

# РАЗРАБОТКА АНСАМБЛЕВЫХ КЛИМАТИЧЕСКИХ ПРОГНОЗОВ В ПРОГРАММНОЙ СРЕДЕ С ОТКРЫТЫМ КОДОМ

---

*Федотова Екатерина Валерьевна  
Клименко Владимир Викторович*

*Лаборатория глобальных проблем энергетики НИУ "МЭИ", Москва*

*Институт физики атмосферы РАН, Москва*

# FREE & OPEN SOURCE SOFTWARE

Программное обеспечение с открытым кодом стало ведущим трендом последнего десятилетия

## Microsoft takes over GitHub as it expands its open source plan

is a major contributor to open source.

## It's an open-source world: 79 percent of companies run open-source software

Black Duck Software and North source software in businesses e it worth a darn.



By Steven J. Vaughan-Nichols for Mobile  
Topic: CXO

JUL 11, 2017 @ 07:00 AM

## Open Source Is The New Normal In Data and Analytics

Forbes



**Forbes Technology Council**

Elite CIOs, CTOs & execs offer firsthand insights on tech & business. [FULL BIO](#) ▾

Opinions expressed by Forbes Contributors are their own.



POST WRITTEN BY

**Scott Gnau**

CTO at [Hortonworks](#), enabling the next generation data architecture and driving technology vision in the enterprise.



**Scott Gnau**, Forbes Councils

You hear a lot these days about how the growing deluge of digital data is changing the nature of how nearly every business operates fundamentally.



SHARE >

# FREE & OPEN SOURCE SOFTWARE

---

Для исследователей использование открытого программного обеспечения имеет целый ряд преимуществ:

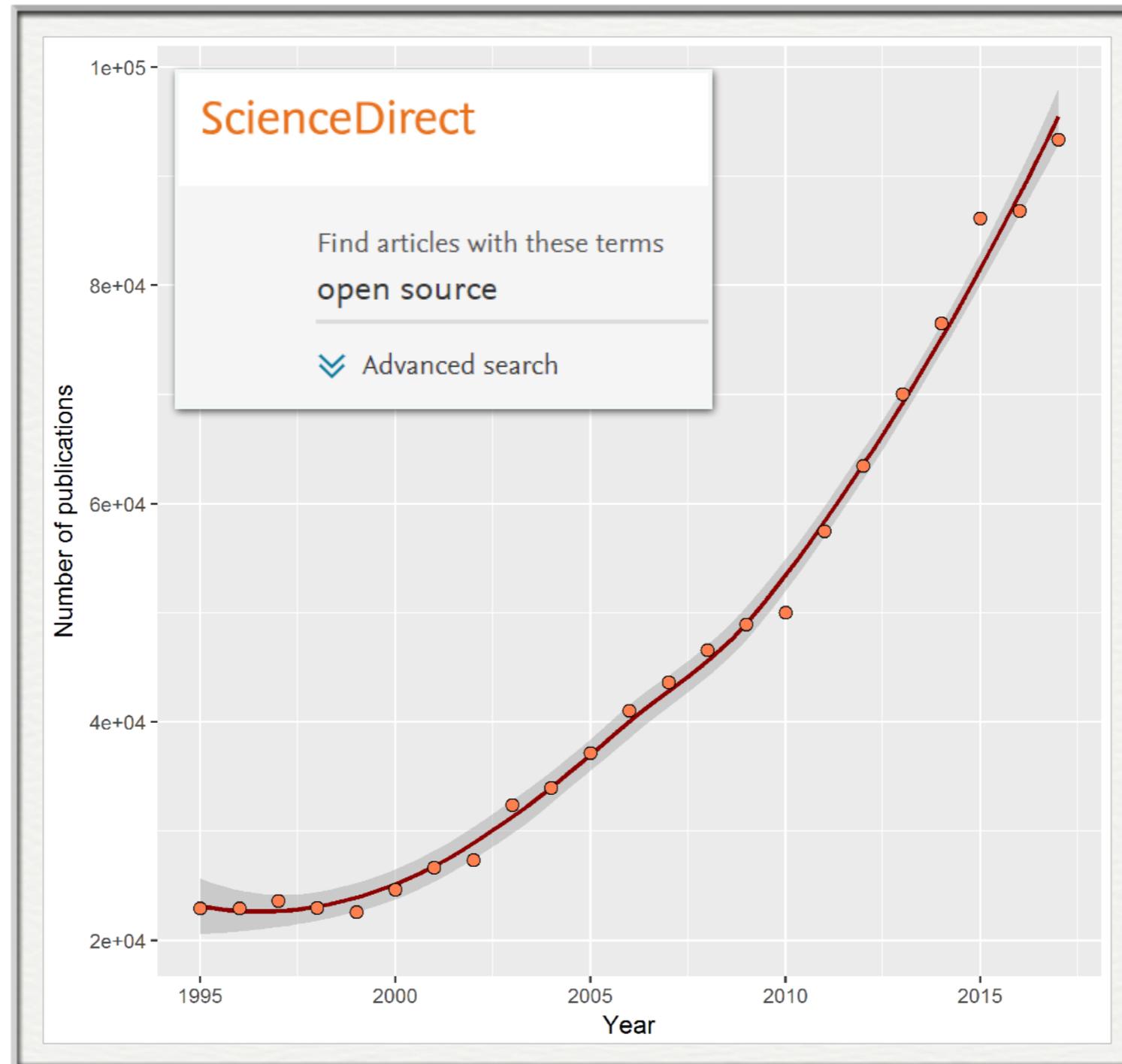
- обеспечение воспроизводимости исследований
- повышение видимости работы
- доступ к самым современным технологиям при минимальных финансовых затратах

