

Quantifying the predictability of the timing of jökulhlaups from Merzbacher Lake, Kyrgyzstan

Jonathan KINGSLAKE,* Felix NG

*Department of Geography, University of Sheffield, Sheffield, UK
E-mail: jonngs@bas.ac.uk*

ABSTRACT. Glacier-dammed lakes can yield subglacial outburst floods (jökulhlaups) repeatedly. Predicting flood timing is crucial for hazard mitigation, but incomplete understanding of flood-initiation physics makes this challenging. Here we examine the predictability of the timing of jökulhlaups from Merzbacher Lake, Kyrgyzstan, using five flood-date prediction models of varying complexity. The simplest model, which offers a benchmark against which the other models are compared, assumes that floods occur on the same date each year. The other four models predict flood dates using a flood-initiation threshold approach and incorporate weather forcing (approximated by the output of two climate reanalyses) behind the meltwater input to the lake; the most complex of these models accounts for a moving subglacial water divide beneath the glacier that dams the lake. Each model is optimized against recorded flood dates to maximize its prediction ability. In terms of their flood prediction ability, our two best models are those that assume a variable outburst threshold governed by the rate of meltwater input to the lake and the rate of lake-level rise. They excel over the simplest and most complex models and correctly predict flood dates to within ± 20 days 57.4% of the time. We also quantify the impact of weather uncertainty on prediction success. Our findings can inform practical flood-forecasting schemes and future investigations of flood-initiation physics.

1. INTRODUCTION

Subglacial outburst floods, or *jökulhlaups*, originating from ice-dammed lakes are some of the most dramatic and hazardous glaciological phenomena. With peak discharges often exceeding $10^3 \text{ m}^3 \text{ s}^{-1}$, they can pose significant danger to downstream settlements and environments, a threat likely exacerbated by population growth and climatic warming (e.g. Björnsson, 2004; Barnett and others, 2005; Ng and others, 2007). While an ability to predict the timing and magnitude of these floods can enable mitigation of their consequences, this problem has remained largely untackled, despite much research into their underlying physics (Nye, 1976; Spring and Hutter, 1981; Clarke, 1982, 2003; Evatt and others, 2006; Fowler, 2009; Kingslake and Ng, 2013), how they initiate (Thórarinnsson, 1953; Glen, 1954; Liestøl, 1956; Fowler, 1999; Jóhannesson, 2002; Flowers and others, 2004; Walder and others, 2005, 2006; Huss and others, 2007; Werder and Funk, 2009), their geomorphic impact (Walder and Driedger, 1994; Cenderelli and Wohl, 2003; Roberts and others, 2003; Russell and others, 2006; Carrivick, 2007; Burke and others, 2009) and their influence on the flow dynamics of glaciers (Anderson and others, 2005; Walder and others, 2006; Magnússon and others, 2007, 2010; Sugiyama and others, 2007; Mayer and others, 2008; Riesen and others, 2010; Bartholomäus and others, 2011; Kingslake and Ng, 2013) and ice sheets (Bell and others, 2007; Sergienko and others, 2007; Stearns and others, 2008; Sergienko and Hulbe, 2011).

Recently, Ng and Liu (2009) put forward a mathematical theory for understanding the long-term timing pattern of jökulhlaups. Their model represents the ice-dammed lake as a threshold system that drains when it reaches a threshold

water depth, which they initially kept constant. They found that in such a nonlinear system the lake yields an irregular sequence of flood dates with predictable timing characteristics when the annual total volume of meltwater feeding the lake differs from the volume needed to trigger it into a flood. They used their threshold model to analyse the recorded date sequence of 51 jökulhlaups from Merzbacher Lake, Kyrgyzstan, and showed that the variable meltwater input to the lake caused by weather interacts with the threshold water depth to produce flood sequences that resemble the Merzbacher Lake flood record. Like the recorded flood dates, the simulated flood dates are focused on the warmest part of the year when the lake fills with meltwater more rapidly than at other times. This analysis was conducted with a constant outburst threshold, but Ng and Liu (2009) went further and found empirical evidence suggesting that Merzbacher Lake's threshold is not constant, but varies from flood to flood and is partially correlated with the rate of lake-level rise.

If certain aspects of jökulhlaup timing sequences can be explained using the outburst threshold concept, it may be possible to use it to predict individual flood dates. We explore this in this paper and take steps towards operational flood forecasting. Like Ng and Liu (2009), we focus on the timing of floods (which could inform predictions of flood size) and use the Merzbacher Lake system as an example (Fig. 1a). How reliable, or hopeless is our best prediction of the next flood? Is operational forecasting feasible given our current knowledge of how jökulhlaups initiate? How complex should the forecast models be? We consider such questions, which should interest policy-makers as well as populations exposed to flood risks. Accordingly we set out to establish how well simple threshold models can predict the Merzbacher flood dates and, in so doing, develop measures of predictability that can be applied to other jökulhlaup lakes. Our results can serve as a benchmark for future forecasting efforts.

*Present address: British Antarctic Survey, Natural Environment Research Council, Cambridge, UK.

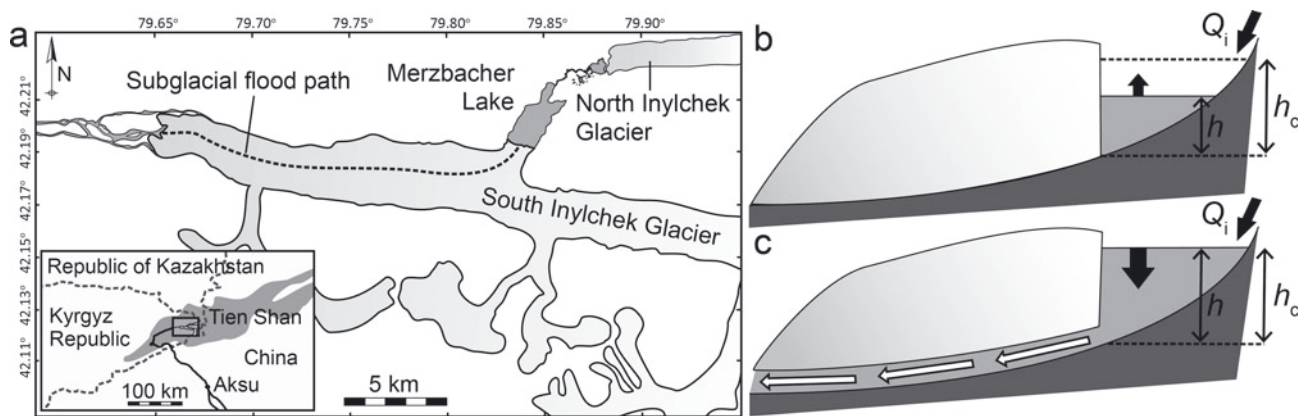


Fig. 1. The Merzbacher jökulhlaup system. (a) Map of Merzbacher Lake and North and South Inylchek Glaciers; inset shows the system's location in the Tien Shan. (b, c) Schematic lake and glacier cross-sections along a possible subglacial flood path (b) while the lake is filling and (c) when lake water depth h reaches the threshold h_c and a flood initiates.

Because successful flood prediction requires accounting for the variability in the lake outburst threshold highlighted by the work of Ng and Liu (2009) and by earlier studies (e.g. Clague and Mathews, 1973; Walder and Costa, 1996; Björnsson, 2003; Ng and Björnsson, 2003), a hierarchy of assumptions for this threshold is examined below. The assumptions are motivated by different hypotheses of how they depend on environmental conditions, such as the weather near the lake. Our 'threshold assumptions' range from the predominantly empirical to those based more strongly on physical grounds, and are presented and studied in order of increasing complexity.

The paper is organized as follows. In Section 2, we outline Ng and Liu's (2009) model of lake filling and draining and explain how it is used to predict flood dates in our study. Section 3 introduces Merzbacher Lake, its record of jökulhlaup dates and a melt equation needed in the Ng–Liu model for estimating this lake's water supply. The flood dates are used with the melt equation to reconstruct flood volumes, whose distribution gives us a probabilistic handle on the size and timing of future floods. In Sections 4 and 5, we use the Ng–Liu model with different threshold assumptions to evaluate how well this model predicts the Merzbacher flood dates. After considering how to quantify mismatch between predicted and observed flood dates, we optimize each version of the model for prediction success. In this exercise, the choices of threshold behaviour, weather-data source and assumptions of future weather lead to multiple sets of prediction results, whose performances are compared and discussed. Significantly, we show that, in terms of their flood prediction ability, models that assume a constant outburst threshold are not only inferior to more complex models that allow the threshold to vary, but also inferior to a naïve model that assumes floods occur on the same date each year.

2. FLOOD-DATE PREDICTION WITH THE THRESHOLD MODEL

2.1. The Ng–Liu model

In Ng and Liu's (2009) model, the jökulhlaup lake, which has volume V and water depth h , fills in response to meltwater input at a rate Q_i , and drains suddenly and completely in a flood when a threshold water depth, h_c , is

reached (Fig. 1b and c). With t denoting time, their model equations are

$$\frac{dV}{dt} = Q_i \quad \text{when } h(V) < h_c, \quad (1a)$$

$$V \rightarrow 0 \quad \text{when } h(V) = h_c, \quad (1b)$$

where the function $h(V)$ represents lake geometry. These equations generate sawtooth-shaped filling and draining cycles in the lake level that are irregular when Q_i depends on the weather. By assuming a constant h_c and estimating Q_i from daily air temperature with the sub-model described later (Section 3.3), Ng and Liu (2009) simulated Eqn (1) for Merzbacher Lake through 1956 to 2005 to obtain model flood dates, which they compared with the observed flood dates in this period (Table 1). Motivated by concepts from nonlinear dynamics, they analysed the simulated and observed date sequences with time-delay maps and demonstrated how the sequences resembled one another.

Ng and Liu (2009) also investigated the possibility of a variable h_c . They integrated Eqn (1a) for each period between successive floods to reconstruct the flood volumes and the long-term lake-level history. They interpreted inter-flood variability in the reconstructed flood volumes as a manifestation of corresponding variability in Merzbacher Lake's outburst threshold. They then plotted each reconstructed flood's outburst threshold against the mean lake-filling rate during the week that preceded the corresponding flood which they extracted from their reconstructed lake-level history. This revealed a negative correlation between the two variables, which suggests that lake-filling rate plays a role in determining the outburst threshold.

2.2. Our prediction strategy

In this paper, we use Ng and Liu's threshold model for the purpose of predicting the next flood date from the date of the last (known) flood, and specifically for *hindcasting* the observed Merzbacher flood dates in Table 1. Our study period spans the first and the last flood dates and has 19 006 days on which we could ask when the next flood will be. We explore the effect of a variable outburst threshold on prediction success and seek the best assumption for h_c . Most of our simulations use Eqns (1), but one set of simulations use a modified model with a threshold different from h_c that mimics a mobile subglacial seal. A detailed explanation of this threshold is given in Section 4.3.

