Hainzl, S. (2016): Rate-Dependent Incompleteness of Earthquake Catalogs. - *Seismological Research Letters*, *87*, 2A, pp. 337–344.

DOI: http://doi.org/10.1785/0220150211

THE SEISMOLOGICAL SOCIETY OF AMERICA
400 Evelyn Ave., Suite 201
Albany, CA 94706-1375
(510) 525-5474; FAX (510) 525-7204
www.seismosoc.org

Rate-Dependent Incompleteness of Earthquake Catalogs

by Sebastian Hainzl

ABSTRACT

Important information about the earthquake generation process can be gained from instrumental earthquake catalogs, but this requires complete recordings to avoid biased results. The local completeness magnitude M_c is known to depend on general conditions such as the seismographic network and the environmental noise, which generally limit the possibility of detecting small events. The detectability can be additionally reduced by an earthquake-induced increase of the noise level leading to short-term variations of M_c, which cannot be resolved by traditional methods relying on the analysis of the frequency–magnitude distribution. Based on simple assumptions, I propose a new method to estimate such temporal excursions of M_c solely based on the estimation of the earthquake rate resulting in a high temporal resolution of M_c. The approach is shown to be in agreement with the apparent decrease of the estimated Gutenberg–Richter *b*-value in high-activity phases of recorded data sets and the observed incompleteness periods after mainshocks. Furthermore, an algorithm to estimate temporal changes of M_c is introduced and applied to empirical aftershock and swarm sequences from California and central Europe, indicating that observed *b*-value fluctuations are often related to rate-dependent incompleteness of the earthquake catalogs.

Online Material: MATLAB code to compute rate-dependent completeness magnitude, and figures showing variation in the estimated M_c values in equalized time scale.

INTRODUCTION

A detailed analysis of instrumental earthquake catalogs is fundamental for a systematic characterization and deeper understanding of earthquake nucleation and interaction on different temporal and spatial scales. However, partial incompleteness of the data can significantly bias the results and their interpretations. The completeness magnitude M_c defines the magnitude threshold, above which an occurring earthquake is almost certainly reported in the corresponding catalog. The local M_c-value depends on many general factors, such as the number and quality of the sensors and their geometrical configuration, as well as the environmental noise level. These factors limit the detectability of earthquakes at any time and define the absolute minimum of M_c. However, the detectability is additionally reduced by an earthquake-related increase of the noise level, which further limits the possibility of detecting small events. This can result in short-term variations of M_c, which, if ignored, can lead to spurious results and misinterpretations (Kagan, 2004; Hainzl, 2013). Thus, a temporal resolution of M_c is necessary to get unbiased parameter estimations of seismicity models such as the epidemic-type aftershock sequence (ETAS) model (Ogata, 1988; Zhuang *et al.*, 2012) and to evaluate variations of statistical parameters such as the *b*-value (e.g., Imoto, 1991; Hainzl and Fischer, 2002; Mallika *et al.*, 2013; Schurr *et al.*, 2014; Huang and Beroza, 2015) or Omori *c*-value (e.g., Shcherbakov *et al.*, 2004; Narteau *et al.*, 2009; Davidsen *et al.*, 2015). Both are important for physical interpretations and seismic-hazard estimations.

The catalog completeness can be visually analyzed by appropriate data plotting (Agnew, 2015). For a quantitative estimation of the local M_c-value, a number of statistical techniques have been developed, which can be categorized into two groups: (1) estimations solely based on catalog information (see Mignan and Woessner, 2012, and references therein) and (2) methods based also on additional information, whether measurements of noise levels (Peng *et al.*, 2007) or picking information of seismic waves (Schorlemmer and Woessner, 2008). While the second approach requires additional detailed informations such as phase data, stations, and network attenuation relations, the first approach is easy to apply and is based on the analysis of the frequency–magnitude distribution, for which most methods assume the validity of the Gutenberg–Richter relation. These methods include the goodness-of-fit test (Wiemer and Wyss, 2000), tests of the *b*-value stability (Cao and Gao, 2002), an entire-magnitude-range approach (Ogata and Katsura, 1993; Woessner and Wiemer, 2005), the maximum-curvature technique (Wiemer and Wyss, 2000), and median-based analysis of the segment slope (Amorèse, 2007). They all require a large number of earthquakes (usually a few hundred) to get stable estimates, which limits its spatial and temporal resolution.

However, M_c can vary on very short temporal scales, as observed directly after the occurrence of major earthquakes (Kagan, 2004), for which empirical analyses show that the duration of this incompleteness is largest for the smallest magni-

doi: 10.1785/0220150211

tudes that can be recorded by the seismic network and is shortest for the largest observed events. For the $M \geq 6$ mainshocks in southern California that have occurred since 1985, Helmstetter *et al.* (2006) found that the completeness threshold varies as a function of the mainshock magnitude M and the time t (in days) after an earthquake according to

$$M_c(M, t) = M - 4.5 - 0.75 \log(t) \quad (1)$$

and $M_c(M, t) \geq 2$. The limited detectability after mainshocks might be partially attributed to the long duration of the mainshock coda wave, but it is clear that, without this direct mainshock effect, events cannot be correctly picked if the activity level of smaller magnitude earthquakes is sufficiently high that their seismic waveforms start to overlap. This leads to a maximum earthquake rate that still can be completely recorded, as derived in the Theoretical Approach section in more detail. In agreement with this theory, empirical catalogs are found to show an apparent b-value decrease for high seismicity rates, indicating incomplete recordings (see Rate-Dependent Apparent b-Value Decrease section). Furthermore, the empirical equation (1) can be derived based on this approach, and b-values of events above the estimated M_c-level are found to be almost stable during aftershock sequences. The method has the advantage that it is equally applicable to swarm activity that is not associated to any mainshock. I show an example in the Incompleteness during Earthquake Swarms section, demonstrating that the completeness magnitude also increases significantly within high-activity phases in this case and that some spurious b-value fluctuations disappear if those variations are considered.

