

УДК 630*1+57.087.1

ПОДХОД К МЕТА-АНАЛИЗУ ПЕРЕМЕННЫХ В ЛОГИСТИЧЕСКИХ РЕГРЕССИОННЫХ МОДЕЛЯХ СМЕРТНОСТИ ДЕРЕВЬЕВ С ИСПОЛЬЗОВАНИЕМ ГОДИЧНЫХ ПРИРОСТОВ КОЛЕЦ.

А. В. Качаев

Сибирский Федеральный университет

660041, Красноярск, просп. Свободный, 79

E-mail: avkachaev@gmail.com

АННОТАЦИЯ

В результате проведенного мета-анализа, определена предметная модель данных текущих дендрохронологических исследований смертности деревьев с использованием логистических регрессионных моделей. Исходные данные предметной области определяются, как временные ряды приростов деревьев и их производные. Уточняется понятие переменной роста, как оконной статистики. Определена полная система окон с выделением в ней подсистемы вложенных и непересекающихся окон. Определена операция замыкания формирования всех логистических регрессионных моделей в рамках предметной области данных. Текущие научные исследования моделей смертности с использованием логистической регрессии при вычислении переменных роста используют подсистему вложенных окон. Предлагается новый подход при формировании переменных роста с применением подсистемы непересекающихся окон. Программная реализация предметной модели данных разработана на базе языка программирования R.

Ключевые слова: *meta analysis, dendrochronology, logistic regression; tree mortality, tree growth, growth variables.*

ВВЕДЕНИЕ

Изменения абиотических и биотических факторов естественным образом связаны с процессами, происходящими в лесных экологических системах. Смертность (усыхание) деревьев является важным элементом динамики популяций для поддержания биологического и структурного разнообразия в лесных экосистемах (Liang et al., 2016).

Актуальность исследуемой темы определяется увеличением процента площадей усыхающих деревьев в связи с изменением климатических условий окружающей среды (Замолотчиков и Краев, 2018). В последние два десятилетия в России отмечается увеличение площадей погибших насаждений по сравнению с предыдущим периодом (Малахова и Лямцев, 2014; Павлов, 2015; Сарнацкий, 2012).

Исследуемый объект обладает эмерджентными свойствами, что определяет сложность в построении предметных моделей экосистем лесов (Розенберг, 2011).

Процессы усыхания изучаются в различных направлениях: построение математических моделей (Monserud, 1976), использование непараметрических статистик (Павлов, 2015), дистанционного зондирования (Харук и др., 2016), вероятностных методов с использованием логистической регрессии и приростов древесных колец растений (Wycoff et al., 2000; Bigler and Bugmann, 2003; Das et al., 2007) и т.д.

Дендрохронология в своих исследованиях использует прирост годовых колец, который является интегральным показателем, регистрирующим влияния абиотических и биотических факторов в древесных растениях (Шиятов и др., 2000).

Анализ работ, использующих логистическую регрессию с использованием прироста древесных колец, представлен в разделе **объект исследования**.

Далее приведем несколько работ, исследующих процессы усыхания древесных растений, зависящих от морфологических признаков древесных растений (размер кроны, высота, диаметр и т.д.), почвенных, топографических, биотических и других факторов.

Scott (Scott et al., 1999) изучают влияние гидрологического режима на рост Тополя дельтовидного (*Populus deltoides* подвид *monilifera*) в прибрежных районах восточной части Колорадо, США. При построении логистической регрессионной модели использовались следующие измерения: объем живой кроны, рост радиальных стеблей, годовой прирост ветвей из 689 деревьев Тополя дельтовидного. Так устойчивое снижение уровня грунтовых вод, превышающее 1м, для Тополя дельтовидного, произрастающего в средних аллювиальных песках приводило к усыханию кроны и росту стеблей и влечет 88% смертности деревьев в течение трехлетнего периода.

В работе Павлова (2015) выявлены причины массового усыхания хвойных (*Pinus sibirica* Du Tour, *Picea obovata* Ledeb., *Abies sibirica* Ledeb., *Pinus sylvestris* L., *Larix gmelinii* (Rupr) Kuzen., *Abies nephrolepis* (Trautv. Ex Maxim.) Maxim., *Pinus koraiensis* Siebold & Zucc.) лесов Сибири и Дальнего Востока за период 1996-2014 гг. Биологическая устойчивость хвойных деревьев была снижена неблагоприятным сочетанием климатических факторов в сочетании с определенными почвенными и топографическими факторами, что дало рост корневым патогенным организмам и определило куртинное (мозаичное) усыхание древесных растений. Для анализа влияния переменных абиотических и биотических факторов на процесс усыхания древесных растений используются дендрохронологические методы. Для выявления связей между факторами используются методы непараметрической статистики (коэффициент корреляции Спирмена и Кендалла и др.). Количество дендрохронологического материала не приведено в статье.

В работе (Харук и др., 2016) анализировались причины усыхания кедровников (*Pinus sibirica*) Прибайкалья (хр. Хамар-Дабан) методами дистанционного зондирования и дендрохронологии. Установлено, что начиная с 1980-х гг. наблюдалось снижение величины индекса прироста ($R^2 = 0,69$) и снижение индекса сухости климата SPEI ($R^2 = 0,72$). В середине 2000-х гг. возрастание засушливости привело к разделению деревьев кедровника на две группы: “выживших” и “усыхающих”. Пространственное распределение указанных групп обусловлено различными орографическими факторами (экспозиция, крутизна, высота над уровнем моря и др.) приводящими к дефициту влаги. Индекс прироста деревьев тесно связан с индексом сухости июня ($r^2 = 0,55$). Наряду с водным стрессом, усыхающие деревья подвергались воздействию стволовых вредителей и фитопатогенов. Установлено, что первопричиной усыхания кедровников является водный стресс, обусловленный возрастанием засушливости климата. В целом в пределах хребта Хамар-Дабана сильно поврежденные и усыхающие насаждения (> 50 % усыхающих и усохших деревьев) составляют 8-10 % общей площади темнохвойных.