**NB Free sources are not like free beer (c)**

Принципиальное отличие парадигмы открытого кода — вовлечение пользователя в процесс разработки

# FREE & OPEN SOURCE SOFTWARE

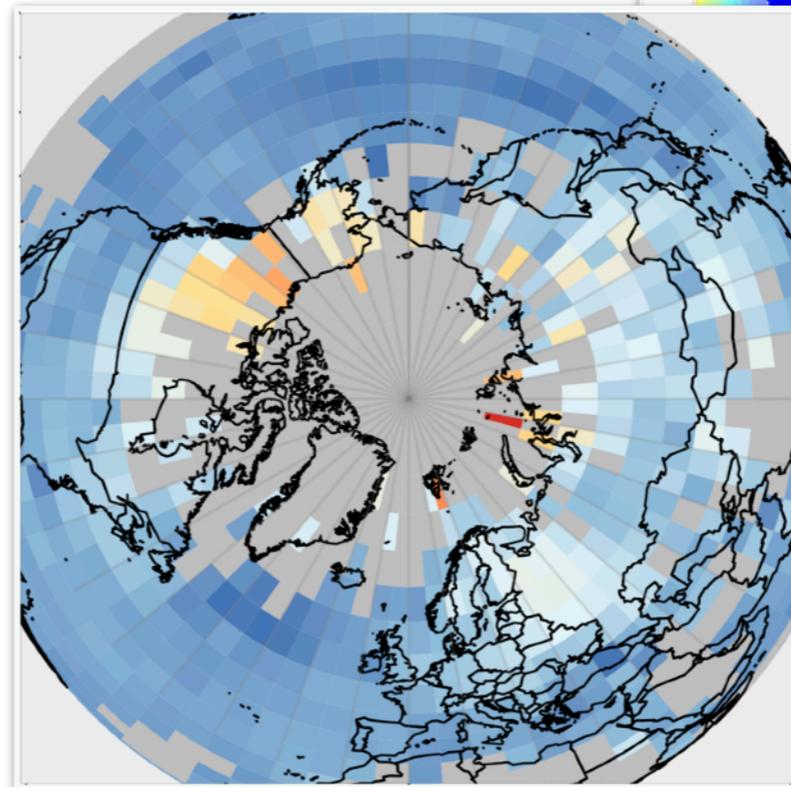
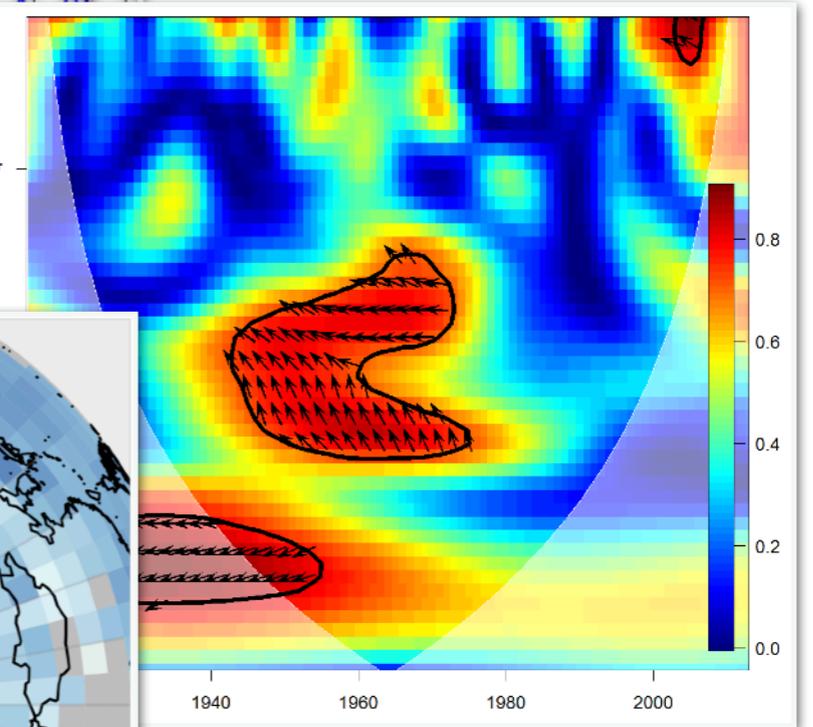
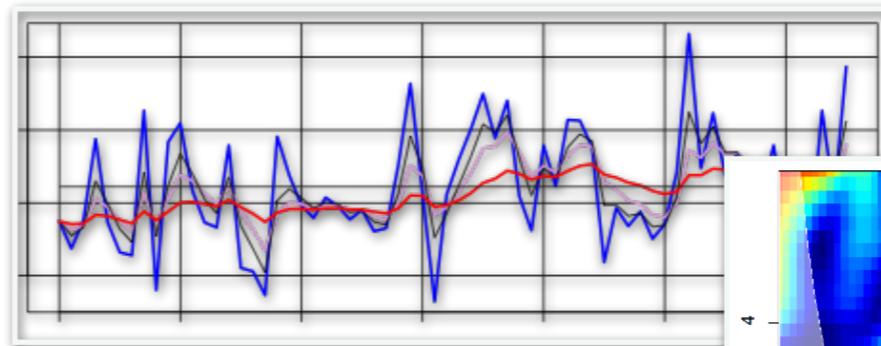
Программное обеспечение с открытым кодом стало ведущим трендом последнего десятилетия