Table 1. Flood record of Merzbacher Lake for the period 1956–2008, showing the dates of peak discharge of 54 floods and the measured volumes of 19 floods. Of these 19 flood volumes, the 13 shown in bold are considered more reliable. Entries 1–51 come from Ng and Liu (2009), and entries 52–54 from Liu Shiyin (personal communication, 2009). The table omits a small subsidiary flood which emanated from the lake when it was nearly empty around 4 August 2005, which followed closely the flood in entry 51 and probably occurred because the ice dam did not ‘re-seal’ completely after the main flood. The reconstructed flood volumes are calculated from the two temperature forcings T_{NCEP} and T_{ERA} using the method described in Section 3.4

Flood index	Date of flood peak	Measured flood volume	Reconstructed flood volume T_{NCEP}	Reconstructed flood volume T_{ERA}	Flood index	Date of flood peak	Measured flood volume	Reconstructed flood volume T_{NCEP}	Reconstructed flood volume T_{ER}
	dd/mm/yyyy	10^8 m^3	10^8 m^3	10^8 m^3		dd/mm/yy	10^8 m^3	10^8 m^3	10^8 m^3
1	02/07/1956		–		28	21/08/1982	1.74 ± 0.14	1.95 ± 0.07	1.89 ± 0.08
2	03/09/1956		0.98 ± 0.14		29	22/08/1983		1.48 ± 0.06	1.35 ± 0.09
3	07/09/1957		1.75 ± 0.07		30	26/08/1984		1.71 ± 0.07	1.69 ± 0.07
4	24/11/1958	1.71 ± 0.09	2.07 ± 0.10	1.96 ± 0.08	31	15/08/1985		1.45 ± 0.06	1.44 ± 0.07
5	19/09/1959	1.44 ± 0.19	1.64 ± 0.06	1.59 ± 0.07	32	01/08/1986		1.36 ± 0.06	1.34 ± 0.07
6	16/07/1961		2.80 ± 0.12	2.94 ± 0.11	33	19/08/1987		1.77 ± 0.07	1.67 ± 0.08
7	06/06/1963	2.99 ± 0.17	3.01 ± 0.12	3.10 ± 0.10	34	12/12/1988	2.10 ± 0.14	2.03 ± 0.08	2.01 ± 0.08
8	18/09/1963	0.95 ± 0.09	1.14 ± 0.07	1.13 ± 0.06	35	31/08/1989	1.28 ± 0.15	1.20 ± 0.05	1.20 ± 0.06
9	28/09/1964	1.40 ± 0.19	1.71 ± 0.09	1.49 ± 0.09	36	10/08/1990		1.39 ± 0.06	1.39 ± 0.06
10	01/09/1965	1.42 ± 0.13	1.69 ± 0.06	1.70 ± 0.08	37	24/07/1991		1.31 ± 0.06	1.39 ± 0.07
11	03/08/1966		1.48 ± 0.06	1.44 ± 0.06	38	04/08/1992		1.65 ± 0.07	1.75 ± 0.07
12	13/12/1966		0.74 ± 0.04	0.66 ± 0.04	39	22/08/1993		1.53 ± 0.08	1.66 ± 0.08
13	13/09/1967	1.82 ± 0.06	1.54 ± 0.07	1.40 ± 0.06	40	24/07/1994		1.19 ± 0.05	1.48 ± 0.06
14	21/08/1968	1.64 ± 0.22	1.51 ± 0.06	1.34 ± 0.06	41	18/07/1995		1.42 ± 0.06	1.56 ± 0.07
15	20/08/1969		1.59 ± 0.07	1.49 ± 0.07	42	05/12/1996	2.84 ± 0.09	2.29 ± 0.09	2.33 ± 0.09
16	31/07/1970	0.86 ± 0.20	1.45 ± 0.07	1.25 ± 0.07	43	31/07/1997		1.26 ± 0.07	1.19 ± 0.06
17	17/08/1971	1.33 ± 0.19	1.64 ± 0.09	1.90 ± 0.07	44	27/07/1998		1.74 ± 0.07	1.77 ± 0.07
18	04/10/1972	1.61 ± 0.10	1.70 ± 0.11	1.93 ± 0.09	45	19/07/1999		1.72 ± 0.07	1.53 ± 0.07
19	02/09/1973		1.30 ± 0.06	1.45 ± 0.06	46	27/07/2000		2.04 ± 0.08	1.85 ± 0.07
20	08/08/1974		1.25 ± 0.06	1.37 ± 0.07	47	31/07/2001		1.88 ± 0.07	1.73 ± 0.07
21	12/09/1975		1.72 ± 0.07	1.80 ± 0.08	48	06/08/2002		1.83 ± 0.07	1.60 ± 0.07
22	24/08/1976	1.31 ± 0.27	1.25 ± 0.07	1.33 ± 0.07	49	26/07/2003		1.54 ± 0.07	
23	24/05/1978	2.51 ± 0.35	2.31 ± 0.15	2.38 ± 0.10	50	09/08/2004		1.95 ± 0.07	
24	09/08/1978		0.79 ± 0.05	0.91 ± 0.07	51	15/07/2005		1.52 ± 0.07	
25	27/05/1980	2.36 ± 0.18	2.33 ± 0.12	2.38 ± 0.12	52	30/07/2006		2.10 ± 0.08	
26	12/09/1980	1.51 ± 0.06	1.28 ± 0.10	1.17 ± 0.07	53	11/07/2007		1.89 ± 0.07	
27	15/07/1981		0.91 ± 0.09	1.01 ± 0.08	54	15/07/2008		2.07 ± 0.09	

In a typical prediction run, we situate ourselves on a particular day D (Fig. 2); and, starting with an empty lake on the day after the previous known flood (e.g. one of the floods in Table 1), we integrate Eqn (1a) forward on a daily time-step to fill the lake, until h , found from the lake volume via $h(V)$, reaches h_c . Thus the next flood date is predicted, and its mismatch in timing from the actual (observed) next flood forms a prediction error, E (Fig. 2). As D marches forward in time past a known flood date, the start date of the integration is renewed. In each set of prediction runs assuming specific threshold behaviour and a specific temperature data source, we apply this procedure for every day in the study period to obtain the same number of predictions (and hence the same number of errors) as the length of the period and summarize the errors into an overall measure of mismatch. Two measures considered later are the root-mean-square (RMS) error and the fraction of predictions that are accurate to a fixed number of days.

2.3. ‘Real’ and ‘Simple’ schemes of weather forcing

When we perform this exercise to quantify the predictability of Merzbacher Lake’s floods, we know the complete past weather forcing – from archived daily temperature data – for deriving Q_i in Eqn (1a). (These data are outlined in Section 3.3.) However, we put ourselves in the position of a forecaster in the past and assume that, on any day D on

which a prediction is made, he or she knew the past weather but not the future weather, which must be estimated to run the model. We simulate this scenario by the procedure shown in Figure 2a. For all days up to and including D , we use archived daily temperature data for finding Q_i . But for all days after D (until the model lake reaches its outburst threshold), we find Q_i from a temperature *forecast*, taken to be the multi-year mean of the archived daily temperature on that calendar date (Fig. 2a). Although such a forecast could be made using sophisticated methods (e.g. regional climate models), this procedure is the easiest that incorporates weather uncertainty into our predictions; we call it the ‘Real’ weather-forcing scheme. As D approaches the next flood, more of the lake-filling period is simulated with known weather forcing, so the predicted flood date or ‘flood-date hindcast’ varies with D . In Section 5.3, this idea allows us to study whether predictions improve as we approach the next flood.

Later, in Section 5.2, we consider a second scheme of weather forcing, called the ‘Simple’ scheme, to assess the impact of weather uncertainty on flood predictability. The previous procedure is followed, but we deliberately assume that the forecaster knew the future as well as the past weather; thus, archived daily temperature data are used to derive Q_i throughout each lake-filling period (Fig. 2b). In this case, flood-date hindcasts calculated on different days D

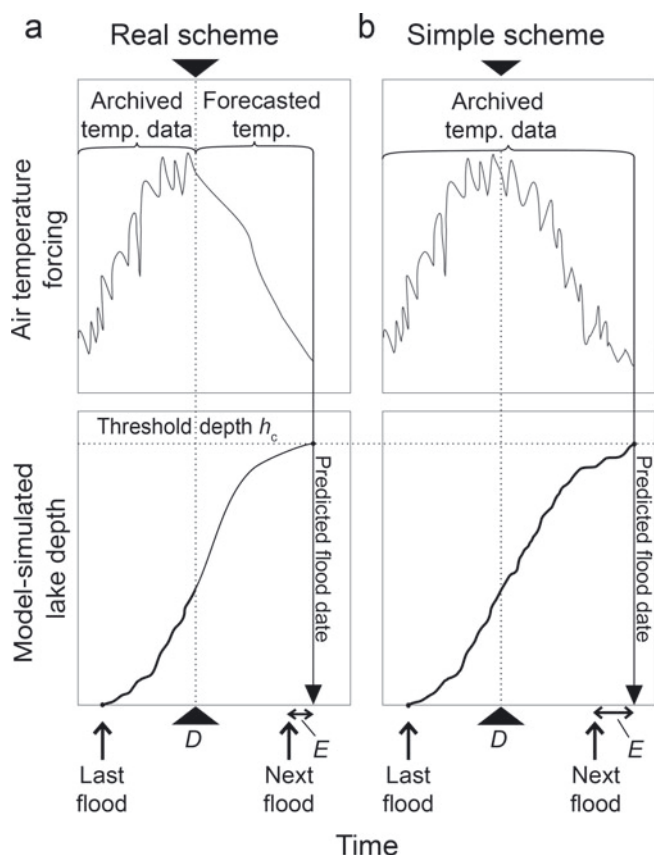


Fig. 2. (a) ‘Real’ scheme and (b) ‘Simple’ scheme of application of weather forcing in our models. In both (a) and (b), upper and lower plots show, respectively, the history of the temperature forcing used to calculate the lake water supply Q_i and the corresponding history of the simulated lake depth h . (a) The Real scheme calculates Q_i using archived daily temperature data on all days before the day on which the prediction is made, D , and using forecasted temperature on all days after D . (b) The Simple scheme calculates Q_i using archived daily temperature data throughout the prediction run. In both schemes, a flood is predicted when h reaches the outburst threshold h_c , and E denotes the prediction error.

that are bracketed by the same pair of observed floods will be identical.

3. STUDY SITE AND DATA SOURCES

3.1. The Inylchek glaciers and Merzbacher Lake

South Inylchek Glacier and North Inylchek Glacier together form the largest glacier system in the Tien Shan, central Asia (Fig. 1a). The southern glacier stretches for a distance of 50 km from its accumulation area by the ~ 7000 m high peaks of Khan Tengri and Podeba to its debris-covered terminus at ~ 2900 m a.s.l. Although located in Kyrgyzstan, its runoff flows into China to feed that country’s fifth-largest river, the Tarim, which provides a vital water supply to oases around the Taklamakan Desert. Glacial meltwater contributes at least 35% of the Tarim’s total runoff (Aizen and Aizen, 1998), and this proportion is predicted to rise over the next few decades (Aizen and others, 2007).

Merzbacher Lake forms behind the ice dam made by South Inylchek Glacier across the valley occupied by North Inylchek Glacier (Fig. 1a). The lake fills typically to about 80–100 m depth before draining subglacially, producing jökulhlaups with durations of about a week and peak

discharges of up to $1500 \text{ m}^3 \text{ s}^{-1}$ (Liu, 1992; Mavlyudov, 1997; Ng and Liu, 2009). Besides being a hazard, these floods represent a waste of valuable water resources (Shen and others, 2007). The Inylchek river is also a candidate for hydroelectric projects (Ng and others, 2007; Mamatkanov and Deng, 2011). These reasons necessitate reliable forecasting of the floods.

3.2. Flood-date record

Merzbacher Lake is chosen for study because of its long and comprehensive jökulhlaup record and because its outbursts recur on a roughly regular basis (once every year on average, doubling or missing in some years; Ng and Liu, 2009), so that attempts to hindcast their dates may have some chance of success. Table 1 lists the lake’s 54 flood dates from 1956 to 2008. All but the last three dates are taken from Ng and Liu (2009), who compiled them from hydrological measurements at Xiehela gauging station near Aksu, China (Fig. 1a), and earlier publications. Those authors also derived the volumes of 19 floods by subtracting an estimated base-flow component from the area of flood hydrographs (Ng and others, 2007; Ng and Liu, 2009). Of these flood volumes, 13 (bold in Table 1) are considered more reliable than the remaining 6 (Ng and Liu, 2009) and are used to calibrate our melt sub-model below. The last three dates in Table 1 come from Liu Shiyin (personal communication, 2009).

