

Математическая модель фильтрации ошибок отчетных данных вылова

Кобзарь Иван Сергеевич

Камчатский государственный технический университет

Магистрант

Проценко Игорь Григорьевич

Камчатский государственный технический университет

*Заведующей кафедрой Информационных систем, доктор технических наук,
профессор*

Аннотация

В качестве инструмента оценки качества данных о вылове, представляемых рыболовными компаниями, предлагается использование технологии, основанной на применении авторегрессионных моделей, что позволяет уменьшить доверительный интервал ошибок отчетности и проводить эффективный контроль качества поступающих отчетных данных.

Ключевые слова: рыболовные участки, рыбный промысел, информационная рыбопромысловая система, промысловая отчетность, стохастическая авторегрессионная модель.

Mathematical model of error filtering catch reporting data

Kobzar Ivan Sergeevich

Kamchatka State Technical University

Student

Protsenko Igor Grigorievich

Kamchatka State Technical University

*Head of the Department "Information systems", doctor of technical science,
Professor*

Abstract

The article is devoted to the problem of the quality of reporting in industrial fishing. As a tool for assessing the quality of catch data provided by fishing companies, it is proposed to use a technology based on the use of a wide class of stochastic autoregressive models. The assessment is based on a comparison of the predictive value of the catch made according to the model and a similar indicator contained in the submitted fishing report. In the process of comparison, based on newly received data, the coefficients of the model are refined. The use of a mathematical model adapted to the received reporting data makes it possible to reduce the confidence interval of reporting errors and carry out effective quality

control of incoming reporting data, which, ultimately, will provide users of the fishing information system with high-quality analytical information about the catch.

Keywords: fishing areas, fishing, information fishing system, fishing reporting, stochastic autoregressive models.

Введение

Основным объектом промышленного рыболовства на рыболовных участках Дальневосточного бассейна является семейство лососевых. Право на осуществление вылова у пользователей водных биологических ресурсов (ВБР) – рыбопромышленных предприятий – возникает на основании разрешения на добычу (вылов) водных биоресурсов, которое выдается федеральным органом по рыболовству или его бассейновыми управлениями по охране, воспроизводству рыбных запасов и регулированию рыболовства либо органом исполнительной власти субъекта РФ.

Вести промысел пользователь может только на рыболовном (рыбопромысловом, сокращенно РПУ) участке, указанном в разрешении, применять при этом только разрешенные орудия лова (ставные или закидные невода, реже ставные сети) и осуществлять добычу по графику в определенные для промысла дни. В некоторых случаях в разрешениях на добычу указывается лимит вылова, который можно сделать на данном РПУ.

Выполнение ограничений и условий промысла, содержащихся в выданном разрешении, контролируется мониторингом использования водных биоресурсов путем непрерывного контроля за деятельностью на РПУ. Решение задачи контроля обеспечивается путем изучения промысловой отчетности, поступающей от пользователей участков. Показатели отчетности, наряду с другими независимыми данными, используются для оценки состояния ВБР, уровня освоения квот вылова, интенсивности промысла и его влияния на состояние рыбного запаса.

Промысловая отчетность представляет собой формализованный отчет о результатах промысловой деятельности на участке, формируемый по результатам вылова за отчетные сутки. Эти отчеты составляются и формализуются в соответствии со стандартами и требованиями к промысловой отчетности, которые регламентируются Приказами Госкомрыболовства РФ и Федерального агентства по рыболовству.

Каждый отчет представляет собой структурированную по специальным правилам написания последовательность информационных блоков, состоящих из сгруппированных по категориям показателей. Все показатели отчета – это система кодов и идентификаторов. Состав показателей промысловой отчетности в полной мере характеризует работу участка. В связи с вышеизложенным, при разработке структуры и содержания отчетности был выделен ряд основных показателей, способных в совокупности передать качественную и количественную оценку результатов деятельности конкретного участка. Поэтому главными принципами её построения были обозначены логичность, единство и взаимосвязь

показателей.