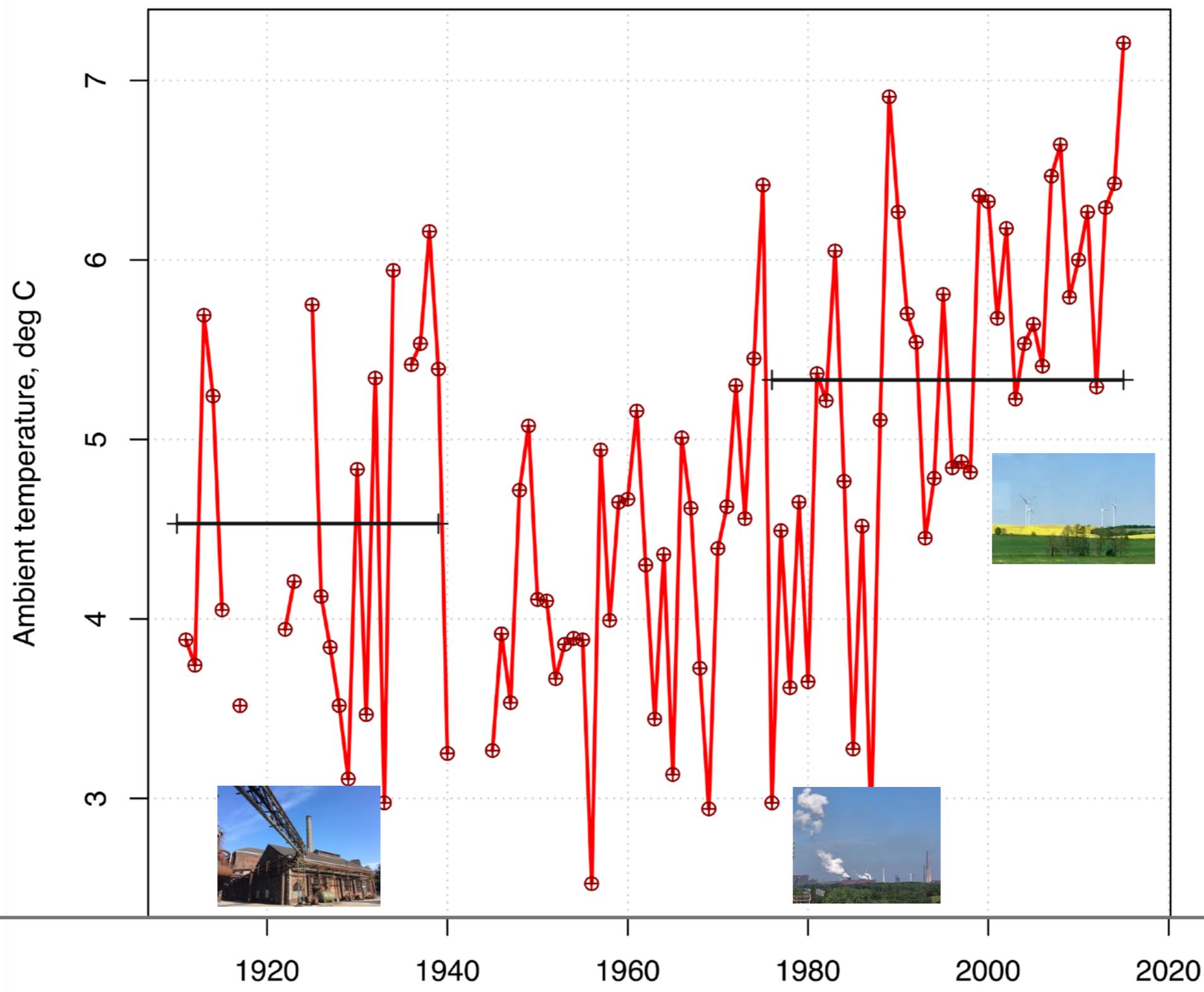
# FREE & OPEN SOURCE SOFTWARE

---

R — лучший на сегодня язык для статистических расчетов



Задача: оценка долговременных тенденций климата



# БАЗА ДАННЫХ

## Проект CMIP5: результаты моделирования

- формат netCDF
- масштаб — глобальный
- временное разрешение?

Интерес представляют временные масштабы порядка десятилетий

**Project** 1 ▲

type to filter... Sort by ▼

Show all

IPCC-AR5\_CMIP5 (IPCC As... 12155

**Topic Name** 1 ▲

type to filter... Sort by ▼

Show all

air\_temperature 12155

**Keywords** 133 ▼

**Aggregation** 4 ▲

type to filter... Sort by ▼

Show all

<input type="checkbox"/> mon	6358
<input type="checkbox"/> day	4468
<input type="checkbox"/> 6hr	1044
<input type="checkbox"/> 3hr	285

