Санкт-Петербургский Государственный Университет

Кафедра информационно-аналитических систем

# Автоматическое Выделение Фаз Ментальной Активности

Курсовая работа студентки 341 группы

Дубатовка Алины

Научный руководитель

Д.ф.-м.н., проф. Б.А. Новиков

Санкт-Петербург

2014

## Введение

В современном мире огромная часть данных производится не человеком, а различными устройствами, такими как датчики, приборы и камеры. Помимо того, что эта информация генерируется в неограниченных объёмах и динамически изменяется во времени, она также требует различной обработки в условиях реального времени, которая не может быть осуществлена при помощи только человеческих ресурсов. Примером может служить выделение в них фаз с различными заранее заданными свойствами. Анализ реальных потоковых данных затрудняется не только большими объёмами, непрерывным изменением и необходимостью онлайн-обработки, но часто ещё и наличием шума, мешающего анализировать поведение, выделять тренды и осуществлять проверку критериев фаз. Одной из таких задач является выделение фаз ментальной активности путём анализа графика частоты сердечных сокращений во время различных умственных нагрузок. В данной работе предложены несколько методов для выделения фаз и проведено их сравнение с целью выявления наиболее подходящих для данной задачи.

## Постановка Задачи

Данные можно разбить на 2 основных этапа, границы которых нам известны, поскольку известны моменты получения заданий испытуемыми и моменты прекращения выполнения заданий:

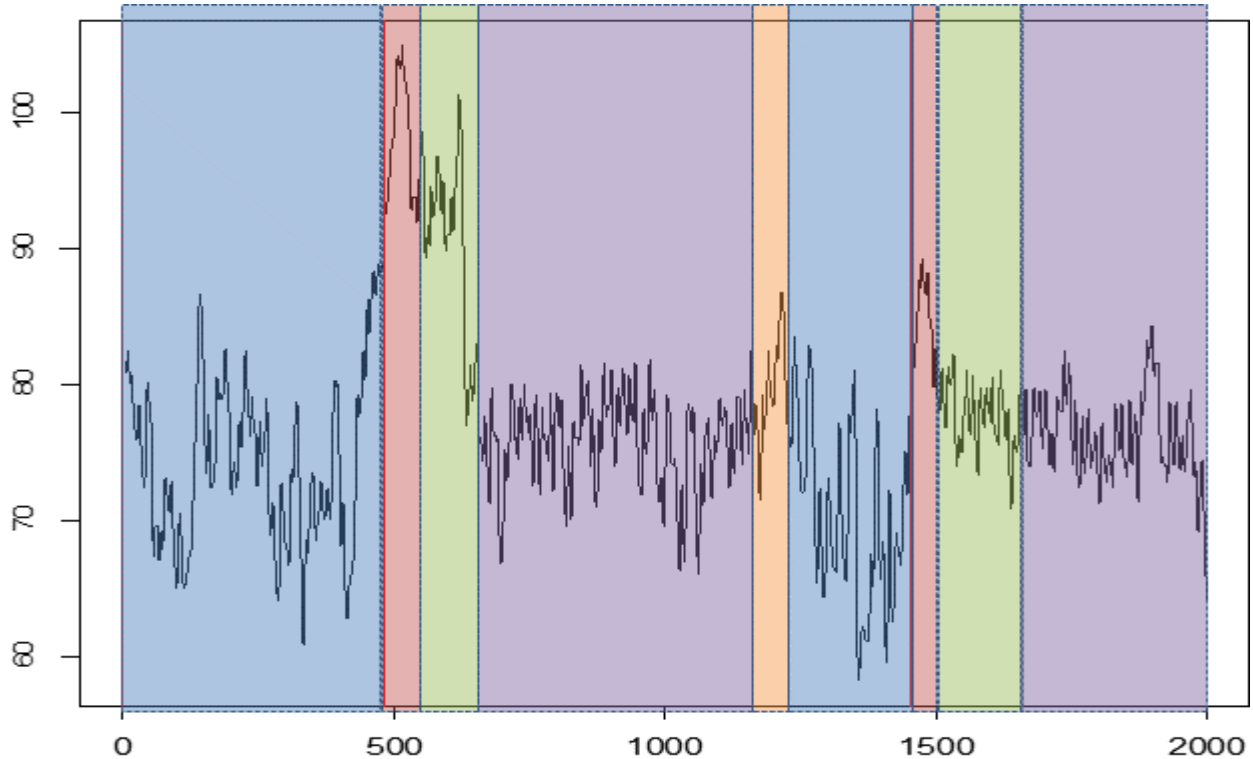
- Выполнение задания
- Покой

В имеющихся данных предполагается выделять следующие фазы (интервалы):

- Отдых (синий)
- Мобилизация (красный)
- Стабилизация (зеленый)
- Работа (фиолетовый)
- Восстановление (оранжевый)

Важно отметить, что фазы мобилизации, стабилизации и работы составляют этап выполнения задания, а фазы восстановления и отдыха – этап покоя.

На картинке приведен график ЧСС одного из испытуемых, на котором примерно обозначены интервалы времени, соответствующие фазам.



Ниже перечислены некоторые свойства этих фаз, по которым мы могли бы их различать:

- Отдых (синий)
    - Относительно низкая средняя ЧСС (по сравнению с другими фазами)
    - Относительно большая вариация ЧСС (по сравнению с другими фазами)
    - Заканчивается в момент получения задания (с возможной краткосрочной задержкой)
- Мобилизация (красный)
    - Стабильный подъем ЧСС во времени (заметный по сравнению с обычными колебаниями)
    - Начинается в момент получения задания (с возможной краткосрочной задержкой)
    - Среднее значение ЧСС выше, чем на фазе отдыха
    - Вариация ЧСС ниже, чем на фазе отдыха
- Стабилизация (зеленый)
    - Снижение ЧСС во времени
    - Среднее значение ЧСС выше, чем на фазе отдыха
    - Вариация ЧСС ниже, чем на фазе отдыха
- Работа (фиолетовый)
    - Среднее значение ЧСС выше, чем на фазе отдыха
    - Вариация ЧСС ниже, чем на фазе отдыха
    - Заканчивается в момент прекращения работы над заданием (с возможной краткосрочной задержкой)
- Восстановление (оранжевый)
    - Снижение ЧСС во времени
    - Среднее значение ЧСС выше, чем на фазе отдыха
    - Вариация ЧСС ниже, чем на фазе отдыха
    - Начинается в момент прекращения работы над заданием (с возможной краткосрочной задержкой)

## Описание данных

Данными для данной работы являются показания ЧСС (частоты сердечных сокращений) трех групп испытуемых, выполняющих по очереди 5 заданий с перерывами на отдых: новички, новички после обучения и эксперты. Показания снимались 4 раза в секунду, а всё испытание длилось 1375 секунд. Ниже приведена таблица временных рамок нагрузки.

| Начало | Окончание | Тип |
|---|---|---|
| 0 | 120 | Покой |
| 120 | 290 | Нагрузка тип1 |
| 290 | 363 | Покой |
| 363 | 533,5 | Нагрузка тип2 |
| 533,5 | 606 | Покой |
| 606 | 776,5 | Нагрузка тип3 |
| 776,5 | 849 | Покой |
| 849 | 1020 | Нагрузка тип4 |
| 1020 | 1092 | Покой |
| 1092 | 1262,5 | Нагрузка тип5 |
| 1262,5 | 1375 | Покой |

Первые эксперименты по автоматическому выделению фаз были проведены на этапах выполнения заданий, поскольку фазы мобилизации, стабилизации и работы являются более значимыми с точки зрения практического использования.

В качестве данных приведённых ниже экспериментов использовались показания ЧСС экспертов со 110-й по 300-ю секунду (10 секунд до начала выполнения первого задания и 10 секунд после окончания первого задания). Границы задания (120-я и 290-я секунды) на последующих графиках обозначены зелёными прямыми.

