Kurdish Studies
September 2018
Volume: 06, No: 02, pp 286-309
ISSN: 2051-4883 (Print) | ISSN 2051-4891 (Online)
www.KurdishStudies.net

**DOI:** 10.67145/ks.v6i2.4146

# Exact pruned dynamic programming tests the documented periodisation of the Turkey–PKK conflict in geocoded event data, 1989–2016

**Luis Eduardo Muñoz Guerrero***

*Facultad de Ingenierías, Universidad Tecnológica de Pereira, Risaralda, Colombia, lemunozg@utp.edu.co, ORCID: 0000-0002-9414-6187

**Abstract**
**Periodisations of the conflict between Turkey and the Kurdistan Workers' Party rest on qualitative turning points. Periodisation is cast as penalised optimal segmentation, a shortest-path problem, solved exactly by pruned dynamic programming on 1,461 weeks of event data. The optimum recovers five of six documented turning points, but a pre-specified circular-shift test cannot distinguish this from chance (p = 0.123). Post hoc, simulations show low power, lower with the seasonal cycle than without it; modelling or removing that cycle lowers p to 0.006–0.010, and a greedy solver would have reported a valid positive; none survives correction for multiple testing.**

***Keywords: Kurdish conflict; UCDP GED; change-point detection; shortest path; seasonality***

## 1. Introduction

The armed conflict between the Turkish state and the Kurdistan Workers' Party, Partiya Karkerên Kurdistanê (PKK), is usually told as a sequence of phases. In the account of the International Crisis Group, PKK attacks inside Turkey began in 1984 and gradually escalated; most of the casualties fell in a second phase, from 1991 to 1999; the most intense fighting ended in 1999, the year of Öcalan's capture, and a PKK ceasefire then held until 2004 (International Crisis Group 2011). In March 2013 Öcalan's letter calling for a ceasefire was read out at Newroz in Diyarbakır and the PKK declared a unilateral ceasefire (International Crisis Group 2014), which collapsed in July 2015 (Göksel 2015). Such periodisations are built from political events and from the judgement of observers.

The quantitative studies of the conflict cited here used its records for other questions: how voters in two general elections of the 1990s responded to the deaths of soldiers and police (Kibris 2011), and why militants joined the insurgency, from biographical records of militants and interviews with their relatives (Tezcür 2016). Closer to the question here, an academic account reads the phases of the conflict against yearly counts of violent incidents (Ünal 2016), and Öcalan's own periodisation of the PKK's strategy has been compared with the resolutions of PKK congresses and the characteristics of its violence (Ünal 2014). None of these studies estimates change points from an event record and tests a documented periodisation against them. This article does so with a test fixed before the analysis.

Formally, a periodisation is a partition of the time axis into contiguous intervals, and the partition that minimises a penalised cost is a shortest path on a directed acyclic graph (DAG) over the time points, an instance of the shortest-route problem (Bellman 1958); dynamic programming solves it exactly with a number of operations quadratic in the length of the series (Jackson et al. 2005). Pruning the dynamic program keeps the search exact at a lower expected cost (Killick, Fearnhead, and Eckley 2012; Maidstone et al. 2017), whereas binary segmentation, a greedy alternative, only approximates the optimum and can misplace the changes (Maidstone et al. 2017).

The article makes four contributions. First, it formulates the periodisation of the event record as penalised segmentation with a minimum segment length; makes explicit the delay that this length imposes on the pruning rule of the Pruned Exact Linear Time (PELT) method, whose authors note that a larger minimum length is easily implemented (Killick, Fearnhead, and Eckley 2012), and proves that the delayed rule returns the same segmentation as the unpruned recursion while immediate pruning does not; checks the optimum with the classical integrality of the equivalent flow program and, post hoc, with a dual certificate; and, post hoc, extends the dynamic program to find exactly the cheapest segmentation that recovers any given set of turning points (Section 4). Second, it solves the problem exactly on the weekly record of the Uppsala Conflict Data Program (UCDP) for the Turkey–PKK dyad and reports the resulting segments. Third, it tests the alignment of the estimated change points with six documented turning points by a circular-shift (rotation) test fixed before the analysis; the result is null, post hoc simulations show that the test had little power, and post hoc analyses that model or remove the annual cycle lower the p-value of the alignment, although none survives the correction for multiple testing (Section 6). Fourth, it measures the cost of exact search against optimal partitioning, an integer-programming solver and the greedy heuristic, and shows, post hoc, that the choice of solver changes the outcome of the same test in 25 of the 81 specifications of the multiverse.

## 2 Background

Periodising the conflict. Crisis Group texts periodise the conflict into an initial escalation, a second phase in the 1990s with most of the casualties, a lull after 1999, the resumption of violence after 2004 and the peace process of 2013–2015 (International Crisis Group 2011; International Crisis Group 2014; Göksel 2015). An academic account divides the same years

into five phases, a confrontation that ended with the PKK's military defeat (1984–1994), a soft, stable and self-serving stalemate (1994–1999), a unilateral ceasefire with international political recognition (1999–2003), a second such stalemate (2003–2007) and a mutually perceived stalemate from 2007,
and dates further unilateral ceasefires in 1993, 2009 and 2010 (Ünal 2016). Öcalan's own retrospective classification of the PKK's strategic periods has been compared with the resolutions of PKK congresses and the characteristics of PKK violence, which point, particularly after 1994, to pragmatic shifts rather than planned periods (Ünal 2014), and academic work asks why the PKK remobilised its armed forces when opportunities for a peaceful settlement had widened (Tezcür 2010). The ceasefire of March 2013 was the ninth declared by the PKK since 1993 (International Crisis Group 2014). Comparative work sets the Kurdish movement inside a broader regional 'state of violence', comparing the 1980s with the periods that followed 2003 and 2011 (Bozarslan 2014). Monthly casualties of PKK attacks recorded in the Global Terrorism Database for 1995–2010 peak in July–September and decline sharply in winter (Ciftci and Kula 2015). Quantitative analysis has a precedent in this journal, where census and survey data were used to map language shift among Kurds in Turkey (Zeyneloglu, Sirkeci, and Civelek 2016).
Event data. UCDP data define an armed conflict by a threshold of 25 battle-related deaths a year, lower than the 1,000 deaths of the Correlates of War project (Gleditsch et al. 2002). Its Georeferenced Event Dataset (GED) disaggregates organised violence into events, each with at least one fatality, a date and a location (Sundberg and Melander 2013); version 17.1 defines an event as an incident of lethal force between organised actors, or against civilians, at a specific place and date (Croicu and Sundberg 2017). Media-based event data can carry reporting bias (Weidmann 2016), and the quality of conflict event datasets differs, which matters for subnational analyses (Eck 2012). For other conflicts, phases have been identified in event data by statistical techniques rather than by human coders (Schrodt and Gerner 1997), change points in daily civilian casualties have been found close to state-building events (Spirling 2007), and violence data have been segmented by time-constrained hierarchical clustering (Murtagh, Spagat, and Restrepo 2011).
Exact segmentation. Optimal partitioning (OP) searches the exponentially large space of partitions of n points in time O(n2) and determines the number of segments (Jackson et al. 2005). The PELT method adds a pruning step to the dynamic program and, when the number of changes grows linearly with the number of observations, runs in expected linear time (Killick, Fearnhead, and Eckley 2012); functional pruning is empirically faster still, and its efficiency is robust to the number of changes (Maidstone et al. 2017). The whole solution path over a continuous range of penalties can be computed (Haynes, Eckley, and Fearnhead 2017); the choice of penalty has been studied for Gaussian change-point models (Zhang and Siegmund 2007) and for Poisson and negative binomial counts (Cleynen and Lebarbier 2014); and implementations are available as software (Killick and Eckley 2014). Greedy binary segmentation is fast but approximate (Maidstone et al. 2017), and it can miss closely spaced changes (Fryzlewicz 2014).

## 3 Data and pipeline

Evidence base. By design, every dataset and every source used here was published on or before 31 December 2017: the event data are UCDP GED version 17.1, released in 2017, and the literature stops at that date.
The computations themselves were run in 2026 on a desktop machine with an eight-core processor and 16 GB of memory, in a single thread; the software is described without version numbers, which are recorded in the lockfile of the repository.
Events. GED 17.1 contains 135,181 events of organised violence from 1989 to 2016 (Croicu and Sundberg 2017). The conflict coded as Turkey:Kurdistan contains a single dyad, the government of Turkey against the PKK, with 4,364 events, 4,293 located in Turkey and 71 in Iraq (Table 1); the source codes the opposing side as the PKK throughout. Each event has a start and an end date and a code for the precision of its date. Events dated to the day, to a span of two to six days or to a known week (4,289 events) are assigned to the central day of their interval. The 75 events dated only to a span of more than a week and at most a month, or of more than a month and at most a year, cannot be placed in a week; they are excluded from the primary series and included in a sensitivity series that spreads every event uniformly over its date interval. The calendar from 1 January 1989 to 31 December 2016 has 10,227 days, exactly 1,461 weeks of seven days starting on Sundays, and yt is the number of events in week t.
Scope of the record. The dyad covers state-based violence only. GED records a further 422 events of one-sided violence by the PKK against civilians and 46 non-state events between the PKK and Kurdish parties in Iraq; the one-sided events enter a post hoc sensitivity analysis. The best estimates of the dyad sum to 28,370 deaths for 1989–2016, of the same order as the more than 30,000 deaths since 1984 that Crisis Group reports (International Crisis Group 2011). Of these, GED counts 22,129 as deaths of PKK members, 5,428 of the state forces, 354 of civilians and 459 of unknown status, and the events of the primary series fall in 47 provinces of Turkey and governorates of Iraq, most often Şırnak, Hakkari and Diyarbakır (Figure 1). The series counts events rather than deaths because the death estimates are lumpy: the 75 events dated only to a span of more than a week carry 3,265 of the deaths, and 132 dyad events have a low estimate above the best one and 263 a best estimate above the high one (538 and 1,218 in the whole dataset). A log-transformed death series enters the sensitivity analysis.

**Table 1. Composition of the event record used in the analysis: UCDP GED 17.1 events of the Turkey–PKK dyad by country and date precision, the weekly series built from them and the two negative-control series.**

| Quantity | Value |
|---|---|
| Events in UCDP GED 17.1 (all conflicts) | 135,181 |
| Turkey–PKK dyad (conflict Turkey:Kurdistan) | 4,364 |
| located in Turkey / in Iraq | 4,293 / 71 |

| Quantity | Value |
|---|---|
| date known to the day | 3,842 |
| within 2–6 days / week known | 420 / 27 |
| within a month / a year only (excluded)a | 63 / 12 |
| Events in the weekly series | 4,289 |
| Deaths, best estimate (all dyad events) | 28,370 |
| Weeks / weeks with at least one event | 1,461 / 940 |
| Maximum / mean events per week | 26 / 2.94 |
| Control: Philippines:Government events | 1,567 |
| Control: India:Kashmir events | 5,763 |

[a] Included in the sensitivity series, which spreads each event uniformly over its date interval; they carry 3,265 deaths.

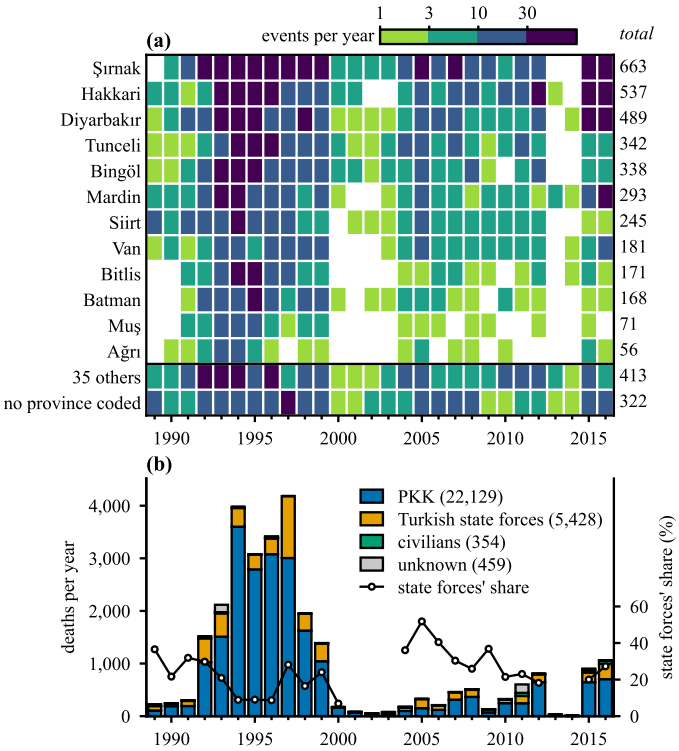



**Figure 1. Where and who in the event record (post hoc).**

(a) Events of the primary series by province (the first-level administrative unit coded in GED) and year: the twelve provinces with most events, the other provinces of Turkey and governorates of Iraq together, and the events without a province; white cells have no event, and the classes are logarithmic. Row totals at the right. (b) Deaths per year in all events of the dyad by the side of those killed, as GED records them (best estimates; side A is the government of Turkey and side B the PKK), and the share of the state forces in the deaths of each year with at least a hundred deaths (right axis).