# ЛОГИКА R

---

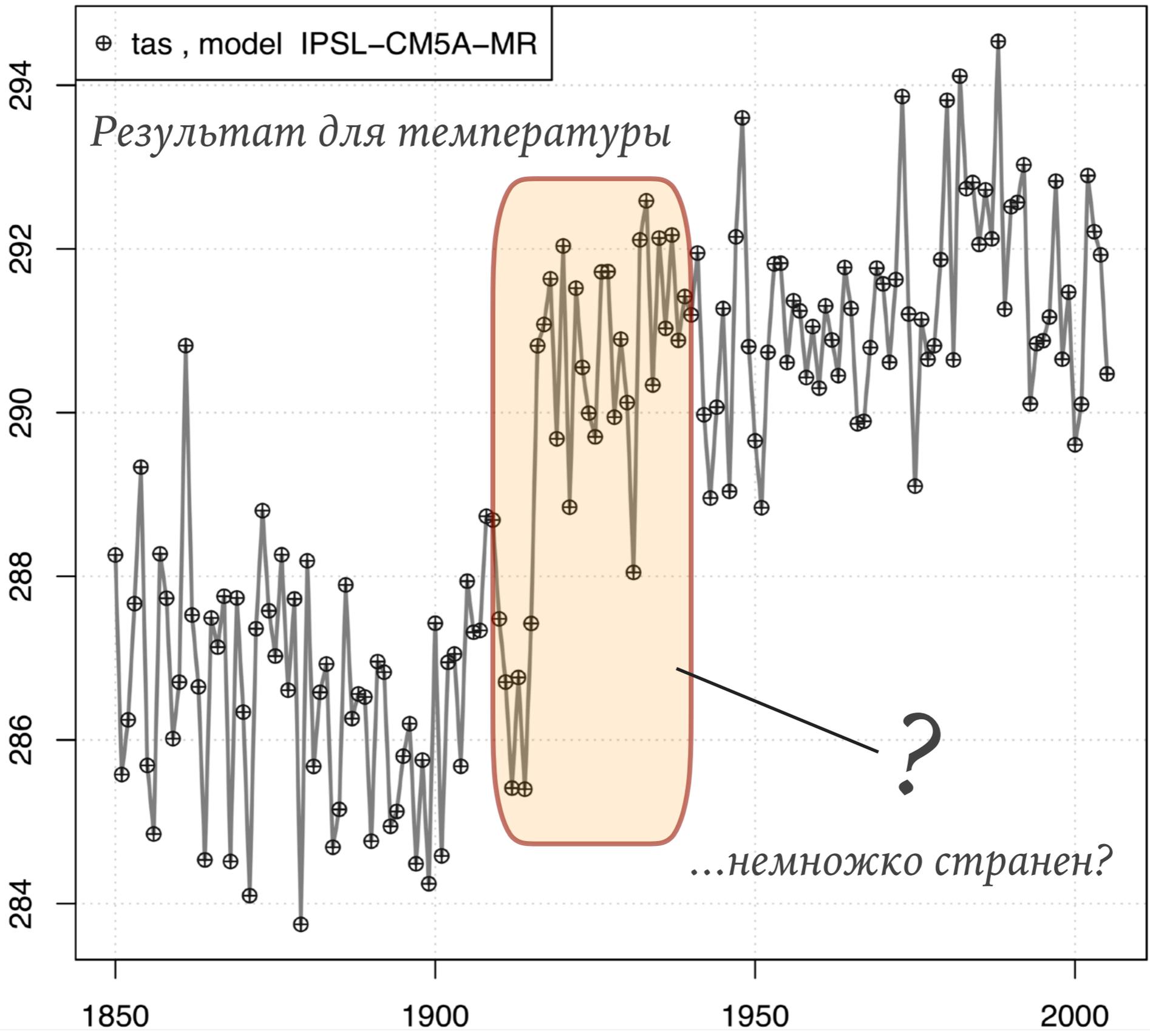
## *Classic R, ::raster, ::RCMIP5*

- 1) *определить интересующие индексы (например, все записи для весны)*
- 2) *по значениям индексов выбрать элементы массива*
- 3) *применить все необходимые функции к выбранному под-массиву*

```
# 1) set a processed month
month_nL <- 3L

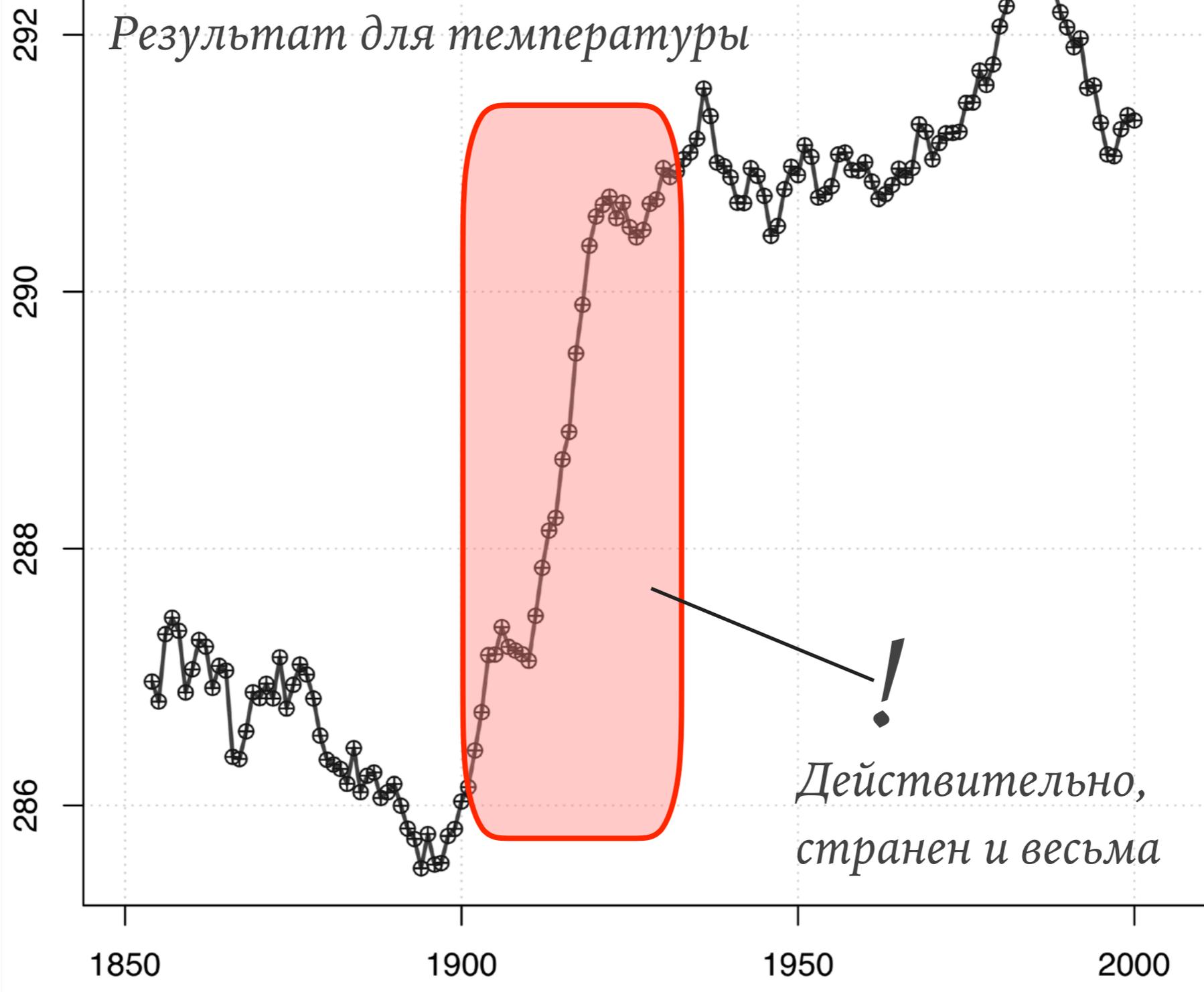
# 2) subset the data
months_nc <- as.numeric(format(time_nc_date, "%m"))
nc_dates_slctd <- time_nc_date[which(months_nc %in% month_nL)]
T_3D_array_slctd <- T_3D_array_to_process[ , , which(months_nc %in% month_nL)]

# 3) make calculations
T_time_series_slctd <- sapply(function(i) mean(T_3D_array_slctd[, , i]),
  X = seq(along.with = T_3D_array_slctd[1, 1, ]))
```