## Описание используемых методов

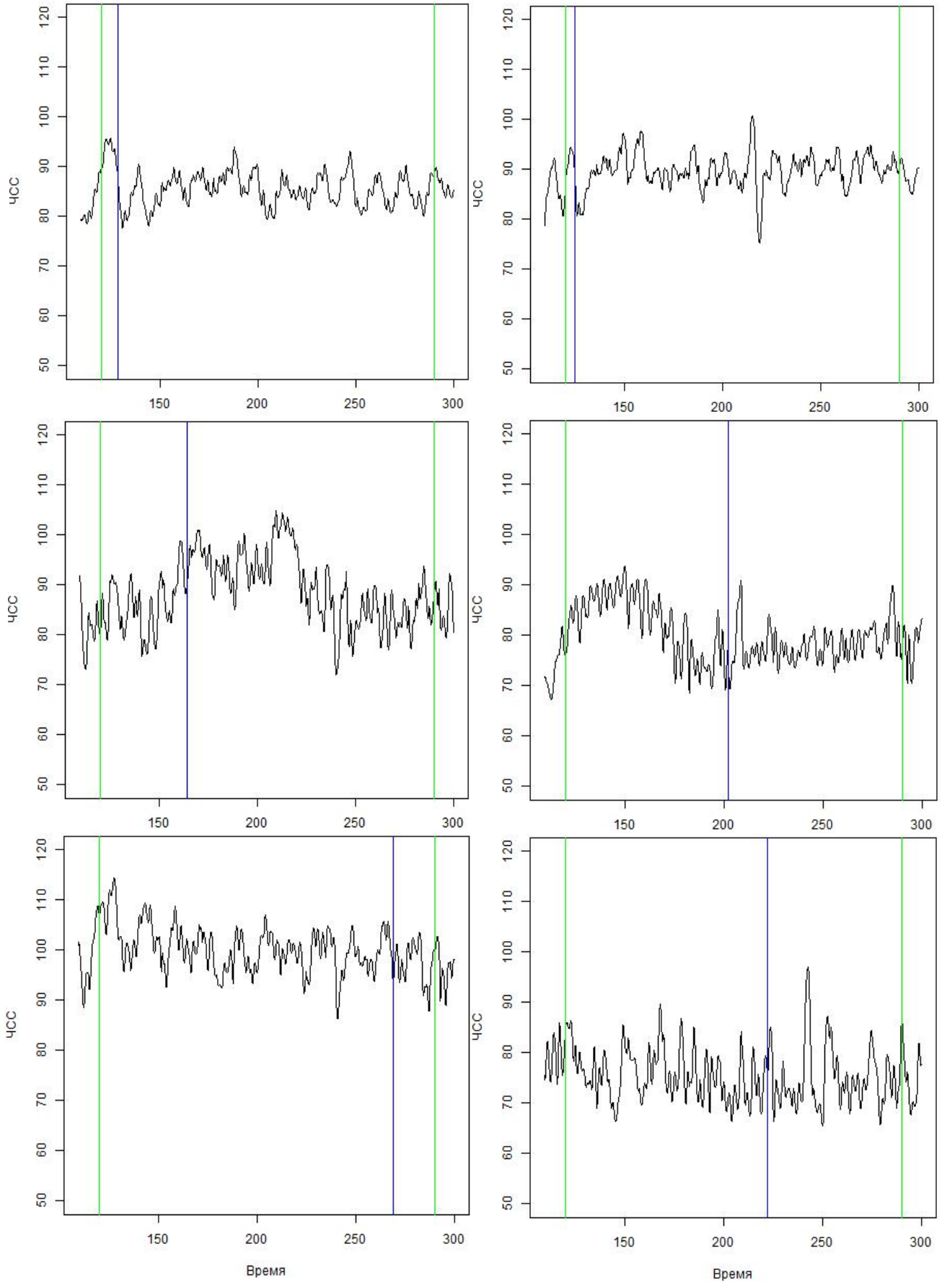
Поскольку выделение границ фаз мобилизация-стабилизация и стабилизация-работа происходит независимо, можно выделять их отдельно и для каждой из них попробовать применить свои методы. В данной работе мы попробовали сравнить результаты различных методов для выделения каждого вида разрезов с целью выявления наиболее эффективных в каждом случае. Некоторые из методов похожи, другие специфичны для каждого вида разрезов, поскольку учитывают свойства выделяемых фаз. Поскольку в данных нет разметки «правильной» разбивки этапа выполнения задания на фазы, качество работы методов оценивалось на глаз и примеры результатов каждого показаны на нескольких картинках.

### Выделение границы фаз мобилизации и стабилизации

В этом разделе описаны эксперименты по поиску только границы фаз мобилизации и стабилизации. Обработка начиналась с точки минимума на отрезке со 110-й по 120-ю секунду (временной интервал, предшествующий началу задания) в предположении, что мобилизация начинается с точки минимума на отрезке перед началом задания. Предполагаемый разрез обозначен синей прямой.

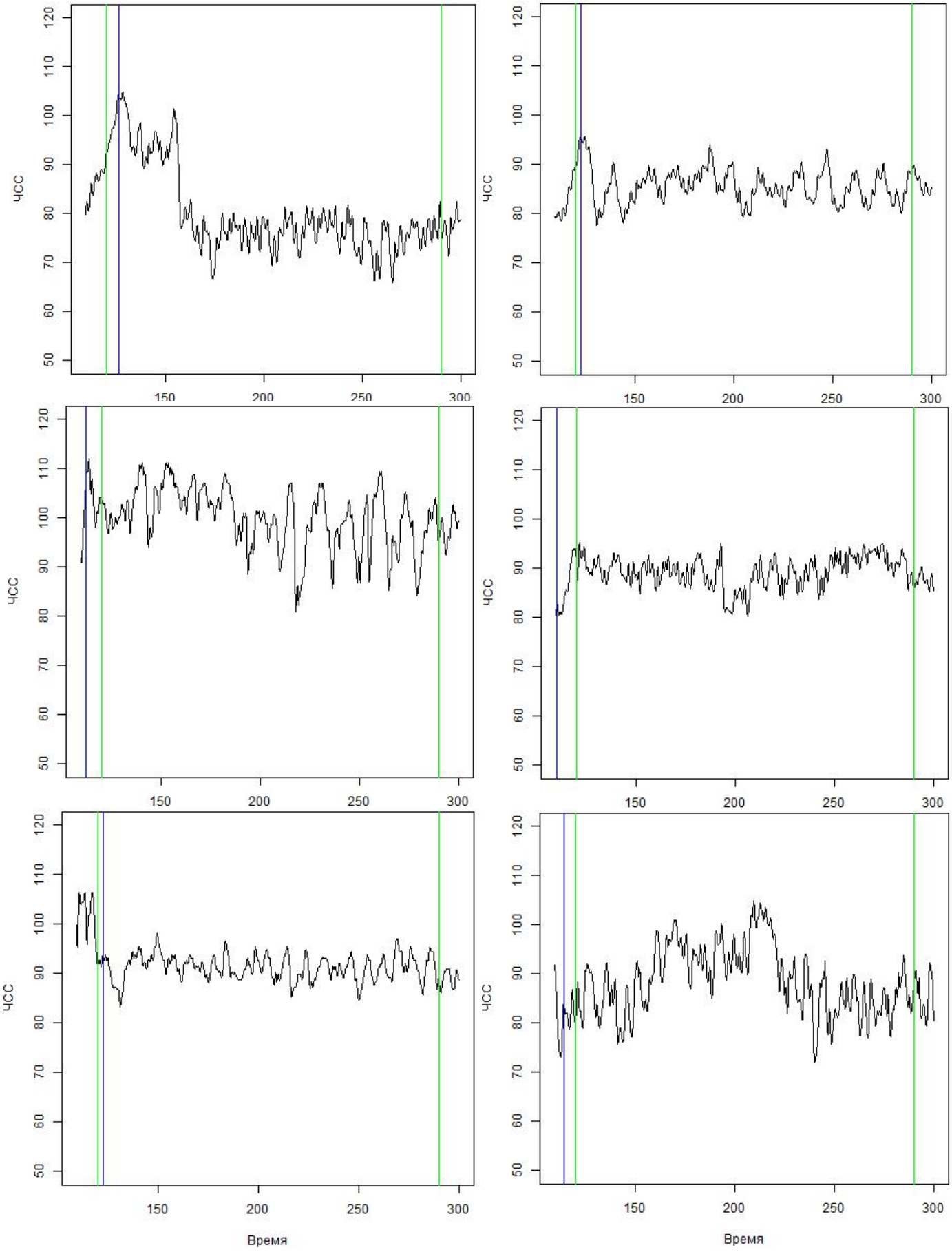
### Склеивание интервалов, выделенных с помощью алгоритма changepoints