**Pipeline.** Figure 2 shows the stages. Each result file records the commit, the tag of the analysis plan, the checksum of the data, the random seed and the hardware, and the document is assembled from those files, so that no number in the text is typed by hand. Python with NumPy and pandas carries the analysis (van der Walt, Colbert, and Varoquaux 2011; McKinney 2010), and the practices follow published rules for reproducible computational work (Sandve et al. 2013; Wilson et al. 2017).
**Variance stabilisation.** Weekly counts are overdispersed and their variance grows with their mean. The square-root transform in Equation (1) brings the variance of a Poisson count with a moderate mean close to one, by the delta method. The scale of the remaining noise is estimated from first differences, a standard device for error variance (Hall, Kay, and Titterington 1990).

Through the median absolute deviation (MAD) of the differences divided by 2, as in published change-point practice (Fryzlewicz 2014); the MAD makes the estimate robust to the jumps themselves. On the weekly series σ= 1.154.

$$z_t = 2\sqrt{y_t + \frac{3}{8}}, \qquad \hat{\sigma} = \frac{\operatorname{med}|\Delta z - \operatorname{med}(\Delta z)|}{\sqrt{2}\,\Phi^{-1}(3/4)} \qquad (1)$$

where Φ−1(3/4) is the upper quartile of the standard normal distribution and Δz the vector of first differences of z.

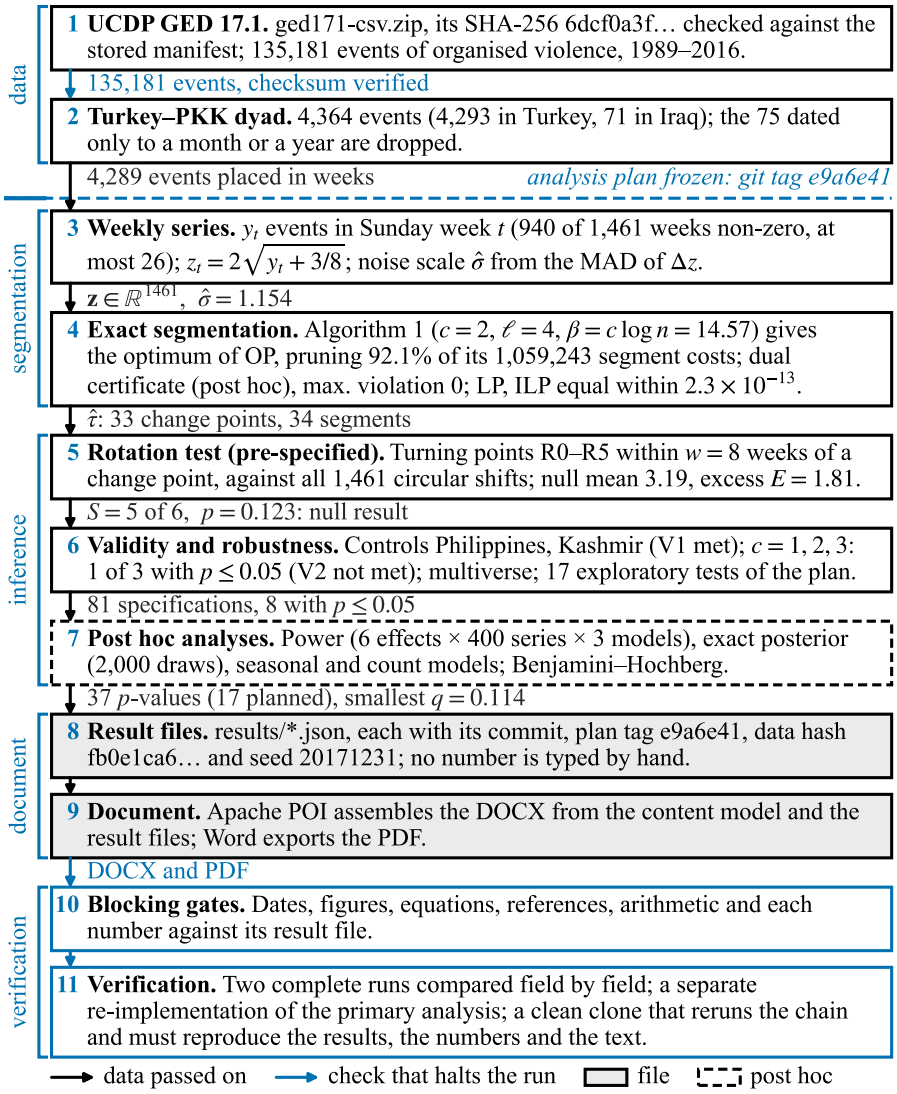



**Figure 2. Analysis pipeline**.

Boxes are the eleven stages in run order, numbered in blue, and the label on each arrow is the size of the data passed to the next stage. The blue brackets on the left group the stages into phases; the dashed blue line marks the git tag that froze the analysis plan before the primary run (before it, only the yearly event counts and the data audit were inspected). Blue arrows and outlines mark the checks that stop the run when they fail: the SHA-256 checksum of the archive against the stored manifest, the blocking gates and the verification. Grey boxes are files; the dashed box holds the post hoc analyses, which are outside the frozen plan. The data hash recorded in every result file is the SHA-256 of the checksum manifest.

## 4 Formulation and algorithms

Problem (P). A segmentation of the stabilised series $z_1, \ldots, z_n$ is a set of change points $\tau = (\tau_1 < \cdots < \tau_K)$, with $\tau_0 = 0$ and $\tau_{K+1} = n$; the segment $k$ covers the weeks $\tau_k + 1, \ldots, \tau_{k+1}$ and has length at least $\ell$. Table 2 lists the notation.

The cost of a segment is its within-segment sum of squares in units of $\hat{\sigma}^2$, Equation (2), and problem (P) asks for the segmentation that minimises the penalised cost of Equation (3), where every change costs $\beta$. The penalty is of Schwarz type, $\beta = c \log n$ with $c = 2$ in the primary analysis, because each change adds a location and a level; the minimum length $\ell = 4$ weeks keeps single extreme weeks from forming segments of their own.

**Table 2. Notation of the segmentation problem, the algorithms and the alignment test.**

| Symbol | Meaning | Domain |
|---|---|---|
| $n$ | number of weeks, 1 January 1989 to 31 December 2016 | 1,461 |
| $y_t$, $z_t$ | events in week $t$; stabilised value | $\mathbb{N}$; $\mathbb{R}_{\geq 0}$ |
| $\hat{\sigma}$ | noise scale from first differences | $\mathbb{R}_{>0}$ |
| $C(s,t)$ | cost of the segment $s+1,\ldots,t$ | $\mathbb{R}_{\geq 0}$ |
| $\beta$, $\ell$ | penalty per change; minimum length | $c \log n$; weeks |
| $\tau$, $K$ | change points and their number | $(K+1)\ell \leq n$ |
| $F(t)$ | optimal penalised cost of weeks $1..t$ | $\mathbb{R}$ |
| $R$, $Q$ | live and admissible candidates at time $t$ | $\subseteq \{0,\ldots,t\}$ |
| $[a_j, b_j]$, $w$ | documented interval $j$ of $m$; tolerance | weeks |
| $S$, $E$ | intervals recovered; excess over null | $\{0..m\}$; $\mathbb{R}$ |
| $p$, $q$ | rotation $p$-value; Benjamini–Hochberg adjusted value | $[0,1]$ |

$$C(s,t) = \frac{1}{\hat{\sigma}^2} \sum_{i=s+1}^{t} (z_i - \bar{z}_{s+1:t})^2 \qquad (2)$$

$$F(\tau) = \sum_{k=0}^{K} C(\tau_k, \tau_{k+1}) + \beta K, \qquad \tau_{k+1} - \tau_k \geq \ell \quad (k = 0,\ldots,K) \qquad (3)$$

A shortest path. Build the DAG with vertices $0,\ldots,n$ and an arc $s \to t$ for every admissible segment ($s = 0$ or $s \geq \ell$, and $t - s \geq \ell$), weighted by $C(s,t) + \beta$. Every path from $0$ to $n$ is a segmentation, and its length is the penalised cost plus $\beta$, so (P) is a shortest-path problem (Bellman 1958) on a DAG. Taking the vertices in topological order gives the recursion of Equation (4), which is OP (Jackson et al. 2005). With cumulative sums of the centred $z$ and $z^2$ each segment cost takes constant time, so OP runs in $\Theta(n^2)$ time and $O(n)$ memory; any exact method must read every observation, which gives the lower bound $\Omega(n)$.

$$F(t) = \min_{s \in \{0\} \cup [\ell,\, t-\ell]} \{F(s) + C(s,t) + \beta\}, \qquad F(0) = -\beta \qquad (4)$$

Exact pruning with a minimum length. The quadratic cost satisfies Equation (5): splitting a segment never increases the within-segment sum of squares, so the cost of a union is at least the sum of the costs of its parts. PELT uses this property to discard candidate positions that can never again be the last change point of an optimal solution (Killick, Fearnhead, and Eckley 2012). With a minimum length the discarding has to be delayed, because a position fewer than $\ell$ weeks back cannot yet close a segment. Proposition 1 is the pruning theorem of PELT with that delay made explicit, written as Equation (6); Algorithm 1 implements it.

$$C(s,u) + C(u,t) \le C(s,t), \qquad s < u < t \qquad (5)$$

Proposition 1. Let the cost satisfy Equation (5), as the quadratic cost does on any series; let $F$ be the optimal values of Equation (4); and let $s$ be admissible ($s = 0$ or $s \ge \ell$), with $s \le t - \ell$ and $F(s) + C(s,t) > F(t)$. Then Equation (6) holds: for every $T \ge t + \ell$ the position $t$ is strictly better than $s$ as the last change point before $T$. Proof. Because $T - t \ge \ell$, the position $t$ is admissible for $T$, and by Equation (5) $F(s) + C(s,T) \ge F(s) + C(s,t) + C(t,T) > F(t) + C(t,T)$. ▫

$$\begin{aligned} & F(s) + C(s,t) > F(t), \quad s \le t - \ell \\ \Rightarrow\ & F(s) + C(s,T) > F(t) + C(t,T) \\ & \forall\, T \ge t + \ell \end{aligned} \qquad (6)$$

| **Algorithm 1 Pruned exact segmentation with deferred pruning (PELT with minimum length $\ell$)** | |
|---|---|
| | Input: series $z_1, \dots, z_n$; penalty $\beta$; minimum length $\ell$; scale $\hat{\sigma}$ |
| | Output: optimal change points $\tau^*$ of problem (P) |
| 1: | compute cumulative sums of $z$ and $z^2$, so that $C(s,t)$ costs $O(1)$ |
| 2: | $F(0) \leftarrow -\beta$; $R \leftarrow \{0\}$; $\text{dead}(s) \leftarrow \infty$ for all $s$ |
| 3: | for $t = \ell, \dots, n$ do |
| 4: | $Q \leftarrow \{s \in R : s \le t - \ell,\ \text{dead}(s) > t\}$ |
| 5: | $F(t) \leftarrow \min_{s \in Q} F(s) + C(s,t) + \beta$; $\text{last}(t) \leftarrow$ smallest minimiser |
| 6: | for each $s \in Q$ with $F(s) + C(s,t) > F(t)$ do $\text{dead}(s) \leftarrow \min(\text{dead}(s), t + \ell)$ |
| 7: | $R \leftarrow \{s \in R : \text{dead}(s) > t + 1\} \cup \{t\}$ |
| 8: | trace $\text{last}(\cdot)$ back from $n$ to obtain $\tau^*$ |

Proposition 2. Algorithm 1 returns $F(n)$ and the same change points as OP when both break ties by the smallest position. Proof. By induction on $T$, assume that $F(u)$ is exact for every $u < T$. A position $s$ removed by time $T$ was dominated at some time $t_1$ with $s \le t_1 - \ell$ and $t_1 + \ell \le T$; if $t_1$ was itself removed by time $T$, it was dominated at some $t_2 \ge t_1 + \ell$, and so on. The chain grows by at least $\ell$ at each step and stays at or below $T - \ell$, so it ends at a position $t_k$ that is still in $Q$ at time $T$, and Proposition 1 applied along the chain gives $F(s) + C(s,T) > F(t_k) + C(t_k,T)$. Every removed position is therefore strictly worse at $T$ than a kept one: $Q$ is never empty, no minimiser of Equation (4) is removed, and the smallest minimiser over $Q$ is that of OP. ▫

In floating point the pruning test is applied with a relative tolerance of $1 \times 10^{-9}$, which only prunes less.

Without the delay the rule is not exact: on a short integer series, immediate pruning returns a cost of 123.55 against the optimum 121.77. Property-based tests compare Algorithm 1 with OP and OP with exhaustive search. Over 10,000 random integer series with runs of zeros, where exact ties occur ($n$ from 10 to 200 and $\ell$ from 1 to 6), Algorithm 1 returned the change points of OP in all 10,000; pruning happened in 8,870 of them, and immediate pruning lost the optimum in 592. In the worst case no position is ever pruned and the cost stays quadratic. Within a long segment the live set grows with the length of the segment, so the guarantee of a linear expected cost assumes that the number of changes grows linearly with $n$ (Killick, Fearnhead, and Eckley 2012).