В работе (Cailleret et al., 2016) сравниваются различные методики и оценки взаимосвязей между ростом и смертностью деревьев с использованием прироста годовых древесных колец.

Итогом статьи стала постановка нескольких вопросов, ответы на которые позволят улучшить прогностическую способность полученных статистических моделей и разрабатываемых в будущем. Какую схему следует использовать для выборки живых и усыхающих деревьев? Какой тип переменных роста следует учитывать и какие конкретные показатели следует использовать? Как определить наилучшую длину временного окна для расчета переменных роста? Как построить лучшую логистическую модель с несколькими переменными, учитывая коллинеарность между переменными роста, и как лучше всего определить влияние каждой из них на прогнозируемую вероятность смертности?

Цель данной работы – выполнить мета-анализ (Glass, 2000) текущих научных исследований, в которых разрабатываются вероятностные модели смертности деревьев на основе дендрохронологических данных с использованием логистической регрессии модели. В результате мета-анализа определить модель данных предметной области. Уточнить понятие переменной роста как результат обработки дендрохронологических рядов «оконными» статистическими процедурами. Программными средствами сформировать полный набор логистических регрессионных моделей, из которых по определенным критериям осуществлять поиск моделей имеющих максимальную предсказательную вероятность, как живых, так и усыхающих деревьев.

МАТЕРИАЛЫ И МЕТОДЫ

Объект исследования.

Далее краткий обзор статей с целью определения переменных роста и выделения других типов информационных объектов, служащих исходными данными для логистических регрессионных моделей.

В работе (Wyckoff and Clark 2000) исследуется средний рост живых и усохших деревьев за последние 5 лет (mean_5) на основании байесовского подхода. При моделировании смертности рассматриваются непараметрические методы в сравнении с параметрическими методами. Для моделирования выбраны древесные породы: Клён красный (*Acer rubrum* L.) и Кизил цветущий (*Cornus florida* L.) в районе южного Аппалачи (southern Appalachian, USA).

Ель обыкновенная (*Picea abies* (L.) Karst.), выбранная на трех участках в субальпийских лесах Швейцарии с общим количеством 110 пар живых и усыхающих деревьев, исследуется в работе (Bigler and Bugmann 2003). Переменные логистических регрессионных моделей определялись множеством переменных уровня роста и переменных тренда роста. Переменные уровня роста вычислялись как средние значения BAI (см² / год) за последние 3, 5 или 7

лет дерева (BAI3, BAI5, BAI7), где BAI– это стандартизованный показатель радиального роста - прирост базальной площади. Вторым набором переменных уровня роста- это логарифмы от первого набора: $\log BAI3$, $\log BAI5$, $\log BAI7$. Для характеристики тренда роста деревьев, использовались коэффициенты наклона уравнений локальных линейных регрессий за последние 5, 10, 15, ..., 40 лет BAI (locreg5 , locreg10 , ..., locreg40).

В работе (Bigler and Bugmann, 2004) также исследуется Ель обыкновенная (*Picea abies* (L.) Karst.), выбранная в субальпийских лесах Швейцарии. Дополнительно к переменным роста в статье (Bigler, et al., 2003) добавлены RW_N (средний прирост за лет N) и $\log \text{relBAI}$ (это логарифм отношения приращения базальной площади к общей базальной площади дерева).

В работе (Bigler, et al., 2004a) исследуется Европейская пихта (*Abies alba* Mill.), произрастающая в Словении. Переменные роста следующие: относительный базальный прирост (relbai) и тренд роста за последние 5 лет (locrec5).

Засуха в 1998-1999 годов в Северной Патагонии вызвала высокую смертность Бука– *Nothofagus dombeyi* (coihue), доминирующего в национальном парке Nahuel Huapi (Suarez, et al., 2004). Исследование живых и мертвых деревьев проводились методами дендрохронологического анализа. Один из результатов статьи - построение дискриминантной функции, разделяющей живые и мертвые деревья с использованием среднего прироста ширины годичных колец за последние 5 лет.

Das et al., (2007) исследует смертность Белой пихты – *Abies concolor* (Gord. & Glend.) Lindl и Сосны сахарной– *Pinus lambertiana* Dougl путем разработки логистических моделей с использованием трех показателей роста, полученных из древесных колец за последние N лет: средний рост (avgN), тенденция роста (trendN) и количество резких спадов роста (abruptN), где N принимает значения из интервала от 5 до 40 лет с шагом 5 лет. Построенные логистические модели корректно классифицируют 78.6% погибающих и 83.7%

выживающих деревьев. Место произрастания для двух распространенных видов хвойных - район Сьерра-Невада в Калифорнии.

В работе (Hartmann et al., 2007) анализируются следующие породы деревьев: Белая ель (*Picea glauca* (Moench) Voss), Клен сахарный (*Acer saccharum* Marsh.), Пихта бальзама (*Abies balsamifera* (L.) P. Mill.), Ель черная (*Picea mariana* (P. Mill.) BSP), Ель обыкновенная (*Picea abies* (L.) Karst.). Выбраны четыре места произрастания в провинции Квебека, Канады (Temiscaming, два места в Lower St. Lawrence и Abitibi) и национальный парк в северной Финляндии (Pallas Yllas Tunturi, Finland). Исходные данные были разбиты на три класса по DBH (диаметр дерева на высоте груди): 19,1–29,0 см, 29,1–39,0 см и 39,1–49,0 см. В каждом классе было по 30 живых и 30 мертвых деревьев. Для каждого дерева рассчитан уровень роста – переменная Med_N , это вычисленная медиана для N равного 3, 5 и 10 годам и уровень роста тренда – переменная SLP_N , это угол наклона линии, вычисленной методом линейной регрессии за 3, 5, 10, 25 и 35 лет. В данной работе вычисляются коэффициенты логистических регрессионных моделей от двух переменных Med_N и SLP_N . Модели оцениваются коэффициентом D_{xy} – аналог площади ROC кривой и коэффициентом Nagelkerke R_N^2 (Nagelkerke, 1991).