⊕ tas , model IPSL-CM5A-MR

*Результат для температуры*

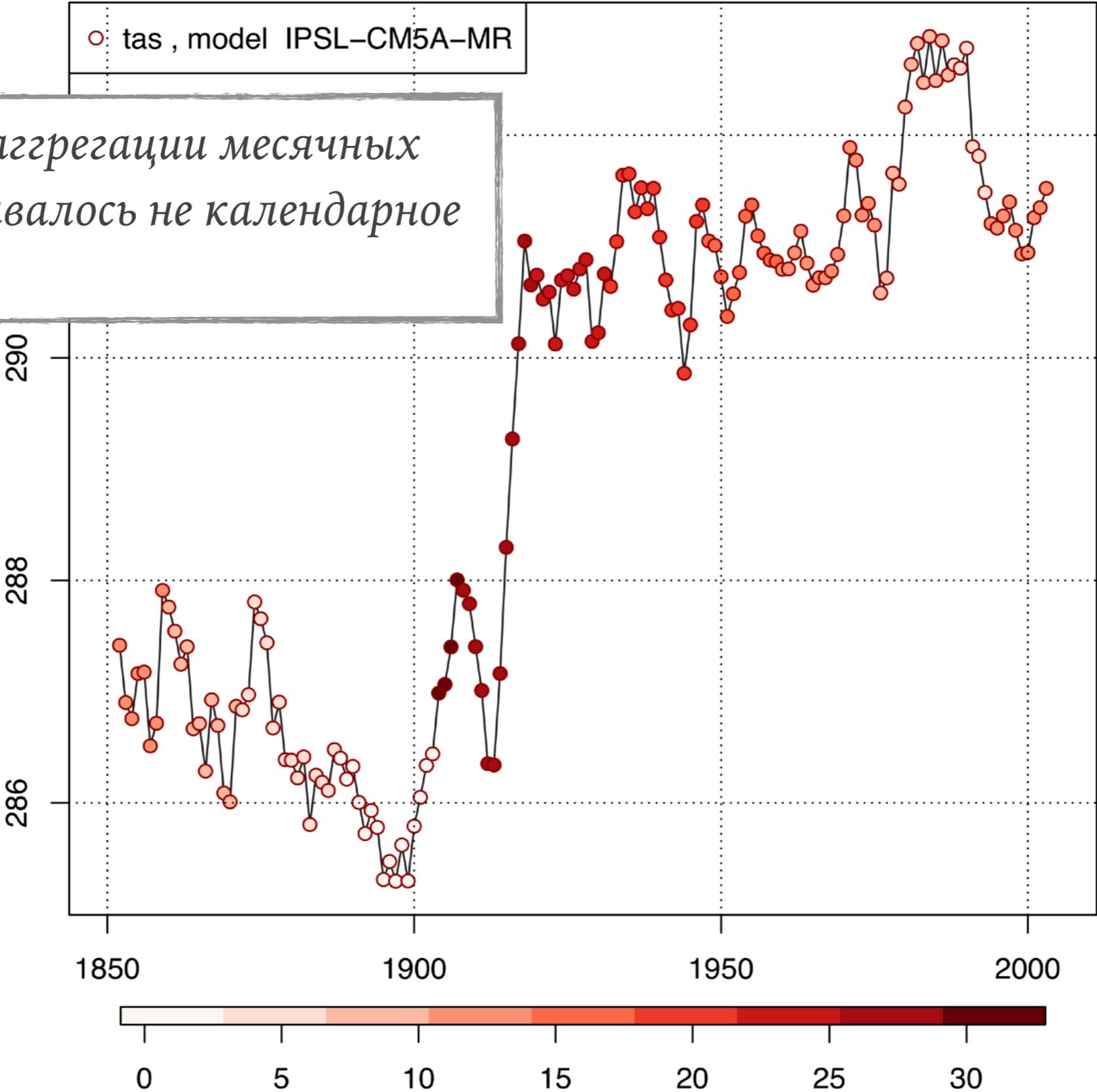


!  
*Действительно,  
странен и весьма*

# Points colors according to the day of the month

○ tas , model IPSL-CM5A-MR

*Проблема: при агрегации месячных значений учитывалось не календарное время*



# РЕШЕНИЕ

## Применить интерполяцию

### ВМЕСТО

04"	"1901-03-04"	"1902-03-04"	"1903-03-04"	"1904-03-03"	"1905-03-03"	"1906-03-03"	"1907-03-03"	"1908-03-02"	"1909-03-02"
02"	"1911-03-02"	"1912-03-01"	"1913-03-01"	"1914-03-01"	"1915-03-01"	"1916-03-31"	"1917-03-31"	"1918-03-31"	"1919-03-31"
30"	"1921-03-30"	"1922-03-30"	"1923-03-30"	"1924-03-29"	"1925-03-29"	"1926-03-29"	"1927-03-29"	"1928-03-28"	"1929-03-28"