Certificates of optimality. The values $F(t)$ returned by Algorithm 1 certify the optimum, in a check added post hoc, by linear-programming duality if they satisfy Equation (7) on every arc, with equality along the returned path. Values from OP satisfy it by construction, so the check is informative for the pruned solver, at the cost of one pass over the arcs, as much as OP itself; on the weekly series it covers 1,059,243 arcs, and the largest violation is 0.0. Problem (P) is also the flow program of Equation (8) over the set $A$ of admissible arcs that end at $n$ or at most $n - \ell$, with $b_0 = 1$, $b_n = -1$ and $b_u = 0$ otherwise.

$$\begin{aligned} & F(t) \le F(s) + C(s,t) + \beta \quad \forall\,(s,t), \\ & F(\tau^*_{k+1}) = F(\tau^*_k) + C(\tau^*_k, \tau^*_{k+1}) + \beta \end{aligned} \qquad (7)$$

$$\begin{aligned} &\min_{x} \sum_{(s,t)\in A} (C(s,t)+\beta)\, x_{st} - \beta \\ &\text{s.t.} \sum_{t:(u,t)\in A} x_{ut} - \sum_{s:(s,u)\in A} x_{su} = b_u, \\ &x_{st} \in \{0,1\} \end{aligned} \tag{8}$$

Remark. The linear-programming (LP) relaxation of Equation (8), with $x_{st} \geq 0$, has an integral optimum equal to $F(n)$; this is the classical integrality of network flow problems. Proof. A feasible unit flow on a DAG decomposes into paths from $0$ to $n$ with weights summing to one, so its cost is a weighted average of path costs, at least that of the shortest path, which is feasible and integral. ▫ Solved by a general solver, the relaxation and the integer linear program (ILP) cross-check the dynamic program up to the tolerance of the solver, with the same segment costs but none of the recursion or pruning code. The bound of Equation (9) on the number of change points, from the cost $C(0,n)$ of no change and the cost $\beta$ of each change, serves only as an internal check.

$$K^* \leq \left\lfloor \frac{C(0,n)}{\beta} \right\rfloor \tag{9}$$

Cost of a given alignment (post hoc). The same dynamic program answers a question that the test cannot: how much worse than the optimum is the best segmentation that recovers a given set of turning points? Adding to every vertex the set $\mathcal{W}$ of widened documented intervals that contain a change point so far gives the recursion of Equation (10), where $h(t)$ is the set of intervals that contain $t$ (empty at $t = n$). The minimum of $G(n,\mathcal{W})$ over the sets with $k$ elements is the cheapest segmentation that recovers exactly $k$ turning points; $s$ ranges over the same positions as in Equation (4). As in the test, a rotated interval that passes the end of the series continues at its start.

The state space grows by a factor of $2^m$, and so does the cost.

$$\begin{aligned} G(t,\mathcal{W}) &= \min_{s,\mathcal{W}'} \{G(s,\mathcal{W}') + C(s,t) + \beta\}, \\ \mathcal{W} &= \mathcal{W}' \cup h(t), \quad G(0,\emptyset) = -\beta \end{aligned} \tag{10}$$

Greedy baseline. Binary segmentation splits the series at the position with the largest reduction of cost in Equation (11), accepts the split if the reduction exceeds $\beta$, and recurses on the two parts; the order of the splits does not change the result, and the cost is at most of order $nK$ segment costs. With the same cost, penalty and minimum length, its penalised cost can be compared with the optimum directly, and it is never lower.

$$\begin{aligned} g(a,b) &= \max_{a+\ell \leq u \leq b-\ell} \Delta C(u), \\ \Delta C(u) &= C(a,b) - C(a,u) - C(u,b) \end{aligned} \tag{11}$$

## 5 Pre-specified alignment test

**Documented turning points**. Before any segmentation was computed, the analysis plan fixed six documented turning points, those that three Crisis Group texts date explicitly, each as an interval at the precision given by its source: the start of the second phase in 1991 and the end of the ceasefire in 2004 at the precision of a year (International Crisis Group 2011); Öcalan's capture in February 1999, his call in August 1999 for the withdrawal of militants, followed by a unilateral ceasefire, and his Newroz letter of 21 March 2013, followed by the PKK's ceasefire (International Crisis Group 2014); and the collapse of the ceasefire in July 2015 (Göksel 2015), labelled R0 to R5 in that order. Everything the plan contains was committed and tagged in the project repository before the primary analysis ran; the tag carries no third-party timestamp. The only data inspected before the tag were the yearly event counts and the descriptive quality statistics of the data audit, and the yearly counts already show the large phases.

**Statistic and null distribution.** Week indices count from zero, so that a change point τk is the first week of the segment it opens. A turning point j with documented interval [aj,bj] is recovered when some change point falls within w weeks of the interval, and S in Equation (12) counts the recovered points. Its reference distribution comes from the n circular shifts of the optimal change points, which keep their number and spacing and move them relative to the documented intervals and to the calendar year. The p in Equation (13) is a complete-enumeration rotation p-value, exact under the assumption that every rotation of the observed change points was equally likely; including the identity keeps it from being zero, as recommended for permutation p-values (Phipson and Smyth 2010). The excess E in Equation (14) measures recovery beyond chance.

$$S(\boldsymbol{\tau}) = \sum_{j=1}^{m} \mathbb{1}\left[\exists k : \tau_k \in [a_j - w,\ b_j + w]\right] \tag{12}$$

$$p = \frac{1}{n} \sum_{\delta=0}^{n-1} \mathbb{1}\left[S((\boldsymbol{\tau}^* + \delta) \bmod n) \geq S(\boldsymbol{\tau}^*)\right] \tag{13}$$

$$E = S(\boldsymbol{\tau}^*) - \frac{1}{n}\sum_{\delta=0}^{n-1} S\left((\boldsymbol{\tau}^* + \delta) \bmod n\right) \qquad 14)$$

**Decision rule.** The primary specification is c=2, l=4 and w=8 weeks on the event series. The result is positive if p≤0.05 and null otherwise. A positive result is valid only if two negative controls, the GED series of the conflicts Philippines:Government and India:Kashmir aligned against the same Turkish intervals, show no alignment (p>0.05, condition V1), and if at least two of the three penalties c∈{1,2,3} give p≤0.05 (condition V2). An excess of at least two turning points was fixed as the smallest difference of practical relevance: a third of the documented periodisation beyond chance. Because the null depends on how many change points there are, the plan also recorded the resolution of the test before any data were segmented: with change points placed at random, about four of the six points would have to be recovered for p≤0.05 when there are ten change points, and all six when there are forty. The plan made no power analysis against a specific alternative; a post hoc one is reported in Section 6.

**Secondary and exploratory analyses**. The plan declared as secondary the recovery of each turning point, the fraction P of change points that fall inside some interval (Equation (15)), the segment means with intervals from a moving-block bootstrap (Künsch 1989) defined in Equation (16), the comparison of the greedy change points τG with the optimum by relative cost gap Δ (Equation (17)) and Hausdorff distance dH (Equation (18)), the exactness checks of Section 4, and the scaling of the three algorithms. Robustness is a multiverse (Steegen et al. 2016) of every combination of c∈{1,2,3}, l∈{2,4,8}, w∈{4,8,13} and three series (events, events spread over their date intervals, and the logarithm of one plus the deaths). The exploratory analyses of the plan are the exact solution path over the penalties c∈[1,4] (Haynes, Eckley, and Fearnhead 2017), a null that permutes half-year blocks of the series and re-segments it, and the dyad without its events located in Iraq.

$$\mathrm{P}(\tau) = \frac{1}{\mathrm{K}}\sum_{\mathrm{k}=1}^{\mathrm{K}} \mathbb{1}\left[\exists \mathrm{j}: \tau_{\mathrm{k}} \in [\mathrm{a_j} - \mathrm{w},\ \mathrm{b_j} + \mathrm{w}]\right] \qquad (15)$$

$$\bar{\mathrm{y}}_{\mathrm{k}}^{(\mathrm{b})} = \frac{1}{\mathrm{L_k}}\sum_{\mathrm{i}=1}^{\mathrm{L_k}} \mathrm{y}_{\mathrm{k,i}}^{(\mathrm{b})}, \qquad \mathrm{b} = 1, \ldots, \mathrm{B} \qquad (16)$$

where $\mathrm{y}_{\mathrm{k,\cdot}}^{(\mathrm{b})}$ concatenates blocks of four consecutive weeks drawn with uniform starting points inside segment $\mathrm{k}$ of length $\mathrm{L_k}$, truncated to $\mathrm{L_k}$ weeks; the interval is formed by the percentiles $2.5$ and $97.5$ over $\mathrm{B} = 2000$ replicates. Segments shorter than three blocks get no interval, because the resampling then has too little support and the interval ends at the estimate.

$$\Delta = \frac{\mathrm{F}(\tau_{\mathrm{G}}) - \mathrm{F}(\tau^*)}{\mathrm{F}(\tau^*)} \qquad (17)$$

$$\mathrm{d_H} = \max\{\mathrm{h}(\tau_{\mathrm{G}}, \tau^*),\ \mathrm{h}(\tau^*, \tau_{\mathrm{G}})\}, \qquad \mathrm{h}(\mathrm{A}, \mathrm{B}) = \max_{\mathrm{a}\in\mathrm{A}} \min_{\mathrm{b}\in\mathrm{B}} |\mathrm{a} - \mathrm{b}| \qquad (18)$$

**Post hoc analyses.** Analyses added after the primary result was known are labelled post hoc wherever they appear; Table 3 lists every analysis with its status. Three examine the power and the sensitivity of the test. Power is estimated on series simulated from a model fitted to the record: the segment levels of the harmonic model below, rotated at random around the calendar, times a factor for each calendar month, with negative binomial noise of size 7.2, estimated by moments around these expected counts. Under the alternative the mean is multiplied by $e^{\pm\eta}$ from the middle of each documented interval onwards, up at R0, R3 and R5 and down at R1, R2 and R4, so that the changes accumulate; the simulation is repeated without the monthly factor, and the series with the monthly factor are also analysed with the harmonic model, from two starting cycles. The annual cycle is modelled by Equation (19), a piecewise-constant level plus a common cycle of two harmonics, with ν= 52.18 weeks in a year and the σ and β of the primary analysis. It is fitted by alternating the exact segmentation of zt minus the cycle with the least-squares fit of levels and cycle, which never increases the penalised cost but ends at a fixed point that need not be the joint optimum; the fit therefore starts from a zero cycle, from the cycle fitted without change points and from 10 random cycles, and keeps the cheapest fixed point. The cost of recovering a given set of turning points comes from Equation (10), with a null that re-optimises the segmentation for every rotation of the intervals.

**Table 3. Every analysis of the article with its status in the frozen plan and its deviation from the plan, if any.**

| **Analysis** | **Status** | **Deviation from the plan** |
|---|---|---|
| Alignment test of the exact optimum ($\mathrm{c} = 2$, $\ell = 4$, $\mathrm{w} = 8$) | confirmatory | as planned |

| **Analysis** | **Status** | **Deviation from the plan** |
|---|---|---|
| Validity conditions V1 and V2 | validity | as planned |
| Recovery per point; fraction P; segment means | secondary | no interval for two short segments |
| Greedy segmentation: cost gap and Hausdorff distance | secondary | as planned |
| Multiverse of 81 specifications | robustness | as planned |
| Exact solution path; block-permutation null; dyad without Iraq | exploratory | as planned |
| Pruning rule of Algorithm 1 | secondary | strict inequality, so that ties are never pruned |
| Benchmark and exact baselines | secondary | OP up to $2^{17}$, not $2^{14}$; repetitions vary by size; baselines timed once |
| Power by simulation; cost of recovering k turning points; exact posterior and the test on its draws | post hoc | added |
| Annual cycle, monthly medians and negative binomial cost, with variants | post hoc | added |
| Greedy alignment and validity; joint multiverse test; other conflicts | post hoc | added |
| Calendar months; whole-year rotations; distance and direction; one-sided violence; R4 in December 2012; ceasefires of 1993, 2009 and 2010 | post hoc | months, R4 date and ceasefires chosen after seeing the data |
| Residual autocorrelation beyond the first lag and spread by segment; dual certificate; GED sources by year | post hoc | added checks |
| False discovery rate over 37 exploratory and post hoc tests | post hoc | the plan's correction, extended to the post hoc tests |

The other post hoc analyses are: the alignment of the greedy segmentation, with its validity conditions and multiverse; the calendar months of the change points, compared by Equation (20) with the rotations of the calendar, where K↑ is the set of upward changes, $\varkappa(\cdot)$ the month of a week and M↑ the months March to May (October to December for downward changes), months chosen after seeing the change points, so that these two p-values overstate the evidence even after correction; a null that rotates the change points only by whole years; the series with the median of each calendar month removed by Equation (21); the raw counts segmented with the negative binomial cost of Equation (22) (Cleynen and Lebarbier 2014), in deviance units, where $L=t-s$, Y is the sum of the counts and $\mu=Y/L$, with a common size r estimated by moments within the segments of the primary optimum (as r grows, the cost differences tend to those of the Poisson cost); the total distance of Equation

(23) from the intervals to the nearest change point, which needs no tolerance; a direction-aware count; a joint rotation test over the multiverse; the exact posterior of the change points of a product-partition model (Fearnhead 2006) by the forward recursion of Equation (24), where Λ(t) is the joint density of z1,…,zt and a change point after week t, λ(s,t) integrates out a normal segment mean with prior mean μ0, the mean of the series, and prior standard deviation σ0, that of the optimal segment means, ω makes segment lengths geometric beyond l with a weekly change probability θ=1/52 (1/104 and 1/26 as alternatives, and σ0 halved or doubled), the last segment takes (1−θ)L−l instead, and the scale is σ; segmentations drawn backwards with π(s|t), the probability that the change point before t is at s, are submitted to the pre-specified test; the dyad with the one-sided violence of the PKK; R4 dated to December 2012, also chosen after seeing the change points; and the same test on other conflicts.