3.3. Lake water supply sub-model and temperature data

Following Ng and Liu (2009), we calculate the rate of meltwater input to the lake, Q_i , in Eqn (1a) by using the temperature-index parameterization

$$Q_i = k(T - T_0)_{0+} + c, \quad (2)$$

where T ($^{\circ}\text{C}$) is air temperature near the lake, k ($\text{m}^3 \text{ d}^{-1} \text{ }^{\circ}\text{C}^{-1}$) quantifies the melt sensitivity to temperature, T_0 ($^{\circ}\text{C}$) is the temperature threshold above which melting occurs, and c ($\text{m}^3 \text{ d}^{-1}$) denotes effective water input to the lake due to calving from the ice dam. k , T_0 and c are assumed constant. The subscript $0+$ means that $T - T_0$ is set to zero whenever this difference is negative.

Ng and Liu (2009) prescribed data for T from the daily surface temperature provided by the US National Centers for Environmental Prediction (NCEP)/US National Center for Atmospheric Research (NCAR) reanalysis project (Kalnay and others, 1996; Kistler and others, 2001). Here we also employ the reanalysed daily surface temperature provided by the European Centre for Medium-Range Weather Forecasts, called ERA-40 (Uppala and others, 2005), in order to study the impact of different weather forcings on our predictions. We call these sources ‘NCEP’ and ‘ERA’ and use T_{NCEP} and T_{ERA} to denote the respective temperature data after these have been interpolated to the coordinates of Merzbacher Lake. Both projects assimilate weather data from multiple sources (e.g. satellites, radiosondes, aircraft, ships, ocean buoys) into a global climate model to reconstruct the state of the atmosphere, but they have different temporal ranges. T_{NCEP} began before 1956 and continues to today and covers our entire study period. T_{ERA} is available from 1 September 1957 to 31 August 2002, so it can be used in our prediction of floods 3–48 only. Like Ng and Liu (2009), we do not use the temperature data from Tien Shan weather station (41.92°N , 78.23°E), Kyrgyzstan, because gaps in this dataset make its application difficult.

Ng and Liu (2009) reported values of c estimated by other authors. They derived the other two sub-model parameters, k and T_0 , by a nonlinear regression that fits Eqn (2) to the 13 reliable flood volumes in Table 1. The idea is that the daily melt volume predicted by Eqn (2), when summed over the known filling period in the run-up to each of these 13 jökulhlaups, should match each observed flood volume, because Merzbacher Lake is seen to drain completely in the floods. We find k and T_0 here by the same method, assuming Ng and Liu's 'typical' value of c , $(0.33 \pm 0.06) \times 10^5 \text{ m}^3 \text{ s}^{-1}$. Table 2 lists the parameter values corresponding to the T_{NCEP} and T_{ERA} temperature forcings.

3.4. Probability distributions of flood volume and timing

Armed with Eqns (1a) and (2), the flood dates in Table 1 can be used to derive an empirical probability distribution of historical flood volumes, which in turn helps us gauge the size and timing of future floods. We do this here in three steps. First we reconstruct the volumes of the floods in Table 1 by integrating Eqn (1a) in the period between each pair of successive flood dates, filling the lake from empty. Two separate reconstructions are made with T_{NCEP} and T_{ERA} as forcings, with Eqn (2) taking the corresponding parameters in Table 2. With the NCEP and ERA forcings, this procedure reconstructs 53 and 45 flood volumes, respectively (Table 1). Figure 3 compares these two sets of results with the 19 observed volumes in Table 1 and with each other. The good agreements shown by these plots support our use of both temperature data sources for flood-date prediction.

Next we form a cumulative probability distribution to represent the likelihood that a flood does not exceed a certain size, a volume V_F . This is done by putting the reconstructed flood volumes in Table 1 in ascending order, counting the number of floods with volumes $\leq V_F$ and turning the number into a fraction of the total, to represent probability. Figure 4a plots the resulting probability distribution against V_F . It shows that historically the mean and the median flood volumes have been similar, $\sim 1.64 \times 10^8 \text{ m}^3$.

Table 2. Values of parameters k and T_0 of the lake water supply sub-model, derived using a nonlinear multivariate regression that fits Eqn (2) to 13 measured lake volumes (Table 1). Parameters are derived separately for the two temperature forcings, T_{NCEP} and T_{ERA}

Temperature forcing	k $10^5 \text{ m}^3 \text{ d}^{-1} \text{ } ^\circ\text{C}^{-1}$	T_0 K	r^2
T_{NCEP}	1.05	2.65	0.877
T_{ERA}	1.13	5.65	0.830

No floods had V_F exceeding $3.2 \times 10^8 \text{ m}^3$, and 83% of them had $1.1 \times 10^8 \text{ m}^3 < V_F < 2.2 \times 10^8 \text{ m}^3$ (the interval between guides A and C in Fig. 4a). This volumetric range, where the cumulative probability rises steeply, highlights the size of most floods.

Finally, we use the distribution in Figure 4a to characterize flood timing. We assume that the probability of a flood having volume $\leq V_F$ is also the probability that the lake filled to this volume should have already outburst. Hence, by calculating how long the lake takes to fill to this volume (from empty), we can determine the probability of a flood having occurred by this time after the date of the last flood. By varying V_F in this calculation, we can also determine how this probability changes with time. Figure 4b shows such an *outburst probability curve* calculated with flood 3 ($t = 1957.7$ years) as the starting time reference. At any point on the curve with probability p , the time value has been found by integrating Eqn (1a) from $t = 1957.7$ years until the lake reaches a volume that equals the V_F value corresponding to the same probability p in Figure 4a.

Since the outburst probability curve combines information from the flood-volume distribution, the last flood date and the weather-dependent lake supply Q_i (which is seasonal, with peaks in summer and troughs in winter), probability curves constructed for other times using other floods in Table 1 (or future floods) as the starting reference

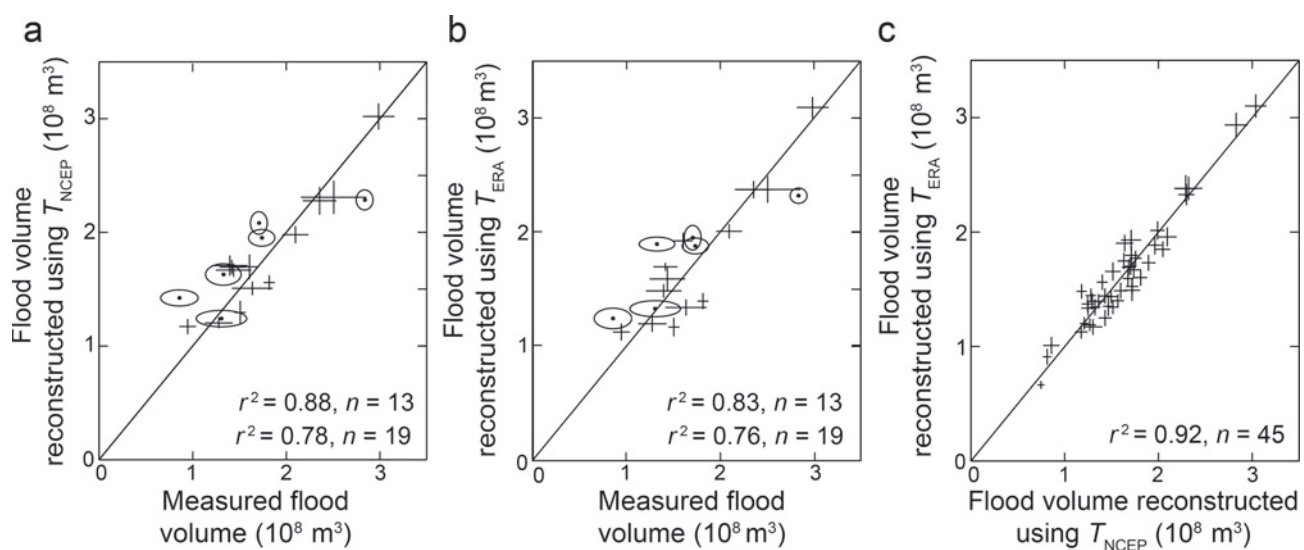


Fig. 3. Comparison of 19 measured flood volumes with flood volumes reconstructed by integrating the lake water supply sub-model forced by (a) T_{NCEP} and (b) T_{ERA} reanalysis temperature data. (c) Comparison between the flood volumes reconstructed using these temperature forcings. In (a) and (b), crosses display these comparisons for the 13 flood volumes used in deriving this sub-model's parameters (Table 2), and ellipses display these comparisons for 6 additional flood volumes (Section 3.3). The vertical and horizontal size of crosses and ellipses indicate the error in the reconstructed and measured flood volumes.

Table 3. Summary of our five flood-date prediction models

No.	Model name	Label	Equation	Tunable parameter (unit)	Parameter meaning
1	Constant Date	CD		Δ (day)	Calendar date on which flood occurs
2	Constant Threshold	CT	$h_c = \text{constant}$	h_c (m)	Constant outburst threshold depth
3	Variable Threshold (Lake-rise rate)	VT_h	Eqn (3)	$h_{0\beta}$ (m) β (d^{-1})	Value of outburst threshold depth if $dh/dt=0$ Sensitivity of threshold on the rate of lake-level rise
4	Variable Threshold (Temperature)	VT_T	Eqn (4)	$h_{0\lambda}$ (m) λ ($m^\circ C^{-1}$)	Value of outburst threshold depth if $T=T_0$ Sensitivity of threshold on temperature
5	Divide Migration	DM	Eqn (5)	h_{ca} (m) α (d^{-1})	Outburst threshold depth of the lake if there is no delay in the divide's arrival at the lake Rate constant of divide migration

will be different. However, Figure 4b illustrates key features of such curves. Its shape derives from the curve in Figure 4a, with the steep rise in outburst probability between A' and B' corresponding to the rise in cumulative probability between A and B in Figure 4a. This rise occurs in 80 days (between 1958.53 and 1958.75 years). This is a manifestation of the ‘focusing effect’ discussed by Ng and Liu (2009, p. 611 and fig. 11), who showed that abundant meltwater supply during summer fills Merzbacher Lake rapidly and concentrates most of its outbursts in this period. In contrast, low supply towards the end of the melt season and through winter causes a slow increase in outburst probability with time. Figure 4b shows that the interval $B'-C'$ (corresponding to the interval B–C in Fig. 4a) is 200 days long but experiences only a 10% increase in outburst probability.

While useful, these results offer predictions whose validity relies wholly on past empirical data and that are probabilistic, unable to tell us each flood's size and timing. For instance, although Figure 4b indicates a high probability of flood 4 occurring by late August ($t \approx 1958.65$, $p = 0.5$) and very high probability of this before early November (guide B' , $p = 0.81$), the actual flood came later, on 24 November. We turn to specific flood-date prediction based on the model next.

4. FLOOD-DATE PREDICTION MODELS: THRESHOLD ASSUMPTIONS

In this section we detail the four threshold assumptions to be used with Eqns (1) and (2) to hindcast the Merzbacher flood dates. Three of them concern the lake depth threshold h_c ; the fourth invokes a mobile subglacial seal, as mentioned before. We label each assumption with an abbreviation (Table 3).