Macalady и Bugmann (2014) изучают влияние засух в периоды 1950-х и 2000-х годов на процессы смертности Сосны колорадской (*Pinus edulis*). При разработке логистических регрессионных моделей использовались архивные образцы прироста древесных колец из четырех мест на юго-западе США. Для моделей были рассчитаны относительные приращения базальной области (Rel-BAI), используя приращения базальной площади (BAI), и индексы приростов колец (RWI), используя значения приростов древесных колец (RW). В качестве переменных роста в моделях рассчитывались средние: прирост RW_N , индекс RWI_N , базальный прирост BAI_N , чувствительность MS_N , стандартное отклонение SD_N и другие показатели, где N принимало значения равное 3, 5, 7, 10, 15, ..., 50 годам. Самые экономные модели обладали высокой

дискриминационной способностью (AUC равен 0.84) и правильно классифицировали 70% деревьев.

Зависимость изменения дефицита воды на увеличение смертности Сосны обыкновенной (*Pinus sylvestris* L.) расположенной на пределе низких высот роста видов древесных растений в Центральной Испании - объект исследования в работе (Gea-Izquierdo et al., 2014). Авторы проанализировали факторы, определяющие состояние здоровья Сосен в сравнении с сопутствующей и более устойчивой к засухе породе Дуба пиренейского (*Quercus pyrenaica* Willd). В работе использовалась порядковая логистическая регрессионная модель. Для вычисления значений переменных роста выбраны дендрохронологические ряды: прирост ширины годичных колец (RW) и прирост базальной площади (BAI). В модели использовались следующие переменные роста: относительный прирост базальной площади (relBAI), средняя чувствительность (ms), количество лет с отрицательными изменениями роста ниже 25% (NGC); количество лет с положительными изменениями роста более 25% (PGC) и другие.

На основании анализа рассмотренных выше статей делаем вывод, что при разработке логистических моделей используются следующие информационные объекты: дендрохронологические ряды (RW, RWI, BAI и другие), статистики (среднее, медиана, индекс прироста и другие), временные окна и алгоритмы логистических регрессионных моделей. Переменные роста определяются выбором статистики и размером окна.

Описание метода.

ПРЕДМЕТНЫЕ ДАННЫЕ ДЛЯ МОДЕЛИРОВАНИЯ УСЫХАНИЯ ДЕРЕВЬЕВ.

При сборе экспериментального материала отметим, что авторы используют важные дендрохронологические принципы: закон лимитирующих факторов (закон минимума Либиха), принцип отбора местообитания и другие (Шиятов, 1973).

Выбор места произрастания древесных растений, как живых, так и усыхающих, определяется однотипными орографическими, эдафическими и другими факторами. При разработке логистической модели в случае, если разброс деревьев по DBH большой, то данные разбивают на две группы и для каждой группы осуществляется поиск переменных роста (Das et al., 2007). На Рис. 2. в самом общем виде изображены взаимосвязи между абиотическими, биотическими факторами и их влияние на рост древостоя.

В экспериментальных данных по каждому дереву выделим точечные и регулярные измерения. К точечным измерениям относятся следующие: тип породы, статус дерева (живое, усыхающее), высота, диаметр на высоте груди и другие характеристики места произрастания. Регулярные измерения проводятся в лабораторных условиях, по кернам взятых с деревьев (Шиятов, 2000; Speer, 2010). По каждому годичному кольцу с керна проводят следующие метрические измерения: ширина годичного кольца, ширина ранней и поздней зоны, параметры плотности колец и т.д. Таким образом, для каждого дерева формируются дендрохронологические ряды (RW, RWI, BAI и другие), на основании этих рядов вычисляются производные временные ряды различными методами стандартизации (Мазепа, 1982; Cook et al., 1990; Шишов и др., 2015). Используя дополнительные измерения дерева можно вычислить базальную площадь прироста (Visser, 1995; Bowman et al., 2013) и т.д. Расширим точечные данные до заголовочных, вводя описание данных об исследователе, полевые записи о месте произрастания древостоя. Вариант программной реализации подобной структуры данных приводится в работе (Качаев, 2017).

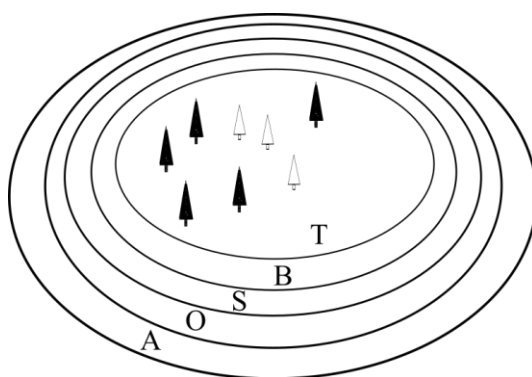


Рис. 2. Диаграмма Эйлера-Венна. Влияние совокупности факторов на рост древесных растений. Где А- абиотические, О- орографические, S- эдафические, В- биотические факторы и Т- место произрастания древостоя (живого и усыхающего)

СТРУКТУРА ВХОДНЫХ ДАННЫХ ДЛЯ ЛОГИСТИЧЕСКОЙ МОДЕЛИ.

Логистическая регрессия (Hosmer and Lemeshow, 2000) – статистическая модель, которая применяется для предсказания вероятности возникновения некоторого события по значениям множества признаков (переменных). Формула логистической регрессии следующая: $P(Y|X) = \frac{1}{1+e^{-f(X)}}$. Где $P(Y|X)$ – вероятность события при векторе признаков $X = (X_1, X_2 \dots X_m)$ принимает значения между 0 и 1, $f(X) = b_0 + \sum_{i=1}^m b_i X_i$ – линейная функция от X_i признаков

Таблица 2. Вид входной таблицы для логистической регрессии.