Эффективное государственное управление рациональным использованием водных биоресурсов, изучением и сохранением рыбных запасов требует соответствующего информационного обеспечения. Для этого была разработана отраслевая система мониторинга (ОСМ) [1], в основе которой лежат данные промысловой отчетности, данные рекомендуемых объемов вылова, сведения об их распределении между промышленниками, разрешительная и прочая справочная информация. Но главными показателями работы отрасли являются, естественно, сведения об уловах.

Мониторинг водных биоресурсов ведется через ОСМ путем непрерывного наблюдения и контроля за разведкой рыбных запасов, ловом рыбы, её обработкой, транспортировкой и хранением. Слежение и контроль за промыслом судов и рыбопромышленных бригад на рыболовных участках, сбор отчетности в рамках ОСМ позволяет получить оценку состояния водных биоресурсов на основе определенных показателей: вылов на усилие, площади промысловых скоплений, изменение размеров популяции и т.д. В свою очередь, на основе этих показателей осуществляется наблюдение за интенсивностью промысла и его влиянием на состояние рыбных запасов, обеспечивается непрерывный контроль за реализацией выделенных квот вылова.

Понятно, что сохранить численность биоресурсов в размерах, необходимых для их воспроизводства и последующего использования в течение длительного срока невозможно без надежной работы ОСМ. В рамках ОСМ формируются аналитические материалы, позволяющие решать широкий комплекс задач по управлению рыболовством: осуществление контроля за промысловой деятельностью участков, отслеживание полноты и достоверности отчетности, поступающей от рыболовных участков.

Вот почему важно, чтобы ОСМ развивалась вместе с совершенствованием системы государственного управления рыболовством и соответствовала современным стандартам информационных технологий.

Разработка автоматизированной системы отчетности промысловых участков является сложной и емкой задачей, которая требует решения большого количества вопросов из различных областей знаний. Основной проблемой использования промысловых данных можно выделить их качество, а именно полноту, достоверность, точность и легитимность. Промысловые отчеты, которые формируются на рыболовных участках, при подготовке и передаче подвержены искажению, а также умышленной фальсификации. Этот фактор резко снижает качество всего информационного ресурса ОСМ, а также ставит под вопрос эффективность его использования в процессе государственного управления рыболовством.

Проблеме повышения качества информационного ресурса ОСМ уделено достаточно много внимания [2,3]. Кроме умышленного искажения поступающих сведений важно отметить высокую волатильность уловов, связанную с состоянием лососевой популяции и неравномерным подходом рыбы к участкам. В связи с этим очень сложно провести оценку величины

возможного улова.

Можно привести пример возникновения неопределенности. Предположим, что среднесуточный вылов группы участков за прошедший период составил 300–500 тонн. Необходимо принять решение по оперативному управлению (ограничение или увеличение объемов вылова) в данном районе на основе величины вылова, который будет в течение следующей недели. Эта величина не определена, однако можно оценить значение вылова с большой долей уверенности, как лежащие в интервале тех же 300–500 тонн.

Интуитивно ясно, что вероятность того, что вылов будет равен нижней (300 тонн) или верхней (500 тонн) границе меньше, чем вероятность того, что вылов будет равен 400 тоннам. Следовательно, если разбить шкалу значений вылова на интервалы, то каждому такому интервалу можно сопоставить степень уверенности в том, что будущий вылов примет именно это значение. Величина прогнозируемого вылова зависит от множества случайных факторов. Это приводит к тому, что до сих пор нет системы или метода, которые бы позволили бы дать ему точную оценку. Однако, получив наилучшую в смысле достоверности оценку и точность этой оценки, можно приступать к процедурам выработки решения.