Сначала алгоритм changepoints, применённый к частичным разностям, вычисленным для сглаженных данных (усредненных по окну длины 10), разбивает исходные данные на несколько отрезков (в данном случае, не более чем на 10). Затем на полученных отрезках высчитывается коэффициент линейной регрессии для несглаженных данных. Далее все левые отрезки с положительным коэффициентом линейной регрессии склеиваются и строится предполагаемая граница фаз мобилизации и стабилизации.



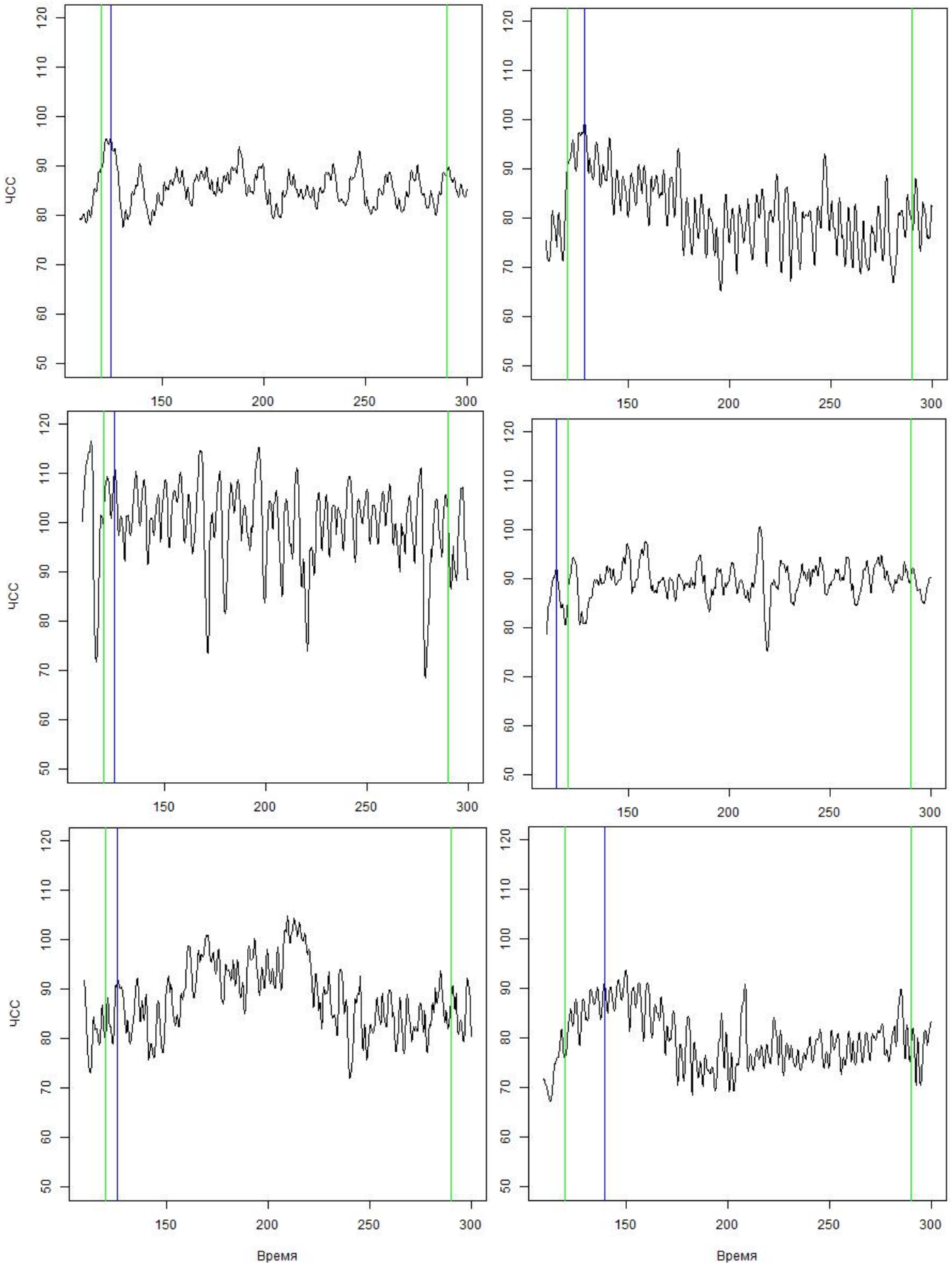
### Склеивание интервалов, выделенных с помощью поиска локальных экстремумов

В следующем эксперименте разбивка на отрезки происходила в точках локальных экстремумов. Здесь под локальным экстремумом понимается экстремум по симметричному «окну» некоторой длины. В данном случае использовалось «окно» длины 21 (10 точек слева и 10 – справа). Затем отрезки также склеивались на основе коэффициента линейной регрессии. На графиках ниже показан результат работы метода на несглаженных данных.