		X					
		X ₁	X ₂	.	.	.	X _m
Y	y ₁	X _{1,1}	X _{1,2}	.	.	.	X _{1,m}
	y ₂	X _{2,1}	X _{2,2}	.	.	.	X _{2,m}

	y _n	X _{n,1}	X _{n,2}	.	.	.	X _{n,m}

Листинг 1. Формирование входной таблицы для логистической регрессии.

```
# пример входной таблицы для логистической регрессии.
# рассмотрим вариант для трех переменных:
# x1,x2,x3
# 2020.06.26 Alex Kachaev
#
```

```
data.log<- data.frame(x1,x2,x3)
model<- gme("y~x1+x2+x3", data= data.log, family = binomial)
#
```

Выделим три предметных множества при формировании переменных в логистических регрессионных моделях смертности деревьев: типы дендрохронологических рядов, статистики и система различных «окон».

Для описания переменных для логистических регрессионных моделей, введем следующие множества и обозначения.

Пусть $D_n = \{\mathbf{d}_i | i \in [1, n]\}$ множество исходных дендрохронологических рядов. Обозначим через $length(\mathbf{d}_i)$ функцию длины ряда \mathbf{d}_i , она определяет возраст i дерева. Заметим, каждое древесное кольцо имеет год роста, тогда для упрощения индексации пронумеруем их от 1 до возраста дерева. Переход от индекса к году роста каждого древесного кольца не вызовет затруднений.

При построении вероятностных моделей смертности деревьев, авторы Das et al., (2007) анализируют информацию о приросте за последние 40 лет, в работе (Cailleret et al., 2016) 50 лет. Далее будем рассматривать дендрохронологические ряды за последние p лет из D_n .

Пусть $D_{n,p} = \{d_{i,p}, d_{i,p-1}, \dots, d_{i,1} | i \in [1, n]\}$ множество дендрохронологических рядов с индексацией с «хвоста», где p — длина ряда, i — номер ряда, n — количество исследуемых деревьев. Отметим, что обратный переход индексации ряда от 1 до p , к обозначению конкретных возрастных дат исходного ряда не вызовет затруднений. Представим $D_{n,p}$, в виде таблицы 3.

Таблица 3. Таблица дендрохронологических рядов за последние p лет.

	p	p-1	.	.	.	2	1
\mathbf{d}_1	$d_{1,p}$	$d_{1,p-1}$.	.	.	$d_{1,2}$	$d_{1,1}$
\mathbf{d}_2	$d_{2,p}$	$d_{2,p-1}$.	.	.	$d_{2,2}$	$d_{2,1}$

\mathbf{d}_n	$\mathbf{d}_{n,p}$	$\mathbf{d}_{n,p-1}$.	.	.	$\mathbf{d}_{n,1}$	$\mathbf{d}_{n,1}$

Выберем вариант представления $D_{n,p}$ на языке программирования R (Кабаков, 2014) как тип данных `data.frame()`. `Dnp<- data.frame($\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_n$)`, где \mathbf{d}_i - вектор (дендрохронологический ряд) для $i \in [1, n]$

Обозначим через $TD = \{td \mid td \in \{RW, RWI, BAI, et al. \}\}$ множество различных типов измерений по приросту годичных колец и производные с использованием различных методов стандартизации.

С каждым дендрохронологическим рядом свяжем «статус» $Y = \{y_1, y_2, \dots, y_n \mid y_i \in \{0,1\}, i \in [1, n]\}$. «Статус» 1 обозначает дерево живое и 0-усыхающее.

ОКНА. СИСТЕМЫ ОКОН

«Окном»¹ будем называть любую конечную последовательность элементов дендрохронологического ряда.

Поскольку окна находятся во взаимном однозначном соответствии с последовательностью натуральных чисел, нумерующих элементы дендрохронологического ряда, то естественным образом эту последовательность также можно называть окном.

Определение. Окно $w_{l,r} = \{l, l-1, \dots, r+1, r\}$ это последовательность чисел, от l до r по убыванию и $l > r$.

Другим обозначением окна будем считать символическую запись $[l, r]$.

Определение. Размер окна определяется количеством чисел в окне.

Очевидно, что размер окна $w_{l,r}$ равен $l - r + 1$.

Приведем два варианта обозначения окна: $w_{l,r}$ и $w_{c,s}$, где l, r левый и правый индекс, c индекс центра и s размер окна. Приведем обозначения окон

¹ Далее левую и правую кавычки(«») опускаем.

на Рис. 2.: $w_{2,6,1}$, $w_{3,8,1}$, $w_{4,11,9}$, $w_{5,15,5}$, заметим, что окно w_1 можно обозначить по-разному $w_{1,2,3}$ или $w_{1,3,1}$. Далее будем использовать только вариант обозначения окон – $w_{l,r}$.

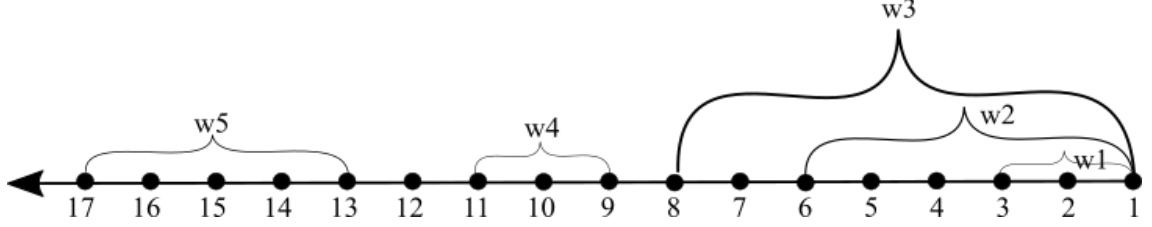


Рис. 2. Варианты расположения вложенных и непересекающихся окон².

