые годы и в терминальный год промысла будут равны.

#### Приложение 4. Уравнения ФК для когортной модели

Основная особенность ФК состоит в том, что для каждого момента времени оценка состояния системы образуется как средневзвешенное от прогноза этого состояния  $N_t^+$  и поступившего наблюдения затем  $y_t$ .

$$\hat{N}_t = w_t N_t^+ + (1-w_t)y_t. \quad (\text{П4.1})$$

Здесь рассматриваются уравнения ФК применительно к специальной задаче с одной скалярной переменной состояния  $N_t$  и множеством переменных наблюдения за индексами состояния  $I_{i,t}$ ,  $i = 1, 2, 3$ . Из требования статистической эффективности, т.е. несмещенности и состоятельности, предъявляемого к оценке, выводится специальное представление для статистических весов (Михеев, 2014):

$$w_t = \frac{\prod_i s_i^2}{D_t}, \quad (\text{П4.2})$$

где  $\prod_i s_i^2 = (s_1 s_2 s_3 \dots)^2$ ;  $s_i^2$  – дисперсия ошибок в наблюдениях из  $i$ -го источника;  $D_t$  – дис-

персия обновляющего процесса, т.е. процесса корректировки прогнозных оценок. Если для простоты ограничиться случаем трех источников наблюдения, то указанная дисперсия определяется следующим образом (Михеев, 2014):

$$D_t = P_t^+ \left[ (H_{i,t} s_j s_k)^2 + (H_{j,t} s_j s_k)^2 + (H_{k,t} s_i s_j)^2 \right] + (s_i s_j s_k)^2, \quad (\text{П4.3})$$

где  $i, j, k = 1, 2, 3$ ;  $i \neq j$ ;  $i \neq k$ ;  $j \neq k$ .

В уравнении (П4.3)  $H_{i,t}$  – это переходный коэффициент между наблюдением за индексом состояния из  $i$ -го источника и самим состоянием:

$$I_{i,t} = H_{i,t} \hat{N}_t + \varepsilon_{i,t}, \quad (\text{П4.4})$$

а  $\varepsilon_{i,t}$  – гауссовый шум с дисперсией  $s_j^2$ ,  $i = 1, 2, 3$ . Переменную  $y_t$  из уравнения (П4.1) можно рассматривать как масштабированный индекс запаса:  $y_t = \frac{I_{i,t}}{H_{i,t}}$ .

Если использовать специальную переменную, известную как коэффициент усиления Калмана,

$$K_{i,t} = \frac{P_t^+ H_{i,t} (s_j s_k)^2}{D_t}, \quad (\text{П4.5})$$

то формулу (П4.2) можно переписать в виде тождества:

$$\sum_i K_{i,t} H_{i,t} + w_t \equiv 1, \quad (\text{П4.6})$$

где  $i = 1, 2, 3$ .

Формула (П4.6) с очевидностью вытекает из вышеприведенных формул для любых  $i, j, k = 1, 2, 3$ ;  $i \neq j$ ;  $i \neq k$ ;  $j \neq k$ :

$$K_{i,t} H_{i,t} D_t = P_t^+ H_{i,t}^2 (s_j s_k)^2;$$

$$w_t D_t = (s_i s_j s_k)^2;$$

$$D_t = P_t^+ \left[ (H_{i,t} s_j s_k)^2 + (H_{j,t} s_j s_k)^2 + (H_{k,t} s_i s_j)^2 \right] + (s_i s_j s_k)^2 = \sum_i K_{i,t} H_{i,t} D_t + w_t D_t = \left( \sum_i K_{i,t} H_{i,t} + w_t \right) D_t.$$

В качестве начального значения для скорректированной оценки запаса  $\hat{N}_1$  берется

соответствующая оценка запаса, полученная с помощью когортной модели  $\hat{N}_1 = N_1$ .

Начальное значение дисперсии ошибки оценки вычисляется из параметров фильтра:

$$P_1 = \frac{\prod_i s_i^2}{\left[ (H_{i,1} s_j s_k)^2 + (H_{j,1} s_i s_k)^2 + (H_{k,1} s_i s_j)^2 \right] + \prod_i s_i^2}.$$

Далее вычисления осуществляются в цикле по  $t = 1, T+1$ , где  $T$  — период лет наблюдений за запасом. Каждая итерация начинается с прогноза запаса  $N_t^+$  и вычисления дисперсии ошибки прогноза  $P_t^+$ :

$$N_{t+1}^+ = s \hat{N}_t - s^{1-\Delta} C_t + R'_{t+1}, \quad (\text{П4.7})$$

$$P_{t+1}^+ = s^2 P_t + \sigma_N^2, \quad (\text{П4.8})$$

где  $R'_t$  — скорректированные численности пополнения (приложение 3).