Если речь идет о сложных процессах, нельзя рассчитывать на получение удовлетворительной детерминированной модели. Однако, стоит учитывать и то, что принципиальным недостатком детерминированных моделей систем является отсутствие эффективного метода сравнения различных возможных моделей, которые могут быть разработаны на основе одних и тех же эмпирических данных. Существует ряд объективных причин, в силу которых, по истечении определенного количества времени, отличия между моделируемым и реальным процессами становятся настолько сильными, что состояние математической системы более не будет отражать основных характеристик исследуемых явлений. К числу таких причин в первую очередь следует отнести конечное количество пространства состояний, ошибки начальных данных, входных воздействий и параметров модели. Важно заметить, что ошибки начальных данных и входных воздействий носят случайный характер.

Неопределенность значения прогнозируемого вылова можно охарактеризовать с помощью понятий расплывчатости, вводя, такие термины как "норма вылова" и "доверительный интервал", которые определяются степенью близости к норме (близко, около, меньше, гораздо меньше и т.п.).

Итак, в адрес контролирующего государственного органа поступил промысловый отчет от РПУ, в котором указано значение вылова определенного объекта промысла. Пусть уже имеются достоверные отчетные данные предыдущих отчетов по этому участку и по другим близко расположенным участкам. Идея заключается в том, что можно построить математическую модель, а затем на основе модели и имеющихся в наличии данных дать прогноз на одни сутки вперед и сравнить это прогностическое значение со значением вылова в поступившем отчете.

Основными данными в ОСМ являются результаты промысловых операций, выраженные в объемах выловленной рыбы – сведения, которые предполагается ежесуточно передавать с рыболовных участков. Процесс обработки входных данных включает в себя первичную обработку, в процессе которой выполняется преобразование формата исходных данных во внутренний формат информационной системы, комплексный анализ этой информации, выполняющий синтаксический и семантический разбор данных и подготовку аналитического материала, ввода потока в базу данных и его рассылку другим пользователям.

Материалы и методы

Отчетную величину улова на участке – x_t за указанные сутки t , можно рассматривать как сумму двух элементов: первой части, являющейся функцией всех значений до выбранных суток $t - 1$, и случайной составляющей w_t :

$$x_t = f(\tilde{x}_{t-1}, a, u_t) + w_t \quad (1)$$

В этом уравнении \tilde{x}_{t-1} – это совокупность данных вылова x_t , полученных до момента времени t , то есть $\tilde{x}_{t-1} = (x_{t-1}, x_{t-2}, \dots)$; a – вектор неизвестных коэффициентов модели; u_t – некоторая интегральная функция отчетных значений уловов с близлежащих участков; w_t – шум, последовательность случайных величин; f – детерминированная функция.

Ввод в уравнение шума w_t отражает допущенные погрешности при случайном или умышленном искажении величины улова, а также поиске неизвестного вида функции f .

Главное для построения эффективной статистической модели оценки улова – это получение оценщиков, на основе использования методов несмещенного оценивания и максимального правдоподобия, исходя из выборочных значений вылова, и изучению их асимптотических свойств. Асимптотические свойства оценщиков служат обоснованием для статистических выводов, получаемых при выборках большего объема.

Если привести уравнение (1) в векторно-матричный вид, то система уравнений M для m -мерного векторного процесса $x_t = [x_{1,t}, \dots, x_{m,t}]^T$ будет выглядеть следующим образом:

$$M: x_{i,t} = z_{i,t-1}^T a_i + w_{i,t} \quad i = 1, \dots, m \quad (2)$$

где $a_i - n_i$ – мерный вектор;

$a = [a_1^T, \dots, a_m^T]$ – n_0 -мерный вектор $n_0 = \sum_{i=1}^m n_i$;

$w_t = w_{i,t} = [w_{1,t}, \dots, w_{m,t}]^T$ – случайные гауссовы возмущения типа "белого шума";

$z_{i,t}^T$ – имеет размерность n_i и составляется в виде функции из компонент векторов $x_t, \dots, x_{t-m1}, u_t, \dots, u_{t-m2}$;

$\mathbf{u}_t = u_{i,t}[u_{1,t}, \dots, u_{n,t}]^T T$ – наблюдаемый входной вектор.

