

METHODOLOGY

Open Access



Continuous estimates of heat emission at Mt. Ruapehu using the Unscented Kalman Smoother

Yannik Behr^{*}, Steve Sherburn and Tony Hurst

Abstract

Volcanic lakes often capture a significant amount of volcanic heat emission and thus provide a unique opportunity to monitor changes inside the volcano. We present a Bayesian inversion method to automatically infer changes in volcanic heat emission over time at the base of a volcanic lake from lake monitoring data using a non-linear Kalman Smoother. Our method accounts for the, sometimes large, uncertainties in observations and the underlying physics-based model to generate probabilistic estimates of heat emission. We verify our results using a synthetic test case and then estimate the daily heat input rate into Mt. Ruapehu's Crater Lake, New Zealand, between 2016 and 2022. Time-frequency analysis of the heat input rate shows dominant periods of heating cycles ranged between 100 - 250 days. The period between 2017 and 2020 was dominated by shorter cycles and greater-than-average heat input rate which points to changes in the magmatic heat supply and the hydrothermal system during this time.

Keywords Crater lakes, Non-linear Kalman Smoother, Volcanic heat emission

Introduction

The amount and rate of heat emitted by a volcano over time provides important clues about volcanic processes at depth. For example, a sudden rise in the heat output rate combined with increased microgravity signals may result from the emplacement of fresh magma at shallow depth (Brown et al. 1991). A slow decrease in the heat emission rate together with a high carbon-sulfur ratio of the emitted gas may point to the formation of a hydrothermal seal (Christenson et al. 2010).

Volcanic lakes help with estimating heat emissions. Often, the majority of a volcano's heat and gas output passes through the crater lake. Crater lakes act thus like a filter, integrating heat and gas input from the vents entering the lake.

Monitoring changes in temperature, water mass, and ion concentration in these lakes can, therefore, serve as proxies for changes in gas- and steam-release by the hydrothermal system and, ultimately, the melt zone beneath the volcano.

Ruapehu Crater Lake (Te Wai ā-moe in Te Reo Maori and referred to as RCL from hereon) on Mt. Ruapehu, an active andesitic stratovolcano in the center of New Zealand's North Island (Fig. 1), is one such example with records of observations dating back to the 19th century (Friedlander 1898). Mt. Ruapehu showed eruptive activity on at least 603 days since 1830 with many phreatic and phreato-magmatic eruptions and two major magmatic episodes (Scott, 2013; Historic Eruptive Activity Mt. Ruapehu, 2022). Although the latest magmatic event between June 1995 to November 1997 and a subsequent dam collapse changed the shape and volume of the lake, it has been largely unchanged since then and the lake's bathymetry is well known from times when it was empty.

*Correspondence:

Yannik Behr

y.behr@gns.cri.nz

GNS Science, 114 Karetoto Rd, 3384 Taupo, New Zealand



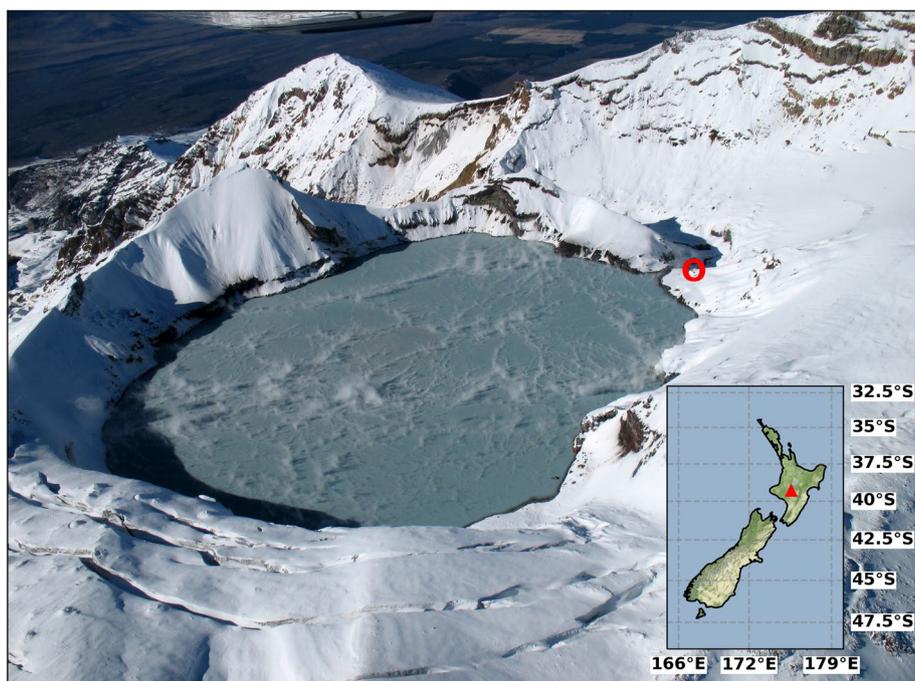


Fig. 1 Image of Mt. Ruapehu's crater lake from 2008 (©Visual Media Library, GNS Science, ID 6277, photographer Karen Britten). The red circle marks the location of the lake outlet. The location of Mt. Ruapehu is shown as a red triangle on the inset map

First attempts to quantify the amount of heat entering the lake date back to 1966 (Dibble, 1966). Since 2016 water temperature and lake level have been monitored continuously and water samples for ion concentration measurements have been taken on an approximately monthly basis. This enables us to continuously estimate heat input into the lake and, in principle, automatically detect changes as soon as they occur.