Для последующих моментов времени скорректированные оценки и дисперсии их ошибки рассчитываются по формулам:

$$\hat{N}_t = w_t N_t^+ + \sum_i K_{i,t} I_{i,t}, \quad (\text{П4.9})$$

$$P_t = w_t P_t^+. \quad (\text{П4.10})$$

Входящие в формулы (П4.9) и (П4.10) величины определяются с помощью формул (П4.2) — (П4.5).

Из формул (П4.1), (П4.4), (П4.6) и (П4.9) следует, что для каждого  $i$ -го источника информации в год  $t$  могут быть установлены следующие статистические веса:

$$w'_{i,t} = \frac{K_{i,t} H_{i,t}}{\sum_i K_{i,t} H_{i,t}}, i = 1, 2, 3. \quad (\text{П4.11})$$

Действительно, анализ формул ФК и особенно формулы (П4.9) показывает, что вклад в скорректированную оценку  $\hat{N}_t$  каждого из наблюдений  $y_t = \frac{I_{i,t}}{H_{i,t}}$  дается величиной  $K_{i,t} H_{i,t}$ . Нормировка суммы весов к единице и приводит к окончательному результату в виде формулы (П4.11). Естественной интерпретацией этой формулы является формальное выражение информационной значимости  $i$ -го источника информации в год

$t$ . В этой связи рассматриваемые веса могут быть использованы в ФП, состоящих из сумм вкладов от различных источников информации ((П3.8) и (П3.9)).

Кроме того, если придерживаться приведенной интерпретации, то следует формулу (П4.9) рассматривать как определение статистически эффективно взвешенного индекса запаса:

$$y_{w_t} = \sum_i w'_{i,t} y_{i,t}. \quad (\text{П4.12})$$

### Приложение 5. Обработка пропущенных значений в ФК

Практический интерес представляет задача получения оценок фильтра из источников с пропущенными значениями. Пусть  $I_{i,t} = 0$ ;  $i = 1, 2, 3$ . Покажем, что если задать  $H_{i,t} = 0$ , то соответствующие нулевые значения на оценки не влияют.

В рассматриваемом случае согласно уравнениям (П4.2), (П4.3) и (П4.5) имеем:

$$\begin{aligned} D_t &= P_t^+ \left[ (H_{i,t} s_j s_k)^2 + (H_{j,t} s_i s_k)^2 + (H_{k,t} s_i s_j)^2 \right] + \\ &\quad + (s_i s_j s_k)^2 = \\ &= P_t^+ \left[ (H_{j,t} s_i s_k)^2 + (H_{k,t} s_i s_j)^2 \right] + (s_i s_j s_k)^2 = \\ &= s_i^2 \left\{ P_t^+ \left[ (H_{j,t} s_k)^2 + (H_{k,t} s_j)^2 \right] + (s_j s_k)^2 \right\}, \\ w_t &= \frac{(s_i s_j s_k)^2}{D_t} = \\ &= \frac{s_i^2 (s_j s_k)^2}{s_i^2 \left\{ P_t^+ \left[ (H_{j,t} s_k)^2 + (H_{k,t} s_j)^2 \right] + (s_j s_k)^2 \right\}} = \\ &= \frac{(s_j s_k)^2}{P_t^+ \left[ (H_{j,t} s_k)^2 + (H_{k,t} s_j)^2 \right] + (s_j s_k)^2}; \\ K_{i,t} &= \frac{P_t^+ H_{i,t} (s_j s_k)^2}{D_t} = 0; \\ K_{j,t} &= \frac{P_t^+ H_{j,t} (s_i s_j)^2}{D_t} = \\ &= \frac{s_i^2 P_t^+ H_{j,t} s_j^2}{s_i^2 \left\{ P_t^+ \left[ (H_{j,t} s_k)^2 + (H_{k,t} s_j)^2 \right] + (s_j s_k)^2 \right\}} = \end{aligned}$$

$$= \frac{P_t^+ H_{j,t} s_k^2}{P_t^+ \left[ (H_{j,t} s_k)^2 + (H_{k,t} s_j)^2 \right] + (s_j s_k)^2}.$$

Следовательно, если  $I_{i,t} = 0$  и  $H_{i,t} = 0$ , то  $w_i$  и  $K_{j,t}$  от  $s_i$  не зависят. Иначе гово-

ря, если информация из  $i$ -го источника для момента времени  $t$  отсутствует, то его влияние на результат для этого момента времени элиминируется приравнением к нулю самого наблюдения и соответствующего элемента  $H_{i,t}$ .