Определение. Обозначим $\mathbf{W}_1^p = \{w_{l,r} | p \geq l > r \geq 1\}$ - множество окон всех размеров для ряда длины p .

Предложение. $|\mathbf{W}_1^p| = \frac{p!}{(p-2)!2!}$.

Док-во. Во множестве \mathbf{W}_1^p для каждого окна $[l, r]$ поменяем местами числа l, r таким образом, пары чисел будут множеством сочетаний из p по два.

Формула числа сочетаний из p по два равна $C_p^2 = \frac{p!}{(p-2)!2!}$. ■³

Определение. Обозначим через $\mathbf{W}_2^p = \{(w_{l,r}, w_{h,k}) | w_{l,r}, w_{h,k} \in \mathbf{W}_1^p\}$ множество произвольных пар окон.

Определение. Два окна $w_{l,r}$ и $w_{h,k}$ с условием $[l, r] \cap [h, k] = \emptyset$ будем называть *непересекающимися*.

Предложение. Пусть $w_{l,r}$ и $w_{h,k}$ непересекающие окна, тогда либо $r > h$ или $k > l$.

Множество непересекающихся пар окон обозначим $\ddot{\mathbf{W}}_2^p = \{(w_{l,r}, w_{h,k}) | w_{l,r} \cap w_{h,k} = \emptyset, p \geq l, k \geq 1\}$.

Определение. Окно $w_{h,k}$ вложено в окно $w_{l,r}$, если $w_{h,k} \subseteq w_{l,r}$.

Предложение. Окно $w_{h,k}$ вложено в окно $w_{l,r}$ тогда и только тогда, когда $l \geq h$ и $k \geq r$.

² Определения вложенных и непересекающихся окон будут даны позже.

³ Знак ■ - обозначающий конец доказательства.

Определение. Обозначим множество окон $\dot{\mathbf{W}}_1^p = \{w_{l,1} | p \geq l \geq 1\}$, где правый индекс окна равен 1.

Предложение. $|\dot{\mathbf{W}}_1^p| = p - 1$.

Определение. Обозначим множество пар окон $\dot{\mathbf{W}}_2^p = \{(w_{l,1}, w_{r,1})\}$, где $w_{l,1} \in \dot{\mathbf{W}}_1^p$ и $w_{r,1} \in \dot{\mathbf{W}}_1^p$. Очевидно, что пары окон обладают свойством вложенности.

Предложение. Мощность $|\dot{\mathbf{W}}_2^p|$ равна $(p - 1)^2$.

В работе (Cailleret et al., 2016) перечисляются переменных роста, и используемые окна обозначаются W_N , для краткости второй индекс окна опускают, так как он всегда равен 1. Заметим что окно $W_N \approx w_{N,1}$ и таким образом $W_N \in \dot{\mathbf{W}}_1^p$ по определению.

Построим систему списков (list оператор в языке программирования R) из окон, которые будут служить итераторами (Кабаков, 2014) для формирования входных таблиц данных (data.frame()) и вычисления логистических регрессионных моделей с использованием функции glm() (Кабаков, 2014).

Определение. Обозначим через \mathbf{I}^p множество чисел от 1 до $\frac{p!}{(p-2)!2!}$

Множество \mathbf{I}^p вводится для индексации множества окон $\dot{\mathbf{W}}_1^p$. Очевидно по определению \mathbf{I}^p и $\dot{\mathbf{W}}_1^p$ следует, что $|\mathbf{I}^p| = |\dot{\mathbf{W}}_1^p|$.

Пусть множество \mathbf{I}^p упорядочено, а $\dot{\mathbf{W}}_1^p$ лексически упорядочено, очевидно, что существует взаимно однозначная функция $\mathbf{o}: \mathbf{I}^p \leftrightarrow \dot{\mathbf{W}}_1^p$, которая каждому $n \in \mathbf{I}^p$ ставит в соответствие окно $w_{l,r} \in \dot{\mathbf{W}}_1^p$ или в других обозначениях $n \leftrightarrow [l, r]$. Задать функцию \mathbf{o} возможно таблично.

Пусть $\mathbf{C}_{\mathbf{I}^p}^k$ или $\mathbf{C}_{|\dot{\mathbf{W}}_1^p|}^k$ (другой вариант обозначения) множество сочетаний векторов (кортежей) длины k из множества \mathbf{I}^p .

Определение. Пусть $\tilde{\mathbf{w}}_{\mathbf{i}^k}^k = (w_{\mathbf{o}(i_1)}, w_{\mathbf{o}(i_2)}, \dots, w_{\mathbf{o}(i_{k-1})}, w_{\mathbf{o}(i_k)})$ список окон длины k , где вектор (кортеж) $\mathbf{i}^k = (i_1, i_2, \dots, i_{k-1}, i_k) \in \mathbf{C}_{\mathbf{I}^p}^k$ где $i_j \in \mathbf{I}^p$ и $w_{\mathbf{o}(i_j)} \in \dot{\mathbf{W}}_1^p$ для $1 \leq j \leq k$.

Определение. $\mathbf{W}_k^p = \{\tilde{\mathbf{w}}_{i^k}^k | i^k \in \mathcal{C}_{\bar{\mathbf{I}}^p}^k\}$ – множество списков окон длины k .

По определению $|\mathbf{W}_k^p| = C_{\bar{\mathbf{I}}^p}^k$, где $C_{\bar{\mathbf{I}}^p}^k$ – число сочетаний из $\bar{\mathbf{I}}^p$ по k и $C_{\bar{\mathbf{I}}^p}^k = \frac{\bar{\mathbf{I}}^p!}{(\bar{\mathbf{I}}^p - k)!k!}$.

Рассмотрим объединение множества списков окон \mathbf{W}_l^p для $l \in [1, p]$. Обозначим это объединение – $\mathbf{W}^p = \bigcup_{l=1}^p \mathbf{W}_l^p$.