We typically infer the heat input from a mass and energy balance calculation: the total mass and energy flow into and out of the lake has to account for the observed changes in lake temperature, water level and ion concentration (Hurst and Dibble 1981; Hurst et al. 1991; Stevenson 1992; Fournier et al. 2009; Scott 1994). At RCL, and probably at most other volcanic lakes, this is an underdetermined problem, that is the number of unknown parameters in the mass and energy balance model exceeds the number of independent observations. Hence, it is necessary to constrain some model parameters using prior knowledge. Further, some observations and input parameters come with large uncertainties and are sampled irregularly. The mass and energy balance model itself also comes with considerable uncertainty as it is only a rough approximation of the physical processes within the crater lake.

One way to handle these difficulties is to make educated guesses about unconstrained parameters, interpolate between irregularly sampled data, and ignore uncertainties. We will call this the deterministic method. While it can lead to reasonable results, they have the following problems:

- 1 It is not possible to quantify our confidence in the accuracy of the results.
- 2 Comparing results to other models becomes very difficult.

To illustrate the second point let us imagine that we make an estimate of intruded magma volume based on measured gas flux. Estimating the heat flux based on this magma volume using, for example, the methods described in Stevenson (1992), will most likely not match the heat flux estimated with the deterministic method. The reason for that is that gas flux, the assumptions on amount of exsolved gas from a magma body, and the mass and energy balance model all carry large uncertainties. Instead of trying to find a single estimate of magma volume that fits all observations perfectly we better try to find a range of magma volumes that are compatible with our models and observations within their uncertainties.

The result would be a probabilistic model that can provide lower and upper bounds on magma volume.

In the following, we propose a probabilistic (Bayesian) inference approach of the time-varying heat input rate which naturally handles these difficulties and transparently propagates uncertainties of observations and prior assumptions into posterior uncertainties.

Probabilistic inference method

Physical model

The mass and energy balance model can be written in terms of three coupled ordinary differential equations (Hurst et al. 1991; Stevenson 1992). The rate of lake water mass change (\dot{M}) is the difference between the rate of steam input through the vent system beneath the lake (\dot{M}_i), the water inflow rate through melt and precipitation (\dot{M}_s), the outflow rate through overflow (\dot{M}_o), and the rate of mass loss through evaporation (\dot{M}_e):

$$\frac{dM}{dt} = \dot{M}_i + \dot{M}_s - \dot{M}_o - \dot{M}_e \quad (1)$$

The energy content of the lake can be described with the notion of enthalpy, H , which is defined relative to a reference enthalpy. In the following we take the enthalpy of the lake at temperature $T = 0^\circ\text{C} = T_0$ as reference $H_0 = H(T_0)$. At constant pressure, the enthalpy of the lake can then be written as a function of T , T_0 , M and the specific heat of water, c_w :

$$H(T) = H_0 + \int_{T_0}^T c_w M \tau d\tau \quad (2)$$

For all following equations we take $c_w = \text{constant}$, such that the enthalpy of lake water at temperature T can be written as $H(T) = H_0 + c_w M T$.

The rate of enthalpy change of the lake water can then be defined as a function of the rate of energy input from volcanic sources (\dot{Q}_i), the rate of energy gain due to solar irradiation from short-wavelength radiation (\dot{Q}_r), the rate of energy loss through outflow/seepage ($c_w T \dot{M}_o$), the rate of surface energy losses due to evaporation (forced and free convection), sensible heat, and long-wavelength radiation (\dot{Q}_e), and the rate of energy loss from heating incoming surface water ($c_w T \dot{M}_s$):

$$\frac{dH}{dt} = \frac{d(c_w M T)}{dt} = c_w T \dot{M} + c_w M \dot{T} = \dot{Q}_i + \dot{Q}_r - c_w T \dot{M}_o - \dot{Q}_e - c_w T \dot{M}_s \quad (3)$$

Because RCL is at an altitude of ~ 2500 m and big parts of its catchment are covered by snow and ice all year round we assume the temperature of incoming surface water to be at 0°C . Several empirical and theoretical equations exist to describe the surface losses and we refer

the reader to previous publications for more detailed discussions (Stevenson 1992; Hurst et al. 2015). There is no consensus on the best quantitative model of evaporation from warm lakes and, in fact, the most suitable equation may depend on the location of and available observations from the lake. Appendix A describes the equations used in this study in detail which include the recent development on forced (i.e. wind-induced) convection (Sartori 2000). It is important to note here that Q_e is a function of T and the wind speed above the lake.

Following from Eq. 3 the temperature change $\frac{dT}{dt}$ can then be written as:

$$\frac{dT}{dt} = \frac{1}{c_w M} (\dot{Q}_i + \dot{Q}_r - c_w T \dot{M}_o - \dot{Q}_e - c_w T \dot{M}_s) - \frac{T}{M} \dot{M} \quad (4)$$

Finally, the change in ion concentration expressed in total ion amount (X) of, for example, Mg^{2+} can be defined as:

$$\frac{dX}{dt} = -\dot{M}_o \frac{X}{M} \quad (5)$$

This equation is only true under the assumption that X only changes due to dilution by the inflow of meltwater and steam and there is no re-supply between eruptive episodes. Equation 5 gives an estimate of the average outflow over periods on the order of months but does not capture shorter term variations in the outflow. In the following we will also only account for water outflow at the outlet and ignore seepage through the lake floor. This approximation is supported by the observation that amongst the various rivers at Mt Ruapehu only the one fed directly by RCL's outlet bears RCL's chemical signature (B. Scott, 2022, GNS Science, pers. commun.).