### ИСПОЛЬЗОВАТЬ

[3,]	"1892-04-03"	"1893-04-03"	"1894-04-03"	"1895-04-03"	"1896-04-04"	"1897-04-04"	"1898-04-04"
	[,50]	[,51]	[,52]	[,53]	[,54]	[,55]	[,56]
[1,]	"1899-02-03"	"1900-02-03"	"1901-02-03"	"1902-02-03"	"1903-02-03"	"1904-02-03"	"1905-02-02"
[2,]	"1899-03-04"	"1900-03-04"	"1901-03-04"	"1902-03-04"	"1903-03-04"	"1904-03-03"	"1905-03-03"
[3,]	"1899-04-04"	"1900-04-04"	"1901-04-04"	"1902-04-04"	"1903-04-04"	"1904-04-03"	"1905-04-03"
	[,57]	[,58]	[,59]	[,60]	[,61]	[,62]	[,63]
[1,]	"1906-02-02"	"1907-02-02"	"1908-02-02"	"1909-02-01"	"1910-02-01"	"1911-02-01"	"1912-02-01"
[2,]	"1906-03-03"	"1907-03-03"	"1908-03-02"	"1909-03-02"	"1910-03-02"	"1911-03-02"	"1912-03-01"
[3,]	"1906-04-03"	"1907-04-03"	"1908-04-02"	"1909-04-02"	"1910-04-02"	"1911-04-02"	"1912-04-01"
	[,64]	[,65]	[,66]	[,67]	[,68]	[,69]	[,70]
[1,]	"1913-01-31"	"1914-01-31"	"1915-01-31"	"1916-02-29"	"1917-02-28"	"1918-02-28"	"1919-02-28"
[2,]	"1913-03-01"	"1914-03-01"	"1915-03-01"	"1916-03-31"	"1917-03-31"	"1918-03-31"	"1919-03-31"
[3,]	"1913-04-01"	"1914-04-01"	"1915-04-01"	"1916-04-30"	"1917-04-30"	"1918-04-30"	"1919-04-30"
	[,71]	[,72]	[,73]	[,74]	[,75]	[,76]	[,77]
[1,]	"1920-02-28"	"1921-02-27"	"1922-02-27"	"1923-02-27"	"1924-02-27"	"1925-02-26"	"1926-02-26"
[2,]	"1920-03-30"	"1921-03-30"	"1922-03-30"	"1923-03-30"	"1924-03-29"	"1925-03-29"	"1926-03-29"
[3,]	"1920-04-29"	"1921-04-29"	"1922-04-29"	"1923-04-29"	"1924-04-28"	"1925-04-28"	"1926-04-28"
	[,78]	[,79]	[,80]	[,81]	[,82]	[,83]	[,84]
[1,]	"1927-02-26"	"1928-02-26"	"1929-02-25"	"1930-02-25"	"1931-02-25"	"1932-02-25"	"1933-02-24"

# РЕШЕНИЕ

---

```
# 1) construct an interpolating function
# @ind_to_interp is a three-column matrix of indices to interpolate from
# @regular_date is a list data sequences to interpolate to
# @param_3D is a 3D array to process
▼ Month_Approx <- function(i, indices_to_int, regular_date, param_3D) {
  tau_intp_from <- time_nc_date[indices_to_int[i, ]]
  dates_seq_into <- regular_date[[i]]
  T_3D_array_intp_from <- param_3D[ , , indices_to_int[i, ]]
  T_area_avr_intp_from <- lapply(function(j) mean(param_3D[ , , indices_to_int[i, j]]),
    X = seq(along.with = indices_to_int[i, ]))
  appr_res <- approx(x = tau_intp_from, y = T_area_avr_intp_from, xout = regular_date[[i]])
  T_res <- mean(appr_res$y)
  return(T_res)
}

# 2) set a data sequence
first_month_day <- ymd(paste(as.numeric(format(nc_dates_slctd, "%Y")), month_nL, 1, sep = "-"))
last_month_day <- ymd(paste(as.numeric(format(nc_dates_slctd, "%Y")),
  (month_nL + 1), 1, sep = "-")) - 1
dates_seq_intp_to <- lapply(function(k) seq.Date(from = first_month_day[k],
  to = last_month_day[k], by = 1), X = 1:length(nc_dates_slctd))

# 3) interpolate the monthly-aggregated value for all selected years
▼ T_monthly_intpd <- sapply(function(i) Month_Approx(i, indices_to_int = ind_to_interp,
  regular_date = dates_seq_intp_to, param_3D = T_3D_array_to_process),
  X = 1:length(nc_dates_slctd))
```