4.1. Constant Date (CD) model, and measures of prediction error

We first consider a naïve prediction model, called the Constant Date (CD) model, that does not implement an outburst threshold nor simulate filling of the lake, unlike the Ng–Liu model. The CD model postulates that the flood occurs on the same date each year, the motivation being that most Merzbacher flood dates cluster in the months July–September (see histogram in fig. 2 of Ng and Liu, 2009). Although this model neglects environmental influence on the system, it sets a benchmark for assessing other prediction models: those failing to match its performance are practically useless. In introducing it here, we also consider how to quantify prediction success.

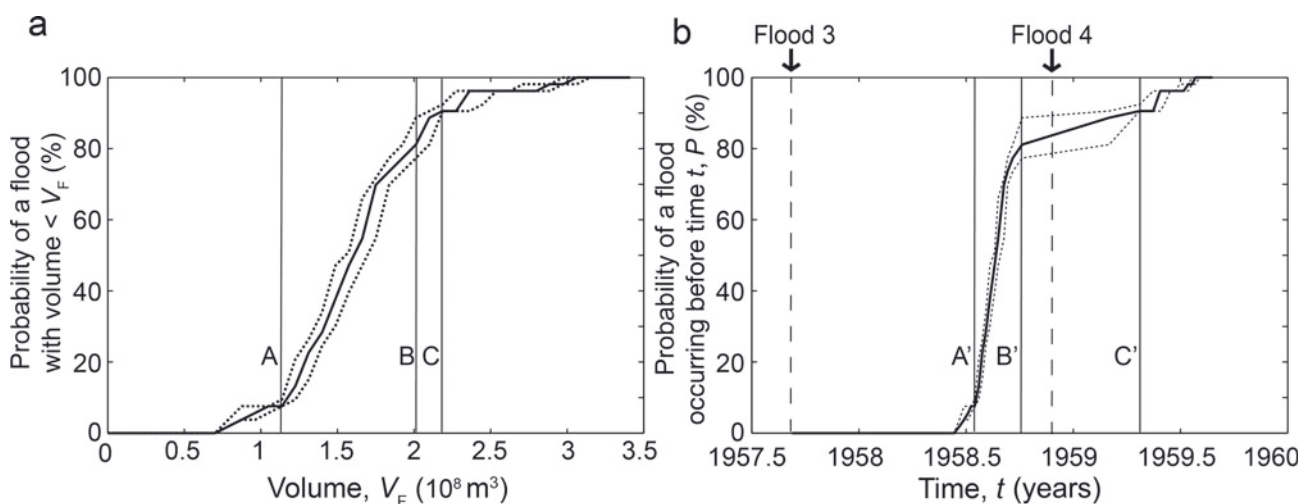


Fig. 4. Probability distributions of flood volumes and timing. (a) Percentage of jökulhlaups from Merzbacher Lake whose reconstructed volumes are $<V_F$. (b) Probability of a flood having occurred before time t , given the modelled lake water supply between floods 3 and 4. The results in both (a) and (b) derive from modelling where T_{NCEP} temperature data have been used to calculate the lake water supply. Dotted lines indicate upper and lower bounds on the volumes and probabilities based on the errors in Figure 3. The dates of floods 3 and 4 (7 September 1957 and 24 November 1958) are marked by the vertical dashed lines. A, A' , B, B' , C and C' label the vertical guides described in Section 3.4.

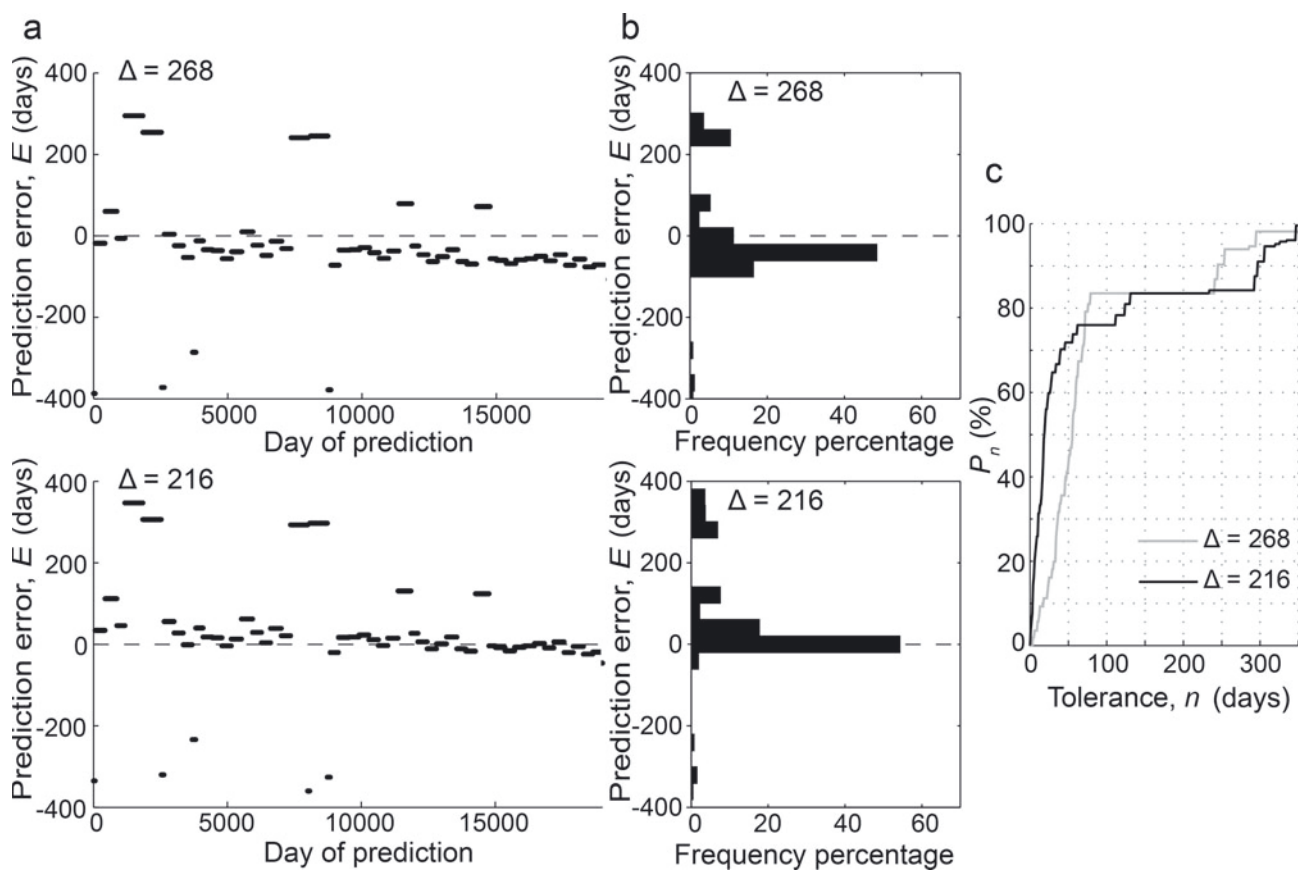


Fig. 5. Prediction errors obtained by running the Constant Date (CD) model with two different values of the model parameter, $\Delta = 268$ days and $\Delta = 216$ days. (a) Prediction errors for each day of the study period, (b) error frequency histograms and (c) the percentage of prediction errors within the tolerance n .

The CD model assumes a fixed calendar date Δ (expressed in day of year) for all floods, where Δ is chosen by the forecaster to optimize prediction success. Using two trial values, $\Delta = 268$ (25 September in non-leap years) and $\Delta = 216$ (4 August in non-leap years), we have used this model to hindcast the next flood date on every day of the study period. Each of these two sets of runs yields 19 006 hindcasts. Figure 5a shows their prediction errors, E , found by differencing each hindcast and the known date of the next flood (we define E to be positive if the hindcast is early). Although the errors span a large range, most of them cluster within 100 days. The outlier errors are associated with floods that occurred after unusually long or short lake-filling periods.

A straightforward quantifier of the errors in Figure 5a is their RMS, and we can find the best Δ that minimizes the RMS. But the RMS is not the only measure of prediction performance, nor necessarily the best measure, as it emphasizes outliers. Other measures may be more desirable from a forecaster's viewpoint. For example, a good forecasting model may not be one with the lowest errors overall but one that gives the greatest number of 'successful' forecasts, i.e. forecasts near the actual flood dates. The histograms in Figure 5b show the percentage frequency of the errors in the two trial runs. Their central bars, each 40 days wide, show that if we define successful hindcasts as those having errors within ± 20 days ($|E| \leq 20$), then the run assuming $\Delta = 216$ yields five times more successful hindcasts (54.3%) than the run assuming $\Delta = 268$ (11.2%). This percentage offers a measure of prediction success, and we

call it P_{20} (the subscript indicates the tolerance). In contrast to the RMS, P_{20} emphasizes hindcasts that are relatively accurate over those that are out by a long way. It is a useful measure for the CD hindcasts because they miss some floods by over a year.

We proceed to optimize the CD model by making prediction runs with different Δ values and calculating the corresponding RMS and P_{20} . The results (Fig. 6) show that the RMS is minimized at 121.9 days when $\Delta = 268$ (then $P_{20} = 11.2\%$), whereas P_{20} is maximized at 54.3% when $\Delta = 216$ (then RMS = 133 days). These Δ choices are optimal in their own right; each choice optimizes one measure at the expense of the other, since RMS and P_{20} quantify different kinds of prediction success. In Section 5, these measures and the approach described here are used to optimize all the prediction models before we evaluate their performance.

The tolerance of 20 days in the P_{20} measure is acceptable because jökulhlaup forecasts with such accuracy could be useful, but the tolerance choice depends on what one perceives as accurate. A general measure P_n considers a tolerance of $\pm n$ days, and this motivates another way of quantifying prediction errors. In Figure 5c we track the percentage frequency of successful hindcasts for different tolerance; their window of admittance widens as n increases. The two curves relate the errors from the trial runs of Figure 5a and rise monotonically. The perfect result would be zero errors for all hindcasts, plotting at 100% across the graph; in reality, we seek a curve lying as close to the top-left corner of the graph as possible. Figure 5c shows that the run that optimizes P_{20} ($\Delta = 216$; black curve) excels over the run

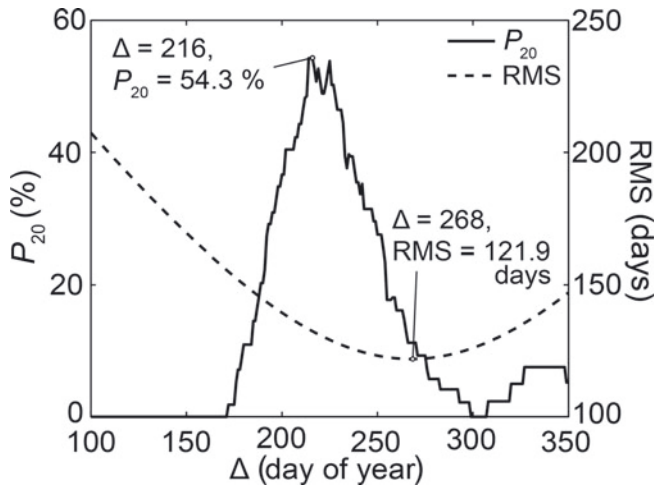


Fig. 6. Dependence of two measures of prediction success, P_{20} and RMS, on the parameter Δ in the CD model. P_{20} is maximized with $\Delta = 216$ and RMS minimized with $\Delta = 268$. (The results in Figure 5 are based on these parameter values.)

that optimizes RMS ($\Delta = 268$; grey curve) for tolerances of tens of day. For $n > 72$ days, the latter run performs better on two occasions, but such tolerance is too large to be useful.