The main challenge in solving the mass and energy balance equations for \dot{Q}_i arises from the non-linear relationship between Q_e and T as shown in Fig. 2.

Inference

The objective of the inference, or inverse, problem is to estimate the rate at which heat is entering the lake through the hydrothermal system below (\dot{Q}_i) from the given observations and the mass and energy balance model described in Section 2.1. Let $y_{1,t}$ be the time series of t observations of T , M , X , and \dot{M}_o and \mathbf{x}_k be the n

model parameters described in Section 2.1 at time step k . Note that \mathbf{x}_k also includes T , M , X and \dot{M}_o . We distinguish here between the state of the lake and observations of this state (T , M , X , \dot{M}_o) which are to some degree uncertain. Solving the inverse problem probabilistically

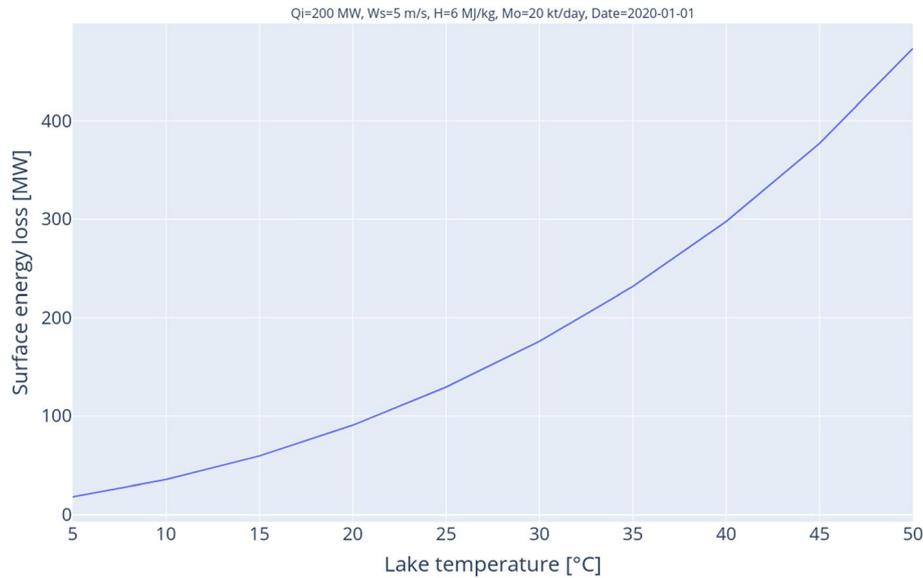


Fig. 2 Surface energy loss (Q_e) with respect to lake water temperature (T). The model parameters that have been held fixed are shown in the title

means that we are looking for a probability distribution that can best explain our observations taking into account the mass and energy balance model and all relevant uncertainties. More concisely we are looking for:

$$\mathbf{x}_k \sim p(\mathbf{x}_k | \mathbf{y}_{1:t}) \quad (6)$$

which means that the variable \mathbf{x}_k has the probability distribution $p(\mathbf{x}_k | \mathbf{y}_{1:t})$. The latter is the conditional probability of \mathbf{x}_k given all observations until time t .

Solving Eq. 6 using standard strategies like Markov Chain Monte Carlo is computationally very expensive as the number of variables grows exponentially with the length of the time series and the number of model parameters. Alternatives are Bayesian Filtering and Smoothing algorithms of which the Kalman Filter and Smoother are the closed form solution for linear models (Kalman 1960; Rauch et al. 1965; Särkkä 2013). Note that to infer \mathbf{x} at time step k we use all available observations, that is $t \leq k$ and $t > k$. This is called a Bayesian or Kalman Smoother. For the Kalman Filter only $t \leq k$ are used.

We assume that the states of the lake form a Markov chain, meaning that the state of the lake at time step k , \mathbf{x}_k , only depends on the state at adjacent time steps:

$$p(\mathbf{x}_k | \mathbf{x}_{1:k-1}, \mathbf{y}_{1:k-1}) = p(\mathbf{x}_k | \mathbf{x}_{k-1}) \quad (7)$$

and

$$p(\mathbf{x}_{k-1} | \mathbf{x}_{k:t}, \mathbf{y}_{k:t}) = p(\mathbf{x}_{k-1} | \mathbf{x}_k) \quad (8)$$

With this assumption and using *Bayes' rule*, Eq. 6 can be rewritten (Särkkä 2013) as:

$$p(\mathbf{x}_k | \mathbf{y}_{1:t}) = p(\mathbf{x}_k | \mathbf{y}_{1:k}) \int \left[\frac{p(\mathbf{x}_{k+1} | \mathbf{x}_k) p(\mathbf{x}_{k+1} | \mathbf{y}_{1:t})}{p(\mathbf{x}_{k+1} | \mathbf{y}_{1:k})} \right] d\mathbf{x}_{k+1} \quad (9)$$

with

$$p(\mathbf{x}_{k+1} | \mathbf{y}_{1:k}) = \int p(\mathbf{x}_{k+1} | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{y}_{1:k}) d\mathbf{x}_k \quad (10)$$

$$p(\mathbf{x}_k | \mathbf{y}_{1:k}) = \frac{1}{z_k} p(\mathbf{y}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) \quad (11)$$

$$z_k = \int p(\mathbf{y}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) d\mathbf{x}_k \quad (12)$$

$$p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k | \mathbf{x}_{k-1}) p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) d\mathbf{x}_{k-1} \quad (13)$$

where we made use of the *Chapman-Kolmogorov* equation:

$$p(\mathbf{x}_3 | \mathbf{x}_1) = \int p(\mathbf{x}_3 | \mathbf{x}_2) p(\mathbf{x}_2 | \mathbf{x}_1) d\mathbf{x}_2$$