The exploratory p-values of the plan and every post hoc p-value given individually form one family of 37 tests, 17 from the plan (the 15 non-primary segmentations of the solution path, the block null and the series without Iraq), adjusted by the Benjamini–Hochberg procedure (Benjamini and Hochberg 1995) as in Equation (25), where p(1)≤…≤p(N) are the N ordered p-values; q is given for every post hoc p≤0.05.

Under positive dependence the procedure controls the false discovery rate; its modification for any dependence (Benjamini and Yekutieli 2001) raises the smallest q to 0.479, and within the plan's 17 tests alone the smallest q is 0.204.

Not corrected are the validity conditions, which belong to a decision rule (the segmentations of V2 enter only as points of the path); the secondary fraction P, descriptive in the plan; the calibration on other conflicts; and variants summarised by a count or a range.

$$z_t = \mu_{k(t)} + s_t + \varepsilon_t,$$

$$s_t = \sum_{h=1}^{2} \left( a_h \cos \frac{2\pi h t}{\nu} + b_h \sin \frac{2\pi h t}{\nu} \right) \tag{19}$$

$$p_{\uparrow} = \frac{1}{n} \sum_{\delta=0}^{n-1} \mathbb{1}\left[ U_\delta \geq U_0 \right], \tag{20}$$

$$U_\delta = \sum_{k \in \mathcal{K}_{\uparrow}} \mathbb{1}\left[ \kappa((\tau_k^* + \delta) \bmod n) \in M_{\uparrow} \right]$$

$$z_t^a = z_t - \left( \mathrm{med}_{u:\,\kappa(u)=\kappa(t)}\, z_u - \mathrm{med}_u\, z_u \right) \tag{21}$$

$$C_{NB}(s,t) = 2rL \log\left(1 + \frac{\mu}{r}\right) - 2Y \log \frac{\mu}{r+\mu} \tag{22}$$

$$D(\tau) = \sum_{j=1}^{m} \min_k \mathrm{dist}\left(\tau_k, [a_j, b_j]\right) \tag{23}$$

$$\Lambda(t) = \sum_{s \in \{0\} \cup [\ell,\, t-\ell]} \Lambda(s)\, \lambda(s,t)\, \omega(t-s),$$

$$\Lambda(0) = 1, \qquad \omega(L) = \theta\,(1-\theta)^{L-\ell},$$

$$\lambda(s,t) = \int \mathcal{N}(\mu;\, \mu_0, \sigma_0^2) \times \prod_{i=s+1}^{t} \mathcal{N}\left(z_i;\, \mu, \hat{\sigma}^2\right) d\mu, \tag{24}$$

$$\pi(s \mid t) \propto \Lambda(s)\, \lambda(s,t)\, \omega(t-s)$$

$$q_{(i)} = \min\left\{1,\ \min_{j \geq i} \frac{N\, p_{(j)}}{j}\right\} \tag{25}$$

## 6 Results

The optimal periodisation. With β= 14.57 the exact optimum has 33 change points and penalised cost 1,748.57. Algorithm 1 and OP return the same change points and cost, and the dual check, added post hoc, confirms the optimum on every arc. An independent re-implementation from the frozen plan, the data description and the documented intervals, without access to the analysis code, reproduced the scale, change points, cost and p-value to at least nine significant digits. Figure 3 shows the series and the 34 segments. The residuals have a standard deviation of 0.93 in units of σ and a lag-one autocorrelation of 0.020 within segments, and follow the normal quantiles up to the steps left by the discrete counts, with slightly heavier tails. Post hoc, the pooled within-segment autocorrelation leaves the band ±1.96/nh, with nh the number of within-segment pairs h weeks apart, at 2 of the first twelve lags but at 27 of the first 60: it is negative near half a year and positive near a year, 0.141 at one year and up to 0.191 (Figure 4). The residuals keep part of the annual cycle, and their standard deviation ranges from 0.38 to 1.55 in units of σ across segments.

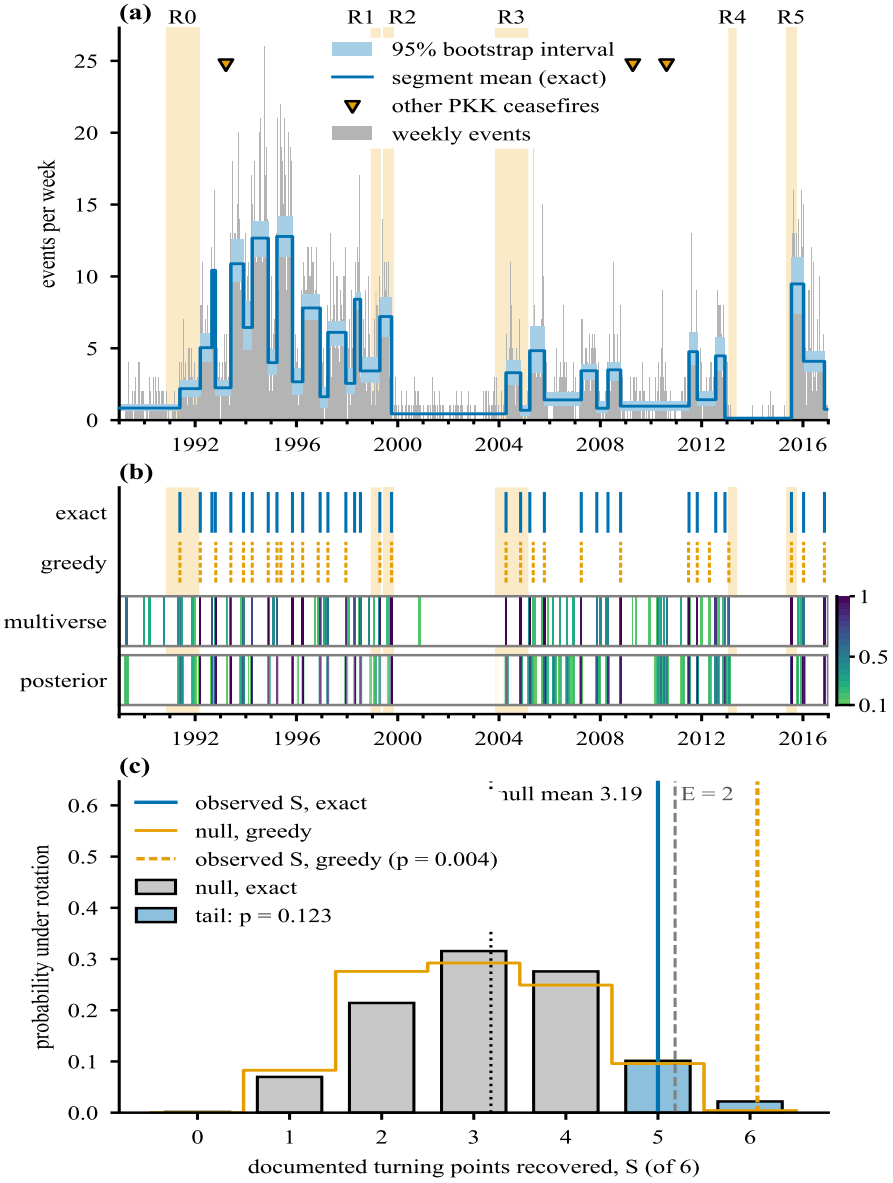



**Figure 3. Exact segmentation of the weekly event record and the pre-specified alignment test.**
(a) Weekly events (grey), mean of each segment of the optimal segmentation (blue) with its 95% moving-block bootstrap interval (shaded), and documented turning points R0–R5, shaded with the tolerance w; triangles, the further PKK ceasefires of 1993, 2009 and 2010 (post hoc). (b) Change points of the exact optimum and of greedy binary segmentation over the same shading, and two strips (post hoc): the share of the distinct segmentations of the multiverse (combinations of penalty, minimum length and series that give different change points; the tolerance does not change them) with a change point within two weeks of each week, and the exact posterior expected number of change points within two weeks of each week; darker for higher values up to one, white below one tenth, as in the key at the right. (c) Circular-shift null distributions of S over all n rotations, for the exact solution (bars; the blue tail at or above the observed S is its one-sided p) and for the greedy one (outline), with the observed S of each; the dotted line is the null mean and the dashed line the relevance threshold, two turning points above it. Each p uses the null of its own solution, and the greedy test is post hoc.

Alignment with the documented turning points. The optimum recovers five of the six turning points (Table 4). The start of the second phase in 1991, the capture of Öcalan, the withdrawal call of 1999, the end of the ceasefire in 2004 and the collapse of July 2015 each have a change point within the tolerance of their interval; the change point closest to the Newroz letter of 21 March 2013 falls 15 weeks earlier, in early December 2012. Five recoveries are not rare under the circular-shift null: the two windows a year wide are hit by 74% of the rotations each and the other four by 37% to 44%, so the null mean is 3.19 recovered points. The exact one-sided p is 0.123, and the excess E= 1.81 is below the threshold of two fixed in the plan. Only six of six would have reached significance, which happens in 32 of the 1,461 rotations. The primary result is therefore null: the alignment cannot be distinguished from that of rotated change points, and the test cannot rule out an excess of two. The fraction P of change points that fall inside some interval, 18.2%, is likewise close to its null mean of 14.9% (p= 0.383).

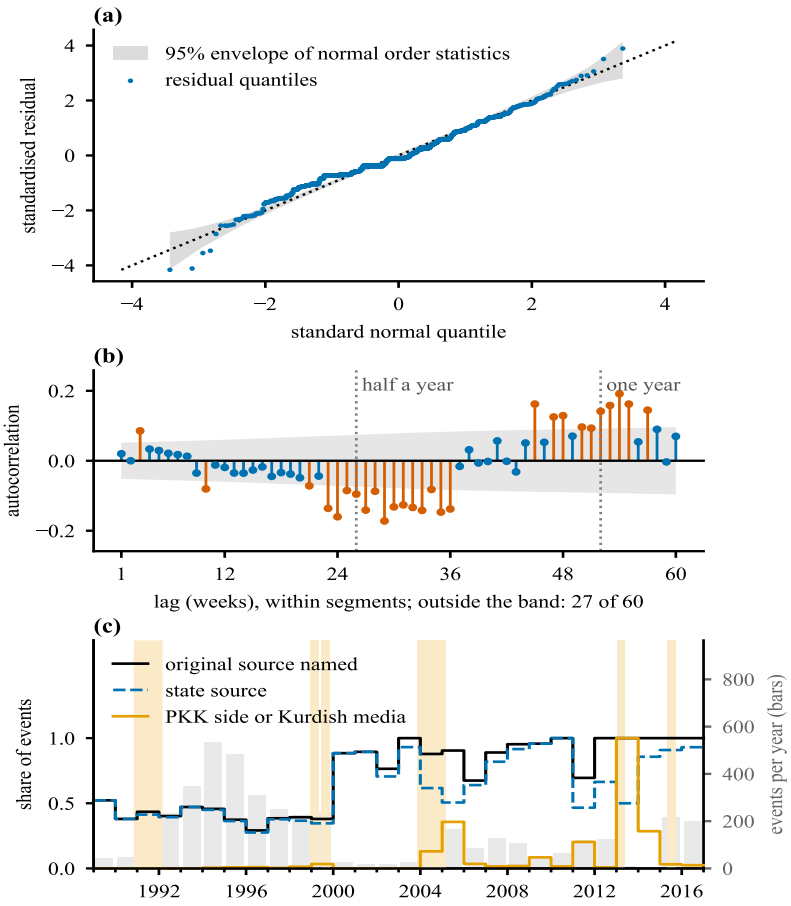



**Figure 4. Diagnostics of the primary segmentation and of the record.**

(a) Quantiles of the standardised residuals against the standard normal, with the pointwise 95% envelope of normal order statistics (grey); the steps in the centre are ties of the transformed counts. (b) Autocorrelation of the residuals within segments by lag, up to sixty weeks (post hoc beyond the first lag); the grey band is ±1.96/nh, with nh the number of pairs of weeks h apart in the same segment; lags outside it are in vermilion, and dotted lines mark half a year and one year. (c) Share of the events of the primary series in each year that name an original source in GED, that cite a state source and that cite the PKK, its armed wing or Kurdish media, classified by keywords (post hoc), with the events of each year as grey bars (right axis); shaded, the documented intervals widened by w.