Other measures of prediction performance could be used. For instance, in order to facilitate flood mitigation and emergency evacuation, one might favour forecasts that are early as opposed to late and employ a one-sided definition of P_n based on $0 \leq E \leq n$. Generally, a suite of measures offers more complete characterization of the prediction errors. The forecaster must decide which measures are more important in a given scenario and weigh their outcomes accordingly for decision-making.

4.2. Depth-threshold assumptions

In the real system, weather influences how fast the lake fills, and glacio-hydrological factors govern the outburst condition of a flood, so we expect prediction schemes incorporating these environmental factors to excel over the CD model. The Ng–Liu model is the simplest of such schemes: Eqn (1) has an adjustable outburst threshold h_c , and Eqn (2) captures the weather dependence of lake filling. We posit three assumptions for the behaviour of h_c to be used with this model.

4.2.1. Constant Threshold (CT)

The simplest assumption, called Constant Threshold (CT), is that h_c is constant. Figure 7 shows an example where it is used with the Ng–Liu model to hindcast flood 3 in the ‘Real’ scheme (Section 2). In this run, which assumes $h_c = 86$ m and

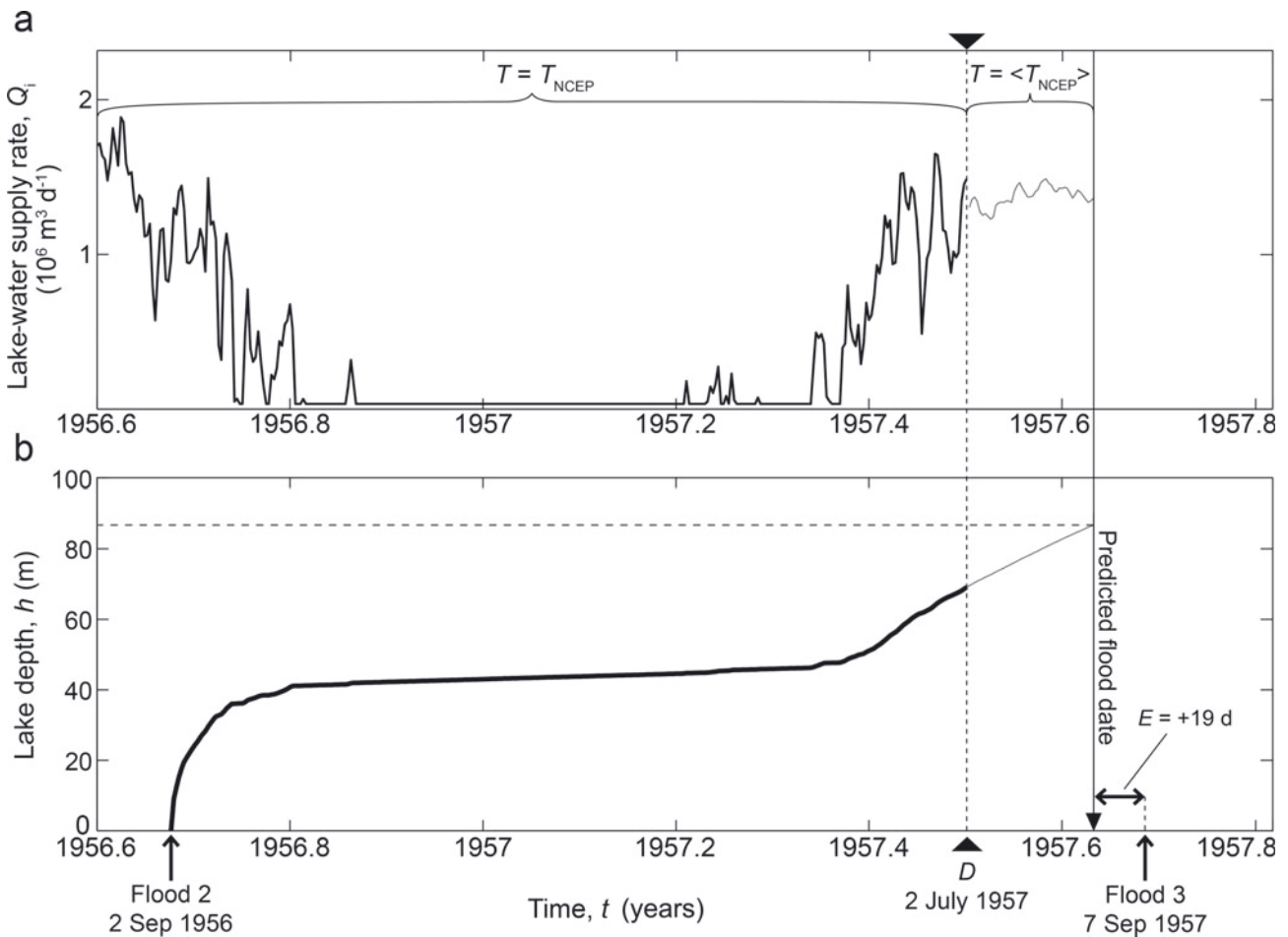


Fig. 7. Example of a Constant Threshold (CT) model prediction run made with the ‘Real’ weather-forcing scheme. On day D (2 July 1957), the forecaster attempts to predict the date of flood 3 (7 September 1957) by simulating histories of (a) lake-water supply rate and (b) lake depth. The lake-water supply model is forced with T_{NCEP} daily temperature data on those days preceding D , and forced with the multi-year mean value of T_{NCEP} afterwards. With the outburst threshold $h_c = 86$ m, the flood-date hindcast is found to be 19 August 1957, and the corresponding prediction error is $E = +19$ days.

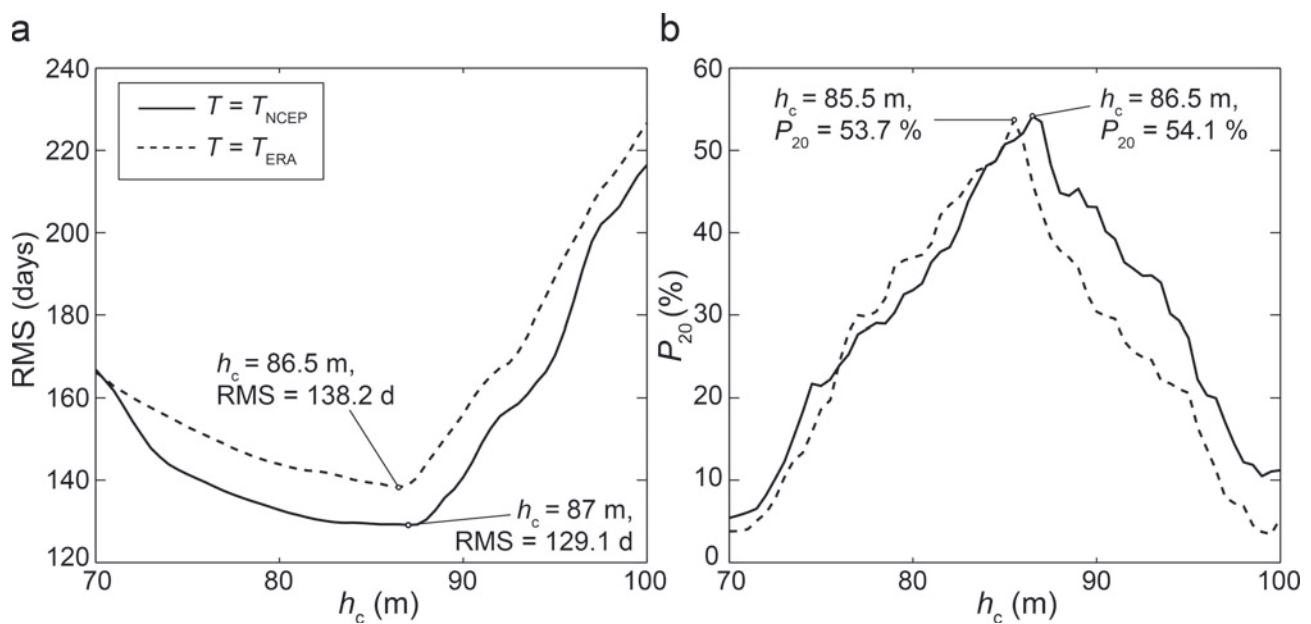


Fig. 8. Dependence of the prediction success of the Constant Threshold (CT) model, as quantified by (a) RMS error and (b) P_{20} , on the threshold value h_c . Solid and dashed lines correspond to results obtained using the T_{NCEP} and T_{ERA} weather forcings respectively.

that the forecaster made the prediction on 2 July 1957 (302 days after flood 2), the flood hindcast is 19 August 1957, 19 days before the actual flood 3. We have made many sets of prediction runs like this to hindcast the Merzbacher flood dates, by adopting different thresholds in the range $70 < h_c < 100$ m and T_{NCEP} or T_{ERA} as weather forcing. Each set of runs used a fixed combination of these inputs to produce next-flood hindcasts on every day of the study period, and we compiled the prediction errors into the RMS and P_{20} as described in Section 4.1. Figure 8 shows how these error measures vary with h_c . For all values of h_c in the imposed range, prediction runs using the NCEP forcing yield lower RMS than runs using the ERA forcing, but the optimal values of h_c are similar: 87 m with NCEP and 86.5 m with ERA (Fig. 8a). P_{20} is maximized when $h_c = 86.5$ m with the NCEP forcing and when $h_c = 85.5$ m with the ERA forcing (Fig. 8b). The value of h_c that optimizes prediction performance thus falls within a narrow range (85.5–87 m). However, with the CT assumption, the lowest achievable RMS is 129.1 days and the highest achievable P_{20} is 54.1%. This performance is worse than that of the CD model (121.9 days, 54.3%). Thus, while using a threshold allows changing weather to be accounted by the model, a fixed threshold does not improve the flood-date predictions in the ‘Real’ scheme.

4.2.2. Variable thresholds (VT_h and VT_T)

Historical changes in h_c , shown by differences between the reconstructed flood volumes in Section 3.4, suggest that one might improve the forecasts by using a threshold that responds to environmental conditions. As discussed in Section 2, this idea originates from Ng and Liu (2009), who reconstructed the lake-level history at Merzbacher Lake to extract the h_c of each flood and the rate of lake-level rise (dh/dt) before it initiated. They found a negative correlation between these two quantities (see their fig. 15b) and interpreted it as being a result of pressure coupling between the lake and the subglacial hydraulics beneath the ice dam.

Motivated by this empirical finding, we explore two formulations of a variable threshold. The first formulation,

which we label VT_h , is the linear model proposed by Ng and Liu (2009):

$$h_c = h_{0\beta} + \beta \frac{dh}{dt}, \quad (3)$$

where $h_{0\beta}$ and β are constants. When using this with Eqn (1) to predict flood dates, we calculate h_c using the exponential moving average of dh/dt (with smoothing constant, $S=0.25$; Brown, 2004). Ng and Liu (2009) determined $h_{0\beta} = 102$ m and $\beta = -58 \text{ d}^{-1}$ for Merzbacher Lake from their correlation.

The second variable-threshold formulation, which we label VT_T , assumes control on h_c by surface temperature rather than the lake-level rise rate. The idea is that meltwater produced at the surface and penetrating to the glacier bed influences the subglacial outburst hydraulics. We suppose

$$h_c = h_{0\lambda} + \lambda(T - T_0)_{0+} \quad (4)$$

where $h_{0\lambda}$ and λ are constants. h_c is calculated using the exponential moving average of $(T - T_0)_{0+}$ (again with $S=0.25$) and the same temperature data, weather-forcing scheme and parameters as used in the lake water supply model in Eqn (2). Since dh/dt depends on T via melt generation and lake filling, Eqn (4) is a more general threshold description than Eqn (3) in the sense that it encapsulates meltwater control on h_c via both the lake pressure and supraglacial water injection to the glacier bed.