Equations 9 to 13 describe a recursive way to compute Eq. 6 in which we only need to know \mathbf{x}_0 , $p(\mathbf{x}_k | \mathbf{x}_{k-1})$, and $p(\mathbf{y}_k | \mathbf{x}_k)$. Equation 9 and 11 are also known as the Bayesian Smoothing and Bayesian Filtering equation, respectively (see Appendix B for more details on the derivation). If $p(\mathbf{x}_k | \mathbf{x}_{k-1})$ and $p(\mathbf{y}_k | \mathbf{x}_k)$ are linear transformations with normally distributed errors, the Kalman filter (Kalman 1960) and Kalman smoother (Rauch et al. 1965)

represent closed form solutions. In our case, these transformations are:

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}) = f(\mathbf{x}_{k-1}) + \mathbf{q}_{k-1} \quad (14)$$

$$p(\mathbf{y}_k|\mathbf{x}_k) = \mathbf{x}_k[1:m] + \mathbf{r}_k \quad (15)$$

where $f(\mathbf{x}_{k-1})$ is the non-linear physical model described in Section 2.1 propagated forward in time using a fourth-order Runge-Kutta scheme. If we assume the errors, \mathbf{r}_k and \mathbf{q}_k , to be normally distributed ($\mathbf{q}_k \sim N(\mathbf{0}, \mathbf{Q}_k)$, $\mathbf{r}_k \sim N(\mathbf{0}, \mathbf{R}_k)$), we can compute an approximate solution to Eq. 9 using the Unscented Kalman Smoother (Van der Merwe, 2004; Särkkä, 2013), a non-linear Kalman Smoother.

We estimate \mathbf{R}_k from the daily variations in the observations, but we do not know \mathbf{x}_0 and \mathbf{Q}_k . Setting \mathbf{x}_0 to an arbitrary value within physically reasonable bounds, but large uncertainty works well in practice and has very little influence on the results. While \mathbf{Q}_k can be formally optimized we treat it here as a design parameter that we modify such that the observations are matched within their error bounds and the inferred heat input rate varies smoothly. The latter is motivated by the fact that the physical model described in Section 2.1 is not able to model short-term transient changes in lake volume, temperature, and ion concentration. These can be caused by rain or increased meltwater influx, partial collapse of the lake walls, or the influx of ion-enriched fluids. The underlying assumption of a perfectly mixed lake typically does not apply in these situations.

The ability to account for these epistemic uncertainties explicitly is a particular strength of the Bayesian Filtering and Smoothing equations. It allows us to gain valuable insights even from very simple models.

Results

Synthetic test

To evaluate the inference scheme we design a synthetic test, consisting of a very large cylinder, the same volume as RCL. We keep the wind speed and surface water inflow constant at 4.5 m/s and 10 kt/day, respectively. To model the outflow we assume the outlet to be a pipe 0.2 m² in cross-section and then use Bernoulli's equation to model the water outflow based on the water level in the cylinder above the pipe. We assume that the incoming heat (\dot{Q}_i) follows a combination of a third-order polynomial and a sinusoidal function and that the enthalpy of the incoming steam is 2.8 MJ/kg which corresponds to the enthalpy of saturated steam at hydrostatic pressures similar to those at the bottom of the lake (Mayhew and Rogers 1978). To the resulting synthetic observations we add uncertainties similar to those estimated from real observations

(see Section 3.2). As with real data, average and standard deviation of all values within a day are calculated. The standard deviation is taken as a proxy for the uncertainty in observations (\mathbf{R}_k). To simulate the disparity between sampling intervals of different data streams we randomly remove most of the observations of total ion amount (X), water outflow rate (\dot{M}_o), and water inflow rate (\dot{M}_s). This dataset is then inverted using the Unscented Kalman Smoother to recover the original heat input rate.

Figure 3 shows the result of the probabilistic inference of heat input compared to the true input of the synthetic model. The observations fit well within their error bounds, but we see some discrepancies between the original and the recovered heat input rates, especially for very steep changes in the input rate. This is the result of preferring a smooth solution as already mentioned in Section 2.2.

It is noteworthy that the mass outflux rates are recovered quite accurately despite providing very few synthetic observations.

Real data

Temperature at RCL is measured by an integrated-circuit temperature device (Texas Instruments LM35) located roughly 1.9 m below the lake's current overflow level. The water level sensor is a hydrostatic pressure sensor (First Sensor KTE/KTU/KTW6000...CS Series) and it is co-located with the temperature sensor. Both sensors have a sampling interval of 15 minutes and transmit data several times a day via satellite to GNS Science, which monitors New Zealand's geohazards through its GeoNet program (GeoNet Home 2022). Due to the type of satellite connection only about 30% of the temperature and level measurements arrive at GeoNet's datacenter; the rest of the data is lost. Water samples are taken manually 8-12 times a year from which subsequently, amongst many other analyses, concentrations of Mg^{++} and Cl^- are determined at GNS Science's Geothermal Analytical Laboratory. We will focus here on Mg^{++} concentration as it shows less deviation from the assumption of constant dilution than Cl^- . Because of the sparse sampling, new information on Mg^{++} concentration is available approximately once per month.