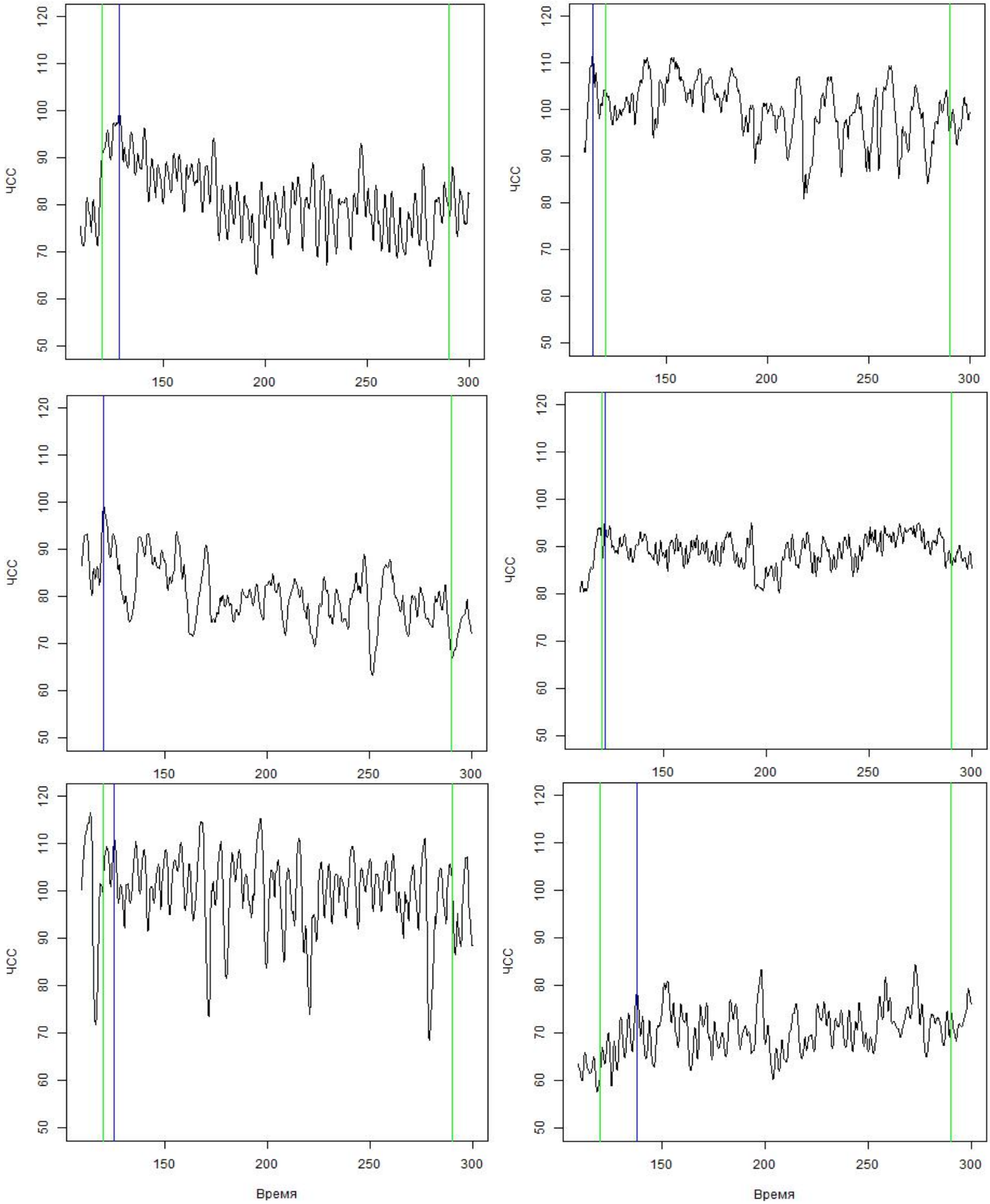
### Линейная регрессия на всех префиксах

В эксперименте был посчитан коэффициент линейной регрессии на всех префиксах несглаженных данных. Предполагаемый разрез мобилизация-стабилизация проведён на правой границе префикса, коэффициент линейной регрессии на котором больше 0.6 и значение ЧСС на правом конце которого максимально.



### Линейная регрессия на префиксах, заканчивающихся в локальных экстремумах

Также можно считать коэффициент линейной регрессии на префиксах, заканчивающихся в точках экстремума (экстремум такой же, как описан выше). Предполагаемый разрез мобилизация-стабилизация проведён на правой границе префикса, коэффициент линейной регрессии на котором больше 0.6 и значение ЧСС на правом конце которого максимально. На картинках ниже показан результат работы метода на несглаженных данных.

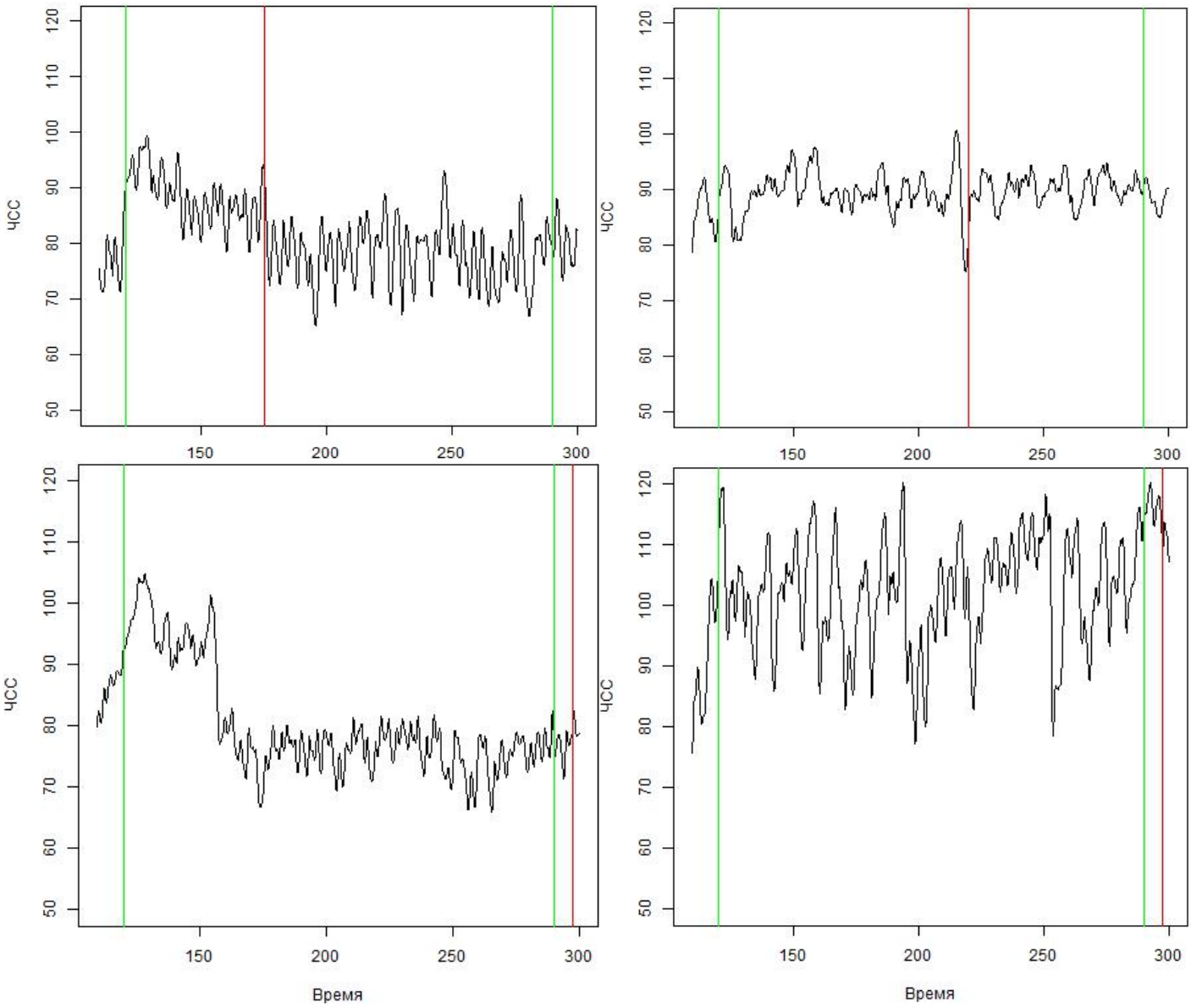


## Выделение границы фаз стабилизации и работы

В этом разделе описаны эксперименты по поиску только границы фаз стабилизации и работы. Границы задания на графиках обозначены зелёными прямыми, предполагаемый разрез – красным.