When each formulation described here (VT_h or VT_T) is used with Eqns (1a) and (2) to predict flood dates, we optimize the model in the same way as before, by minimizing RMS and maximizing P_{20} , but do so by searching over the corresponding two-parameter space (of $h_{0\beta}$ and β , or $h_{0\lambda}$ and λ).

4.3. Threshold based on subglacial water-divide migration (DM)

Our final threshold assumption treats flood-initiation physics in more detail than any of the previous assumptions. In a novel study, Fowler (1999) extended Nye’s (1976) equations of jökulhlaup mechanics to model the establishment of a subglacial water divide or ‘seal’ under the ice dam in the period between floods. He envisaged that as the lake fills,

Table 4. Results from nine sets of model runs using the ‘Real’ weather-forcing scheme. For each of our five prediction models, model parameters are optimized to minimize RMS and maximize P_{20} . The CT, VT_h , VT_T and DM models are optimized separately for each temperature forcing, T_{NCEP} and T_{ERA} . Displayed in columns 3–8 are the optimal model parameters and the RMS and P_{20} they yield

Model	Temperature forcing	Optimized for RMS			Optimized for P_{20}		
		Optimal parameters	RMS	P_{20}	Optimal parameters	RMS	P_{20}
Constant Date (CD)	–	$\Delta = 268$ days	121.9	11.2	$\Delta = 216$ days	132.4	54.3
Constant Threshold (CT)	T_{NCEP}	$h_c = 87.0$ m	129.2	51.7	$h_c = 86.5$ m	129.3	53.3
	T_{ERA}	$h_c = 86.5$ m	138.3	45.9	$h_c = 85.5$ m	139.1	53.6
Variable Threshold (Lake-rise rate) (VT_h)	T_{NCEP}	$h_{0\beta} = 80.0$ m, $\beta = 90$ days	112.6	15.3	$h_{0\beta} = 77.0$ m, $\beta = 30$ days	123.0	57.4
	T_{ERA}	$h_{0\beta} = 78.0$ m, $\beta = 90$ days	121.4	19.6	$h_{0\beta} = 75.5$ m, $\beta = 34$ days	133.9	57.5
Variable Threshold (Temperature) (VT_T)	T_{NCEP}	$h_{0\lambda} = 76.0$ m, $\lambda = 3.0$ m °C ⁻¹	109.6	11.3	$h_{0\lambda} = 76.5$ m, $\lambda = 0.8$ m °C ⁻¹	123.5	57.4
	T_{ERA}	$h_{0\lambda} = 76.0$ m, $\lambda = 4.0$ m °C ⁻¹	118.0	11.6	$h_{0\lambda} = 76.5$ m, $\lambda = 0.9$ m °C ⁻¹	133.6	57.3
Divide Migration (DM)	T_{NCEP}	$h_{c\alpha} = 78.0$ m, $\alpha = 0.015$ d ⁻¹	117.2	27.1	$h_{c\alpha} = 84.5$ m, $\alpha = 0.170$ d ⁻¹	127.6	55.0
	T_{ERA}	$h_{c\alpha} = 78.0$ m, $\alpha = 0.015$ d ⁻¹	127.9	22.2	$h_{c\alpha} = 82.5$ m, $\alpha = 0.080$ d ⁻¹	136.6	54.5

the seal migrates towards the lake in response to changing subglacial water pressure and initiates the flood when it reaches the lake. The rate of seal migration depends on the rate of lake-level rise and the conditions of channelized subglacial drainage (e.g. hydraulic gradients associated with glacier geometry). Faster filling of the lake (due to greater meltwater input Q_i) results in a higher lake ‘highstand’ by the time the seal arrives at the lake to start a flood. Although this theory has not been validated by subglacial observations, it highlights mechanistic links between environmental factors and flood initiation.

Fowler’s (1999) equations are not used here because the basal topography of South Inylchek Glacier and the attendant hydraulic potential gradients are poorly known. However, we mimic his seal dynamics by a lower-order model that parameterizes seal migration as a function of the lake water pressure:

$$\frac{dY}{dt} = -\alpha \left[Y - \left(\frac{h_{c\alpha} - h}{h_0} \right) \right]. \quad (5)$$

In this ‘Divide Migration’ (DM) model, h is the lake depth as before, Y measures the dimensionless distance of the seal from the lake, h_0 is the ice-dam flotation depth (estimated at 108 m by Ng and others, 2007), and $h_{c\alpha}$ and α are constants. The outburst condition $h = h_c$ in the Ng–Liu model is replaced by $Y = 0$, and we modify Eqn (1) to

$$\begin{aligned} \frac{dV}{dt} &= Q_i \quad \text{when } Y > 0, \\ V &\rightarrow 0 \quad \text{when } Y = 0. \end{aligned} \quad (6)$$

During filling, this model describes coupled evolution of lake volume and seal position, with the seal migrating towards the target position $(h_{c\alpha} - h)/h_0$, which itself moves. The constant α controls the migration rate, and $h_{c\alpha}$ is the theoretical outburst depth of the lake if there were no delay in the seal’s arrival at the lake. The water supply Q_i is calculated by Eqn (2) as before.

Equations (5) and (6a) reproduce the behaviour of Fowler’s (1999) model. In each prediction run, we integrate them by the finite-difference method, starting with $V = Y = 0$ on the day after the last flood. Low lake level initially locates the target down-glacier from the lake, so the seal migrates into $Y > 0$. But filling of the lake relocates the target up-glacier from the dam eventually (when $(h_{c\alpha} - h)/h_0 < 0$), attracting the seal back towards the lake after some delay. As with the

models in Section 4.2, we optimize the success of this DM model by tuning α and $h_{c\alpha}$ over their parameter space.

5. FLOOD-DATE PREDICTION EXPERIMENTS: RESULTS AND DISCUSSION

Table 3 summarizes our five prediction models. (For convenience we use the word ‘model’ to refer to the three versions of the Ng–Liu model with different thresholds, as well as to the CD and DM models.) In this section we analyse their performance in prediction runs made to hindcast the Merzbacher flood dates. In total, 18 sets of runs were made. The first 9 sets use the ‘Real’ scheme for weather forcing and consist of 1 set of runs for the CD model and 2 sets of runs each for the CT, VT_h , VT_T and DM models (one set with NCEP forcing, the other with ERA forcing). In each set of runs, model parameters were tuned to optimize prediction success separately in terms of RMS and P_{20} . Table 4 shows the results, listing in its columns the optimal model parameters and the RMS and P_{20} achieved when each of these is optimized. The other 9 sets of runs are structured identically but use the ‘Simple’ scheme for weather forcing. Table 5 shows the corresponding results.

Several matters guide the following analysis of these results: (1) model performance and the floods’ predictability, (2) influences of model complexity and weather forcing on prediction success and (3) the impact of weather uncertainty on the predictions. We cover these matters in Sections 5.1 and 5.2. In Section 5.3 we ask how the reliability of predictions varies with time, and this leads us to suggest an ensemble prediction strategy that uses multiple prediction models to derive forecasts.

5.1. Prediction performance of the models

Since RMS and P_{20} are fundamentally different quantifiers of the prediction errors, a model optimized for RMS must yield lower RMS than the same model optimized for P_{20} , and a model optimized for P_{20} must yield higher P_{20} than the same model optimized for RMS. Table 4 confirms this expectation (compare the RMSs in column 4 with those in column 7, and the P_{20} ’s in column 8 with those in column 5). Accordingly we focus our comparisons here on columns 4 and 8. These show RMS around 120 days and P_{20} of 50–60%. The RMS values are not useful quantifiers of our models’

Table 5. Results from nine sets of model runs using the ‘Simple’ weather-forcing scheme, where archived weather is used throughout each prediction (see Section 2). The layout is identical to that of Table 4

Model	Temperature forcing	Optimized for RMS			Optimized for P_{20}		
		Optimal parameters	RMS	P_{20}	Optimal parameters	RMS	P_{20}
Constant Date (CD)	–	$\Delta = 268$ days	121.9	11.2	$\Delta = 216$ days	132.4	54.3
Constant Threshold (CT)	T_{NCEP}	$h_c = 83.0$ m	129.8	43.5	$h_c = 87.5$ m	137.7	50.6
	T_{ERA}	$h_c = 84.0$ m	139.4	52.6	$h_c = 84.0$ m	139.4	52.6
Variable Threshold (Lake-rise rate) (VT_h)	T_{NCEP}	$h_{0\beta} = 76.0$ m, $\beta = 99$ days	108.9	21.1	$h_{0\beta} = 76.5$ m, $\beta = 24$ days	126.1	55.6
	T_{ERA}	$h_{0\beta} = 78.5$ m, $\beta = 118$ days	120.3	11.6	$h_{0\beta} = 74.0$ m, $\beta = 52$ days	131.0	62.2
Variable Threshold (Tem- perature) (VT_T)	T_{NCEP}	$h_{0\lambda} = 74.5$ m, $\lambda = 3.6$ m °C ⁻¹	108.1	12.2	$h_{0\lambda} = 75.0$ m, $\lambda = 1.0$ m °C ⁻¹	123.1	56.7
	T_{ERA}	$h_{0\lambda} = 70.5$ m, $\lambda = 5.0$ m °C ⁻¹	117.5	12.3	$h_{0\lambda} = 79.5$ m, $\lambda = 0.8$ m °C ⁻¹	131.8	58.4
Divide Migration (DM)	T_{NCEP}	$h_{c\alpha} = 79.0$ m, $\alpha = 0.02$ d ⁻¹	116.0	25.7	$h_{c\alpha} = 85.5$ m, $\alpha = 0.15$ d ⁻¹	127.9	51.3
	T_{ERA}	$h_{c\alpha} = 80.0$ m, $\alpha = 0.02$ d ⁻¹	126.4	21.1	$h_{c\alpha} = 79.0$ m, $\alpha = 0.06$ d ⁻¹	134.5	53.5

performances as they are disappointingly large and dominated by outlier errors (Section 4.1). Thus, although Table 4 provides findings for both measures, forecasters of the Merzbacher floods might choose P_{20} as the main criterion for identifying the best model.

Considering the results in both columns 4 and 8 in Table 4, the overall pattern is that the Variable Threshold models (VT_h and VT_T) perform best, followed by the DM model, then the CT model. This pattern is robust for each temperature forcing. Also, the CD model performs better than the CT model (as noted before), implying that a fixed threshold does not enhance prediction accuracy. Compared with the other models, the CD model performs nearly as well as the DM model but consistently worse than both Variable Threshold models. The latter models yield $P_{20} = 57.3$ – 57.5% , meaning that we can hindcast the floods accurately to within 20 days $\sim 57.4\%$ of the time. We have made additional comparisons with the generalized measure P_n (defined in Section 4.1), which show that the Variable Threshold models surpass all other models in the practical tolerance range $n \leq 20$. These results support Ng and Liu’s (2009) finding that dh/dt plays a role in flood-initiation physics. However, in our VT_h model the optimal β -values are positive, not negative as found by those authors; our threshold h_c needs to increase with dh/dt for successful prediction. One possible explanation of this sign difference is the fact that the correlation analysis of Ng and Liu (see the description leading up to Eqn (3) in Section 4.2) and our fitting of flood hindcasts to observed flood dates do not amount to the same optimization constraint for β .