To estimate uncertainties of temperature and water level observations, we compute their daily averages and standard deviations. Uncertainties of Mg^{++} concentrations are estimated directly when two samples were taken on the same day. Otherwise, their uncertainty is taken to be the median of directly estimated uncertainties from when multiple samples were taken.

Lake outflow (\dot{M}_o) is measured once or twice per year, and we further know from field observations that the

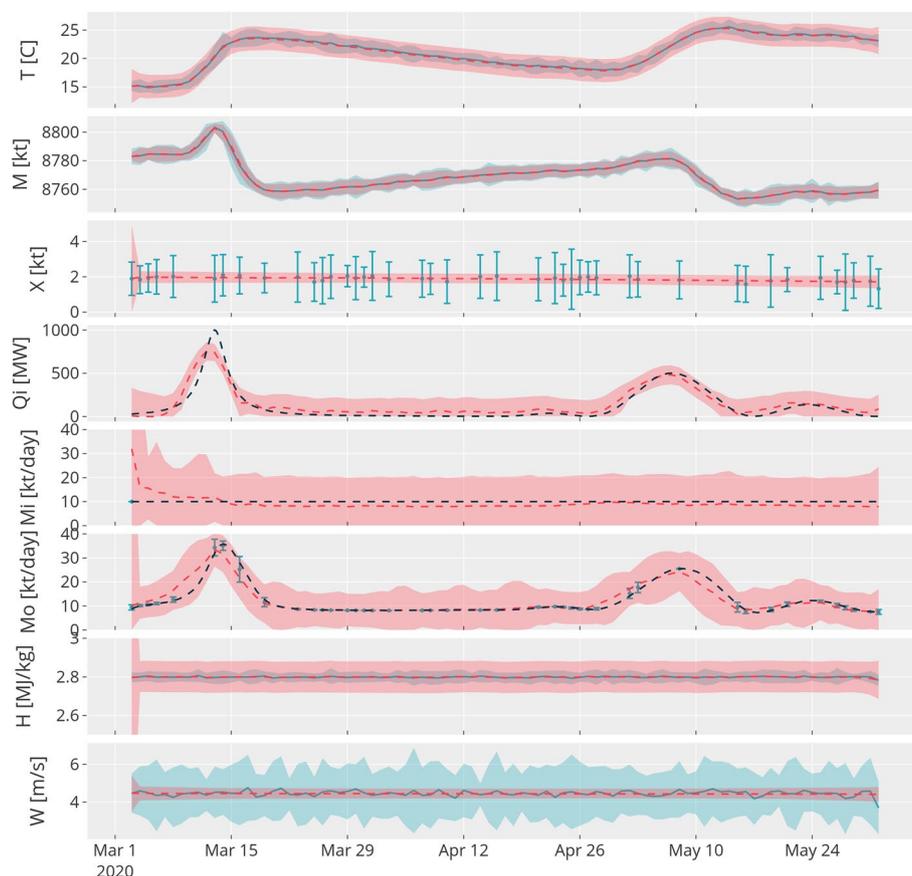


Fig. 3 Inversion result for the synthetic test described in Section 3.1. Turquoise lines and markers represent synthetic observations; red dashed lines are the inversion results and black dashed lines are the true input to the synthetic test. Error bars and shaded areas represent ± 3 standard deviations

lake is not overflowing beneath a certain water level. We fit a sigmoid function to these observations assuming a maximum outflow rate of 250 l/s. We assume the uncertainty in \dot{M}_o as the 95 percent confidence interval of the curve-fitting.

As there is no permanent weather station close to RCL we assume the wind speed to be 5 m/s. This was the average wind speed derived from a temporary deployment of a weather buoy on RCL and adjacent permanent weather stations (Hurst et al. 2012).

Figure 4a shows the heat input rate (\dot{Q}_i) for RCL for lake measurements between 4 March 2016 and 1 February 2022. During this time \dot{Q}_i reached a maximum of 753 MW and was on average 165 MW. The average standard deviation was 34 MW and generally increased for lower values of \dot{Q}_i . To put this in context, currently installed geothermal power plants in New Zealand have a combined output of ~ 1000 MW.

To also investigate changes in periodicity we computed the continuous wavelet transform of \dot{Q}_i using a complex Morlet wavelet (Fig. 4b). We chose the wavelet transform

over the more common windowed Fourier transform for its ability to detect non-stationary signals (Torrence and Compo 1998). Figure 4c shows the global wavelet spectrum which is an average of the wavelet analysis in Fig. 4b over the whole analysis period. The global wavelet spectrum can be interpreted similarly to the Power Spectral Density. Figure 4c shows that periodicity is dominated by periods between 100 and 250 days. Which period is dominant at any one time changes significantly, and we will expand on this further in the following section.

Discussion and conclusion

Inferring the heat input rate (\dot{Q}_i) into a crater lake is an important part of estimating the total energy budget of a volcanic system (Brown et al. 1989). Similar to a volcano's gas budget, changes in the heat input rate can either indicate changes in the magmatic heat source or in the hydrothermal system.