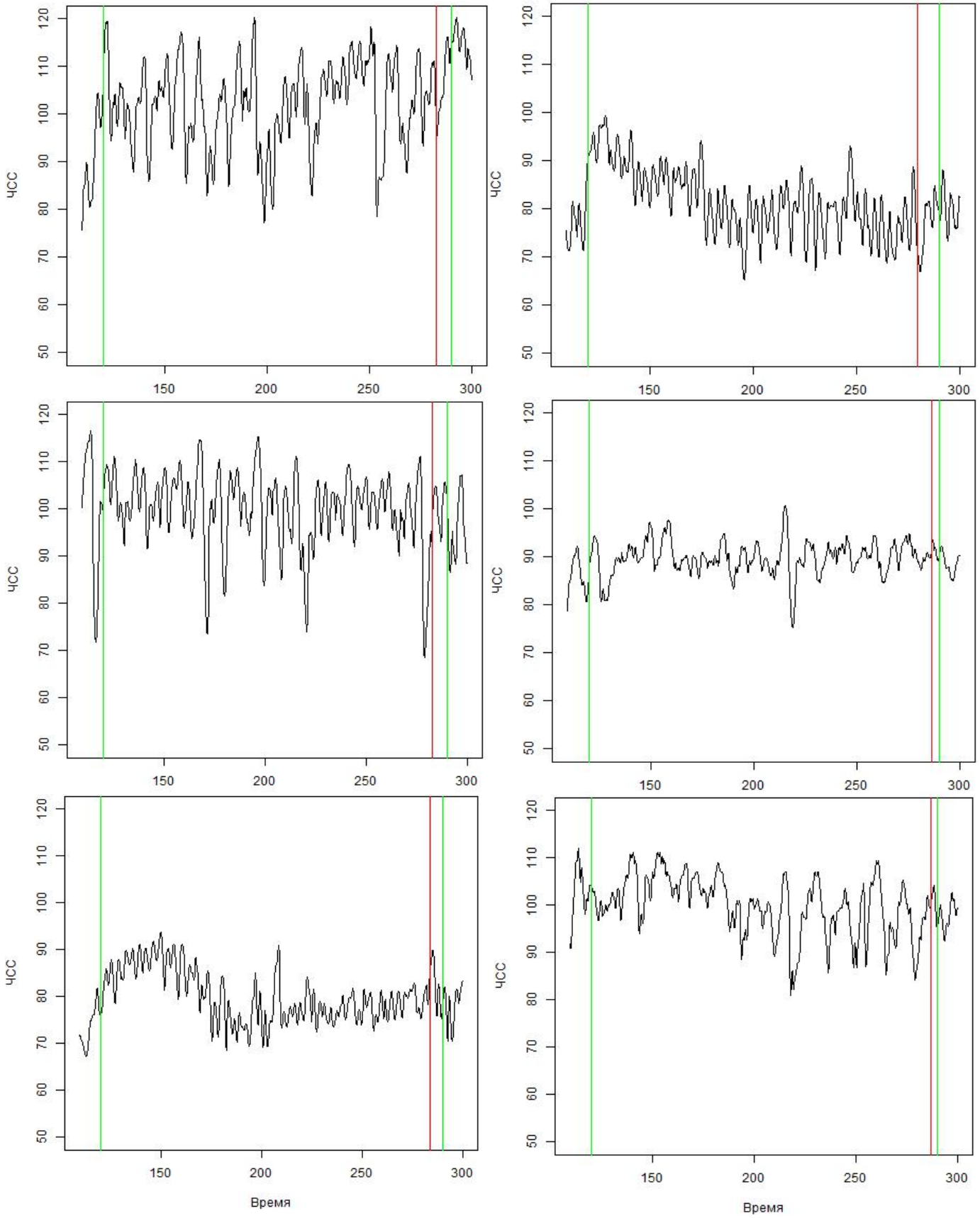
### Склеивание интервалов, выделенных с помощью алгоритма changepoints

Сначала алгоритм changepoints, применённый к частичным разностям, вычисленным для сглаженных по окну длины 10 данных, разбивает исходные данные на несколько отрезков (в данном случае, не более чем на 10). Затем на полученных отрезках высчитывается коэффициент линейной регрессии для несглаженных данных. Все правые отрезки с коэффициентом линейной регрессии, не превышающим по модулю 0.1, склеиваются, и строится предполагаемая граница фаз стабилизации и работы.



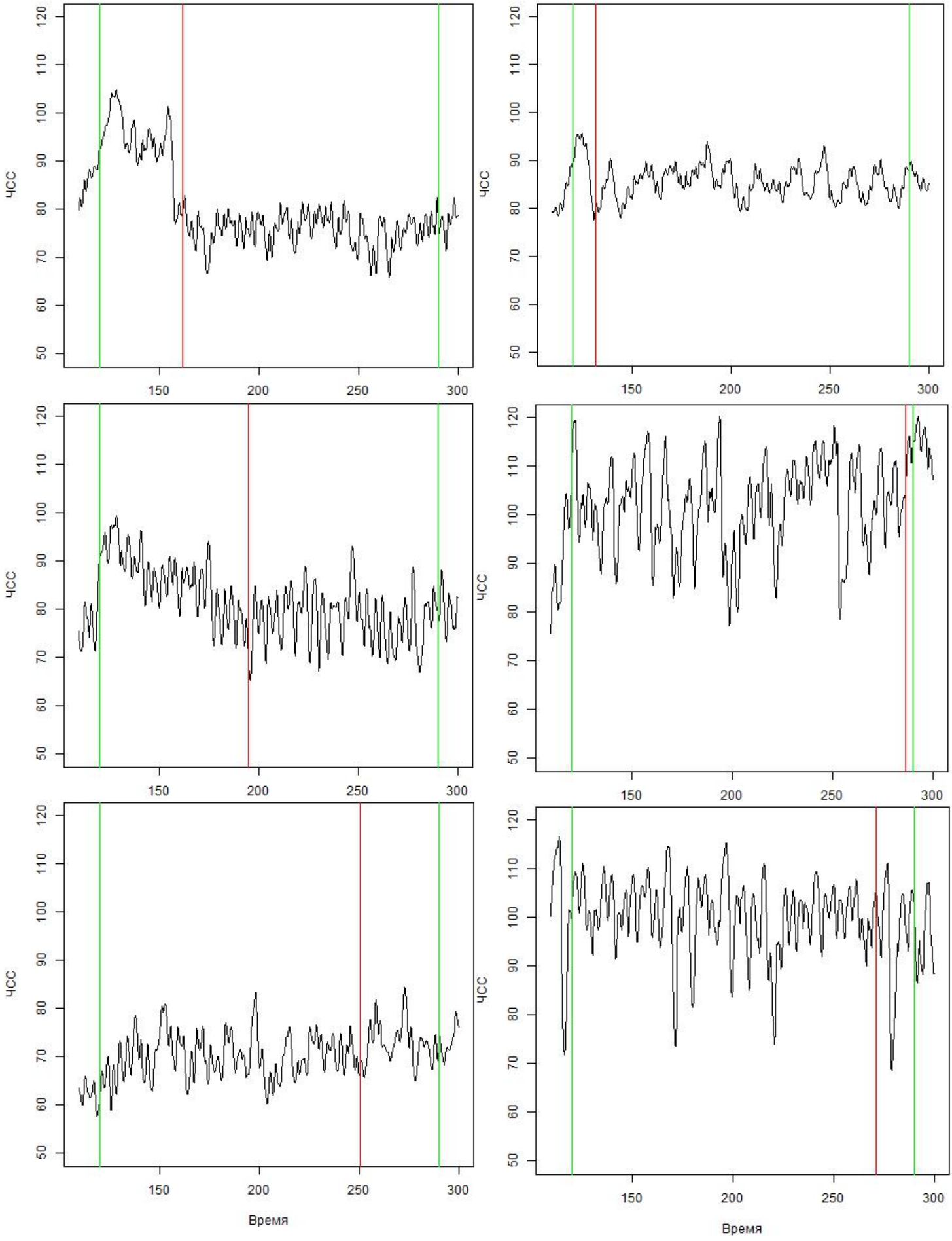
### Склеивание интервалов с концами в точках локальных экстремумов

Этот метод отличается от предыдущего тем, что разбиение на отрезки происходит в точках локальных экстремумов. Здесь под локальным экстремумом понимается экстремум по симметричному «окну» некоторой длины. В данных экспериментах использовалось «окно» длины 21 (10 точек слева и 10 – справа). Далее отрезки также склеивались на основе коэффициента линейной регрессии, как и в предыдущем эксперименте.