Компонентами $[x_{1,t}, \dots, x_{m,t}]^T$ являются уловы разных объектов промысла, например, за сутки в общем улове на РПУ присутствовали горбуша, кета, голец и т.д.

Таким образом, имеем класс моделей (2) и требуется найти значения компонент вектора \mathbf{a}_i .

Естественно, за счет подбора вида функции f и вектора коэффициентов модели \mathbf{a}_i , возможно, минимизировать влияние шума:

$$\sigma_{i,t}^2 = \sum_{j=1}^{t\Sigma} w_{i,j}^2 = \sum_{j=1}^{t\Sigma} (x_{i,j} - z_{i,j-1}^T \mathbf{a}_i)^2 \rightarrow \min \quad (3)$$

При этом минимизация (3) для вектора \mathbf{a}_i даст выражение в виде [4]:

$$\mathbf{a}_i(t) = [\sum_{j=1}^t \mathbf{z}_{i,j-1}, \mathbf{z}_{i,j-1}^T]^{-1} [\sum_{j=1}^t \mathbf{z}_{i,j-1} x_{i,j}] \quad (4)$$

Важное преимущество метода оценивания на основе ограниченной информации состоит в том, что он позволяет находить состоятельную оценку \mathbf{a}_i даже в условиях отсутствия знания $\sigma_{i,t}^2$. Данный метод позволяет осуществлять декомпозицию задачи оценивания на ряд автономных задач оценивания для отдельных уравнений системы, что существенно упрощает задачу в вычислительном аспекте.

Алгоритм контроля входных отчетных данных выглядит следующим образом. По модели конкретного промыслового участка делается прогноз на один шаг вперед по времени. Теперь есть 2 значения улова: фактическое $x_{i,t}$, переданное в отчете, и спрогнозированное $x_{i,t|t-1}$. Именно это значение сравнивается с поступившим на эту дату значением вылова в отчете. Если расхождение укладывается в доверительный интервал $\sigma_{i,t}^2$, то значит, поступивший отчет, скорее всего достоверен и нет причин подозревать пользователя ВБР в его искажении. Если же расхождение велико, то это прямой сигнал к проведению анализа поступивших данных на предмет их ошибочности или умышленного искажения.

Расчет по формуле (4) требует необходимости запоминания всей совокупности наблюдений $[x_{i,j}; j = 1, \dots, t]$ до момента времени t . При мониторинге, т.е. непрерывном отслеживании динамики показателей, характеризующих процесс, необходим пересчет оценки $\mathbf{a}_{i,t-1}$ в новые оценки $\mathbf{a}_{i,t}$ с использованием данных новейших наблюдений $x_{i,t}$. Проведение подобных расчетов требует выполнения, растущего с числом t количества суммирований, что закономерно приведет к росту затрат на поиск и отбор записей наблюдений $[x_{i,j}, j = 1, 2, \dots, t]$ при компьютерной реализации.

Для решения этой проблемы можно воспользоваться алгоритмом, принадлежащим семейству алгоритмов фильтрации Калмана и реализующем процедуру вычислений оценки коэффициентов модели в реальном масштабе

времени [4]. Суть алгоритма, реализующего Калмановскую фильтрацию, заключается в согласовании наблюдений и физических моделей.

Если ввести в рассмотрение матрицу $S_{i,t}$:

$$S_{i,t} = \sigma_{i,t}^2 [\sum_{j=1}^t z_{i,j-1} z_{i,j-1}^T]^{-1} \quad (5)$$

Оценку вектора коэффициентов модели при этом можно записать в виде:

$$\mathbf{a}_{i,t} = \mathbf{S}_{i,t} [\sum_{j=1}^t z_{i,j-1} x_{i,j}] \quad (6)$$

При этом алгоритм вычисления оценки $\mathbf{a}_{i,t}$, исходя из описанных соотношений, представится в следующем виде:

$$\begin{aligned} S_{i,t}^{-1} &= \sum_{j=1}^t z_{i,j-1} z_{i,t-1}^T = z_{i,t-1} z_{i,t-1}^T + \sum_{j=1}^{t-1} z_{i,j-1} z_{i,j-1}^T = \\ &= z_{i,t-1} z_{i,t-1}^T + \mathbf{S}_{i,t-1}^{-1} \end{aligned} \quad (7)$$