**Table 4. Documented turning points and their recovery by the optimal segmentation: documented date at the precision of its source, distance to the nearest optimal change point, whether it falls within the tolerance, how often a rotated set of change points would hit it, and the exact posterior probability of a change point in the widened interval, with and without the annual cycle (post hoc).**

| | Documented turning point | Date | Gap [a] | it[b] | Null[c] | Post.[d] | No cycle[e] |
|---|---|---|---|---|---|---|---|
| 0 | second phase begins (International Crisis Group 2011) | 1991 | 0 | yes | 74% | 0.99 | 0.99 |

(b)

| | Documented turning point | Date | Gap [a] | it[b] | Null[c] | Post.[d] | No cycle[e] |
|---|---|---|---|---|---|---|---|
| 1 | Öcalan captured and jailed (International Crisis Group 2014) | February 1999 | 7 | yes | 44% | 0.97 | 0.88 |
| 2 | withdrawal call, then ceasefire (International Crisis Group 2014) | August 1999 | 5 | yes | 44% | 0.99 | 0.99 |
| 3 | PKK ends the ceasefire (International Crisis Group 2011) | 2004 | 0 | yes | 74% | 0.99 | 0.82 |
| 4 | Newroz letter, then ceasefire (International Crisis Group 2014) | 21 March 2013 | 5 | no | 37% | 0.51 | 0.99 |
| 5 | ceasefire collapses (Göksel 2015) | July 2015 | 0 | yes | 44% | 1.00 | 1.00 |

[a] Weeks between the documented interval and the nearest optimal change point.
[b] Gap within the tolerance w of the primary specification.
[c] Share of the circular shifts of the optimal change points that hit the widened interval.
[d] Exact posterior probability of at least one change point in the widened interval (post hoc), rounded down.
[e] The same, rounded down, with the annual cycle of the harmonic model subtracted from the series.

**Validity and robustness.** Condition V1 holds: aligned against the Turkish intervals, neither control series shows alignment ($p = 0.155$ for the Philippines and 0.739 for Kashmir; Table 5). Condition V2 does not: with $c=1$ there are 44 change points and $p = 0.211$, and only $c=3$, which leaves 20 change points, gives $p \leq 0.05$ (0.024). Across the 81 specifications of the multiverse, 8 (9.9%) reach $p \leq 0.05$ and 23.5% reach the relevance threshold (Figure 5); most change points of the optimum recur, within two weeks, in at least half of the distinct segmentations of the multiverse (Figure 3b). All significant specifications use the primary tolerance of eight weeks. Post hoc, a joint rotation of all 81 specifications gives $p = 0.051$ for the number of significant ones (8 against a null mean of 1.09) and $p = 0.068$ for their median excess. Two of the three exploratory analyses of the plan give the same verdict as the primary test: without its events in Iraq the dyad has 37 change points and recovers 5 turning points ($p = 0.163$), and the null that permutes half-year blocks and re-segments the series gives $p = 0.270$ from 199 permutations, although the block boundaries add change points (median 45 against 33), which makes that null harder to beat. As a post hoc calibration, the same test with the Turkish intervals was applied to the 29 other state-based conflicts of GED 17.1 with at least 500 events in their weekly series (22 of them, with mostly empty weeks, have a MAD of the differences of zero and take the standard deviation of the differences as scale): their median p is 0.639; none of the 29, including the two controls, reaches $p \leq 0.05$, and none reaches the p of the dyad (Figure 6a). On unrelated series the test is conservative, as its simulated size below also shows.

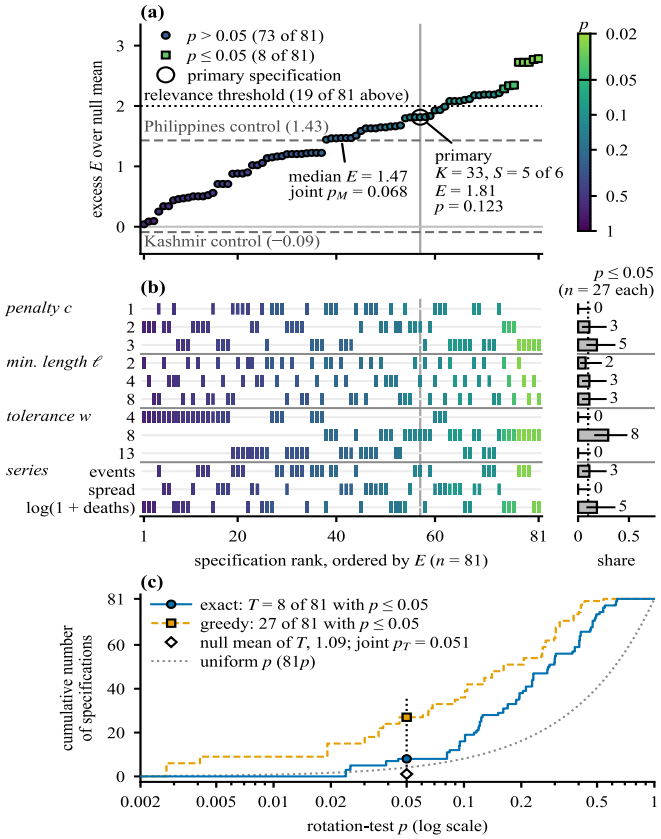



**Figure 5. Specification curve of the 81 specifications of the multiverse.**

(a) Excess E of recovered turning points over the null mean for each specification, in increasing order (ties by p); colour gives the p-value of the circular-shift test on a logarithmic scale (bar at right, with a black mark at p=0.05), squares mark p≤0.05, and the ring and the vertical grey line in (a) and (b) mark the pre-specified primary specification. The dotted line is the relevance threshold E=2, labelled with the number of specifications above it; the dashed lines give the excess of the two negative-control conflicts, computed at the primary specification only.

The label at the median specification gives the median excess and the p-value of its joint rotation test (post hoc). (b) Analytic choices of each specification: one tick in the row of each of its options, coloured as its point in (a); c, penalty multiplier; l, minimum segment length and w, tolerance, both in weeks; series: weekly events, events spread over their date intervals (spread) and the logarithm of one plus the deaths. Right: share of the 27 specifications with each option that give p≤0.05, with their count and a 95% Wilson interval, a descriptive guide only because the specifications share their data; the dotted line is the share over all 81. (c) Cumulative number of specifications with a p-value at or below each value of the horizontal axis, for the exact segmentation (solid) and for the greedy segmentation of the same specifications (dashed, post hoc). At p=0.05 the circle gives T, the number of specifications with p≤0.05, the square the same count for the greedy segmentation, and the diamond the mean of T under a joint rotation of all 81 specifications, with the p-value of that joint test (post hoc); the grey dotted curve is the count expected if the p-values were uniform, an upper bound on the null expectation for valid p-values.

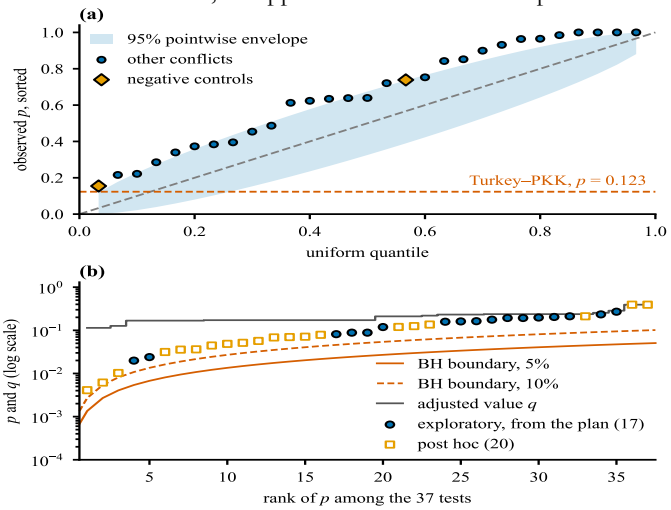



**Figure 6. Calibration on other conflicts and the family of tests (post hoc).**

(a) Sorted p-values of the pre-specified test with the Turkish intervals on the 29 other state-based conflicts of GED 17.1 with at least 500 events in their weekly series, against the quantiles of the uniform distribution, with its 95% pointwise envelope; diamonds, the two negative controls; dashed line, the p-value of the dyad. (b) The 37 p-values of the family, sorted, against the Benjamini–Hochberg boundaries at levels of five and ten per cent, with their adjusted values q of Equation (25) (step line); circles, exploratory tests of the plan; squares, post hoc tests. The three smallest are those of the greedy solver, of the series without its calendar-month medians and of the harmonic model.

(b)

**Table 5. Alignment under the validity conditions, the exploratory and the post hoc analyses: number of change points K, recovered turning points S, mean of S over all circular shifts and one-sided exact p.**

| Segmentation | Rolea | | S | Null mean | p |
|---|---|---|---|---|---|
| Primary (exact, $c = 2$) | C | 33 | 5 | 3.19 | 0.123 |
| Exact, $c = 1$ | V2 | 44 | 5 | 3.54 | 0.211 |
| Exact, $c = 3$ | V2 | 20 | 5 | 2.28 | 0.024 |
| Control: Philippines | V1 | 12 | 3 | 1.57 | 0.155 |
| Control: Kashmir | V1 | 19 | 2 | 2.09 | 0.739 |
| Dyad without Iraq | E | 37 | 5 | 3.35 | 0.163 |
| Greedy segmentation | P | 29 | 6 | 3.01 | 0.004 |
| Annual cycle modelled | P | 15 | 5 | 1.82 | 0.010 |
| Monthly medians removed | P | 17 | 6 | 1.91 | 0.006 |
| Negative binomial cost | P | 37 | 6 | 3.49 | 0.031 |
| With one-sided violence | P | 33 | 5 | 3.19 | 0.125 |

Tolerance w and minimum length l as in the primary specification; p from all circular shifts.
a C confirmatory; V1, V2 validity conditions; E exploratory in the plan; P post hoc.

**Power and sensitivity of the test (post hoc).** In 400 series simulated from the model of Section 5 (each rate within ±4.9 percentage points at 95%), the test rejects 1.8% of the series when the documented dates carry no change, 8.0% when each carries a level change by a factor of 2.7, and at most 28.8% when the factor reaches 20 (Figure 7c). Large changes add change points (median 52), and the segmentation then recovers all six turning points in only 22.3% of the series. Without the monthly factor the same simulation rejects 29.8% at a factor of 2.7 and at most 55.8%. Analysed with the harmonic model below, the series with the monthly factor give a size of 1.8% and a rejection rate of 23.5% at a factor of 2.7, and at most 41.0%. The size of the other post hoc tests was not simulated. Modelling the annual cycle with Equation (19) (Figure 8) leaves 15 change points, which recover 5 turning points, all but R1, with an excess of 3.18 ($p= 0.010$, $q= 0.127$); the 12 starting cycles reach 2 fixed points, each with $S= 5$ and $p\leq 0.011$. The alignment reaches $p\leq 0.05$ in 11 of 15 variants of the number of harmonics, c and w, though not with w=4 nor with one harmonic and w=13. Subtracting the median of each calendar month instead (Equation (21)) leaves 17 change points, which recover all six ($p= 0.006$, $q= 0.114$). On the raw counts, the negative binomial cost of Equation (22) with $r= 8.5$ leaves 37 change points, which recover 6 turning points ($p= 0.031$, $q= 0.168$); over 15 variants of r, including the Poisson limit, and of c, 6 reach $p\leq 0.05$. The dynamic program of Equation (10) shows that the cheapest segmentation recovering all six turning points costs 5.58 more than the optimum, 0.38 of the penalty of one change point (Figure 7a): it moves the change points of 22 July 2012 and 2 December 2012 to 22 April 2012 and 20 January 2013 and

keeps the other 31. For rotated intervals the median extra cost is 27.3, and a fraction 0.049 of the rotations are as cheap as the documented dates (q= 0.172; Figure 7b); with w=4 and w=13 the fractions are 0.120 and 0.136. A null that rotates the change points only by whole years, and so keeps their calendar months, gives p= 0.036 (q= 0.168) at w=8 but 0.393 at w=4 and 0.393 at w=13, with only 28 rotations, of which the value at w=8 is the smallest attainable. With R4 dated to December 2012, when the Crisis Group report records the announcement that the intelligence service was in direct contact with Öcalan (International Crisis Group 2014), the optimum recovers all six turning points (p= 0.036, q= 0.168); the recoding was chosen after seeing the change points, and its validity conditions were not evaluated. The total distance of Equation (23), which needs no tolerance, is 27 weeks against a null mean of 110 (p= 0.057). Integrating over the number and the position of the change points, the exact posterior of a product-partition model (Fearnhead 2006) gives a posterior expected number of 46.1 change points (Figure 3b), puts a change point inside each of five widened intervals with probability at least 0.976, and inside the interval of R4 with probability 0.51 (0.37 to 0.67 over the five priors; Table 4). With the harmonic cycle of Equation (19) subtracted from the series, the probability for R4 rises to 0.99, while those for R1 and R3 fall to 0.88 and 0.82. The expected number of change points inside the intervals, 8.47 against 6.88 under rotation, is not significantly larger (p= 0.211; 0.070 with the cycle subtracted).

Applied to each of 2,000 segmentations drawn from the posterior, the pre-specified test finds S=6 in 49% of them, but with a median of 46 change points, rotations also often reach six, so only 3.9% give p≤0.05 (median p 0.204), and 0.3% to 26.8% across the rates of change and prior spreads, against 75% with the cycle subtracted (Figure 7d). Counting only change points in the direction of the documented phase (rises in 1991, 2004 and 2015, falls in 1999 and 2013) gives S= 4, the largest value any rotation reaches, with p= 0.079. Adding the one-sided violence of the PKK changes the result only slightly (33 change points, S= 5, p= 0.125).