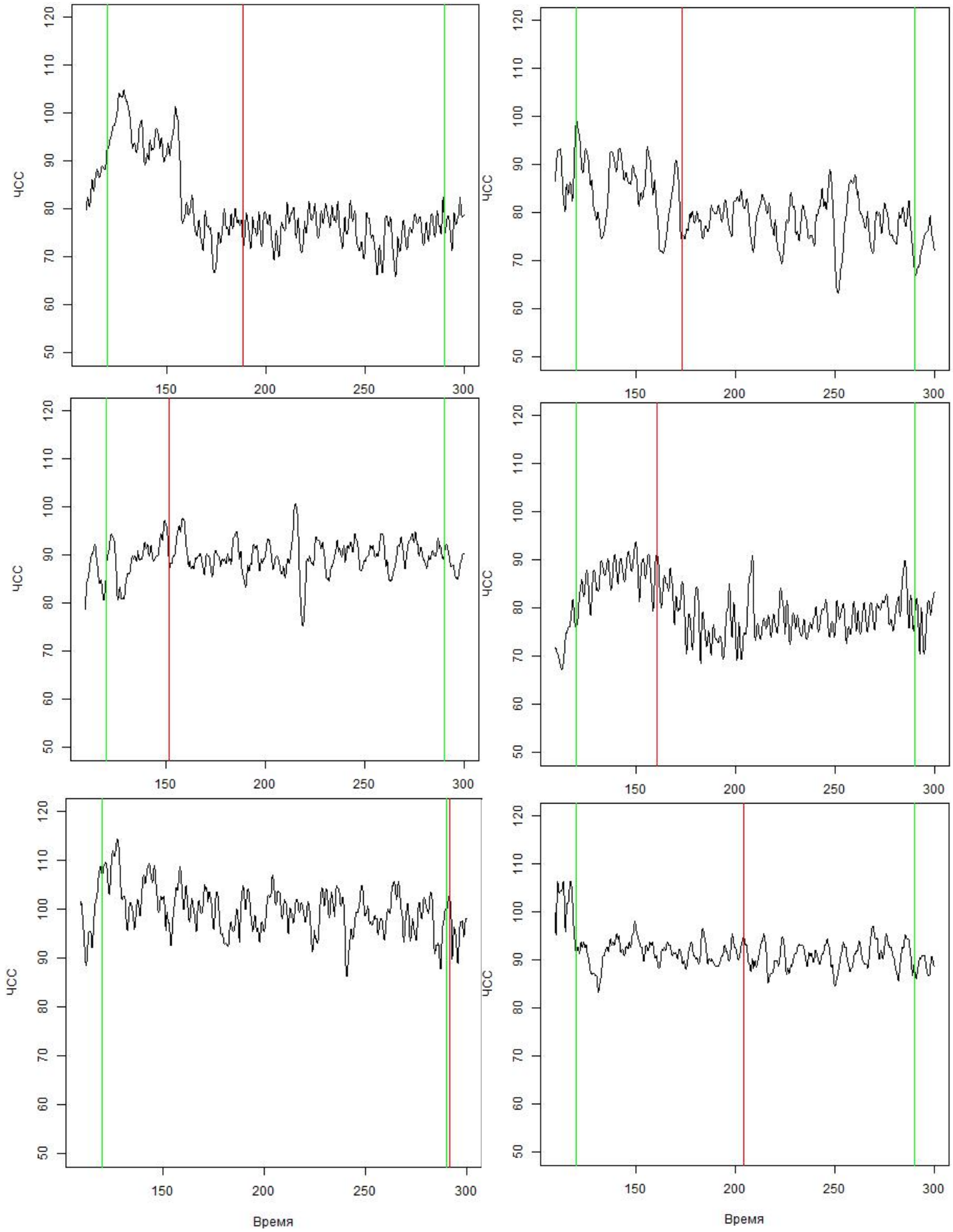
### Линейная регрессия на всех суффиксах

В эксперименте был посчитан коэффициент линейной регрессии на всех суффиксах несглаженных данных. Предполагаемый разрез стабилизация-работа проведён на левой границе суффикса, коэффициент линейной регрессии на котором по модулю минимален.



### Линейная регрессия на суффиксах, заканчивающихся в локальных экстремумах

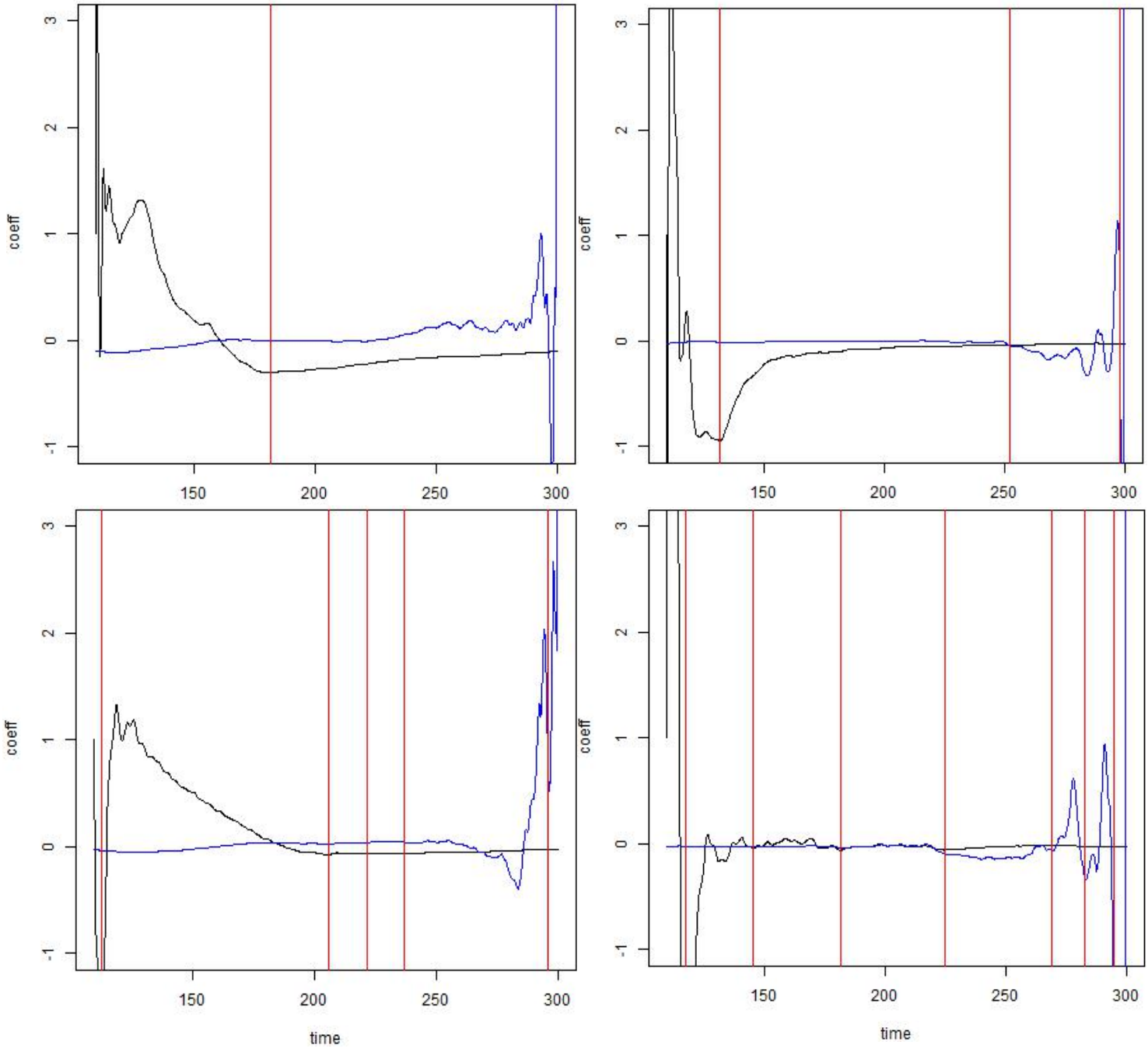
Также можно рассматривать не все суффиксы, а только те, которые заканчиваются в точках локального экстремума (здесь экстремум такой же, как описан выше). Предполагаемый разрез стабилизация-работа проводится на левой границе суффикса, коэффициент линейной регрессии на котором по модулю минимален.