Из этого выражения можно получить соотношение:

$$\mathbf{a}_{i,t-1} = \mathbf{S}_{i,t-1} [\sum_{j=1}^{t-1} z_{i,j-1} x_{i,j}] \quad (8)$$

Тогда выражение для $\mathbf{a}_{i,t}$ в конечном итоге будет выглядеть следующим образом:

$$\mathbf{a}_{i,t} = \mathbf{S}_{i,t} z_{i,t-1} x_{i,t} + \mathbf{S}_{i,t} \mathbf{S}_{i,t-1}^{-1} \mathbf{a}_{i,t-1} \quad (9)$$

Если подставить $S_{i,t}$ выражение, найденное ранее примет вид:

$$\mathbf{a}_{i,t} = \mathbf{S}_{i,t} z_{i,t-1} x_{i,t} + \mathbf{a}_{i,t-1} - \mathbf{S}_{i,t-1} z_{i,t-1} z_{i,t-1}^T \mathbf{a}_{i,t-1} / (1 + z_{i,t-1}^T \mathbf{a}_{i,t-1} z_{i,t-1}) \quad (10)$$

Можно вычислить оценку дисперсии шума $\sigma_{i,t}^2$ в реальном масштабе времени:

$$\sigma_{i,t}^2 = \sigma_{i,t-1}^2 + (x_{i,t} - z_{i,t-1} \mathbf{a}_{i,t-1})^2 \quad (11)$$

Значение $\mathbf{a}_{i,t}$ при этом стремится к квазимоксимальноправдоподобной оценке. Эта оценка, в свое время, приближенно равна $E[\mathbf{a}_{i,t} | x_{i,t}]$, т.е. близка к байесовской оценке, которая основана на всей информации в момент времени t . При этом $S_{i,t}$ имеет смысл матрицы точности оценивания параметров модели, а $\sigma_{i,t}^2$ — это дисперсия ошибок оценивания непосредственно процесса по модели

$$x_{i,t|t-1} = z_{i,t-1} \mathbf{a}_{i,t-1} \quad (12)$$

Практически всегда при вычислениях возможность получения наблюдений по каждой компоненте вектора состояния $x_{i,t}$, $i = 1, 2, \dots, m$ отсутствует. Однако, если предположить, что в некоторый момент времени имеется только часть информации о векторе состояния, а именно $x_{k,t}$, где $k = 1, \dots, l$, причем $l < m$. В этом случае есть возможность проведения уточнения коэффициентов модели только в e -уравнениях. При этом стоит учитывать, что компоненты вектора состояния статистически между собой связаны. Используя эту связь, можно уточнить значения тех составляющих вектора состояния, по которым информация не поступила за счет тех компонент, по которым есть измерение.

Если в момент t имеется линейная оценка $x_{i,t|t-1}$ процесса $x_{i,t}$, которая была получена в результате использования имеющихся на момент времени $t - 1$ данных измерений, и минимизирующая функционал $E\{(x_{i,t|t-1} - x_{i,t})^2\}$, то такая оценка дается условным математическим ожиданием по отношению к процессу $x_{i,t}$:

$$x_{i,t|t} = E\{x_{i,t|t-1} | x_{i,t}\} \quad (13)$$

Из этого уравнения, используя имеющиеся на момент времени $t - 1$ данные измерений, можно получить линейную оценку $x_{i,t|t-1}$ процесса $x_{i,t}$. Оптимальное оценивание состояния $x_{i,t|t}$, с учетом полученных на момент времени t измерений $x_{i,t}$ дается линейной оценкой в виде:

$$x_{i,t|t} = x_{i,t|t-1} + \sum_{k=1}^e G_{i,k} (x_{k,t} - x_{k,t|t-1}) \quad (14)$$

Минимизация такого выражения, в свою очередь, приводит к матричному уравнению Винера-Хопфа относительно $(m \times l)$ -мерной матрицы $G_{i,k}$ решение которого имеет вид:

$$G_{i,k} = P_{i,j,t|t-1} P^{-1}_{k,k,t|t-1} \quad (15)$$

где $P_{i,j,t|t-1} = E\{(x_{i,t} - x_{i,t|t-1})(x_{j,t} - x_{j,t|t-1})\}$ – ковариационная матрица ошибки оценки $x_{i,t|t-1}$.