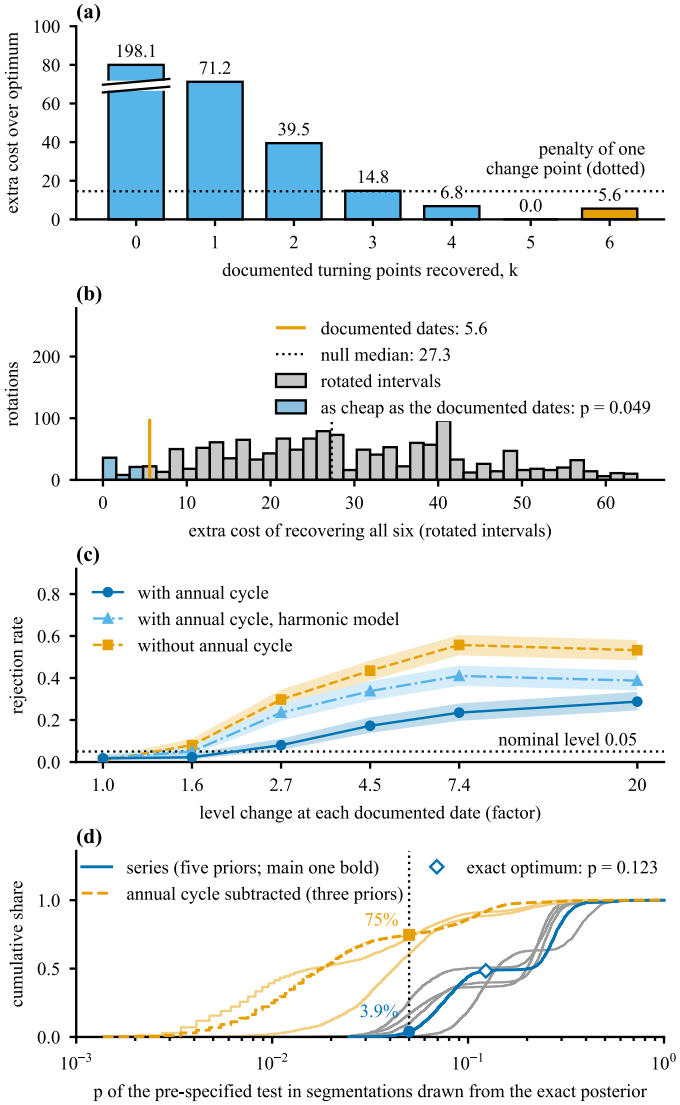



**Figure 7. Post hoc analyses of the alignment.**

(a) Extra penalised cost over the optimum of the cheapest segmentation that recovers exactly k documented turning points, from the exact dynamic program; the optimum recovers five. (b) Extra cost of recovering all six for every rotation of the documented intervals; in blue, the rotations at least as cheap as the documented dates, whose fraction is the one-sided p, and dotted, the null median. (c) Rejection rate of the pre-specified test in series simulated from a model fitted to the record, with

and without its annual cycle, and of the same test on the harmonic model, against the factor eη of the level change at each documented date (η from 0 to 3); bands, 95% Wilson intervals of the simulated rates. (d) Cumulative share of the segmentations drawn from the exact posterior whose pre-specified test gives a p at or below each value, for the series under the five priors (the main one bold) and with the annual cycle subtracted (three priors, dashed); the dots at p=0.05 give the shares of the text and the diamond the p of the exact optimum.

The solver and the outcome of the test. Binary segmentation with the same cost, penalty and minimum length returns 29 change points, 22 of them in the same week as the optimum. Its penalised cost exceeds the optimum by 3.1% (55.0 in units of the cost); every greedy change point lies within 13 weeks of an optimal one, while the optimal change point of 11 November 2007 lies 32 weeks from the nearest greedy one (Hausdorff distance 32). Yet the greedy segmentation recovers all six turning points, and the same test, with the null of the greedy change points, gives p= 0.004 (post hoc, q= 0.114). It would also have passed both validity conditions of the plan: the controls give p= 0.276 and 0.640, and the penalties c=1 and c=3 give 0.038 and 0.030. Under the decision rule of the plan, a greedy periodisation would have been reported as a valid positive. Across the multiverse the two solvers disagree about significance in 25 of the 81 specifications, in 22 of them with only the greedy one significant. The outcome of the test thus depends on which solver is used and, as the cost of alignment shows, on the position of two change points.

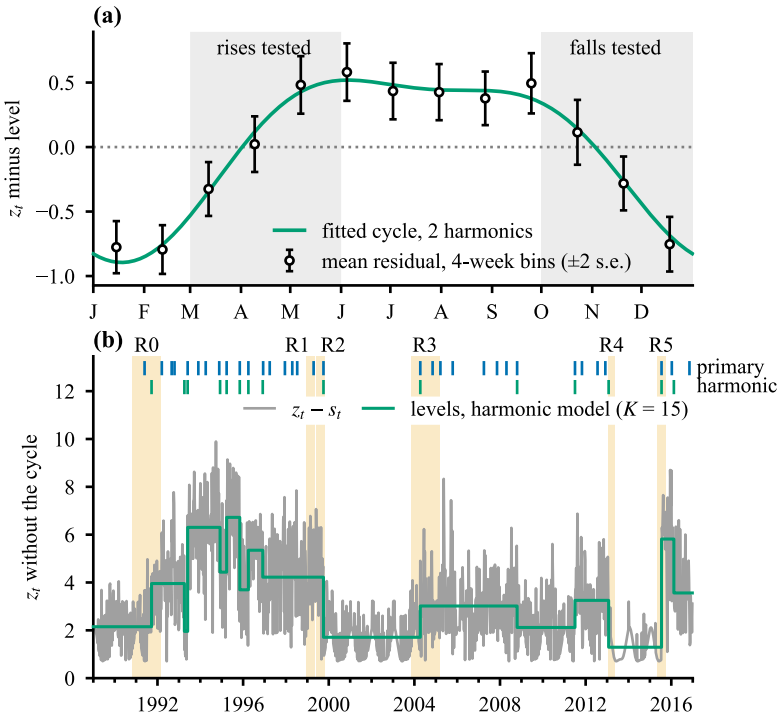



**Figure 8. The harmonic seasonal model of Equation (19) (post hoc).**

(a) Mean of zt minus the level of its segment in the harmonic model, in bins of four weeks of the year, ±2 standard errors, and the common annual cycle st fitted together with the segmentation (green); shaded, the months of the seasonal test, chosen after seeing the change points. (b) zt−st with the levels of the segments of the harmonic model, the documented intervals widened by w, and the change points of the harmonic model (green) and of the primary segmentation (blue, as in Figure 3).

**What the segments describe.** Table 6 lists the segments with their weekly means. The longest runs from October 1999 to April 2004 (236 weeks), during the ceasefire that followed Öcalan's call, and the quietest runs from December 2012 to July 2015. Post hoc, over the whole record, rises and falls alternate in 29 of 32 consecutive pairs of change points, 13 of the 17 rises occur in March, April or May, and 14 of the 16 falls in October, November or December (Figure 9); four of the five change points that recover a documented turning point fall in these months. The rotations of the calendar, which keep the spacing and the alternation of the change points, give p= 0.044 for the rises and 0.071 for the falls (q= 0.172 and 0.172), with months chosen after seeing the change points.

The two documented ceasefires overlap the two long segments of lowest mean, which span 7 springs; 13 of the other 21 springs have a rise. Also post hoc, the PKK's first unilateral ceasefire, of March 1993, falls in a segment of 2.25 events a week that ends with a rise to 10.88 on 30 May 1993, after the killing of unarmed soldiers in Bingöl and the end of the ceasefire (Ünal 2016). The segment from 19 October 2008 to 2 July 2011, which contains the further ceasefires of April 2009 and August 2010 (Ünal 2016), has no change point at either date. Along the exact solution path for c from 1 to 4 (25 solver calls, 16 distinct optimal segmentations, K from 44 to 16), S stays at 5 down to K= 28, drops to four between K= 24 and 22, returns to five at K= 20 and 17 (p= 0.024 and 0.020, q= 0.168 and 0.168) and ends at four (p= 0.088).

**Table 6. The optimal periodisation: start and end of each segment of the exact segmentation, its length and its mean number of events per week with a block-bootstrap interval.**

| From | To | Weeks | Mean | 95% intervala |
|---|---|---|---|---|
| 1 Jan 1989 | 25 May 1991 | 125 | 0.84 | 0.64–1.08 |
| 26 May 1991 | 14 Mar 1992 | 42 | 2.19 | 1.64–2.79 |
| 15 Mar 1992 | 29 Aug 1992 | 24 | 5.04 | 4.08–6.04 |
| 30 Aug 1992 | 17 Oct 1992 | 7 | 10.43 | — |
| 18 Oct 1992 | 29 May 1993 | 32 | 2.25 | 1.69–2.78 |
| 30 May 1993 | 27 Nov 1993 | 26 | 10.88 | 9.62–12.54 |
| 28 Nov 1993 | 2 Apr 1994 | 18 | 6.44 | 4.89–8.28 |
| 3 Apr 1994 | 19 Nov 1994 | 33 | 12.67 | 11.36–13.85 |
| 20 Nov 1994 | 25 Mar 1995 | 18 | 4.00 | 3.17–4.94 |
| 26 Mar 1995 | 4 Nov 1995 | 32 | 12.78 | 11.38–14.22 |
| 5 Nov 1995 | 30 Mar 1996 | 21 | 2.67 | 1.81–3.62 |
| 31 Mar 1996 | 7 Dec 1996 | 36 | 7.81 | 7.00–8.78 |
| 8 Dec 1996 | 29 Mar 1997 | 16 | 1.63 | 0.88–2.31 |
| 30 Mar 1997 | 13 Dec 1997 | 37 | 6.11 | 5.19–6.89 |
| 14 Dec 1997 | 18 Apr 1998 | 18 | 2.56 | 1.83–3.61 |
| 19 Apr 1998 | 11 Jul 1998 | 12 | 8.42 | 7.67–8.92 |
| 12 Jul 1998 | 17 Apr 1999 | 40 | 3.43 | 2.60–4.40 |
| 18 Apr 1999 | 2 Oct 1999 | 24 | 7.21 | 5.79–8.54 |
| 3 Oct 1999 | 10 Apr 2004 | 236 | 0.44 | 0.34–0.54 |
| 11 Apr 2004 | 6 Nov 2004 | 30 | 3.30 | 2.43–4.13 |
| 7 Nov 2004 | 19 Mar 2005 | 19 | 0.68 | 0.32–1.05 |
| 20 Mar 2005 | 15 Oct 2005 | 30 | 4.83 | 3.30–6.57 |
| 16 Oct 2005 | 31 Mar 2007 | 76 | 1.41 | 0.96–1.92 |
| 1 Apr 2007 | 10 Nov 2007 | 32 | 3.44 | 2.81–3.88 |
| 11 Nov 2007 | 19 Apr 2008 | 23 | 0.83 | 0.48–1.30 |
| 20 Apr 2008 | 18 Oct 2008 | 26 | 3.50 | 2.73–4.04 |
| 19 Oct 2008 | 2 Jul 2011 | 141 | 0.97 | 0.67–1.33 |

| From | To | Weeks | Mean | 95% intervala |
|---|---|---|---|---|
| 3 Jul 2011 | 29 Oct 2011 | 17 | 4.76 | 3.82–6.12 |
| 30 Oct 2011 | 21 Jul 2012 | 38 | 1.42 | 0.97–1.97 |
| 22 Jul 2012 | 1 Dec 2012 | 19 | 4.47 | 3.47–5.79 |
| 2 Dec 2012 | 18 Jul 2015 | 137 | 0.14 | 0.04–0.25 |
| 19 Jul 2015 | 9 Jan 2016 | 25 | 9.48 | 7.36–11.32 |
| 10 Jan 2016 | 5 Nov 2016 | 43 | 4.09 | 3.35–4.74 |
| 6 Nov 2016 | 31 Dec 2016 | 8 | 0.75 | — |

Weeks run from Sunday to Saturday; months abbreviated.
a Moving-block bootstrap of the weekly counts within the segment (blocks of four weeks), conditional on the estimated boundaries; none for segments shorter than three blocks.