### Анализ графиков коэффициентов линейной регрессии

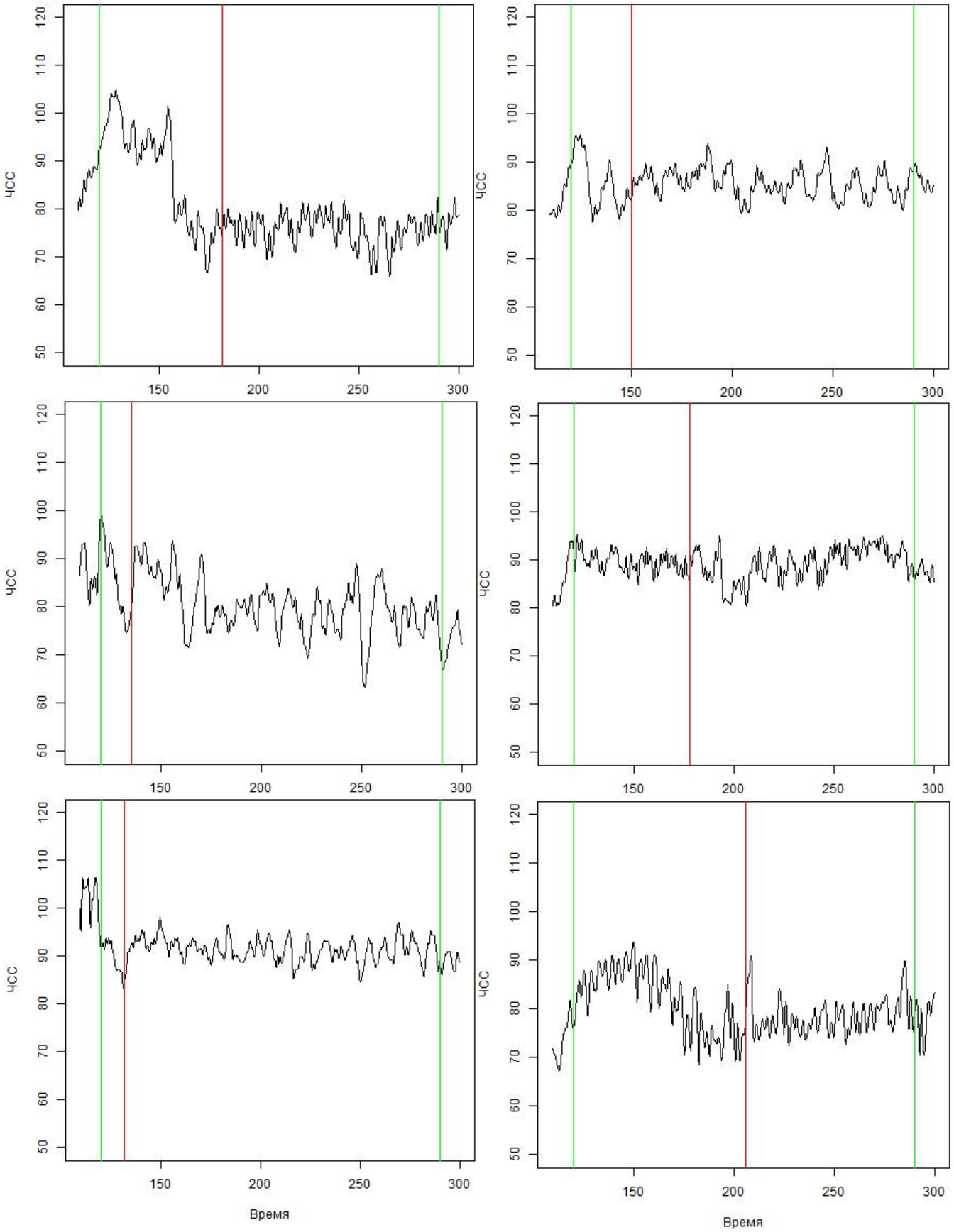
В следующем эксперименте были построены графики коэффициента линейной регрессии для префиксов и суффиксов (чёрный – для префиксов, синий – для суффиксов). Важно отметить, что почти всегда график линейной регрессии для префиксов содержит локальный минимум (иногда несколько), находящийся в левой половине графика (обозначен красной прямой), что наводит на мысли о том, что в этом месте заканчивается убывание в исходных данных, а значит, вероятно, и стадия стабилизации.

Кроме того, в левой половине график коэффициента линейной регрессии практически всегда совпадает с нулём, что даёт основание исключить из рассмотрения коэффициенты линейной регрессии на суффиксах, поскольку оба разреза (границы фаз мобилизации, стабилизации и работы) должны располагаться в левой половине графика (в начале этапа выполнения задания). По всей видимости, это предположение корректно, поскольку фаза стабилизации, как и фаза мобилизации, редко длится больше 30 секунд.



### Локальный минимум коэффициента линейной регрессии, посчитанного на всех префиксах

Исходя из приведённых выше рассуждений, мы попробовали построить границу фаз стабилизации и работы в точке минимума графика коэффициента линейной регрессии на префиксе. В случае, когда локальных минимумов было обнаружено несколько, выбирался самый левый из разрезов, попадающих в интервал выполнения задания (поскольку минимум, найденный до начала работы, скорее всего, предшествует росту ЧСС на фазе мобилизации).