THEORETICAL APPROACH

Seismicity is usually low in most time periods, and earthquakes can be recorded if magnitudes exceed the basic detection threshold, which is related to the general setup of the monitoring system and to the background noise level. The estimation of this basic completeness magnitude M_{c_0} is not a subject of this article and has to be estimated by traditional methods. My approach is only concerned with short-term increases of M_c above this level in periods of high seismic activity, in which numerous small events lead to an increase of the noise level and thus a reduction of the signal-to-noise ratio (SNR) for picking seismic waves in seismograms. Such high seismicity rates occur, for example, during aftershock sequences triggered by earthquakes that are much larger than the basic detection threshold of the seismic network (i.e., the mainshocks). Furthermore, aseismic processes such as slow-slip events (Lohman and McGuire, 2007) and rapid magma (Toda *et al.*, 2002) or fluid intrusion (Miller *et al.*, 2004) can also cause earthquake swarms with high temporal earthquake clustering. Those seismicity phenomena are usually not only temporally, but also spatially, densely clustered and thus usually are recorded by the same seismic stations. In these cases, the reduced SNR of the stations leads to a reduced detection ability. For aftershock activity, temporal increases of the completeness magnitude have been already recognized after major earthquakes (Kagan, 2004; Helmstetter *et al.*, 2006). Although the coda wave of the mainshocks contributes to the SNR decrease, it cannot be the main reason of the observed incompleteness fitted by equation (1), because the durations of coda waves are much shorter than the observed incompleteness intervals. For example, the empirical observation (equation 1) indicates that $m = 2$ events are not completely recorded in the first 4.6 days after an M 7 mainshock. This would imply that, at that time, the coda wave still has an amplitude that is equal to or larger than a local $m = 2$ event in order to explain a reduced SNR, which prevents the detection of $m = 2$ events. This time interval is far larger than the expected duration of the coda wave of an M 7 event, which is on the order of 1000 s (0.01 day) (Bakun and Lindh, 1977).

Here, I conduct a simple theoretical derivation of the incompleteness level following mainshocks. The derivation is solely based on earthquake rates and ignores the magnitude dependence of the coda-wave duration. Besides explaining the observed incompleteness for aftershock sequences, this approach can also be directly applied to earthquake swarms that are not associated with any large earthquake. The basic assumptions of the approach can be summarized as follows:

1. The absolute minimum completeness magnitude M_{c_0} of the seismic network is related to the sensitivity of the recording system in time periods of low seismicity and is assumed to be already estimated by means of traditional methods (e.g., Mignan and Woessner, 2012, and references therein).
2. On short time scales, the earthquake occurrence can be approximated by a stationary Poisson process, which is in agreement with the observation that seismicity is usually well described by the superposition of stationary and nonstationary Poisson models accounting for tectonic background activity, Omori-type aftershock occurrences, and transient aseismic forcing (Langenbruch *et al.*, 2011; Zhuang *et al.*, 2012).
3. The earthquakes are registered by the same stations.
4. The rate of earthquakes with magnitudes $m \geq M_{c_0}$ is given by r.
5. An earthquake cannot be properly distinguished by the seismogram analysis if it occurs less than Δt after the last event of equal or larger magnitude.
6. An earthquake catalog is defined as complete for $m \geq M_{c_0}$ if earthquakes with magnitude M_{c_0} are recorded with a probability p_c.

Based on these assumptions, the probability that an event of magnitude M_{c_0} has no $m \geq M_{c_0}$ predecessor within the precursory time interval Δt (and thus can be recorded) is

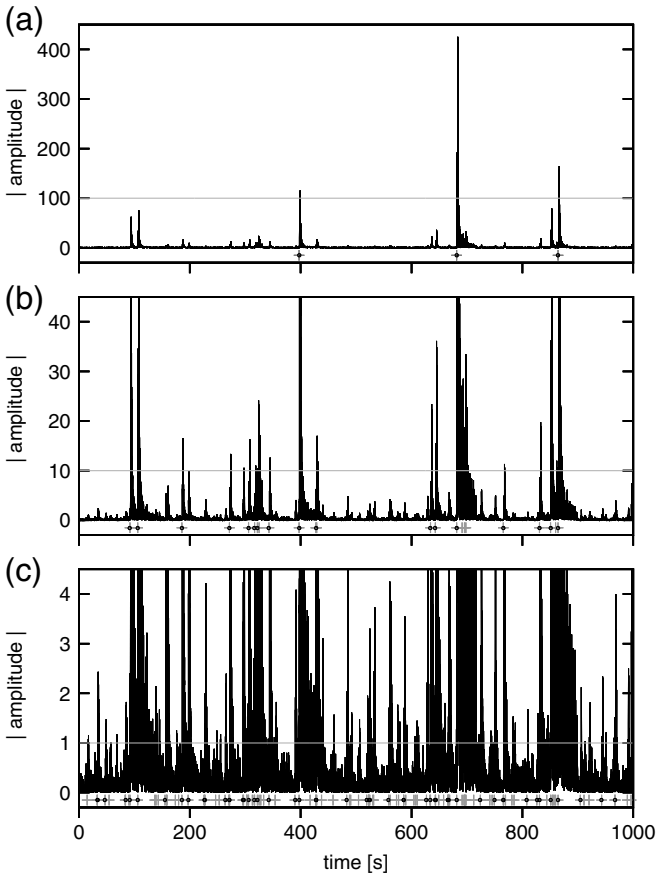
$$p_{\Delta t} = \exp(-r\Delta t), \quad (2)$$

according to the Poisson model. This leads directly to the maximum rate r_{max}, which is related to the probability p_c used for the completeness definition (see point 6),

$$r_{\max} = \frac{-\ln(p_c)}{\Delta t}. \quad (3)$$

Consequently, any time period with a seismicity rate $r \geq r_{\max}$ will be incomplete by definition. In general, Δt (and consequently also $r_{\max}$) will depend on the magnitude level M_{c_0} because of the increase of the coda duration with earthquake magnitude. Although the approach can, in principle, account for this dependence, I ignore its explicit consideration in the following analyses for the sake of simplicity. Furthermore, it should be noted that no spatial component is explicitly introduced in addition to assumption (3). Earthquake clustering on a short time scale is usually dominated by spatially localized earthquakes for which this assumption is fulfilled. However, applications to large spatial scales might lead to an overestimation of $M_c(t)$ in distant regions where simultaneously occurring earthquakes are recorded by independent seismic networks.