GitHub, Inc

Сервисы словарь M

Branch: master New p

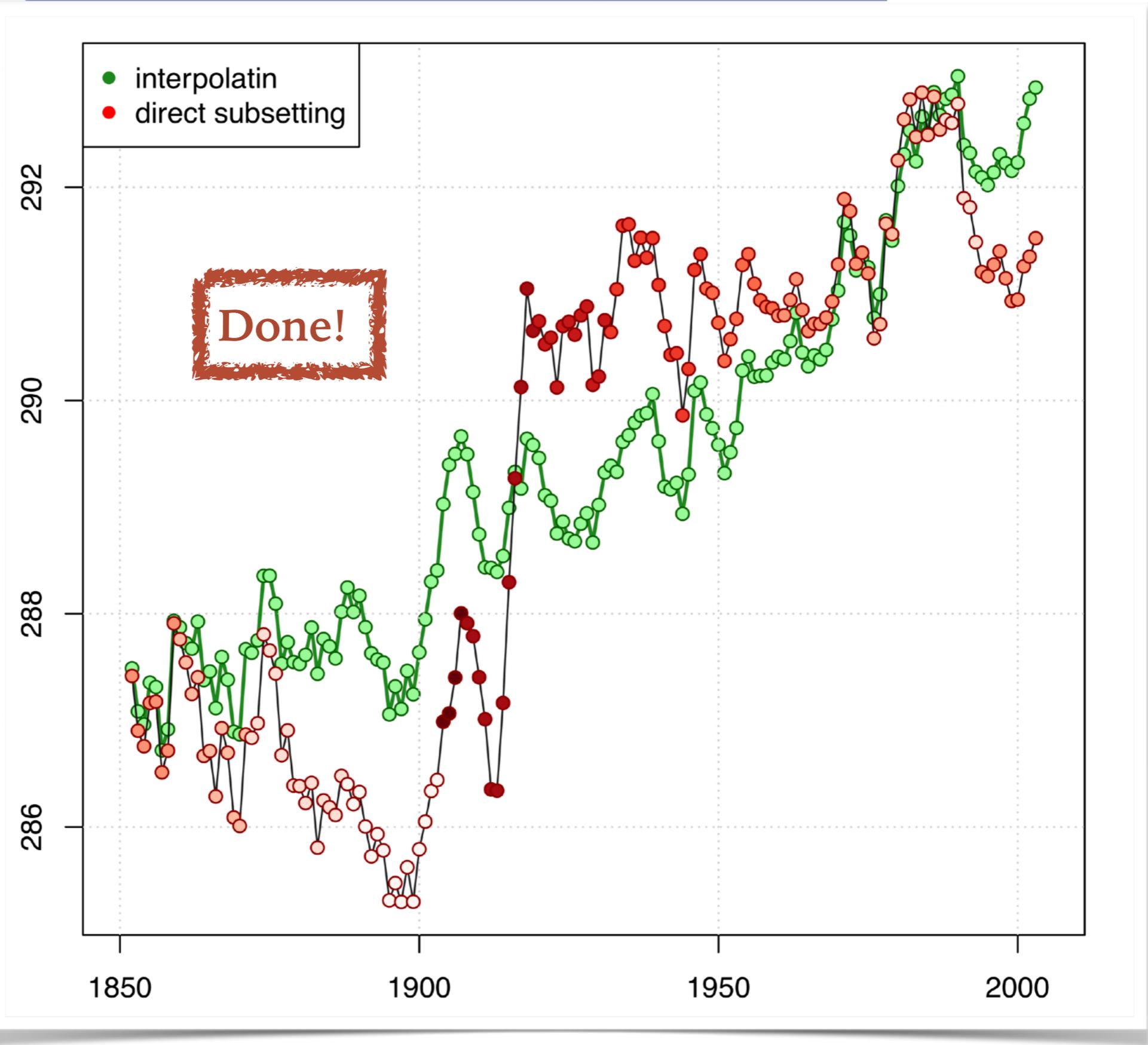
ekatef Minor changes

- 0\_ProcessCMIP5\_Confi
- 1\_ListModels.R
- 2\_SelectSeasons.R
- 3\_Plot\_Calcul\_Field.R
- LICENSE
- ProcessCMIP5\_Script\_v
- README.md

README.md

Set of R-function:  
annual means to

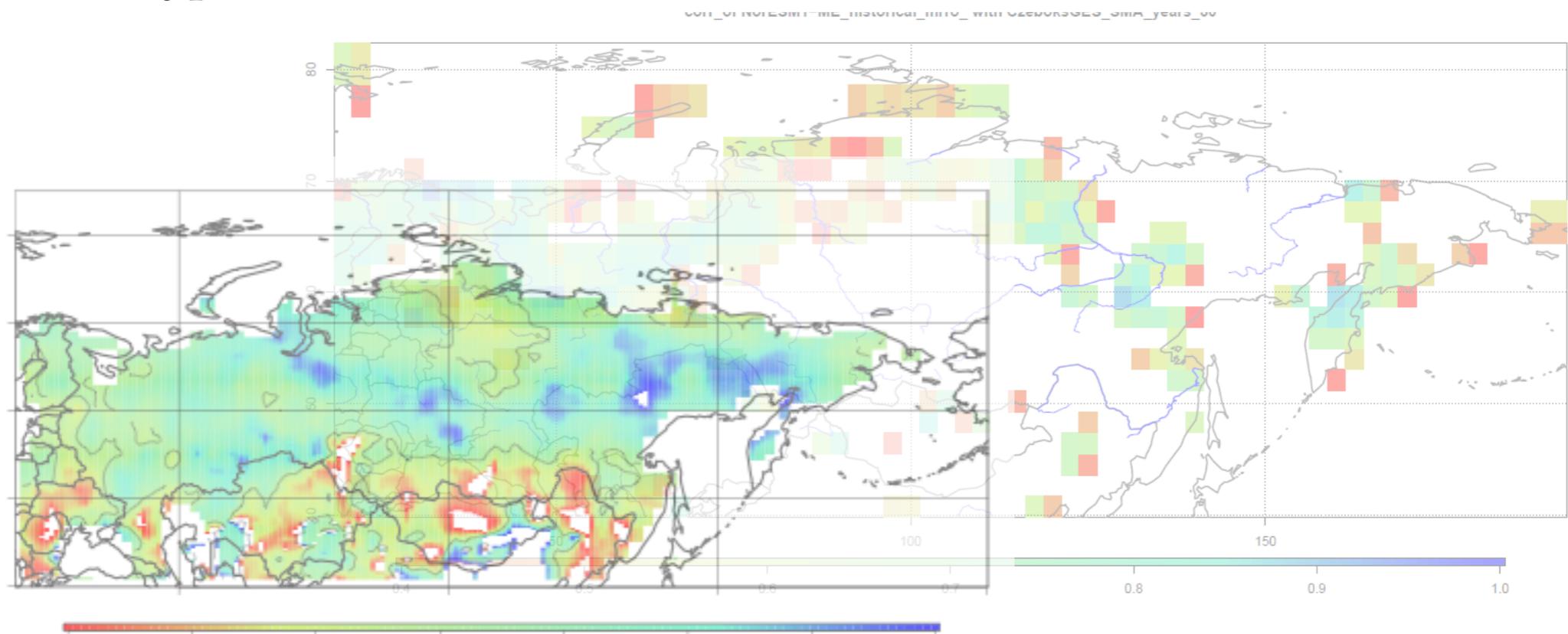
Options:



# ВОЗМОЖНОСТИ

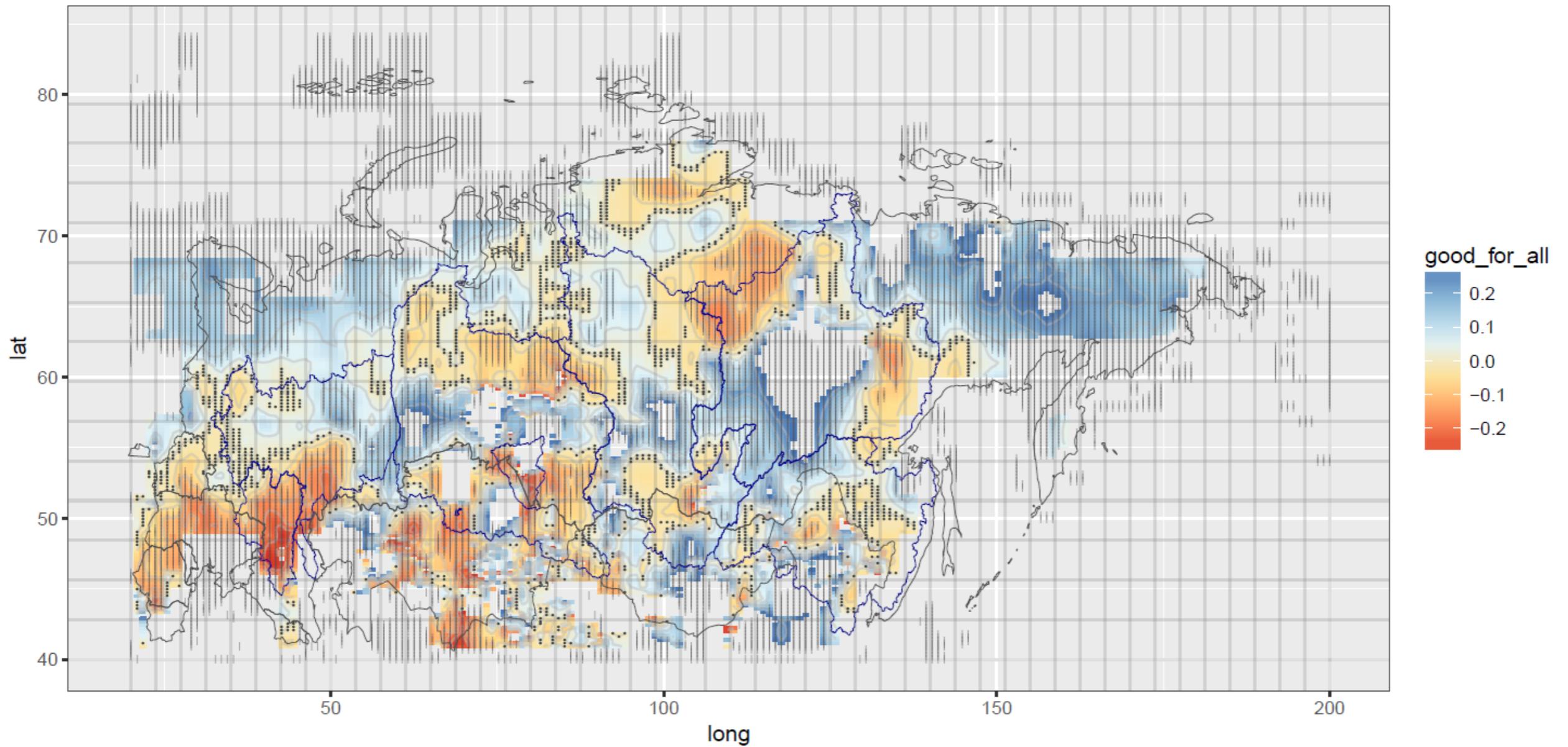
---

- 1) детальное изучение временной структуры ряда
- 2) автоматизация валидационных процедур
- 3) статистическое масштабирование (*downscaling*)
- 4) доступ к современным методам анализа данных (*Rssa::*, *biwavelet::* etc )
- 5) визуализация уровня ГИС



# ПРИМЕР РЕЗУЛЬТАТА

Integral relative change of the runoff  
rcp45: (2045-2054 minus 2007-2016)/2007-2016





*Спасибо за внимание!*

*@Ekatyf*

*nilgepl@mpei.ru*

*ek.v.fedotova@gmail.com*