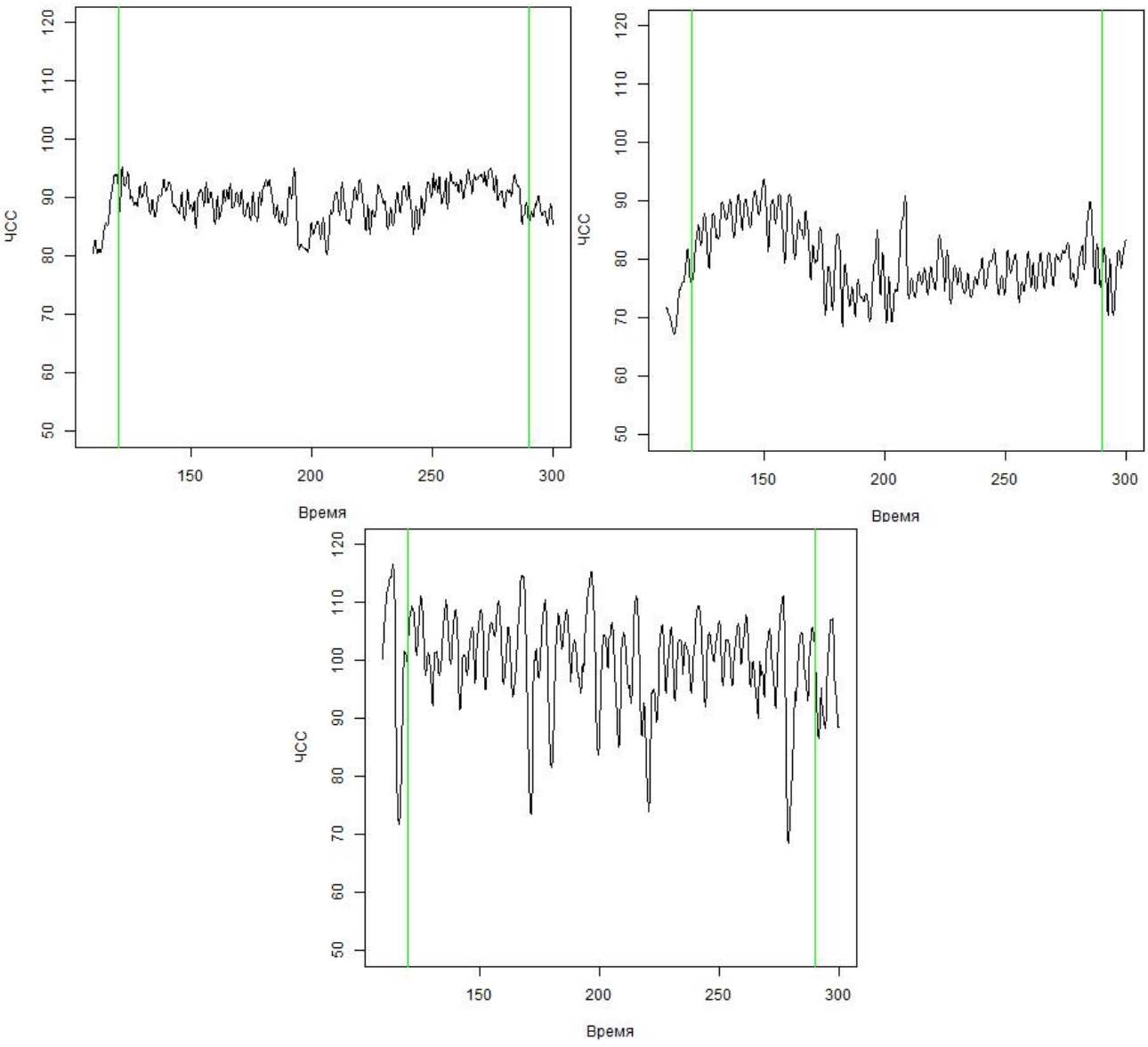
# Результаты

## Выделение границы фаз мобилизации и стабилизации

В приведённой таблице указаны результаты, работы описанных выше подходов на данных ЧСС экспертов (они занумерованы с 1 по 15 в соответствии с исходными данными). Синим обозначены графики, выделить на которых границу мобилизации и стабилизации было непросто (непонятно, где должен проходить разрез). Плюс стоит, если результат метода совпадает с моим видением разреза, минус – нет.

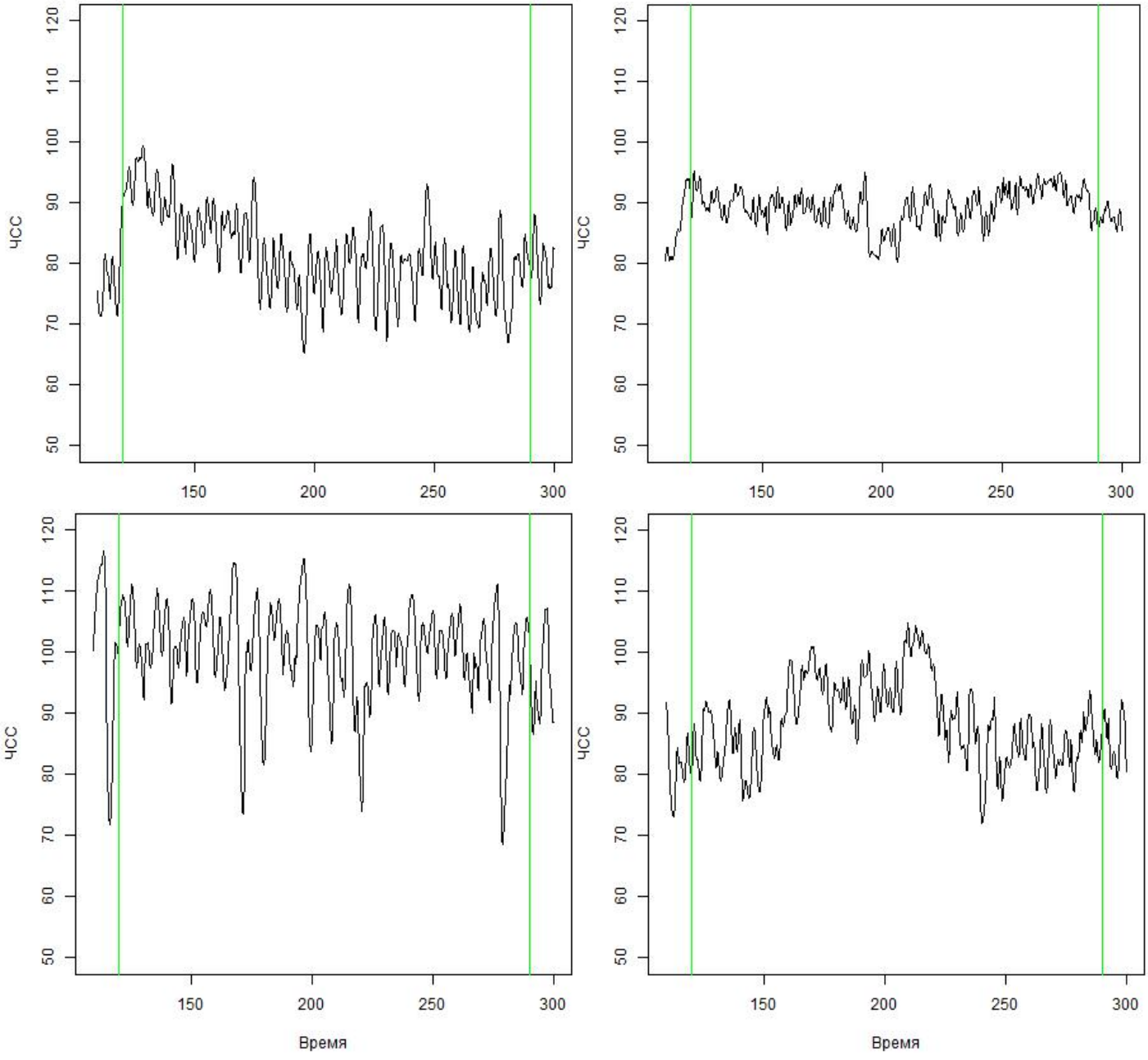
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Интервалы + changepoints** | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| **Интервалы + экстремумы** | + | + | - | - | - | - | - | - | - | + | - | - | - | - | - |
| **Префиксы** | + | + | + | + | + | + | + | + | + | + | + | - | - | + | - |
| **Префиксы + экстремумы** | + | + | + | + | + | + | + | + | + | + | + | - | - | + | - |

Графики, на которых возникли затруднения с ручным выделением границы фаз стабилизации и мобилизации представлены ниже.



## Выделение границы фаз стабилизации и работы

В приведённой таблице указаны результаты, работы описанных выше подходов на данных ЧСС экспертов (они занумерованы с 1 по 15 в соответствии с исходными данными). Красным обозначены графики, выделить на которых границу стабилизации и работы было трудно (не понятно, где подходит разрез). Плюс стоит, если результат метода совпадает с моим видением разреза, минус – если нет, вопрос ставится, когда совсем не понятно, где должна заканчиваться фаза стабилизации.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Интервалы + changepoints** | - | - | + | - | - | - | - | - | - | - | - | - | - | - | - |
| **Интервалы + экстремумы** | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| **Суффиксы** | + | + | + | - | - | - | - | - | - | - | - | - | - | - | - |
| **Суффиксы + экстремумы** | - | - | + | - | - | - | - | - | - | - | - | - | - | - | - |
| **Минимум на префиксах** | + | - | + | - | + | + | + | - | + | ? | + | + | + | ? | + |

Графики, на которых возникли затруднения с ручным выделением границы фаз стабилизации и мобилизации представлены ниже.