Ковариационная матрица $P_{i,j,t|t}$ ошибки оптимальной оценки $x_{i,t|t}$ с учетом предыдущих соотношений равна:

$$P_{i,j,t|t} = P_{i,j,t,t-1} - \sum_{k=1}^t P_{i,k,t|t-1} P_{k,i,t|t-1} P_{k,k,t-1} \quad (16)$$

Для численной реализации на каждом временном шаге процедуры оптимального оценивания, производится модификация этого алгоритма, при которой осуществляется последовательное усвоение каждой компоненты вектора состояния по рекуррентным формулам:

$$h = P_{i,k}^{(r)} / P_{k,k}^{(r)} \quad (17)$$

$$x_i^{(r)} = x_i^{(r-1)} + h[x_k - x_i^{(r-1)}] \quad (18)$$

$$P_{i,j}^{(r)} = P_{i,j}^{(r-1)} - P_{i,k}^{(r-1)} h \quad (19)$$

Индекс i пробегает на одном шаге по времени значения от 1 до m . Начальные условия: $x^{(0)} = x_{t|t-1}$, $P^{(0)} = P_{t|t-1}$.

Обсуждение

Наиболее значимые результаты по увеличению точности получаемых прогнозов и передаваемых сведений, могут быть достигнуты путем комплексной автоматизации формирования и передачи промышленной отчетности в адрес органов государственного управления. Это, в свою очередь, позволит значительным образом приблизиться к решению задачи контроля качества промышленной отчетности. Стоит подчеркнуть, что особую значимость имеет система контроля подготовки и представления промышленных отчетов. При правильной организации этого процесса, возможно существенно изменить результативность и эффективность рыбоохранных мероприятий. Очевидно, что это приведет к росту доказательной базы по браконьерству и нарушениям правил рыболовства. Подобная организация также позволит значительно снизить финансовые затраты на обеспечение функционирования системы промышленной отчетности со стороны рыбопромышленных предприятий и государственных органов контроля.

Заключение

Разработка автоматизированной системы отчетности промышленных участков является сложной и емкой задачей, которая требует решения большого количества вопросов из различных областей знаний. Среди этих вопросов важное место занимают проблемы потоков инструментальных данных, а также информации, подготовленной в ручном режиме, проблемы, связанные с подготовкой, накоплением и передачей информации от объекта мониторинга в центр мониторинга, а также проблемы качества данных и их защиты от искажения.

Без решения ряда этих вопросов, предполагающих исследование и системное проектирование, обеспечить совершенствование и развитие мониторинга рыболовства не представляется возможным.

Библиографический список

1. Проценко И.Г. Информационная система мониторинга рыболовства // Рыбное хозяйство. 2001. Спец. выпуск. С. 3–18.
2. Мониторинг рыболовства. Инструкции и рекомендации экипажам промысловых судов и судовладельцам. / Под общ. ред. д.т.н. И.Г. Проценко. Петропавловск-Камч.: ФГУП «Камчатский центр связи и мониторинга», 2005. 264 с.
3. Кошкарева Л.А. Алгоритм контроля расхода сырца при выпуске рыбной продукции на основе данных промышленной отчетности // Материалы международных научных чтений «Приморские зори – 2005». Владивосток: ТАНЭБ, 2005. 286 с.
4. Кашьяп Р.Л., Рао А.Р. Построение динамических стохастических моделей по экспериментальным данным. М.: Наука, 1983. 197 с.