Reassuringly, Table 4 shows that models incorporating environmental conditions in their determination of the outburst threshold (e.g. the rates of meltwater input to the lake and lake-level rise) do better than the CD model. But does prediction success always improve with model complexity? A model with richer physics and more parameters ought to yield a closer fit between hindcasts and observations, unless it is faulty. Hence we expect model performance to improve in the order: CD, CT, VT_h/VT_T , DM. Our results upset this trend in two ways. Firstly, the DM model performs worse than the VT models. This suggests that either Fowler’s (1999) seal migration does not capture what happens at Merzbacher Lake, or that it does but the DM model incorrectly mimics his mechanism. However, Table 4 shows that the DM model improves upon the CT model in all

cases. We can explain this because these models are related, with the CT model being the limit of the DM model as $\alpha \rightarrow \infty$: then the seal responds infinitely fast to lake-level changes to track the target, and the outburst condition $Y=0$ in Eqn (5) becomes equivalent to $h = h_{c\alpha}$ (constant). Seen in this light, the DM model performs better because it has an extra degree of freedom over the CT model.

Secondly, the benchmark CD model outperforms our CT models, despite the latter models incorporating the physical concept of an outburst threshold and the effect of variable weather conditions on the lake-filling rate. This is significant because, given our lack of understanding of the details of flood-initiation physics, a likely first step for flood forecasters aiming to include simple physics in their models and improve their flood predictions is to assume a constant outburst threshold and simulate the filling of the lake physically (with an equation like our Eqn (2)). Our results show that unless the threshold is allowed to vary with environmental conditions, this approach does not yield better flood hindcasts than simply assuming floods occur on the same day every year.

5.2. Influence of weather forcing on prediction success

How does weather uncertainty influence our predictions? Such uncertainty features in the model runs implementing the ‘Real’ scheme in two ways: (1) via the difference between the NCEP and ERA temperature data, which themselves are estimates of a weather variable, and (2) via the fact that the forecaster does not know the future weather. We assess the influence from both angles.

Considering the data sources first, Table 4 shows that the NCEP forcing yields consistently lower optimal RMS values than the ERA forcing, while the optimal P_{20} values are almost independent of the choice of forcing. Thus, NCEP temperature data seem to be only marginally better as a forcing for predicting the Merzbacher floods. As noted before, the choice of temperature forcing does not change the overall ordering of model performance (CT, DM, VT_h/VT_T) in Table 4.

And what of the need to forecast future weather in the ‘Real’ scheme? We assess the impact of this on prediction performance by comparing Table 4 with Table 5, focusing again on the optimal RMS and P_{20} results in columns 4 and 8. The ‘Simple’ scheme ought to perform better than the

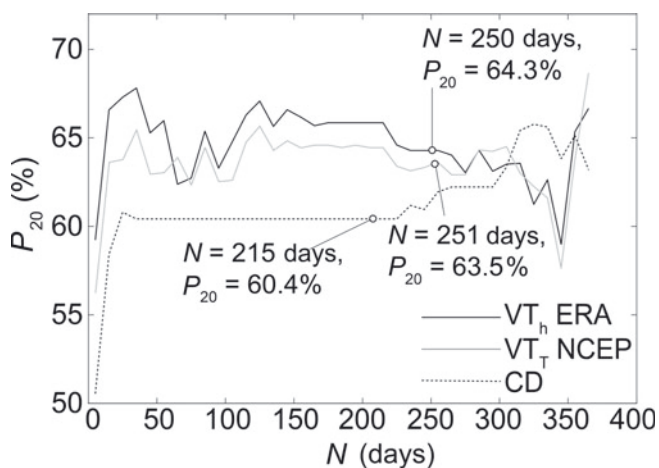


Fig. 9. The variation of prediction success (as quantified by P_{20}) with the predicted time before the next flood, N , for three models: the Constant Date (CD) model, the Variable Threshold model VT_h with T_{ERA} temperature forcing, and the Variable Threshold model VT_T with T_{NCEP} temperature forcing.

'Real' scheme because it uses reanalysis data rather than forecasts of future daily temperature to derive the lake-water supply, and this eliminates a source of uncertainty. Tables 4 and 5 show that only the Variable Threshold models that use the ERA forcing fulfil this expectation. In contrast, the Variable Threshold models that use the NCEP forcing perform worse in the 'Simple' scheme than in the 'Real' scheme. The same is true of the CT and DM models, but the poorer performance of these models suggests that they are limited more by other deficiencies (notably in their threshold assumptions) than by weather uncertainty. It is less clear why the Variable Threshold models with NCEP forcing do not perform better in the 'Simple' scheme. A possible explanation is that short-term weather fluctuations are more accurately represented by T_{ERA} than by T_{NCEP} , even though these temperature forcings produce similar meltwater volumes on seasonal and annual timescales (e.g. Fig. 3c) and have similar multi-year means, so that their flood-prediction performances in the 'Real' scheme are similar (Table 4).

5.3. The forecaster's time frame and ensemble prediction

Having used the RMS and P_{20} of the hindcast errors to evaluate the models, we turn to a different aspect of the forecasting problem. As each prediction is made, the forecaster may wish to know whether its reliability depends on how far the next flood lies in the future. Given the impact of weather uncertainty examined above, a reasonable hypothesis is that predictions close to the flood are more successful than those made far in advance. Here we study this time dependence using hindcast data from our optimized model runs for the two best models in Table 4: VT_h with T_{ERA} temperature forcing, and VT_T with T_{NCEP} temperature forcing.

On the day when each prediction is made, D , the forecaster cannot in fact determine its time difference from the next flood because the flood has not yet occurred. However, the forecaster knows the time difference between D and the predicted flood date. We denote by N this 'predicted time before the next flood'. N is easy to calculate for our hindcast data.

Figure 9 shows how the success of hindcasts (measured with P_{20}) varies with N in the two Variable Threshold models, with the data for N arranged in 10-day bins to suppress noise on the distributions. Results of the CD model are included for comparison. In practice, after predicting a flood date with a given model, the forecaster can look up the corresponding distribution on this plot to learn the probability that the forecast is accurate to within 20 days. We see that both Variable Threshold models excel over the CD benchmark for most of the range in N , and the VT_h model with the ERA temperature forcing performs best overall. The distributions of both Variable Threshold models seem to be consistent with our hypothesis at the beginning of this subsection, as they show a negative (albeit weak) trend. In the 'Real' scheme, prediction runs that forecast an imminent flood (with a small value of N) would have used more archived temperature data and fewer forecasts of future weather for deriving the lake-water supply and hence be more likely to be accurate. The CD model shows no such trend as it does not incorporate weather forcing.

The empirical probabilities in Figure 9 motivate an *ensemble prediction strategy*, as follows. The forecaster predicts the next flood date using each of the models (VT_h , VT_T and CD), then calculates the values of N for these predictions, and reads the corresponding probabilities of prediction success from Figure 9. In the simplest strategy, the forecast having the highest P_{20} is taken as the best guess of the flood date. Consider using this ensemble method on 1 January 1959 (this is 38 days after the date of flood 4, 24 November 1958) to forecast flood 5 (19 September 1959). The CD model predicts the next flood on 4 August 1959, whereas the VT_h and VT_T models predict it on 8 September 1959 and 9 September 1959, respectively. The circles in Figure 9 mark the corresponding values of N (215 days, 250 days, 251 days) and P_{20} (60.4%, 64.3%, 63.5%). In this example, the forecaster learns that the best forecast is the one by the VT_h model, 8 September 1959. The P_{20} of this forecast (64.3%) is much higher than the mean P_{20} of this model in Table 4 (57.5%). In fact, this forecast is successful (accurate to within 20 days) and has a true error of $E=11$ days, while the true errors of the other two forecasts by the VT_T and CD models are 10 days and 46 days, respectively. This ensemble strategy allows the forecaster to choose between models based on empirical experience.

6. CONCLUSIONS AND OUTLOOK

Low-order models that implement an outburst threshold based on the lake-water depth can give useful predictions of the date of jökulhlaups from Merzbacher Lake. Our best models (VT_h and VT_T) assume a variable threshold depth governed by weather. They hindcast observed flood dates successfully to within ± 20 days 57.4% of the time, excelling over a benchmark model that assumes a constant flood date each year. For the Merzbacher Lake–Inylchek Glacier system, these models can be readily incorporated into practical flood-forecasting schemes, and may aid decisions regarding the development of hydropower down-valley and management of the corresponding reservoirs (Mamatkanov and Deng, 2011). The present work complements the theory by Ng and Liu (2009) that addressed mechanisms underlying the long-term timing pattern of multiple floods.

Although we can understand why predictions made far in advance of flood events are generally less reliable (Section

5.3), it is less clear why weather uncertainty impacts differently on the success of model predictions made with NCEP and ERA temperature forcings (Section 5.2). A limitation of our current study is its use of these reanalysis data, which are themselves uncertain estimates of the past weather. By comparing them to meteorological measurements at the lake, future work will evaluate how severely such uncertainty affects jökulhlaup predictability. For example, reanalysis data may prove completely incapable of reproducing short-term (e.g. daily) weather variations, limiting the prediction ability of models that use such data.

An ability to forecast flood dates to within ± 20 days a little more than half of the time is not overwhelmingly successful, and reflects how poorly we understand the physics of flood initiation. However, the relative success of our prediction models may reveal aspects of these physics. For example, assuming a constant outburst threshold – one of the simplest possible physically based assumptions – completely fails to yield useful flood predictions. In contrast, our most successful models allow the threshold to vary with weather conditions. This supports Ng and Liu's (2009) inference that, at Merzbacher Lake, the rate of change of lake depth (dh/dt) influences the outburst threshold through subglacial seal dynamics. Also revealing is the failure of our DM model, designed to mimic the behaviour of Fowler's (1999) model of seal dynamics. How closely our toy divide model mimics Fowler's full thermomechanical model is presently unclear; however, the fact that models with thresholds that are linearly dependent on lake-filling (VT_h) and air temperature (VT_T) outperform the divide model suggests that dh/dt may not influence the threshold in the way predicted by Fowler's (1999) theory. When the bed geometry of Inylchek Glacier along the flood path has been accurately determined, it will be worth revisiting the full model of Fowler (1999) and quantifying its prediction ability.

Future work could investigate several model extensions. First, observations show that after a lake empties in a jökulhlaup, water entering the lake basin may flow directly into the glacier dam as an open subglacial stream for some time before the lake reforms (e.g. Bartholomaeus and others, 2011). The duration of this flow presumably depends on the lake-water input rate and the characteristics of the preceding flood; a large flood might cause significant delay before the lake begins to refill. Such a scenario could be incorporated into our models by considering the open-channel hydraulics. Second, the rate of calving from the ice dam, c , which affects lake-water balance (see Eqn (2)), may vary in time and depend on factors such as the lake-water depth. One could account for this variability in our models using 'calving laws' (e.g. Benn and other, 2007). Third, the ensemble prediction strategy in Section 5.3 can be developed to exploit weighted averages of the predictions from different models. A further valuable extension will be to combine the prediction of flood timing as examined here with thermomechanical modelling of flood-discharge evolution to forecast the duration and magnitude (peak size and volume) of each flood.

ACKNOWLEDGEMENTS

We thank Liu Siyin for providing the last three flood dates in Table 1. J.K. acknowledges the support of a University of Sheffield PhD Studentship. We thank Joseph Walder and an anonymous reviewer for positive and constructive reviews that have helped us clarify and strengthen the paper.

