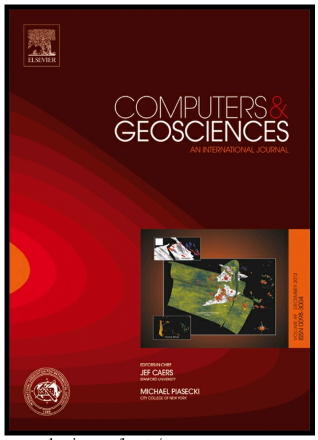
Author's Accepted Manuscript



www.elsevier.com/locate/cageo

CoinCalc – A new R package for quantifying simultaneities of event series

Jonatan F. Siegmund, Nicole Siegmund, Reik V. Donner

PII: S0098-3004(16)30548-9
DOI: http://dx.doi.org/10.1016/j.cageo.2016.10.004
Reference: CAGEO3849

To appear in: *Computers and Geosciences*

Received date: 16 December 2015
Revised date: 11 June 2016
Accepted date: 16 October 2016

Cite this article as: Jonatan F. Siegmund, Nicole Siegmund and Reik V. Donner, CoinCalc – A new R package for quantifying simultaneities of event series *Computers and Geosciences,* http://dx.doi.org/10.1016/j.cageo.2016.10.004

CoinCalc – A new R package for quantifying simultaneities of event series

Jonatan F. Siegmund[a,b], Nicole Siegmund[b,c,d], Reik V. Donner[a]

[a] *Research Domain IV – Transdisciplinary Concepts and Methods, Potsdam Institute for Climate Impact Research, Telegrafenberg A31, 14473 Potsdam, Germany*
[b] *Institute of Earth and Environmental Science, University of Potsdam, Karl-Liebknecht-Straße 24-25, 14476 Potsdam-Golm, Germany*
[c] *Leibniz Centre for Agricultural Landscape Research, Department for Soil Landscape Reseach, Eberswalder Straße 84, 15374 Müncheberg, Germany*
[d] *Institute of Meteorology and Climate Research, Atmospheric Environmental Research (IMK-IFU), Karlsruhe Institute of Technology, Kreuzeckbahnstraße 19, 82467 Garmisch-Partenkirchen, Germany*

Abstract

We present the new R package CoinCalc for performing event coincidence analysis (ECA), a novel statistical method to quantify the simultaneity of events contained in two series of observations, either as simultaneous or lagged *coincidences* within a user-specific temporal tolerance window. The package also provides different analytical as well as surrogate-based significance tests (valid under different assumptions about the nature of the observed event series) as well as an intuitive visualization of the identified coincidences. We demonstrate the usage of CoinCalc based on two typical geoscientific example problems addressing the relationship between meteorological extremes and plant phenology as well as that between soil properties and land cover.

Keywords: event coincidence analysis, R, point processes, extreme events, time series analysis

1. Introduction

In many areas of geosciences, but also other scientific disciplines like neurosciences, there has been a rising interest in inferring information on dynamical interdependencies between different observational series that are not given in the form of continuous or discrete-valued time series, but as sequences of events (e.g., unmarked or marked point processes). Traditional statistical tools like classical (Pearson) correlation analysis are often not directly applicable to such series or of limited explanatory value. While in neurosciences, many methodological developments have been introduced and subsequently applied for studying the statistical interrelationships between event series (e.g., describing sequences of neuronal spiking activity [7, 17, 20]), there have been relatively few attempts to transfer corresponding approaches to geoscientific problems [3, 19, 18].

Event coincidence analysis (ECA) is a recently developed method for studying the statistical interdependency between two event series, which has been originally introduced and applied in a geoscientific context [9, 10, 23, 25, 26]. Unlike correlation analysis, this method exclusively takes the timings of certain well-defined events in two series into account and ignores potentially available other information (e.g., underlying explicit time series values) on the gradual variability of related observables. Therefore, it provides a complementary view on data that are either by definition of binary structure (event/no event) or where only certain values (e.g., extreme events) are expected to result in a specific response of interest. Examples include the timings of natural disasters like earthquakes or floods [10] or cases where strong deviations from "normal" behavior can result in qualitatively different interdependencies between the variables of interest (e.g., ecosystem responses to extreme environmental conditions like droughts, cold spells or volcanic eruptions) [24, 32].

So far, ECA has been successfully applied to studying problems in biogeoscientific [23, 25, 26], socio-ecological [10] and paleoclimatic contexts [9]. The diversity of research questions discussed in the aforementioned publications suggests a wide range of possible future applications. While Rammig et al. [23] and Siegmund et al. [25] used the approach to derive complementary information (beyond classical correlation analysis) by looking at the timing of events in the very tail of the distribution of the underlying continuous variable, the analyses of Donges et al. [9, 10] could not have been conducted using standard tools of classical statistics since they addressed series of explicit events.

This paper introduces CoinCalc, an easy-to-handle implementation of ECA in the open statistical software R. We emphasize that the Comprehensive R Archive Network (CRAN) repository already contains the package CNA for performing an entirely different type of analysis referred to as *coincidence analysis* [1], and that the same term is also used in particle physics [31] in yet another different con-

Email address: jonatan.siegmund@pik-potsdam.de (Jonatan F. Siegmund)