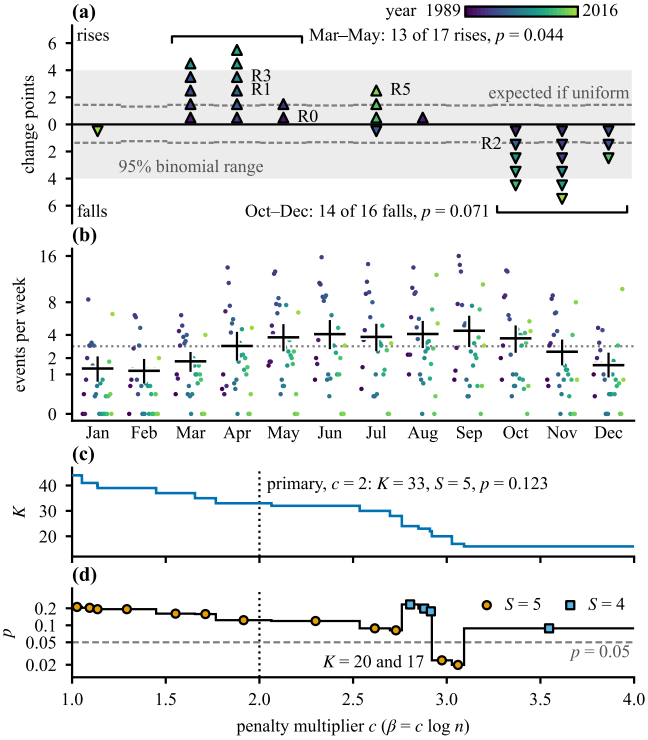



**Figure 9. Seasonal rhythm of the change points (post hoc) and exact solution path (exploratory).**

(a) Calendar month of the 33 change points of the primary segmentation, one triangle each, rises above the line and falls below it, stacked in order of year and coloured by year (scale at top, shared with b); grey band, central 95% range of a binomial count of the 17 rises or 16 falls if every week were equally likely, with its expected count dashed (a descriptive reference, not a test); brackets, months chosen after seeing the change points, with the p-value over all 1,461 rotations of the calendar; R0–R5, change points that recover a documented turning point (R4 is not recovered). (b) Mean events per week in each month of each year from 1989 to 2016 (dots, in order of year within each month) and the mean over the years with its 95% percentile bootstrap interval (crosses; resampling the years with the seed of the analysis); dotted, the mean of all weeks; square-root scale. (c) Number of change points K of each of the 16 optimal segmentations of the exact solution path, over the interval of the penalty multiplier c on which it is optimal. (d) Rotation p-value of each, with the number S of recovered turning points as markers; dashed, p=0.05; dotted, the primary specification.
Exact baselines. The LP relaxation and the ILP of Equation (8), solved with the branch-and-cut solver CBC in one thread, return the dynamic-programming optimum and its change points on every prefix, with no fractional variable and 0 branch-and-bound nodes (Table 7). On the full series the integer program takes 184 s, plus 8.7 s to build it, against 0.017 s for dynamic programming: a cross-check, not a practical method.

**Table 7. Exact baselines on prefixes of the real series: number of change points, number of arcs of the flow program, time of OP, of the LP relaxation and of the ILP, and agreement of the three optima.**

| n | K | Arcs | OP (s) | LP (s) | ILP (s) | Samea |
|---|---|---|---|---|---|---|
| 50 | 0 | 867 | 0.0004 | 0.02 | 0.02 | yes |
| 100 | 0 | 4,192 | 0.0008 | 0.04 | 0.08 | yes |
| 200 | 2 | 18,342 | 0.0018 | 0.16 | 0.53 | yes |
| 400 | 11 | 76,642 | 0.0044 | 0.62 | 3.96 | yes |
| 800 | 18 | 313,242 | 0.0083 | 3.07 | 31.61 | yes |
| 1,461 | 33 | 1,054,884 | 0.0172 | 13.29 | 184.11 | yes |

Times of one run; CBC with its default settings in one thread, without building the model.
a The relaxation is integral and both programs return the optimum and the change points of OP, with objectives within $3 \times 10^{-13}$ of it.

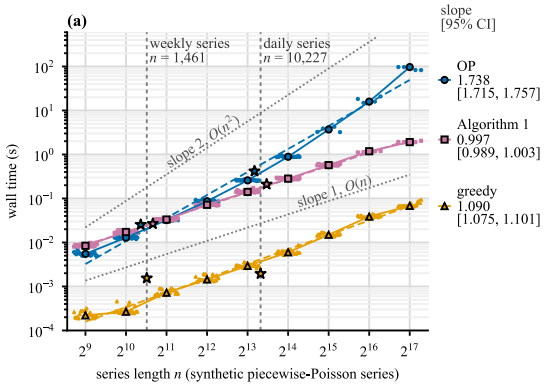
**Measured cost.** Pruning removes 92.1% of the candidate evaluations of OP on the weekly series (83,724 against 1,059,243, with at most 249 live candidates; Figure 10), yet Algorithm 1 is no faster than OP there (0.0267 s against 0.0254 s; Table 8): each of its steps costs 14.5 to 18.1 µs of interpreted work, which the saved arithmetic does not repay at this length.
On the daily series (10,227 days, 36 change points, scale from the standard deviation of the differences) Algorithm 1 takes 0.210 s against 0.419 s, a ratio of 1.99 (bootstrap interval 1.93–2.09). On synthetic series of 29 to 217 points the log-log slopes of Equation (26) are 0.997 for Algorithm 1, 1.090 for the greedy method and 1.738 for OP (Figure 10a). Counted in evaluations, which do not depend on the machine, the slopes of Algorithm 1 and OP are 1.10 and 2.00: the time slope of Algorithm 1 is near one probably because the fixed cost of each step dominates. The local slope of OP rises from 1.23 to 2.61, above two at the largest sizes, where the time per evaluation grows from 6.89 ns at 215 points to 11.29 ns at 217, attributed here to the machine rather than to the algorithm; that of Algorithm 1 stays between 0.69 and 1.12. The ratio of the median times of OP and Algorithm 1 grows from 0.65 at 29 points to 50.71 at 217 (interval 42.92–51.14).

$$\log_{10} T(n) = \alpha + \gamma \log_{10} n + \varepsilon \quad (26)$$

**Table 8. Measured cost of each pipeline stage on the real weekly and daily series: median wall time, interquartile range (IQR) and peak traced memory.**

| Stage | Median (s) | IQR (s) | Traced (MiB)a |
|---|---|---|---|
| Read archive and filter dyad | 0.8118 | 0.7914–0.8360 | 30.39 |
| Build weekly and daily series | 0.0289 | 0.0286–0.0303 | 1.76 |
| OP, weekly | 0.0254 | 0.0244–0.0258 | 0.15 |
| Algorithm 1, weekly | 0.0267 | 0.0252–0.0277 | 0.10 |
| Greedy segmentation, weekly | 0.0015 | 0.0014–0.0018 | 0.11 |
| Circular-shift null | 0.0004 | 0.0004–0.0004 | 0.75 |
| OP, daily | 0.4189 | 0.3993–0.4479 | 1.00 |
| Algorithm 1, daily | 0.2105 | 0.2075–0.2125 | 0.71 |
| Greedy segmentation, daily | 0.0019 | 0.0019–0.0022 | 0.77 |

Median and IQR over 21 repetitions after one warm-up run, one thread, high-performance power plan.
a Peak of the memory traced by tracemalloc, which includes NumPy buffers, in a separate run; resident memory was not measured.



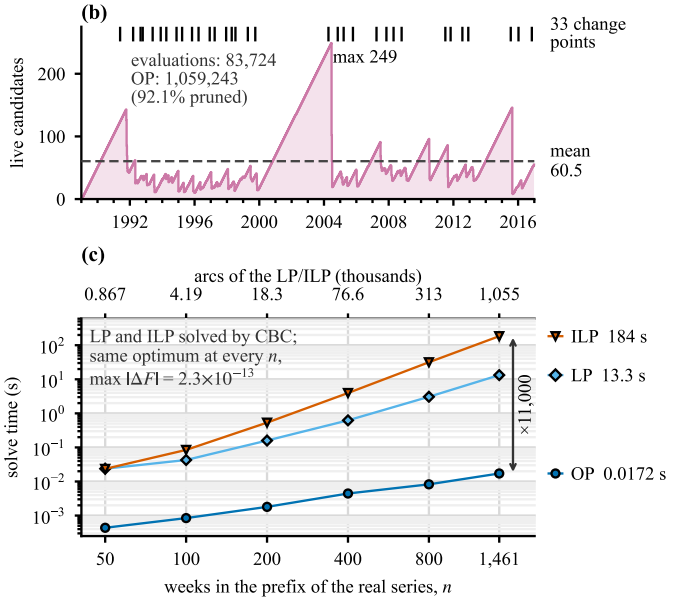



**Figure 10. Measured computational cost.**

(a) Wall time on synthetic piecewise-Poisson series, one series per size, with geometric segment lengths of mean 100 and segment means uniform between 0.5 and 10. Small symbols are single runs (21 per size, fewer at the largest sizes, where one run is slow), spread horizontally around their size in the order in which they were run; large symbols are medians, joined by solid lines, with interquartile ranges as vertical bars, mostly hidden by the symbol. Dashed lines are the least-squares fits of log median time on log n; the slope and its 95% interval are given at the right. Grey dotted lines are guides of slope one, O(n), and two, O(n2). Stars are medians on the real weekly (n= 1,461) and daily (n= 10,227) series, with OP and Algorithm 1 offset horizontally for legibility. (b) Live candidates of Algorithm 1 after each week of the real series (shaded), with the 33 optimal change points as ticks along the top and their mean as a dashed line: the set grows by one candidate per quiet week and collapses within weeks of strong changes; Algorithm 1 evaluated 83,724 candidate segments against 1,059,243 for OP (92.1% pruned). (c) Exact baselines on prefixes of the real series: OP against the LP relaxation and the integer program (ILP), both solved by CBC, one run each; the top axis gives the arcs of the LP network. All three reach the same optimum (largest difference $2.3 \times 10^{-13}$) and the LP optimum is integral at every n; the arrow gives the ratio of ILP to OP time on the full series.

## 7 Discussion

For the history of the conflict. Five of the six documented turning points lie near a change point of the exact optimum, and the two segments of lowest mean overlap the two ceasefires. The pre-specified test does not distinguish this agreement from chance, nor does it for most segmentations drawn from the exact posterior unless the annual cycle is subtracted, and the post hoc simulations show that it would rarely have done so even against large level changes at the documented dates, so its null says little about the periodisation; in the same simulations, power is lower with the monthly factor than without it. Post hoc, most change points mark a rise in March–May or a fall in October–December, and so do four of the five that recover a turning point (months chosen after seeing the change points). Casualties of PKK attacks peak in summer and decline in winter (Ciftci and Kula 2015), and warm weather has been associated with combat fatalities in Afghanistan (Carter and Veale 2013) and with the onset of civil and non-state conflict (Landis 2014). Season and phase may not be separable if the parties time their decisions to the fighting season. Once a common annual cycle is modelled or removed, fewer change points remain and the alignment reaches nominal significance (p= 0.010 and 0.006), but no post hoc result survives the correction for multiple testing (the smallest q is 0.114; Figure 6b). These results are compatible with level changes near the documented turning points once a within-year cycle is allowed for, and with a ceasefire showing as a delayed or missing seasonal rise, as in 1993 and 2009–2010 (post hoc); both are hypotheses for a pre-specified test with a seasonal model, not findings.

The change point that closes the active segment of July–November 2012 falls in the week of 2 December 2012. The Crisis Group report on the peace process dates to that month the announcement that the intelligence service was in direct contact with Öcalan (International Crisis Group 2014), and a fall between October and December is the most common pattern in this record (post hoc, months chosen after seeing the data), so the data cannot tell whether the break anticipates the ceasefire, coincides with the announcement or follows the season. The break of July 2015 falls in the month of the collapse reported at the time (Göksel 2015), and the segment that follows has the highest mean since 1995, although its interval overlaps those of the segments of 1996, 1998 and 1999.

For the engineering of such studies. Exact segmentation of a conflict record is cheap at this scale, even without pruning: OP takes hundredths of a second for 1,461 weeks, and Algorithm 1 a fraction of a second for 10,227 days, where pruning pays (a ratio of 1.99), consistent with the linear expected cost of PELT when changes keep occurring (Killick, Fearnhead, and Eckley 2012). The choice of solver is nevertheless not innocuous for inference: a greedy segmentation would have produced a valid positive under the rule of the plan, and the two solvers disagree about significance in 25 of the 81 specifications of the multiverse (both post hoc). Exactness does not make a segmentation true, but it removes a source of variation that the reader

cannot see, and it allows questions that a heuristic cannot answer, such as the exact price of a given alignment (post hoc here). Compiled implementations and functional pruning (Maidstone et al. 2017) were not compared.

## 8 Threats to validity

**Construct validity.** The series counts lethal events of the government–PKK dyad as GED records them (Sundberg and Melander 2013); it leaves out one-sided violence and non-state fighting, and lethal events are only one observable of a phase. The documented turning points are those that three Crisis Group texts date, at the precision those texts give, a year for two of them; they are one organisation's periodisation, not the historiography of the conflict, and two of them, the end in 2004 of the ceasefire declared in 1999 and the collapse of 2015, are partly defined by a change in violence. An academic account dates the phases differently (Ünal 2016). The coding matters: dating R4 to December 2012 (post hoc) gives $p = 0.036$ ($q = 0.168$).

**Internal validity.** Reporting rates vary with local communication coverage (Weidmann 2016), and reporting in the region may have changed abruptly, for instance under curfews or press closures; abrupt changes of reporting would create change points of their own. GED names the original source of 40% of the events of 1989–1999 and 92% of those of 2000–2016. By keywords (post hoc), 90% of the events that name one cite a state source, while sources of the PKK, its armed wing or Kurdish media appear in 0.8% of the events before 2004 but 13% of those of 2004 and 36% of those of 2005, all of the latter without a state source; the segments after R3 may partly reflect these sources (Figure 4c). Most deaths of the events without a week (73%) fall in 1996–1997. The square-root transform does not stabilise overdispersed counts, and in quiet segments the variance of the transformed counts is well below one, so a single scale over-penalises changes in quiet periods and under-penalises them in active ones. Serial dependence within segments can inflate the number of change points, and an annual cycle can force two changes a year on a piecewise-constant model; the residuals still carry part of the annual cycle (post hoc diagnostics, Figure 4), and their spread differs between segments. The bootstrap intervals are conditional on the estimated boundaries.