For an illustration of the rate dependence of the detection threshold, I calculated a synthetic seismogram, constructed by the superposition of a rescaled local short-period seismogram of an M 1.8 event that occurred 23 August 2011 in western Bohemia, Czech Republic. For that purpose, I performed a Monte-Carlo simulation of a Poisson process with a rate of $r = 10^6$ per day and, for each event, chose a random magnitude between $-2 < m < 4$ from the Gutenberg–Richter distribution with $b = 1$. Thus, the rate of $m \geq 2$ events is 100 per day. The seismogram of each event (the maximum amplitude value) is scaled by its magnitude according to 10^m and superposed to the overall seismogram. Different enlarged views of this synthetic seismogram for the same exemplary time period are shown in Figure 1, in which the occurrences of events in the different magnitude ranges are marked at the bottom. Although the three $m \geq 2$ events ($r = 100$ events/day) could be quite easily picked in the seismogram (Fig. 1a), already some of the $1 \leq m < 2$ ($r = 1000$ events/day) will be missed (Fig. 1b), and a large fraction of the $0 \leq m < 1$ ($r = 10^4$ events/day) are undetectable because of overlapping waveforms (Fig. 1c). This is verified by the application of a standard phase-picking algorithm based on a short-term-average to long-term-average ratio (STA/LTA), namely the algorithm of Earle and Shearer (1994) with standard parameters for short-period seismograms. The fraction of successful detections is found to be 100% for $m \geq 2$, 71% for $1 \leq m < 2$, and 25% for $0 \leq m < 1$ events in the example in Figure 1, where detections are marked by points. Although the detection might be improved by the optimization of the STA/LTA parameters, this example suggests that $r_{\max}$ is in the range between 100 and 1000 events/day.



▲ **Figure 1.** Example of a synthetic seismogram with a rate of 10^6 $m \geq -2$ earthquakes/day, in which the magnitude is defined as $m = \log(A)$. (a) Total amplitude scale with horizontal line indicating the threshold for $m \geq 2$ events; (b) 10-times enlargement of amplitude scale with horizontal line indicating the threshold for $m \geq 1$ events; and (c) 100-times enlargement of amplitude with horizontal line indicating the threshold for $m \geq 0$ events. In each subplot, the occurrences for the corresponding magnitude events are marked on the bottom (crosses), in which a point in the middle of the cross indicates the detection by the short-term average/long-term average-algorithm of Earle and Shearer (1994).

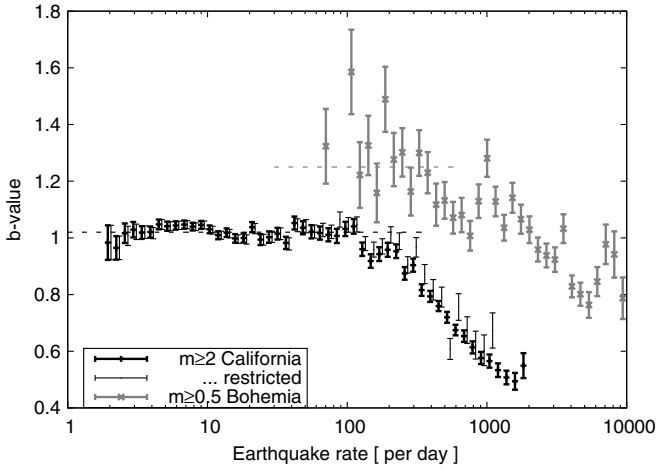
APPLICATIONS

Rate-Dependent Apparent *b*-Value Decrease

To prove this simple model approach, I analyzed real earthquake catalogs and selected magnitudes of earthquakes occurring at different activity levels. Assuming a constant frequency–magnitude distribution, incomplete recordings should express in an apparent decrease of the estimated b-value, because the lowest-magnitude events will be first missed (see Fig. 1 and Hainzl, 2013). For this study, I analyzed the following in particular:

1. the high-resolution southern California catalog (see Data and Resources) containing earthquakes from 1981 to 2011 that occurred in the region extending from Baja California in the south to Coalinga and the Owens Valley in the north (Hauksson *et al.*, 2012), with a cutoff magnitude of $M_{c_0} = 2.0$ (which ensures generally complete recordings leading to 111,980 earthquakes above this threshold); and
2. the combined catalog of the four last large earthquake swarms that occurred in 2000, 2008, 2011, and 2014 in western Bohemia (see Data and Resources), with a cutoff magnitude of 0.5, leading to 11,650 events.

For each earthquake (index i) in those catalogs, I estimate the local rate $\hat{r}_i$ at time t_i by determining the minimum and



▲ **Figure 2.** Estimated b-value as a function of the estimated local earthquake rate (calculated with $N = 10$). The results for the different data sets are slightly offset for better visibility. In the restricted catalog, incomplete times after $M \geq 6$ mainshocks are removed according to equation (1). The horizontal dashed lines refer to $b = 1.02$ and 1.25 and the error bars to ± 1 standard deviation.

maximum time of the nearest N-neighbors in time (including event i) and calculate $\hat{r}_i = (N-1)/(t_{\max} - t_{\min})$. This results in the paired information of magnitude and local rate $(m_i, \hat{r}_i)$ for each event. The magnitude of each event m_i then is sorted to the corresponding logarithmic rate bin that includes $\hat{r}_i$. Finally, I estimate the b-value by the maximum-likelihood method (Aki, 1965; Marzocchi and Sandri, 2003) for rate bins with at least 100 magnitude values according to

$$\hat{b} = \frac{1}{\ln(10)\langle m - (M_c - 0.5\Delta m)\rangle}, \tag{4}$$

in which $\langle ... \rangle$ defines the average value and Δm is the binning interval of reported magnitudes (0.01 for California and 0.1 for the swarm data) and M_c is set to M_{c0}. The corresponding standard deviation is $\sigma_b = \hat{b}/\sqrt{N}$ for N observed earthquakes.

The result is shown in Figure 2 for $N = 10$ but is found to be similar for other values of N. Although completely independent and corresponding to different tectonic regimes and seismicity characteristics, both data sets consistently show a deviation to smaller b-values for rates above ~200 events/day, whereas they show rather constant b-values for smaller rates. This result is in good agreement with the introduced rate-dependent model of incompleteness. To test whether the result for California is solely due to the already recognized incompleteness periods after mainshocks, I removed all incompleteness time intervals after $M \geq 6$ mainshocks according to equation (1). By this restriction, 4031 events are removed. The general pattern, however, is not changed for this reduced data set, and thus it is not dominated by a few major aftershock sequences.