Далее во множестве \mathbf{W}^p введем ограничение на выбор всех окон в списках с правым индексом 1, обозначим полученное множество $\dot{\mathbf{W}}^p = \bigcup_{l=1}^p \dot{\mathbf{W}}_l^p$. Отметим, что при формировании переменных роста для логистических регрессионных моделей, в настоящих исследованиях (Cailleret et al., 2016) используется система окон из множества $\dot{\mathbf{W}}^p$.

Пусть любая пара окон в каждом списке окон множества \mathbf{W}^p будет не пересекающейся. Обозначим полученное множество, как $\ddot{\mathbf{W}}^p$. Далее запишем это условие формально, что любая пара окон не пересекается в списке окон длины k :

$\tilde{\mathbf{w}}_{i^k}^k = (w_{\mathbf{o}(i_1)}, w_{\mathbf{o}(i_2)}, \dots, w_{\mathbf{o}(i_{k-1})}, w_{\mathbf{o}(i_k)})$ где $\forall i_m \& \forall i_n: w_{\mathbf{o}(i_m)} \cap w_{\mathbf{o}(i_n)} = \emptyset$ для $m \neq n$ и $m, n \in [1, k]$.

Очевидно, что $\ddot{\mathbf{W}}^p = \bigcup_{l=1}^{p1} \ddot{\mathbf{W}}_l^p$, где $p1 = p\% / \%2$. Число $p1$ есть целая часть от деления p на 2, при минимальном окне длины 2.

Множества $\dot{\mathbf{W}}^p$ и $\ddot{\mathbf{W}}^p$ будем рассматривать как проекцию или сужение множества \mathbf{W}^p , с целью уменьшить количество расчетных вариантов логистических регрессионных моделей.

ОКОННЫЕ СТАТИСТИКИ. ПЕРЕМЕННЫЕ РОСТА

Далее определим *оконную статистику*.

Определение. Функцию, вычисляющую значение на элементах временного ряда индексированного некоторым окном, будем называть *оконной статистикой*.

Введем обозначение оконной статистики – $St(Ds, w_{l,r})$, где St – идентификатор функции, Ds – временной ряд и $w_{l,r}$ – окно.

Далее на Листинге 2. пример оконной статистики вычисления среднего значения на языке R (Кабаков, 2014).

Листинг 2. Оконная статистика вычисления среднего значения

```
# пример оконной статистики вычисления среднего
# 2020.06.25 Alex Kachaev
St<-function(Ds,w){
  s<-0
  lw<- w[1]-w[2]+1
  t.in<-seq(w[1],w[2],-1)
  for(i in t.in) s<-s+Ds[i]
  return(list(ave=s/lw))
}
```

Определение. Переменной роста будем называть тройку $\langle St, TDs, w_{l,r} \rangle$, где St – идентификатор некоторой оконной статистики, TDs – тип дендрохронологического ряда, $w_{l,r}$ – окно.

Введем правило обозначение переменных роста как конкатенация⁴ трех идентификаторов $\langle St \rangle \langle TDs \rangle \langle w_{l,r} \rangle$. Переобозначим переменные роста из работы (Das et. al., 2007) в соответствии с введенным определением: $aveRW_{w_{N,1}}$ ($aveN^5$), $trendRW_{w_{N,1}}$ ($trendN$) и $abruptRW_{w_{N,1}}$ ($abruptN$).

На Листинге 3. пример вычисления средней переменной роста.

Листинг 3. Вычисления переменной роста.

```
# St- оконная статистики
# 2020.06.08 Alex Kachaev
```

⁴ Операция сцепления строк в информатике.

⁵ Так как в статье используется только ширина годовичного прироста RW , то в обозначения RW опускается и обозначение окна как N , проще, чем $w_{N,1}$ так все окна вложенные.


```

Var.St<-function(Dnp,w){
  lw<-dim(Dnp)[1] # число деревьев
  vec<-as.numeric()
  for(i in 1:lw){
    Ds<- Dnp[i,]; vec<-c(vec, St(Ds,w)) }
  return(list(vec=vec))
}

```

РЕЗУЛЬТАТЫ МЕТА-АНАЛИЗА

Далее построим полную систему переменных роста. Пусть St – множество оконных статистик, Td – множество таблиц измерений по деревьям и W – система окон.

Обозначим Sdw – упорядоченную систему переменных роста, где для каждого $sdw_i \in Sdw$ и $i \in [1:l]$ при $l = |Sdw|$ мы можем восстановить имя оконной статистики, имя таблицы измерений по деревьям и характеристике окна.

Листинг 4. Формирования списка списков переменных роста.

```

# список списков переменных роста.
# Sdw – упорядоченная система переменных роста
# combi(comb,n,j) возвращает следующую перестановку из n по j
# ret.List.Window(Sdw,comb) возвращает список окон
# по перестановки comb
# 2020.06.08 Alex Kachaev
n<-length(Sdw)
LSdw<-NULL
LSdw[[1]]<- Sdw
for(i in 2:k){
  comb<-seq(1, i)
  # формировать список окон длины

```

```

tLSdw<-NULL
tLSdw[[1]]<-ret.List.Window(Sdw,comb)
for(j in 2: choose(n, k)){
  ns<-combi(comb,n, j)
  bl<-ns$bo
  comb<-ns$cm
  tLSdw[[j]]<- ret.List.Window(Sdw,comb)
}
LSdw[[i]]<- tLSdw
}

```

LSdw– список списков переменных роста.

Обозначим через LDf.LSdw– список списков системы табличных данных (типа data.frame()), сформированных по LSdw. Отметим, что $\text{length}(\text{unlist}(\text{LSdw})) = \text{length}(\text{unlist}(\text{LDf.LSdw}))$ и структура списков LSdw и LDf.LSdw– подобны как структура данных.

Далее зафиксируем функции glm(), roc() (Кабаков, 2014). «Обернем» glm() и roc() в процедуру glm.roc(). Назначение glm.roc()– формировать входные данные для glm() и возвращать результаты найденной модели для дальнейшего анализа.