Experiments on analog models of hydrothermal systems suggest that changes in the heat source and the hydrothermal system should also change the length of

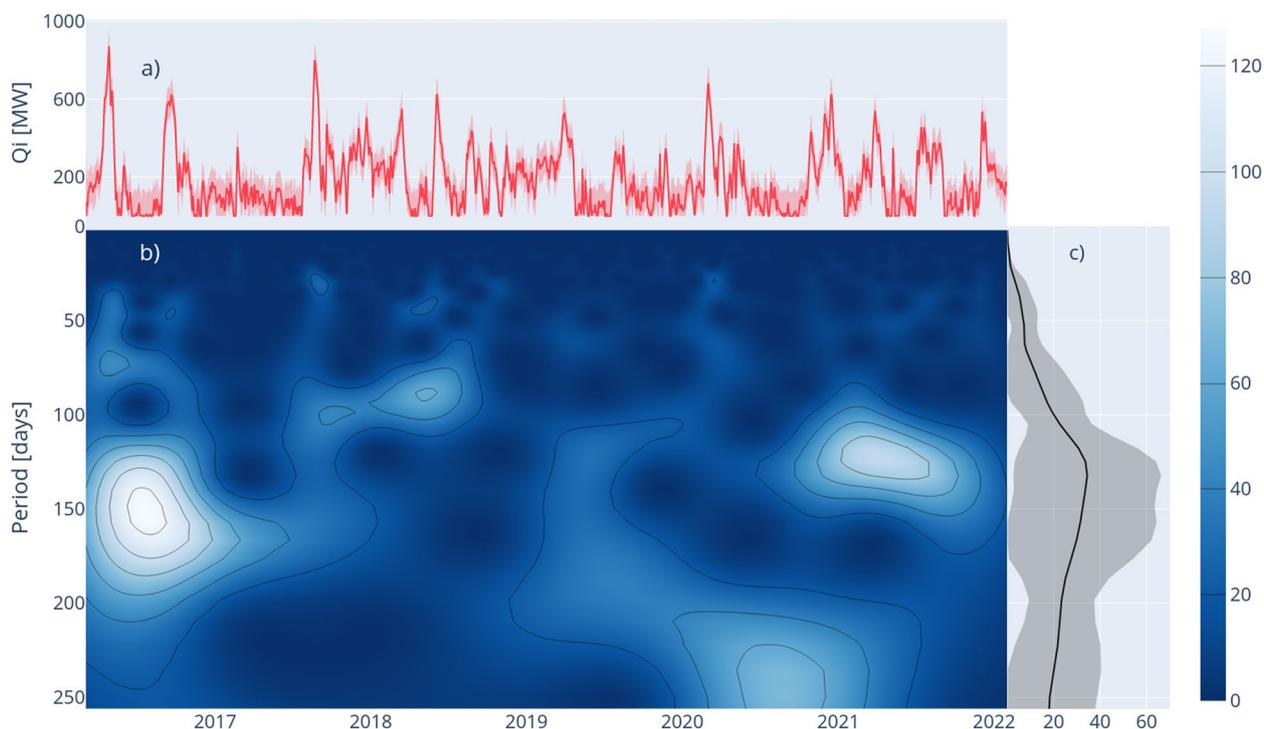


Fig. 4 **a)** Inference results for the heat input rate \dot{Q}_i at RCL between 4 March 2016 and 1 February 2022. The solid line shows the marginal probability and the uncertainty is displayed as shaded region; **b)** Continuous wavelet transform of the time series shown in a; **c)** Power spectral density of the heat input rate shown as a solid line with uncertainty displayed as shaded region

heating cycles (Vandemeulebrouck et al. 2005; Fitzgerald and Woods 1997). Figure 4b shows the heating cycles before mid-2017 and after mid-2019 were dominated by longer periods. From mid-2017 to mid-2019 cycles were mostly of shorter periods which may indicate a change in the volcanic system during this time.

This is further supported by the cumulative difference to the mean for the heat input rate as shown in Fig. 5. This type of graph helps to highlight periods when the heat input rate was above or below the long-term mean. Before 2017 and after 2019 the difference appears to fluctuate more or less around the mean value. After a dip beginning

in 2017 a period of higher than mean values starts towards the end of 2017 and ends early 2020.

To further explore hypotheses like this, the next logical step is to include numerical models of magmatic gas and heat release as well as hydrothermal heat and fluid flow. Our inversion method is well suited for such more complex models as its runtime scales linearly with the number of parameters. Several studies have demonstrated the value of non-linear Kalman Filters for inverse problems with non-linear, high-dimensional numerical models (e.g. White 2018; Huang et al. 2022).



Fig. 5 Cumulative difference to the mean for the heat input rate into RCL (solid line) and its uncertainty (shaded region) calculated from the time-series in Fig. 4a

Previous studies of RCL's mass and energy balance found that they required very high values of enthalpy for the incoming steam to avoid violating physical boundary conditions such as negative values for the inflow rate of meteoric water (\dot{M}_s) (Hurst et al. 1991, 2015). Their explanation was that as rising hot fluids enter the lake, relatively cool lake water flows back into the vent. This would lead to an overall lower addition of mass to the lake and resulting in an enthalpy value that is double the expected for saturated steam at hydrostatic pressures likely to be found at the bottom of the lake. By including uncertainties in our observations, this explanation is still a possibility but no longer a requirement to satisfy the physical constraints. Numerical modelling of hydrothermal processes may also shed further light on the heating processes and whether the lake and the vent form a single or two separate convective systems.

In conclusion, the combination of the mass and energy model, developed in previous studies (Hurst et al. 1991, 2015), with the probabilistic inference method we developed here provides a powerful method to continuously infer the heat input rate into RCL. It allows us to combine disparate data streams, invert them for probabilistic estimates of the heat input rate and thereby keep a finger on Mt. Ruapehu's energy budget.