**External validity.** No other public record known to the author covers lethal events of both sides of this dyad below yearly resolution: the Global Terrorism Database, from which monthly counts of PKK attacks have been built (Ciftci and Kula 2015), selects incidents by a definition of terrorism and leaves out violence by the state. The replication on an independent record is left open, and so is the period after 2016.

Conclusion validity. The plan recorded that the resolution of the test depends on the number of change points but made no power analysis; the post hoc simulation gives a rejection rate of at most 28.8%, and of at most 55.8% without the annual cycle. The null result therefore does not show that the documented periodisation is wrong. It shows that this test, on the event record segmented by this criterion, cannot distinguish the periodisation from chance. Table 9 summarises the threats.

## 9 Reproducibility

Every number in this article is inserted automatically from a result file whose provenance records the commit, the tag of the frozen plan, the checksum of the data and the seed; all result files come from one run on a clean commit, and the build fails if a number has no such source. Two complete runs give identical result files apart from timestamps; the timings are measured once, and a clean clone of the repository downloads the data, reruns the rest of the analysis and rebuilds this document with identical results, numbers and text. Tests and blocking quality gates on dates, figures, equations, references and arithmetic run with every build.

**Table 9. Threats to validity of the study and what was done to mitigate each one.**

| Validity | Threat | Mitigation |
|---|---|---|
| Construct | Lethal events of one dyad, from media-based sources | Deaths and one-sided violence as sensitivity series |
| Construct | Turning points from one organisation, some at year precision | Precision kept; null hit rates shown; recoding as post hoc check |
| Internal | Reporting may change abruptly | Composition of named sources by year, by keywords (post hoc); not mitigated |
| Internal | Seasonality, overdispersion and dependence can inflate the number of segments | Penalty multiverse; seasonal adjustment as post hoc check |
| External | One conflict, one event dataset | No claim beyond this record; replication left open |
| Conclusion | Only a complete recovery could be significant | Resolution recorded in the plan; null reported as such |

## 10 Conclusion

Treating periodisation as an exact optimisation problem allowed the documented periodisation of the Turkey–PKK conflict to be tested on the UCDP record. The exact optimum recovers five of the six documented turning points, but the pre-specified test cannot distinguish this from chance, and post hoc simulations show that it had little power, less with the seasonal cycle of the record than without it. Post hoc, modelling or removing that cycle, recoding one turning point or changing the solver gives nominal $p \leq 0.05$, but none of these results survives the correction for 37 tests; they motivate a confirmatory test rather than replace the null. Exact segmentation costs milliseconds at this scale, can be certified with one pass over its arcs and can be constrained to any set of turning points (both post hoc here), so the choice of a heuristic should be justified rather than assumed. A pre-specified test with a seasonal model, which in the simulations keeps its size but still detects at most 41.0% of the simulated alignments, a complete chronology of declared ceasefires and an independent event record are the next steps.

**Funding**

This research received no external funding.

**Conflict of interest**

The author declares that there is no conflict of interest.

**Ethics statement**

The study analyses publicly available records of lethal events that contain no personal data. It involved no human participants and required no ethics approval.

**Data availability**

UCDP GED version 17.1 is freely available from the UCDP download page (https://ucdp.uu.se/downloads/), with the citations that UCDP requests.

**Code availability**

The analysis code, the frozen analysis plan, the tests and the build of this document are kept in a version-controlled repository that will be made public on publication; until then it is available from the author on request.

## References

1. Bellman, Richard. 1958. "On a routing problem." Quarterly of Applied Mathematics 16 (1): 87–90. https://doi.org/10.1090/qam/102435.
2. Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." Journal of the Royal Statistical Society. Series B (Methodological) 57 (1): 289–300.
3. Benjamini, Yoav, and Daniel Yekutieli. 2001. "The control of the false discovery rate in multiple testing under dependency." The Annals of Statistics 29 (4): 1165–1188. https://doi.org/10.1214/aos/1013699998.
4. Bozarslan, Hamit. 2014. "The Kurds and Middle Eastern 'State of Violence': the 1980s and 2010s." Kurdish Studies 2 (1): 4–13.
5. Carter, Timothy Allen, and Daniel Jay Veale. 2013. "Weather, terrain and warfare: Coalition fatalities in Afghanistan." Conflict Management and Peace Science 30 (3): 220–239. https://doi.org/10.1177/0738894213484054.
6. Ciftci, Irfan, and Sedat Kula. 2015. "The Evaluation of the Effectiveness of Counterterrorism Policies on the PKK-inflicted Violence during the Democratization Process of Turkey." Journal of Terrorism Research 6 (1): 27–42. https://doi.org/10.15664/jtr.1075.
7. Cleynen, Alice, and Emilie Lebarbier. 2014. "Segmentation of the Poisson and negative binomial rate models: a penalized estimator." ESAIM: Probability and Statistics 18: 750–769. https://doi.org/10.1051/ps/2014005.
8. Croicu, Mihai, and Ralph Sundberg. 2017. UCDP Georeferenced Event Dataset Codebook. Codebook Version 17.1. Department of Peace and Conflict Research, Uppsala University. https://ucdp.uu.se/downloads/ged/ged171.pdf.
9. Eck, Kristine. 2012. "In data we trust? A comparison of UCDP GED and ACLED conflict events datasets." Cooperation and Conflict 47 (1): 124–141. https://doi.org/10.1177/0010836711434463.
10. Fearnhead, Paul. 2006. "Exact and efficient Bayesian inference for multiple changepoint problems." Statistics and Computing 16 (2): 203–213. https://doi.org/10.1007/s11222-006-8450-8.
11. Fryzlewicz, Piotr. 2014. "Wild binary segmentation for multiple change-point detection." The Annals of Statistics 42 (6): 2243–2281. https://doi.org/10.1214/14-aos1245.
12. Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg, and Håvard Strand. 2002. "Armed Conflict 1946-2001: A New Dataset." Journal of Peace Research 39 (5): 615–637. https://doi.org/10.1177/0022343302039005007.
13. Göksel, Nigar. 2015. "A new cycle in Turkey-PKK conflict." International Crisis Group analysis republished by Daily Maverick, 12 August. https://www.dailymaverick.co.za/article/2015-08-12-international-crisis-group-a-new-cycle-in-turkey-pkk-conflict/.
14. Hall, Peter, J. W. Kay, and D. M. Titterington. 1990. "Asymptotically optimal difference-based estimation of variance in nonparametric regression." Biometrika 77 (3): 521–528. https://doi.org/10.1093/biomet/77.3.521.
15. Haynes, Kaylea, Idris A. Eckley, and Paul Fearnhead. 2017. "Computationally Efficient Changepoint Detection for a Range of Penalties." Journal of Computational and Graphical Statistics 26 (1): 134–143. https://doi.org/10.1080/10618600.2015.1116445.
16. International Crisis Group. 2011. Turkey: Ending the PKK Insurgency. Europe Report 213. International Crisis Group.

17. International Crisis Group. 2014. Turkey and the PKK: Saving the Peace Process. Europe Report 234. International Crisis Group.
18. Jackson, B., J. D. Scargle, D. Barnes, S. Arabhi, A. Alt, P. Gioumousis, E. Gwin, P. Sangtrakulcharoen, L. Tan, and T. T. Tsai. 2005. "An algorithm for optimal partitioning of data on an interval." IEEE Signal Processing Letters 12 (2): 105–108. https://doi.org/10.1109/lsp.2001.838216.
19. Kibris, Arzu. 2011. "Funerals and Elections: The Effects of Terrorism on Voting Behavior in Turkey." Journal of Conflict Resolution 55 (2): 220–247. https://doi.org/10.1177/0022002710383664.
20. Killick, R., P. Fearnhead, and I. A. Eckley. 2012. "Optimal Detection of Changepoints With a Linear Computational Cost." Journal of the American Statistical Association 107 (500): 1590–1598. https://doi.org/10.1080/01621459.2012.737745.
21. Killick, Rebecca, and Idris A. Eckley. 2014. "changepoint: An R Package for Changepoint Analysis." Journal of Statistical Software 58 (3). https://doi.org/10.18637/jss.v058.i03.
22. Künsch, Hans R. 1989. "The Jackknife and the Bootstrap for General Stationary Observations." The Annals of Statistics 17 (3): 1217–1241. https://doi.org/10.1214/aos/1176347265.
23. Landis, Steven T. 2014. "Temperature seasonality and violent conflict." Journal of Peace Research 51 (5): 603–618. https://doi.org/10.1177/0022343314538275.
24. Maidstone, Robert, Toby Hocking, Guillem Rigaill, and Paul Fearnhead. 2017. "On optimal multiple changepoint algorithms for large data." Statistics and Computing 27 (2): 519–533. https://doi.org/10.1007/s11222-016-9636-3.
25. McKinney, Wes. 2010. "Data Structures for Statistical Computing in Python." In Proceedings of the 9th Python in Science Conference, 56–61. SciPy.
26. Murtagh, Fionn, Michael Spagat, and Jorge A. Restrepo. 2011. "Ultrametric Wavelet Regression of Multivariate Time Series: Application to Colombian Conflict Analysis." IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 41 (2): 254–263. https://doi.org/10.1109/tsmca.2010.2064301.
27. Phipson, Belinda, and Gordon K. Smyth. 2010. "Permutation P-values Should Never Be Zero: Calculating Exact P-values When Permutations Are Randomly Drawn." Statistical Applications in Genetics and Molecular Biology 9 (1). https://doi.org/10.2202/1544-6115.1585.
28. Sandve, Geir Kjetil, Anton Nekrutenko, James Taylor, and Eivind Hovig. 2013. "Ten Simple Rules for Reproducible Computational Research." PLoS Computational Biology 9 (10): e1003285. https://doi.org/10.1371/journal.pcbi.1003285.
29. Schrodt, Philip A., and Deborah J. Gerner. 1997. "Empirical Indicators of Crisis Phase in the Middle East, 1979-1995." Journal of Conflict Resolution 41 (4): 529–552. https://doi.org/10.1177/0022002797041004003.
30. Spirling, Arthur. 2007. "'Turning Points' in the Iraq Conflict." The American Statistician 61 (4): 315–320. https://doi.org/10.1198/000313007x247076.
31. Steegen, Sara, Francis Tuerlinckx, Andrew Gelman, and Wolf Vanpaemel. 2016. "Increasing Transparency Through a Multiverse Analysis." Perspectives on Psychological Science 11 (5): 702–712. https://doi.org/10.1177/1745691616658637.
32. Sundberg, Ralph, and Erik Melander. 2013. "Introducing the UCDP Georeferenced Event Dataset." Journal of Peace Research 50 (4): 523–532. https://doi.org/10.1177/0022343313484347.
33. Tezcür, Güneş Murat. 2010. "When democratization radicalizes: The Kurdish nationalist movement in Turkey." Journal of Peace Research 47 (6): 775–789. https://doi.org/10.1177/0022343310386156.
34. Tezcür, Güneş Murat. 2016. "Ordinary People, Extraordinary Risks: Participation in an Ethnic Rebellion." American Political Science Review 110 (2): 247–264. https://doi.org/10.1017/s0003055416000150.
35. Ünal, Mustafa Coşar. 2014. "Strategist or Pragmatist: A Challenging Look at Ocalan's Retrospective Classification and Definition of PKK's Strategic Periods Between 1973 and 2012." Terrorism and Political Violence 26 (3): 419–448. https://doi.org/10.1080/09546553.2012.728153.
36. Ünal, Mustafa Coşar. 2016. "Is it ripe yet? Resolving Turkey's 30 years of conflict with the PKK." Turkish Studies 17 (1): 91–125. https://doi.org/10.1080/14683849.2015.1124020.
37. van der Walt, Stefan, S. Chris Colbert, and Gael Varoquaux. 2011. "The NumPy Array: A Structure for Efficient Numerical Computation." Computing in Science & Engineering 13 (2): 22–30. https://doi.org/10.1109/mcse.2011.37.
38. Weidmann, Nils B. 2016. "A Closer Look at Reporting Bias in Conflict Event Data." American Journal of Political Science 60 (1): 206–218. https://doi.org/10.1111/ajps.12196.
39. Wilson, Greg, Jennifer Bryan, Karen Cranston, Justin Kitzes, Lex Nederbragt, and Tracy K. Teal. 2017. "Good enough practices in scientific computing." PLOS Computational Biology 13 (6): e1005510. https://doi.org/10.1371/journal.pcbi.1005510.
40. Zeyneloglu, Sinan, Ibrahim Sirkeci, and Yaprak Civelek. 2016. "Language shift among Kurds in Turkey: A spatial and demographic analysis." Kurdish Studies 4 (1): 25–50.
41. Zhang, Nancy R., and David O. Siegmund. 2007. "A Modified Bayes Information Criterion with Applications to the Analysis of Comparative Genomic Hybridization Data." Biometrics 63 (1): 22–32. https://doi.org/10.1111/j.1541-0420.2006.00662.x.