Листинг 5. Формирования списка логистических регрессионных моделей.

```

# фрагмент программы на языке R
ldf<-length(LDf.LSdw)
s<-0
LRM<-NULL
for(i in 1:ldf){
  t.LDf.LSdw<- LDf.LSdw[[i]]
  for(j in 1: length(t.LDf.LSdw)){
    s<-s+1
    LRM[[s]]<-glm.roc(t.LDf.LSdw[[j]])
  }
}

```

}
}

Определение. Семейство данных **LRM** будем называть *замыканием* относительно $glm()$, St – множество оконных статистик, Td – множество таблиц измерений по деревьям и W – система окон.

Предложение. Семейство логистических регрессионных моделей $LRM = \langle St, Td, W, glm() \rangle$ полно.

Док-во. Список моделей **LRM** программно формируется в результате полного перебора $LDf.LSdw$ – табличных данных (см. фрагмент Листинга 5.), таким образом определяется полнота. ■

Модели **LRM**– это подобные структуры данных, сформированные функцией $glm.roc()$, которые в дальнейших исследованиях можно сравнивать и выбирать «наилучшие», например, по значению AUC.

ОБСУЖДЕНИЕ

В результате описания полной системы окон W^p и выделение ней двух подсистем \dot{W}^p и \ddot{W}^p , позволяет нам утверждать, что поиск переменных роста для логистических моделей в текущих научных исследованиях (Cailleret et al., 2017) осуществляется с использованием системы вложенных окон W^p .

Формирование подсистемы непересекающихся окон \ddot{W}^p , предлагает новые варианты формирования переменных роста.

В работе Качаева (2019) приводится пример построения модели с использованием системы непересекающихся окон. В качестве примера выбран Южный бук (*Nothofagus dombeyi*) (Suarez et al., 2004) с данными прироста годовичных колец, находящихся в открытом доступе (Cailleret et al., 2016). В результате расчетов найдено более 100 пар окон. Значения AUC лежат в пределах от 0.77 до 0.82, что относится к разделению от приемлемого до

отличного по классификации, представленной в работе (Hosmer and Lemeshow, 2000).

Определим содержательно разность подходов.

Формирование разных переменных роста с использованием системы вложенных окон за последние N лет в логистических моделях смертности деревьев определяют нахождение линейной дискриминантной функции.

Использование системы непересекающихся окон при формировании логистических моделей смертности определяется анализом переменных роста деревьев на двух или более различных временных участках.

Мы считаем, что обозначенное различие двух подходов в построении логистических регрессионных моделей открывает перспективы использования системы непересекающихся окон.

БЛАГОДАРНОСТИ

Выражаю благодарность д-ру физ. - мат. наук Евгению Лейнартасу за обсуждения и полезную дискуссию в подготовки статьи.

СПИСОК ЛИТЕРАТУРЫ

- Bigler C, Bugmann H* (2004) Predicting the time of tree death using dendrochronological data. *Ecological Applications*, 14:902–914.
- Bigler C, Gričar j, Bugmann H, Čufar K* (2004a) Growth patterns as indicators of impending tree death in silver fir. *Forest Ecology and Management*, 199:183-190
- Bigler, C., & Bugmann, H.* (2003). Growth-dependent tree mortality models based on tree rings. *Canadian Journal of Forest Research*, 33(2), 210–221. <https://doi.org/10.1139/x02-180>
- Bowman, D. M. J. S., R. J. W. Brienen, E. Gloor, O. L. Phillips, and L. D. Prior.* (2013). Detecting trends in tree growth: not so simple. *Trends in Plant Science* 18:11–17.

- Cailleret, M. , Jansen, S. , Robert, E. M., Desoto, L. , ..., Martínez-Vilalta, J. (2017),*
A synthesis of radial growth patterns preceding tree mortality. *Glob Change Biol*, 23: 1675-1690. doi:10.1111/gcb.13535
- Cailleret, M., Bigler, C., Bugmann, H., Camarero, J. J., Cufar, K., Davi, H., ... Martínez-Vilalta, J. 2016.* Towards a common methodology for developing logistic tree mortality models based on ring-width data. *Ecological Applications*, 26(6), 1827–1841. <https://doi.org/10.1890/15-1402.1>
- Cook, E. R., Briffa, K. R., Shiyatov, S. G., Mazepa, V. S., 1990.* Treering standardisation and growth-trend estimation. In: Cook, E.R., Kairiukstis, L.A. (Eds.), *Methods of Dendrochronology: Applications in the Environmental Sciences*. IIASA/Kluwer, Dordrecht, pp. 104 – 123.
- Das A. J., Battles J. J., Stephenson N. L., and van Mantgem P. J. (2007)* The relationship between tree growth patterns and likelihood of mortality: a study of two tree species in the Sierra Nevada. *Canadian Journal of Forest Research*, 37:580–597
- Gea-Izquierdo G, Viguera B, Cabrera M, and Cañellas I. (2014)* Drought-induced decline could portend widespread pine mortality at the xeric ecotone in managed Mediterranean pine-oak woodlands. *Forest Ecology and Management*, 320:70–82
- Glass, G.V. (2000).* Meta-Analysis at 25. <http://www.gvglass.info/papers/meta25.html>
- Hartmann H, Messier C, and Beaudet M (2007)* Improving tree mortality models by accounting for environmental influences. *Canadian Journal of Forest Research*, 37:2106–2114
- Hosmer, D. W., and Lemeshow, S. 2000.* Applied logistic regression. John Wiley & Sons, New York.
- Liang J., Zhou M., Watson J. V., Crowther T. W., Glick H. B., Picard N., Tavani R., Wiser S., Alberti G., Schulze E. D., McGuire A. D., Bozzato F., Pretzsch H., Brandl S., De-Miguel S., Paquette A., Hérault B., Scherer-Lorenzen M., Barrett*