Algorithm to Estimate M_c Based on Catalog Data

The proposed incompleteness approach allows estimation of the time-dependent completeness magnitude $M_c(t)$ for defined $r_{\max}$ by estimations of the local rate. For a given earthquake catalog with earthquake times t_i and magnitudes m_i $(i = 1, ..., Z)$, the following algorithm can be applied:

1. Define the number of neighbors N used for the estimation of the local rate and set the maximum rate $r_{\max}$.
2. Set $i = 1$.
3. Set $M_c(t_i) = M_{c_0}$.
4. Search for the nearest N neighbors in time (including event i) with $m \geq M_c(t_i)$.
5. Determine the minimum and maximum times of these N events $(t_{\min}, t_{\max})$, and calculate the local rate for $m \geq M_c(t_i)$ events by $\hat{r}(t_i) = (N-1)/(t_{\max} - t_{\min})$.
6. If $\hat{r}(t_i) > r_{\max}$, increase $M_c(t_i)$ by a small increment (e.g., 0.01) and continue with step (4).
7. Continue for the next event (index $i + 1$) with step (3).

For a stationary Poisson process with Gutenberg–Richter distributed magnitudes, the standard deviation of $M_c(t) > M_{c_0}$ can be estimated, because the standard deviation of the interevent times is known to be $1/r$ in this case, which leads to $r/\sqrt{N-1}$ as the standard deviation of the rate estimate. Because the rate depends on M_c according to $r = 10^{a-bM_c}$, the standard deviation of M_c becomes $1/[b\ln(10)\sqrt{N-1}]$. For example, the standard deviation of M_c is thus 0.14 in the case of $b = 1$ and $N = 10$.

Incompleteness during Aftershock Sequences

I will now demonstrate that the empirical fitting relation (equation 1) is in agreement with the approach of rate-dependent incompleteness introduced in the Theoretical Approach section. This is first derived theoretically, based on the empirical decay law of aftershock sequences, and then is confirmed for two particular aftershock sequences by means of the estimation of $M_c(t)$, with the algorithm introduced in the Algorithm to Estimate M_c Based on Catalog Data section.

The rate decay of aftershocks triggered by a particular mainshock is usually well described by the modified Omori–Utsu law:

$$R(t) = K_0(t + c)^{-p}, \tag{5}$$

in which t indicates the elapsed time since the mainshock and K_0, c, p are constants (see Utsu *et al.*, 1995, for a review). Considering a Gutenberg–Richter relation for the frequency–magnitude distribution of the sequence, the Omori–Utsu law can be rewritten as function of the cutoff magnitude m; in particular the rate of aftershocks with magnitude larger or equal to m following a mainshock of magnitude M becomes

$$R(t, m) = K10^{b(M-m)}(t + c)^{-p} \tag{6}$$

with sequence-dependent constants K, b, c, and p.

The definition $R(t, M_c) = r_{\max}$ for the completeness magnitude M_c leads to

$$M_c(t) = M - \frac{1}{b}\log\left(\frac{r_{\max}}{K}\right) - \frac{p}{b}\log(t+c) \qquad (7)$$

if $M_c(t) > M_{c_0}$; else $M_c(t) = M_{c_0}$. This has a very similar form to the empirical relation found for California (equation 1). Here, it should be noted that c in equations (5)–(7) refers to the underlying physical delay time and not to estimated c-values often related to incompleteness issues of the operating seismic network. This is obvious because all aftershocks (including those not reported in the catalog) add to the noise level and reduce the detection ability. Detailed aftershock studies showed that this physical c-value is very small, on the order of one to several minutes or even less (Kagan and Houston, 2005; Peng *et al.*, 2006, 2007; Enescu *et al.*, 2007). Thus, the term $\log(t+c)$ in equation (7) will become indistinguishable to the term $\log(t)$ in equation (1) for t larger than a few minutes.

Typical p-values are observed in the range 0.8–1.1 (Hainzl and Marsan, 2008, and references therein) and b scatters around 1; for example, values are found in the range $b = 0.9$–1.2 for California (Kamer and Hiemer, 2015). Therefore, the prefactor p/b is expected to be in the range 0.67–1.22, which includes the value of 0.75 in the empirical equation (1). To constrain the middle term, $\log(r_{\max}/K)/b$, one can use Båth's law (Båth, 1965), stating that the average magnitude difference between a mainshock and its largest aftershock is 1.2, independent of the mainshock magnitude. Assuming that, on average, one aftershock with $m = M - 1.2$ or larger occurs within the aftershock time T, the total number of aftershocks (i.e., the integration of equation (6) from 0 to T) should equal $10^{b(M-1.2-m)}$, leading to

$$K = \begin{cases} 10^{-1.2b}\frac{1-p}{(T+c)^{1-p}-c^{1-p}}, & \text{if } p \neq 1 \\ 10^{-1.2b}/[\ln(T+c)-\ln(c)] & \text{if } p = 1 \end{cases}. \qquad (8)$$

For $p = b = 1$, $c = 1$ minute, and $T = 100$ (1000) days, the empirical value of −4.5 in equation (1) would, for example, correspond to $r_{\max} = 125(109)$ events/day. This $r_{\max}$ value is in agreement with the expectations from Figure 1 and the observed deviation of the b-values for rates above 100–200 events/day (Fig. 2). Furthermore, the corresponding value $r_{\max} = -\ln(p_c)/\Delta t$ transforms to $\Delta t = 73$ s (83 s) for $p_c = 0.9$ and to $\Delta t = 7$ s (8 s) for $p_c = 0.99$ in the case of $T = 100$ days (1000 days). These values are in the range of the typical length of time windows for the determination of the background noise level in earthquake-detection procedures. In STA/LTA algorithms, the proposed value of the LTA time window is, for example, 7 s for short-period seismograms and 30 s for long-period seismograms (Earle and Shearer, 1994).