Appendix A: Surface losses

A.1 Long-wavelength radiation

The net loss through long-wavelength radiation can be written as:

$$Q_{rad} = A\sigma(\epsilon_w T_s^4 - \epsilon_a T_a^4) \quad (16)$$

where

- A is the surface of the lake (in m^2)
- σ is the Stefan-Boltzmann coefficient ($5.67e^{-8} \frac{W}{m^2 K^4}$)
- ϵ_w is the lake emissivity
- T_s is the water surface temperature (here assumed to be 1 °C less than the measured temperature)
- ϵ_a is the effective atmospheric emissivity
- T_a is the effective air temperature (here assumed to be 0.9 °C)

A.2 Evaporation

As proposed by Hurst et al. (2015) we use the equation of Adams et al. (1990) but replace the term for forced convection by that of Sartori (2000):

$$Q_{evap} = [(2.2(T_{sv} - T_{av})^{1/3}(e_s - e_a))^2 + (L(0.00407u^{0.8}X_0^{-0.2})(e_s - e_a)/P_a)^2]^{1/2} \quad (17)$$

where

- T_{sv}, T_{av} are the virtual surface and air temperatures respectively. This correction accounts for the extra buoyancy of water vapor compared to air (see Eq. 18 for details)
- e_s, e_a are the saturation vapor pressure and the ambient air vapor pressure respectively
- L is the latent heat of evaporation
- u is the wind speed
- X_0 is the fetch, or characteristic length, of the lake
- P_a is the atmospheric pressure

$$T_{xv} = \frac{T_x}{(1 - 0.378e_x/P_a)} \quad \text{for } x = s \text{ or } a \quad (18)$$

A.3 Sensible heat

This describes the heat loss due to conduction and convection above a hot lake and according to Stevenson (1992) the ratio between sensible heat, Q_{sh} , and heat loss due to evaporation, Q_{evap} can be written as:

$$\frac{Q_{sh}}{Q_{evap}} = R_{sh} = \frac{\rho c_a (T_s - T_a)}{L (e_s - e_a)} \underbrace{\left[\frac{(T_s - T_a)}{(T_{sv} - T_{av})} \right]^{1/3}}_A \quad (19)$$

where

- ρ is the air density
- c_a is the specific heat of air

We decided to ignore term A as it tends to be very close to 1.

A.4 Putting it all together

The final equation to compute surface heat losses is then:

$$Q_e = Q_{rad} + Q_{evap}(1 + R_{sh}) \quad (20)$$

and the mass-loss due to evaporation is:

$$M_e = Q_{evap}/L \quad (21)$$

Appendix B: Derivation of the Bayesian Filtering and Smoothing equations

In the following we will reiterate the derivation of the Bayesian Filtering and Smoothing equations presented by Särkkä (2013).

B.1 Bayesian Filtering

For the Bayesian Filtering equation we only consider observations up to time step k .

$$p(\mathbf{x}_k | \mathbf{y}_{1:k}) = \frac{1}{z_k} p(\mathbf{y}_k | \mathbf{x}_k, \mathbf{y}_{1:k-1}) p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) \quad (22)$$

$$= \frac{1}{z_k} p(\mathbf{y}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) \quad (23)$$

where:

$$z_k = \int p(\mathbf{y}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) d\mathbf{x}_k$$

and, using the *Chapman-Kolmogorov* equation:

$$p(\mathbf{x}_k | \mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k | \mathbf{x}_{k-1}) p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}) d\mathbf{x}_{k-1}$$

Where we made use of *Bayes' rule* for Eq. 22 and the Markov property (Eq. 7) for Eq. 23.

This is a recursive scheme where $p(\mathbf{x}_k | \mathbf{y}_{1:k})$ depends on the previous step's solution $p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1})$. To start the recursion we have to guess $p(\mathbf{x}_0)$.

B.2 Bayesian Smoothing

Bayesian Smoothing takes into account all observations until time step t . This is sometimes referred to in the literature as *history matching*.

Because the states \mathbf{x}_k form a Markov chain, they are independent of future measurements given \mathbf{x}_{k+1} :

$$p(\mathbf{x}_k | \mathbf{x}_{k+1}, \mathbf{y}_{1:t}) = p(\mathbf{x}_k | \mathbf{x}_{k+1}, \mathbf{y}_{1:k})$$

Combining the Markov property with *Bayes' rule* we can express the probability distribution of \mathbf{x}_k given \mathbf{x}_{k+1} and $\mathbf{y}_{1:t}$ as follows:

$$\begin{aligned} p(\mathbf{x}_k | \mathbf{x}_{k+1}, \mathbf{y}_{1:t}) &= p(\mathbf{x}_k | \mathbf{x}_{k+1}, \mathbf{y}_{1:k}) \\ &= \frac{p(\mathbf{x}_k, \mathbf{x}_{k+1} | \mathbf{y}_{1:k})}{p(\mathbf{x}_{k+1} | \mathbf{y}_{1:k})} \\ &= \frac{p(\mathbf{x}_{k+1} | \mathbf{x}_k, \mathbf{y}_{1:k}) p(\mathbf{x}_k | \mathbf{y}_{1:k})}{p(\mathbf{x}_{k+1} | \mathbf{y}_{1:k})} \\ &= \frac{p(\mathbf{x}_{k+1} | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{y}_{1:k})}{p(\mathbf{x}_{k+1} | \mathbf{y}_{1:k})} \end{aligned}$$