- C. B., Hengeveld G. M. et al.* Positive biodiversity-productivity relationship pre-dominant in global forests. *Science*. 2016. T. 354. № 6309.
- Macalady A. K., and Bugmann H.* (2014) Growth-mortality relationships in piñon pine (*Pinus edulis*) during severe droughts of the past century: shifting processes in space in time. *PlosOne*, 9(5): e92770. doi:10.1371/journal.pone.0092770
- Monserud Robert A.* Simulation of Forest Tree Mortality. *Forest Science*, Volume 22, Issue 4, December 1976, Pages 438–444
- Nagelkerke, N. J. D.* (1991) A Note on a General Definition of the Coefficient of Determination. *Biometrika*, 78, 691–2.
- Scott, M. L., Shafroth, P. B., and Auble, G. T.* 1999. Responses of Riparian Cottonwoods to Alluvial Water Table Declines. *Environ. Manage.* **23**(3): 347–358. doi:10.1007/s002679900191.
- Speer J. H.* (2010) Fundamentals of Tree-Ring Research. University of Arizona Press. p 368.
- Suarez, M., Ghermandi, L., and Kitzberger, T.* (2004). Factors Predisposing Episodic Drought-Induced Tree Mortality in *Nothofagus*: Site, Climatic Sensitivity and Growth Trends. *Journal of Ecology*, 92(6), 954–966. Retrieved April 3, 2020, from www.jstor.org/stable/3599738
- Visser, H.* (1995). Note on the relation between ring widths and basal area increments. *For. Sci.* 41, 297–304.
- Wycoff P. H., Clark J.S.* (2000) Predicting tree mortality from diameter growth: a comparison of maximum likelihood and Bayesian approaches. *Canadian Journal of Forest Research*, 30:156-167
- Антон Г.* Анализ таблиц сопряженности. М.: Финансы и статистика, 1982. – 143 с.
- Замолодчиков Д. Г., Краев Г. Н.* Влияние изменений климата на леса России: зафиксированные воздействия и прогнозные оценки. Устойчивое лесопользование. 2016. № 4 (48). с. 23–31.
- Кабаков Р. И.* R в действии. Анализ и визуализация данных в программе R / пер. с англ. Полины А. Волковой. – М.: ДМК Пресс, 2014. – 588 с.

- Качаев А. В.* Модель описания структуры дендроклиматических данных В сборнике: Региональные проблемы дистанционного зондирования Земли Материалы IV международной научной конференции. Сибирский федеральный университет, Институт космических и информационных технологий. 2017. С. 120–122.
- Качаев А. В.* О выборе переменных в логистических регрессионных моделях усыхания деревьев В сборнике: Лесные экосистемы бореальной зоны: биоразнообразие, биоэкономика, экологические риски Материалы Всероссийской конференции с международным участием. 2019. С. 165–168.
- Мазена В. С.* Метод расчета индексов годичного прироста обобщенного дендроклиматологического ряда. Экология. 1982. № 3. С. 21–27.
- Малахова Е. Г., Лямцев Н. И.* Распространение и структура очагов усыхания еловых лесов Подмосковья 2010–2012 годах // Известия Санкт-Петербургской лесотехнической академии. 2014. Вып. 207. С. 193–201.
- Павлов И. Н.* Биотические и абиотические факторы усыхания хвойных лесов Сибири и Дальнего Востока // Сибирский экологический журнал. – 2015. – Т. 22, № 4. – С. 537–554.
- Розенберг Г. С.* Экология и физика: параллели или сети? (в продолжение дискуссии). Биосфера. 2011. Т. 3. № 3. С. 296–303.
- Сарнацкий В. В.* Зонально-типологические закономерности периодического массового усыхания ельников Беларуси. Труды БГТУ. 2012. № 1. Лесное хозяйство. С. 274–276.
- Харук В. И., Им С. Т., Петров И. А., Ягунов М. Н.* Усыхание темнохвойных древостоев прибайкалья. Сибирский экологический журнал. 2016. Т. 23. № 5. С. 750–760
- Шишов В. В., Тычков И. И., Попкова М. И.* Методы анализа дендроклиматических данных и их применение для территории Сибири: Учебное пособие ФГАОУ ВПО «Сибирский федеральный университет». – Красноярск, 2015 – 210с.

Шиятов С. Г. Дендрохронология, ее принципы и методы // Записки Свердловского отделения ВБО. Свердловск, 1973. Вып. 6. С.53–81.

Шиятов С. Г., Ваганов Е. А., Кирдянов А. В., Круглов В. Б., Мазена В. С., Наурзбаев М. М., Хантемиров Р. М. Методы дендрохронологии. Часть I. Основы дендрохронологии. Сбор и получение древесно-кольцевой информации: Учебно-методич. пособие. Красноярск: КрасГУ, 2000. 80 с.

A NEW APPROACH TO META-ANALYSIS OF VARIABLES OF TREE MORTALITY LOGISTIC REGRESSION MODELS USING ANNUAL GROWTH OF RINGS

A. V. Kachaev

Siberian Federal University

660041, Krasnoyarsk, Svobodny ave., 79

E-mail: avkachaev@gmail.com

Using a meta-analysis, we built a subject model of the data obtained by dendro-chronological tree mortality studies underway with the help of logistic regression models. The initial data of the subject area were tree increment time-series and their derivatives. Growth variable was notionally refined as a window statistic. We determined the complete window system and subsystems of nested and disjoint windows within it, as well as the closure operation of the formation of all logistic regression models for the data of the subject area. Today's tree mortality studies involving logistic regression use a subsystem of nested windows to calculate growth (increment) variables. The novelty of our approach to the development of growth (increment) variables is in the use of a subsystem of disjoint windows. The software implementation of the subject data model was done based on R computer language.

Key words: *meta analysis, dendrochronology, logistic regression; tree mortality, tree growth, growth variables.*