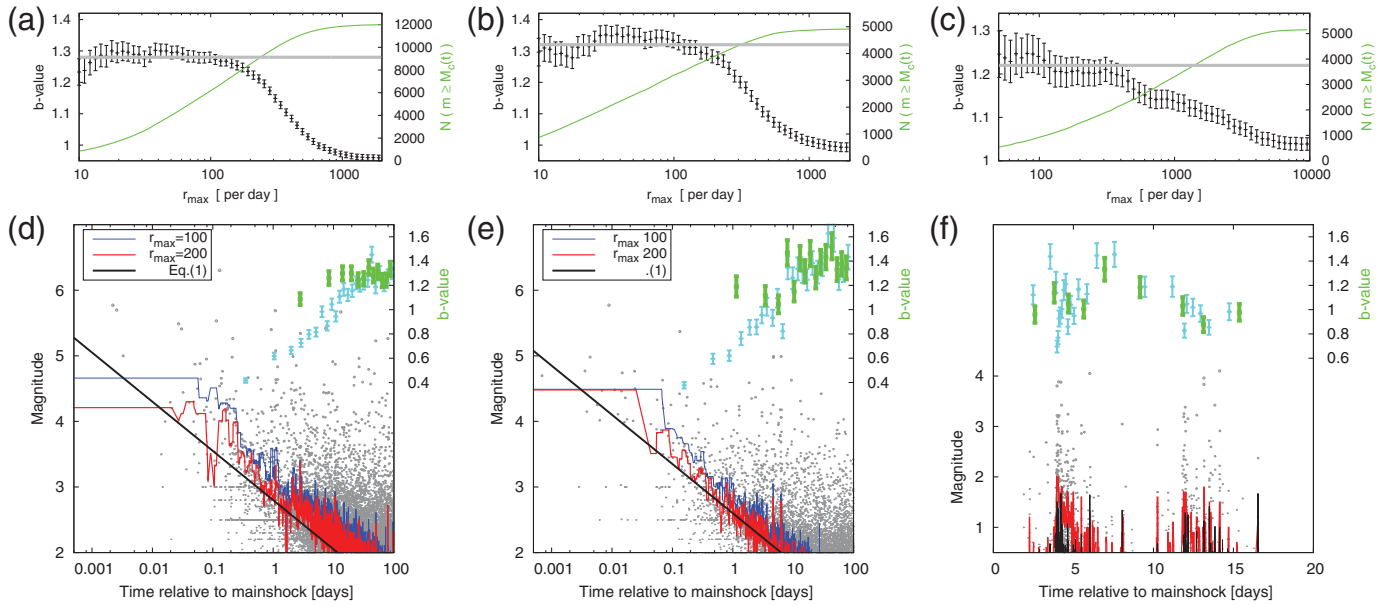
The agreement between the rate-dependent incompleteness estimation and the empirical curve (equation 1) can not only be theoretically demonstrated. I also estimated the incompleteness variations directly using the introduced algorithm (see Algorithm to Estimate M_c Based on Catalog Data section; $N = 10$) for major aftershock sequences in California. For that purpose, I selected events that occurred within the first 100 days following a mainshock of magnitude M and with distances d of less than five times the mainshock rupture length, that is, $d \leq 5 \times 10^{-2.44+0.59M}$ km (Wells and Coppersmith, 1994). In Figure 3, the result is shown for the 1992 M_w 7.3 Landers and the 1999 M_w 7.1 Hector Mine mainshocks. The b-value is found to be stable for rates smaller than approximately 200 events/day in both cases (see Fig. 3a,b). Using this threshold, the estimated $M_c(t)$ has generally the same shape as the empirical incompleteness curve of Helmstetter *et al.* (2006) (see equation 1). However, the results suggest that the completeness magnitudes are slightly underestimated by equation (1), which was derived by the average characteristics of aftershock sequences following $M \geq 6$ mainshocks in southern California. The short-term fluctuations of the $M_c(t)$ estimation are at least partly related to local phases of secondary activation (e.g., secondary aftershock sequences) and thus also contain some important information about M_c variations embedded in the overall decay. Ⓔ This can be also seen in Figures S1 and S2 of the electronic supplement to this article, which show the corresponding equalized time plots for both sequences (Agnew, 2015).

In Figure 3d,e, the b-value is estimated for the $m \geq 2$ aftershocks in successive, nonoverlapping bins of 500 and 200 events, respectively (thin blue error bars). The result shows a systematic apparent increase of the b-value from ~0.4 to 1.3 in both cases. However, small b-values for aftershocks were previously explained by the incomplete recordings (Hainzl, 2013). Indeed, b-values become almost constant (bold green error bars), if only aftershocks above the rate-defined $M_c(t)$ level are considered in equation (4).

Incompleteness during Earthquake Swarms

Swarms are highly clustered earthquakes that cannot be solely attributed to the occurrence of mainshocks and that are often related to transient aseismic forcing, such as magma intrusion (Toda *et al.*, 2002), fluid flow (Miller *et al.*, 2004), or slow-slip events (Lohman and McGuire, 2007). Thus, the main triggering source, equivalent to the mainshock in typical mainshock–aftershock sequences, is not included in the catalog and, consequently, the application of equation (1) will not work. In contrast, the rate-dependent approach to estimate M_c does not require any source information and thus can be straightforwardly applied.

As a test case, I used swarm activity recorded in western Bohemia/Vogtland, central Europe, which has been known for more than 100 years for its episodic occurrence of earthquake swarms. Since 1994, the Nový Kostel area has been monitored by the local seismic network WEBNET (see Data and Resources), which provides high-quality data that has enabled detailed studies of the triggering mechanisms and driving forces of the western Bohemia/Vogtland swarms based on seismicity data. Because of the presence of close-by carbon dioxide emanations and observed correlations of their isotopic content with swarm activity (Braeuer *et al.*, 2007), episodic intrusions of fluid or magma are likely to be one of the driving forces of the observed earthquake clusters. This interpretation was confirmed by analysis of the spatiotemporal seismicity patterns showing clear