REFERENCES

- Aizen VB and Aizen EM (1998) Estimation of glacial runoff to the Tarim River, central Tien Shan. *IAHS Publ.* 248 (Symposium at Merano 1998 – *Hydrology, Water Resources and Ecology in Headwaters*), 191–98
- Aizen VB, Aizen EM and Kuzmichenok VA (2007) Glaciers and hydrological changes in the Tien Shan: simulation and prediction. *Environ. Res. Lett.*, **2**(4), 045019 (doi: 10.1088/1748-9326/2/4/045019)
- Anderson RS, Walder JS, Anderson SP, Trabant DC and Fountain AG (2005) The dynamic response of Kennicott Glacier, Alaska, USA, to the Hidden Creek Lake outburst flood. *Ann. Glaciol.*, **40**, 237–242 (doi: 10.3189/172756405781813438)
- Barnett TP, Adam JC and Lettenmaier DP (2005) Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature*, **438**(7066), 303–309 (doi: 10.1038/nature04141)
- Bartholomaeus TC, Anderson RS and Anderson SP (2011) Growth and collapse of the distributed subglacial hydrologic system of Kennicott Glacier, Alaska, USA, and its effects on basal motion. *J. Glaciol.*, **57**(206), 985–1002 (doi: 10.3189/002214311798843269)
- Bell RE, Studinger M, Shuman CA, Fahnstock MA and Joughin I (2007) Large subglacial lakes in East Antarctica at the onset of fast-flowing ice streams. *Nature*, **445**(7130), 904–907 (doi: 10.1038/nature05554)
- Benn DI, Hulton NRJ and Mottram RH (2007) 'Calving laws', 'sliding laws' and the stability of tidewater glaciers. *Ann. Glaciol.*, **46**, 123–130 (doi: 10.3189/172756407782871161)
- Björnsson H (2003) Subglacial lakes and jökulhlaups in Iceland. *Global Planet. Change*, **35**(3–4), 255–271 (doi: 10.1016/S0921-8181(02)00130-3)
- Björnsson H (2004) Glacial lake outburst floods in mountain environments. In Owens P and Slaymaker O eds. *Mountain geomorphology*. Edward Arnold, London, 165–184
- Brown RG (2004) *Smoothing, forecasting and prediction of discrete time series*. Courier Dover, New York
- Burke MJ, Woodward J, Russell AJ and Fleisher PJ (2009) Structural controls on englacial esker sedimentation: Skeiðarárjökull, Iceland. *Ann. Glaciol.*, **50**(51), 85–92 (doi: 10.3189/172756409789097568)
- Carrivick JL (2007) Hydrodynamics and geomorphic work of jökulhlaups (glacial outburst floods) from Kverkfjöll volcano, Iceland. *Hydrol. Process.*, **21**(6), 725–740 (doi: 10.1002/hyp.6248)
- Cenderelli DA and Wohl EE (2003) Flow hydraulics and geomorphic effects of glacial-lake outburst floods in the Mount Everest region, Nepal. *Earth Surf. Process. Landf.*, **28**(4), 385–407 (doi: 10.1002/esp.448)
- Clague JJ and Mathews WH (1973) The magnitude of jökulhlaups. *J. Glaciol.*, **12**(66), 501–504
- Clarke GKC (1982) Glacier outburst floods from 'Hazard Lake', Yukon Territory, and the problem of flood magnitude prediction. *J. Glaciol.*, **28**(98), 3–21
- Clarke GKC (2003) Hydraulics of subglacial outburst floods: new insights from the Spring–Hutter formulation. *J. Glaciol.*, **49**(165), 299–313 (doi: 10.3189/172756503781830728)
- Evatt GW, Fowler AC, Clark CD and Hulton NRJ (2006) Subglacial floods beneath ice sheets. *Philos. Trans. R. Soc. London, Ser. A*, **364**(1844), 1769–1794 (doi: 10.1098/rsta.2006.1798)
- Flowers GE, Björnsson H, Pálsson R and Clarke GKC (2004) A coupled sheet–conduit mechanism for jökulhlaup propagation. *Geophys. Res. Lett.*, **31**(5), L05401 (doi: 10.1029/2003GL019088)
- Fowler AC (1999) Breaking the seal at Grímsvötn, Iceland. *J. Glaciol.*, **45**(151), 506–516
- Fowler AC (2009) Dynamics of subglacial floods. *Proc. R. Soc. London, Ser. A*, **465**(2106), 1809–1828 (doi: 10.1098/rspa.2008.0488)

- Glen JW (1954) The stability of ice-dammed lakes and other water-filled holes in glaciers. *J. Glaciol.*, **2**(15), 316–318
- Huss M, Bauder A, Werder M, Funk M and Hock R (2007) Glacier-dammed lake outburst events of Gornersee, Switzerland. *J. Glaciol.*, **53**(181), 189–200 (doi: 10.3189/172756507782202784)
- Jóhannesson T (2002) Propagation of a subglacial flood wave during the initiation of a jökulhlaup. *Hydrol. Sci. J.*, **47**(3), 417–434
- Kalnay E and 21 others (1996) The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.*, **77**(3), 437–471 (doi: 10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2)
- Kingslake J and Ng F (2013) Modelling the coupling of flood discharge with glacier flow during jökulhlaups. *Ann. Glaciol.*, **54**(63), 25–31 (doi: 10.3189/2013AoG63A331)
- Kistler R and 12 others (2001) The NCEP/NCAR 50-year reanalysis: monthly means CD-ROM and documentation. *Bull. Am. Meteorol. Soc.*, **82**(2), 247–267 (doi: 10.1175/1520-0477(2001)082<0247:TNNYRM>2.3.CO;2)
- Liestøl O (1956) Glacier dammed lakes in Norway. *Nor. Geogr. Tidsskr.*, **15**(3–4), 122–149 (doi: 10.1080/00291955608542772)
- Liu J (1992) Jökulhlaups in the Kunmalike River, southern Tien Shan Mountains, China. *Ann. Glaciol.*, **16**, 85–88
- Magnússon E, Rott H, Björnsson H and Pálsson F (2007) The impact of jökulhlaups on basal sliding observed by SAR interferometry on Vatnajökull, Iceland. *J. Glaciol.*, **53**(181), 232–240 (doi: 10.3189/172756507782202810)
- Magnússon E, Björnsson H, Rott H and Pálsson F (2010) Reduced glacier sliding caused by persistent drainage from a subglacial lake. *Cryosphere*, **4**(1), 13–20
- Mamatkanov DM and Deng M (2011) Water and hydropower resources of the Sarydjaz-Kumaryk river and prospects of their use. In Tuzova TV and Dzhumanazarova AZ eds. *Proceedings of AASA Regional Workshop on the Roles of Academies of Sciences in Water and Energy Problems in Central Asia and Ways for their Solution, 30 June–2 July 2011, Bishkek, Kyrgyzstan*. National Academy of Sciences (NAS KR), Bishkek, 190–195 <http://www.interacademies.net/File.aspx?id=16818>
- Mavlyudov BR (1997) Drainage of the ice-dammed Mertzbacher Lake, Tien Shan. *Mater. Glytsiol. Issled./Data Glaciol. Stud.* **81**, 61–65
- Mayer C, Lambrecht A, Hagg W, Helm A and Scharrer K (2008) Post-drainage ice dam response at Lake Merzbacher, Inylchek Glacier, Kyrgyzstan. *Geogr. Ann. A*, **90**(1), 87–96 (doi: 10.1111/j.1468-0459.2008.00336.x)
- Ng F and Björnsson H (2003) On the Clague–Mathews relation for jökulhlaups. *J. Glaciol.*, **49**(165), 161–172 (doi: 10.3189/172756503781830836)
- Ng F and Liu S (2009) Temporal dynamics of a jökulhlaup system. *J. Glaciol.*, **55**(192), 651–665 (doi: 10.3189/002214309789470897)
- Ng F, Liu S, Mavlyudov B and Wang Y (2007) Climatic control on the peak discharge of glacier outburst floods. *Geophys. Res. Lett.*, **34**(21), L21503 (doi: 10.1029/2007GL031426)
- Nye JF (1976) Water flow in glaciers: jökulhlaups, tunnels and veins. *J. Glaciol.*, **17**(76), 181–207
- Riesen P, Sugiyama S and Funk M (2010) The influence of the presence and drainage of an ice-marginal lake on the flow of Gornergletscher, Switzerland. *J. Glaciol.*, **56**(196), 278–286 (doi: 10.3189/002214310791968575)
- Roberts MJ, Tweed FS, Russell AJ, Knudsen Ó and Harris TD (2003) Hydrologic and geomorphic effects of temporary ice-dammed lake formation during jökulhlaups. *Earth Surf. Process. Landf.*, **28**(7), 723–737 (doi: 10.1002/esp.476)
- Russell AJ and 6 others (2006) Icelandic jökulhlaup impacts: implications for ice-sheet hydrology, sediment transfer and geomorphology. *Geomorphology*, **75**(1–2), 33–64 (doi: 10.1016/j.geomorph.2005.05.018)
- Sergienko OV and Hulbe CL (2011) ‘Sticky spots’ and subglacial lakes under ice streams of the Siple Coast, Antarctica. *Ann. Glaciol.*, **52**(58), 18–22 (doi: 10.3189/172756411797252176)
- Sergienko OV, MacAyeal DR and Bindshadler RA (2007) Causes of sudden, short-term changes in ice-stream surface elevation. *Geophys. Res. Lett.*, **34**(22), L22503 (doi: 10.1029/2007GL031775)
- Shen Y, Wang G, Chun S, Mao W and Wang S (2007) Response of glacier flash flood to climate warming in the Tarim River Basin. *Adv. Climate Change Res.*, **3**, Suppl. 0051–06, 51–56
- Spring U and Hutter K (1981) Numerical studies of jökulhlaups. *Cold Reg. Sci. Technol.*, **4**(3), 227–244 (doi: 10.1016/0165-232X(81)90006-9)
- Stearns LA, Smith BE and Hamilton GS (2008) Increased flow speed on a large East Antarctic outlet glacier caused by subglacial floods. *Nature Geosci.*, **1**(12), 827–831 (doi: 10.1038/ngeo356)
- Sugiyama S, Bauder A, Weiss P and Funk M (2007) Reversal of ice motion during the outburst of a glacier-dammed lake on Gornergletscher, Switzerland. *J. Glaciol.*, **53**(181), 172–180 (doi: 10.3189/172756507782202847)
- Thórarinnsson S (1953) Some new aspects of the Grímsvötn problem. *J. Glaciol.*, **2**(14), 267–275
- Uppala SM and 45 others (2005) The ERA-40 re-analysis. *Q. J. R. Meteorol. Soc.*, **131**(612), 2961–3212 (doi: 10.1256/qj.04.176)
- Walder JS and Costa JE (1996) Outburst floods from glacier-dammed lakes: the effect of mode of lake drainage on flood magnitude. *Earth Surf. Process. Landf.*, **21**(8), 701–723 (doi: 10.1002/(SICI)1096-9837(199608)21:8<701::AID-ESP615>3.0.CO;2-2)
- Walder JS and Driedger CL (1994) Rapid geomorphic change caused by glacial outburst floods and debris flows along Tahoma Creek, Mount Rainier, Washington, U.S.A. *Arct. Alp. Res.*, **26**(4), 319–327
- Walder JS and 6 others (2005) Fault-dominated deformation in an ice dam during annual filling and drainage of a marginal lake. *Ann. Glaciol.*, **40**, 174–178 (doi: 10.3189/172756405781813456)
- Walder JS and 6 others (2006) Local response of a glacier to annual filling and drainage of an ice-marginal lake. *J. Glaciol.*, **52**(178), 440–450 (doi: 10.3189/172756506781828610)
- Werder MA and Funk M (2009) Dye tracing a jökulhlaup: II. Testing a jökulhlaup model against flow speeds inferred from measurements. *J. Glaciol.*, **55**(193), 899–908 (doi: 10.3189/002214309790152375)

MS received 26 August 2012 and accepted in revised form 18 May 2013