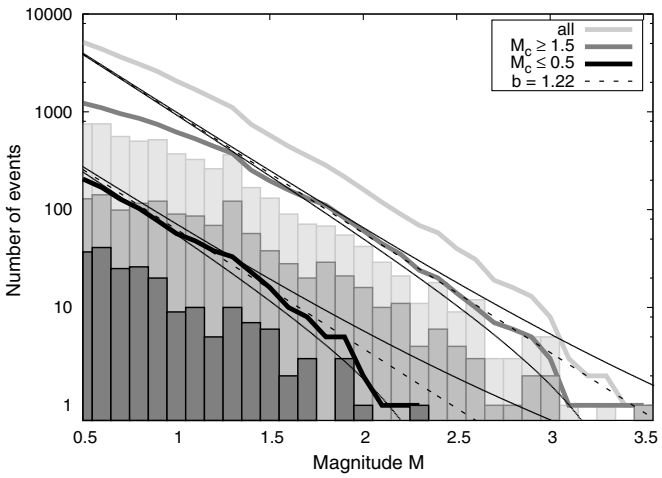
▲ **Figure 3.** Applications of the algorithm (Algorithm to Estimate M_c Based on Catalog Data section with $N = 10$) to estimate the rate-dependent completeness magnitude $M_c(t)$. Aftershock sequences of (a,d) the M_w 7.3 Landers and (b,e) the M_w 7.1 Hector Mine earthquake, as well as (c,f) the year 2011 swarm in western Bohemia. Plots (a–c) show the estimated b-values for $M_c(t)$ calculated for different threshold rates r_{max}, showing stable results for values smaller than approximately 200 events/day, in which horizontal lines refer to (a) $b = 1.28$, (b) $b = 1.32$, and (c) $b = 1.22$. Plots (d–f) show the corresponding $M_c(t)$ estimations for $r_{max} = 100, 200$ events/day in (d,e) and $r_{max} = 300$ events/day in (f), in which the black curve refers to the empirical result of Helmstetter *et al.* (2006) (see equation 1) applied to the mainshock in (d,e), and to all events in (f). Error bars refer to b-values estimated for successive, nonoverlapping 500 events (plot d) and 200 events (plots e,f). The b-values calculated with constant completeness magnitude ($M_c = 2.0$ for California and 0.5 for the swarm) are marked by light blue error bars, and those determined for the estimated $M_c(t)$ (with $r_{max} = 100$ in the aftershock sequences) are shown by green bars. Gray points indicate recorded earthquakes with their magnitudes. In all plots, error bars refer to ±1 standard deviation.

hypocenter migration and reloading (Parotidis *et al.*, 2003; Hainzl *et al.*, 2012).

I applied the new rate-dependent M_c algorithm to the earthquake swarm that occurred in August–September 2011, which was the most intense swarm recorded by the WEBNET so far. The result is shown in Figure 3c,f. Stable b-values of 1.22 ± 0.03 are found for rates below 300 events/day, but b-values start to decrease for larger rates similarly to the results for the aftershock sequences. The slightly higher threshold of 300 events/day, compared to 100–200 events/day for the aftershock sequences, is likely to be related to shorter coda waves associated with the smaller magnitude level M_{c_0}, the local geology (hard rock) known to cause high-frequency recordings with short coda waves, and the spatially strongly clustered seismicity, which is monitored by a dense local seismic network with one station located directly on top of the activity that facilitated event separation. In Figure 3f, the estimated variation of $M_c(t)$ is shown as a red curve for $r_{max} = 300$ events/day and $N = 10$. During the strongest phase of the swarm activity, which lasted more than one day, the estimated M_c-value exceeds 1.5. However, also in other phases of the swarm activity, M_c is significantly larger than the expected overall completeness magnitude of 0.5 estimated by the maximum-curvature method (Wiemer and Wyss, 2000). This also becomes obvious in the equalized time plot (see Ⓔ Fig. S3).

An estimation of $M_c(t)$ based on the superposition of equation (1) relative to precursory events for this swarm shows only a few short-lived excursions from 0.5 (see black curve in Fig. 3f), and only 1.6% of the $m \geq 0.5$ events fall below $M_c(t)$ in this case, whereas this portion is 50.7% in the case of the rate-based estimation of $M_c(t)$.

Similar to the aftershock cases, I estimated the b-value of $m \geq 0.5$ and $m \geq M_c(t)$ earthquakes in successive, nonoverlapping bins of 200 events. The result is shown by error bars in Figure 3f. Although b-values estimated for $M_c(t)$ still show some fluctuations over time, these variations are much smaller than the b-value variations resulting from the analysis, assuming a constant completeness magnitude of 0.5. In particular, the very low values during the highest activity phase do not show up in the case of $M_c(t)$. In Figure 4, the frequency–magnitude distributions are compared between phases in which rate-dependent estimates indicate that a completeness of 0.5 is fulfilled and phases in which $M_c(t)$ is significantly larger, namely $M_c(t) \geq 1.5$. The latter distribution shows a kink at a magnitude of approximately 1.4, with smaller b-value below and larger b-value above this point. In contrast, the entire magnitude range of events that occurred in phases with $M_c(t) \leq 0.5$ can be fitted by $b = 1.22$, the same value as for $m \geq 1.4$ events in the other case. This result confirms the conclusion that the events below 1.4 are incompletely recorded in the high-activity



▲ **Figure 4.** Frequency–magnitude distribution for the year 2011 swarm in western Bohemia. Bold solid lines refer to the cumulative distribution, whereas boxes refer to the number of events in each magnitude bin for the cases of the complete data set (light gray), the events in periods with estimated $M_c \geq 1.5$ (dark gray), and in periods with estimated $M_c \leq 0.5$ (black). For comparison, the two dashed lines refer to Gutenberg–Richter relations with $b = 1.22$ and ± 1 standard deviation (thin solid lines).

phases and that the rate-dependent estimation of $M_c(t)$ seems to work well.

CONCLUSIONS

Based on laboratory experiments, earthquake rates and magnitudes are expected to depend on the stressing state and history (Dieterich, 1994; Scholz, 1998, 2015). Thus, earthquake catalogs contain important information about the state and evolution of the crustal stresses as well as the earthquake nucleation and interaction processes. However, spatiotemporal variations of the incompleteness level of observed earthquake catalogs can prevent or, even worse, distort the results and interpretations of empirical data. This has been known for a long time, and a number of methods have been developed to determine the completeness magnitude M_c of catalogs (see review of Mignan and Woessner, 2012). Though short-term fluctuations of M_c are not resolvable so far because the previously developed methods require a significant number of earthquakes to estimate the completeness value. The proposed approach is able to estimate such short-term exceedance of the basic M_c level. It is based on earthquake rates and needs fewer events and can thus provide a higher resolution of M_c, which might help to discriminate between distorted and robust seismicity features. For example, b-values can be estimated by means of equation (4) for time-dependent $M_c(t)$. The results of this study show that, in this case, only some minor b-value variations during swarm sequences remain, whereas very low b-values during the swarm and aftershock activity are shown to result from incomplete recordings due to high activity rates. Thus, observations showing an apparent anticorrelation between b-values and activity level for particular data sets might be suspicious and should be tested in more detail. However, the incompleteness issue is not only important for b-value calculations, but also for the determination of the activity level. In particular, seismicity models such as the ETAS model (Ogata, 1988; Zhuang *et al.*, 2012) or the rate- and state-dependent friction model (Dieterich, 1994) are now widely used to fit and forecast the evolution of earthquake sequences. Short-term variations of the completeness magnitude are known, for example, to bias the ETAS-parameter estimations, leading particularly to an underestimated scaling exponent of the aftershock productivity with mainshock magnitude (Hainzl *et al.*, 2013), which can lead to strongly underestimated aftershock forecasts for future large earthquakes. The high resolution of $M_c(t)$ can now be used to avoid biased parameter estimations and biased model tests, which can be done by fitting and comparing the corrected model forecast rate $10^{-b(M_c(t)-M_{c_0})}R(t)$ (assuming a Gutenberg–Richter distribution) only to observed $m \geq M_c(t)$ earthquakes, instead of using $R(t)$ for all $m \geq M_{c_0}$ events. These applications might help to clarify open questions concerning the predictive power of statistical seismicity parameters and models.

DATA AND RESOURCES

The high-resolution southern California catalog of Hauksson *et al.* (2012) has been downloaded from the Southern California Earthquake Data Center (http://scedc.caltech.edu/research-tools/alt-2011-dd-hauksson-yang-shearer.html, last accessed January 2016). The swarm catalogs were provided by Tomas Fischer. WEBNET can be accessed at http://www.ig.cas.cz/en/structure/observatories/west-bohemia-seismic-network-webnet (last accessed January 2016).

ACKNOWLEDGMENTS

I am grateful to Zhigang Peng and two anonymous reviewers for their helpful suggestions. Furthermore, I would like to thank my colleagues, Flaminia Catalli, Torsten Dahm, Luigi Passarelli, Olga Zakharova, and Gert Zöller, for critical reading and valuable comments. Finally, I am also grateful to Tomas Fischer for providing the swarm data sets.

REFERENCES

Agnew, D. C. (2015). Equalized plot scales for exploring seismicity data, *Seismol. Res. Lett.* **86,** 1412–1423.

Aki, K. (1965). Maximum likelihood estimate of b in the formula $\log N = a - bM$ and its confidence limits, *Bull. Earthq. Res. Inst.* **43,** 237–239.

Amorèse, D. (2007). Applying a change-point detection method on frequency–magnitude distributions, *Bull. Seismol. Soc. Am.* **97,** 1742–1749.

Bakun, W. H., and A. G. Lindh (1977). Local magnitudes, seismic moments, and coda durations for earthquakes near Oroville, California, *Bull. Seismol. Soc. Am.* **67** no. 3, 615–629.

Båth, M. (1965). Lateral inhomogeneities in the upper mantle, *Tectonophysics* **2,** 483–514.

Braeuer, K., H. Kampf, U. Koch, S. Niedermann, and G. Strauch (2007). Seismically induced changes of the fluid signature detected by a multi-isotope approach (He, CO_2, CH_4, N^{-2}) at the Wettinquelle,

Bad Brambach (central Europe), *J. Geophys. Res.* **112,** no. B04307, doi: 10.1029/2006JB004404.
Cao, A. M., and S. S. Gao (2002). Temporal variations of seismic *b*-values beneath northeastern Japan island arc, *Geophys. Res. Lett.* **29,** no. 48, 1–3.
Davidsen, J., C. Gu, and M. Baiesi (2015). Generalized Omori–Utsu law for aftershock sequences in southern California, *Geophys. J. Int.* **201,** 965–978.
Dieterich, J. H. (1994). A constitutive law for rate of earthquake production and its application to earthquake clustering, *J. Geophys. Res.* **99,** 2601–2618.
Earle, P. S., and P. M. Shearer (1994). Characterization of global seismograms using an automatic-picking algorithm, *Bull. Seismol. Soc. Am.* **84,** no. 2, 366–376.
Enescu, B., J. Mori, and M. Miyazawa (2007). Quantifying early aftershock activity of the 2004 Mid-Niigata Prefecture earthquake (M_w 6.6), *J. Geophys. Res.* 112**,** no. B04310, doi: 10.1029/2006JB004629.
Hainzl, S. (2013). Comment on "Self-similar earthquake triggering, Båth's law, and foreshock/aftershock magnitudes: Simulations, theory, and results for southern California" by P. M. Shearer, *J. Geophys. Res.* **118,** doi: 10.1002/jgrb.50132.
Hainzl, S., and T. Fischer (2002). Indications for a successively triggered rupture growth underlying the 2000 earthquake swarm in Vogtland/NW-Bohemia, *J. Geophys. Res.* **107,** 2338.
Hainzl, S., and D. Marsan (2008). Dependence of the Omori–Utsu law parameters on main shock magnitude: Observations and modeling, *J. Geophys. Res.* **113,** no. B10309, doi: 10.1029/2007JB005492.
Hainzl, S., T. Fischer, and T. Dahm (2012). Seismicity-based estimation of the driving fluid pressure in the case of swarm activity in western Bohemia, *Geophys. J. Int.* **191,** 271–281.
Hainzl, S., O. Zakharova, and D. Marsan (2013). Impact of aseismic transients on the estimation of aftershock productivity parameters, *Bull. Seismol. Soc. Am.* **103,** 1723–1732, doi: 10.1785/0120120247.
Hauksson, E., W. Yang, and P. Shearer (2012). Waveform relocated earthquake catalog for southern California (1981 to June 2011), *Bull. Seismol. Soc. Am.* **102,** 2239–2244.
Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2006). Comparison of short-term and time-independent earthquake forecast models for southern California, *Bull. Seismol. Soc. Am.* **96,** no. 1, 90–106.
Huang, Y., and G. C. Beroza (2015). Temporal variation in the magnitude-frequency distribution during the Guy-Greenbrier earthquake sequence, *Geophys. Res. Lett.* **42,** doi: 10.1002/2015GL065170.
Imoto, M. (1991). Changes in the magnitude–frequency *b*-value prior to large ($M \geq 6.0$) earthquakes in Japan, *Tectonophysics* **193,** no. 4, 311–325.
Kagan, Y. Y. (2004). Short-term properties of earthquake catalogs and models of earthquake source, *Bull. Seismol. Soc. Am.* **94,** no. 4, 1207–1228.
Kagan, Y. Y., and H. Houston (2005). Relation between mainshock rupture process and Omori's law for aftershock moment release rate, *Geophys. J. Int.* **163,** no. 3, 1039–1048.
Kamer, Y., and S. Hiemer (2015). Data-driven spatial *b* value estimation with applications to California seismicity: To *b* or not to *b*, *J. Geophys. Res.* **120,** doi: 10.1002/2014JB011510.
Langenbruch, C., C. Dinske, and S. A. Shapiro (2011). Interevent times of fluid induced earthquakes suggest their Poisson nature, *Geophys. Res. Lett.* **38,** L21302, doi:10.1029/2011GL049474.
Lohman, R. B., and J. J. McGuire (2007). Earthquake swarms driven by aseismic creep in the Salton trough, California, *J. Geophys. Res.* **112,** no. B04405, doi:10.1029/2006JB004596.
Mallika, K., H. Gupta, D. Shashidhar, N. P. Rao, A. Yadav, S. Rohilla, H. V. S. Satyanarayana, and D. Srinagesh (2013). Temporal variation of *b* value associated with M ~ 4 earthquakes in the reservoir-triggered seismic environment of the Koyna–Warna region, western India, *J. Seismol.* **17,** no. 1, 189–195.
Marzocchi, W., and L. Sandri (2003). A review and new insights on the estimation of the *b*-value and its uncertainty, *Ann. Geophys.* **46,** no. 6, 1271–1282.
Mignan, A., and J. Woessner (2012). Estimating the magnitude of completeness for earthquake catalogs, *Community Online Resource for Statistical Seismicity Analysis*, doi: 10.5078/corssa-00180805; also available at http://www.corssa.org (last accessed December 2015).
Miller, S. A., C. Collettini, L. Chiaraluce, M. Cocco, M. Barchi, and B. J. P. Kaus (2004). Aftershocks driven by a high-pressure CO_2 source at depth, *Nature* **427,** 724–727.
Narteau, C., S. Byrdina, P. Shebalin, and D. Schorlemmer (2009). Common dependence on stress for the two fundamental laws of statistical seismology, *Nature* **462,** 642–646.
Ogata, Y. (1988). Statistical models for earthquake occurrence and residual analysis for point processes, *J. Am. Stat. Assoc.* **83,** 9–27.
Ogata, Y., and K. Katsura (1993). Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogues, *Geophys. J. Int.* **113,** 727–738.
Parotidis, M., E. Rothert, and S. A. Shapiro (2003). Pore-pressure diffusion: A possible triggering mechanism for the earthquake swarms 2000 in Vogtland/NW-Bohemia, central Europe, *Geophys. Res. Lett.* **30,** 2075, doi: 10.1029/2003GL018110.
Peng, Z. G., J. E. Vidale, and H. Houston (2006). Anomalous early aftershock decay rate of the 2004 M_w 6.0 Parkfield, California, earthquake, *Geophys. Res. Lett.* **33,** no. 17, L17307, doi: 10.1029/2006GL026744.
Peng, Z. G., J. E. Vidale, M. Ishii, and A. Helmstetter (2007). Seismicity rate immediately before and after main shock rupture from high-frequency waveforms in Japan, *J. Geophys. Res.* **112,** no. B3, B03306, doi: 10.1029/2006JB004386.
Scholz, C. H. (1998). Earthquakes and friction laws, *Nature* **391,** 37–42.
Scholz, C. H. (2015). On the stress dependence of the earthquake *b* value, *Geophys. Res. Lett.* **42,** 1399–1402, doi: 10.1002/2014GL062863.
Schorlemmer, D., and J. Woessner (2008). Probability of detecting an earthquake, *Bull. Seismol. Soc. Am.* **98,** 2103–2117.
Schurr, B., G. Asch, S. Hainzl, J. Bedford, A. Hoechner, M. Palo, R. Wang, M. Moreno, M. Bartsch, Y. Zhang, *et al.* (2014). Gradual unlocking of plate boundary controlled initiation of the 2014 Iquique earthquake, *Nature* **512,** 299–302.
Shcherbakov, R., D. L. Turcotte, and J. B. Rundle (2004). A generalized Omori's law for earthquake aftershock decay, *J. Geophys. Res.* **31,** L11613, doi: 10.1029/2004GL019808.
Toda, S., R. S. Stein, and T. Sagiya (2002). Evidence from the AD 2000 Izu islands earthquake swarm that stressing rate governs seismicity, *Nature* **419,** no. 6902, 58–61.
Utsu, T., Y. Ogata, and R. S. Matsu'ura (1995). The centenary of the Omori formula for a decay of aftershock activity, *J. Phys. Earth* **43,** 1–33.
Wells, D. L., and K. J. Coppersmith (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement, *Bull. Seismol. Soc. Am.* **84,** 974–1002.
Wiemer, S., and M. Wyss (2000). Minimum magnitude of complete reporting in earthquake catalogs: Examples from Alaska, the western United States, and Japan, *Bull. Seismol. Soc. Am.* **90,** 859–869.
Woessner, J., and S. Wiemer (2005). Assessing the quality of earthquake catalogues: Estimating the magnitude of completeness and its uncertainty, *Bull. Seismol. Soc. Am.* **95,** no. 2, 684–698.
Zhuang, J., D. Harte, M. J. Werner, S. Hainzl, and S. Zhou (2012). Basic models of seismicity: Temporal models, *Community Online Resource for Statistical Seismicity Analysis*, doi: 10.5078/corssa-79905851 also available at http://www.corssa.org (last accessed December 2015).

Sebastian Hainzl
GFZ German Research Centre for Geosciences
Telegrafenberg
14473 Potsdam, Germany
hainzl@gfz-potsdam.de

Published Online 10 February 2016