The joint distribution of \mathbf{x}_k and \mathbf{x}_{k+1} given $\mathbf{y}_{1:t}$ can be expressed as:

$$\begin{aligned} p(X_K, X_{K+1} | Y_{1:t}) &= p(X_K | X_{K+1}, Y_{1:t}) p(X_{K+1} | Y_{1:t}) \\ &= p(\mathbf{x}_k, \mathbf{x}_{k+1} | \mathbf{y}_{1:t}) = p(\mathbf{x}_k | \mathbf{x}_{k+1}, \mathbf{y}_{1:t}) p(\mathbf{x}_{k+1} | \mathbf{y}_{1:t}) \end{aligned}$$

Combining the joint and the conditional distributions and using again the *Chapman-Kolmogorov* equation we get the Smoothing Equation (Eq. 9) that solves Eq. 6:

$$p(\mathbf{x}_k | \mathbf{y}_{1:t}) = p(\mathbf{x}_k | \mathbf{y}_{1:k}) \int \left[\frac{p(\mathbf{x}_{k+1} | \mathbf{x}_k) p(\mathbf{x}_{k+1} | \mathbf{y}_{1:t})}{p(\mathbf{x}_{k+1} | \mathbf{y}_{1:k})} \right] d\mathbf{x}_{k+1}$$

Note that Eq. 9 is again computed recursively where $p(\mathbf{x}_k | \mathbf{y}_{1:t})$ is the solution to the previous smoothing step. The recursion starts with the final step of the Bayesian Filtering recursion.

Abbreviations

R _{CL}	Ruapehu Crater Lake
\dot{M}	Lake water mass change
\dot{M}_i	Steam input
$\dot{M}_{i,s}$	Melt water and precipitation inflow
\dot{M}_o	Lake overflow
H	Enthalpy
T	Lake temperature
c_w	Specific heat of water
Q_i	Energy input from volcanic sources
Q_r	Solar irradiation
Q_e	Surface energy loss
X	Total ion amount in the lake

Acknowledgements

The authors would like to thank Nico Fournier for providing feedback and ideas throughout the project, Sophie Pearson-Grant for her constructive and insightful feedback on this manuscript, and Yuri Taran and one anonymous reviewer for their helpful comments. GNS Science's volcano monitoring group has been an invaluable team to test and improve this work. Graphs have been produced using the open-source graphing libraries Plotly (<https://plotly.com/python/>), matplotlib (<https://matplotlib.org/>) and cartopy (<https://scitools.org.uk/cartopy/docs/latest/>). Kalman Smoothing was implemented using the FilterPy library (<https://github.com/rllabbe/filterpy>).

Authors' contributions

Y.B. wrote the main manuscript text, produced graphs, and implemented the Kalman Smoothing. S.S. worked on data collection and interpretation. T.H. developed and implemented the mass and energy balance model. All authors reviewed the manuscript.

Authors' information

All authors are affiliated with GNS Science. TH is Emeritus Volcano Geophysicist, SS is Senior Volcano Geophysicist and YB is Volcano Science Operations Specialist.

Funding

This project was funded in parts by the Strategic Science Investment Funding to GNS Science within the Data Science program and by the GeoNet program.

Availability of data and materials

The source code for this work can be accessed at <https://github.com/yanni/buhr/pumahu>. All data used in this study is provided free of charge by the GeoNet program (<https://www.geonet.org.nz>).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 19 April 2022 Accepted: 28 October 2022

Published online: 07 February 2023

References

- Adams EE, Cosler DJ, Helfrich KR (1990) Evaporation from Heated Water Bodies: Predicting Combined Forced Plus Free Convection. *Water Resour Res* 26(3):425–435. <https://doi.org/10.1029/WR026i003p00425>
- Brown G, Rymer H, Dowden J, Kapadia P, Stevenson DS, Barquero J, Morales LD (1989) Energy Budget Analysis for Poás Crater Lake: Implications for Predicting Volcanic Activity. *Nature* 339(6223):370–373. <https://doi.org/10.1038/339370a0>
- Brown GC, Rymer H, Stevenson D (1991) Volcano Monitoring by Microgravity and Energy Budget Analysis. *J Geol Soc* 148(3):585–593. <https://doi.org/10.1144/gsjgs.148.3.0585>
- Christenson BW, Reyes AG, Young R, Moebis A, Sherburn S, Cole-Baker J, Britten K (2010) Cyclic processes and factors leading to phreatic eruption events: Insights from the 25 September 2007 eruption through Ruapehu Crater Lake. *New Zealand. J Volcanol Geotherm Res* 191(1):15–32. <https://doi.org/10.1016/j.jvolgeores.2010.01.008>
- Dibble RR (1966) Volcanic Seismology. In: Thompson BN, Kermode LO, Ewart A (eds) *New Zealand Volcanology, Central Volcanic Region*, 1st edn. Department of Scientific and Industrial Research, New Zealand, pp 93–98
- Fitzgerald SD, Woods AW (1997) The Stability of Two-Phase Geothermal Reservoirs. In: *Proceedings of the 22nd Workshop on Geothermal Reservoir Engineering*. Stanford, p 3
- Fournier N, Witham F, Moreau-Fournier M, Bardou L (2009) Boiling Lake of Dominica, West Indies: High-temperature Volcanic Crater Lake Dynamics. *J Geophys Res Solid Earth* 114(2):1–17. <https://doi.org/10.1029/2008JB005773>
- Friedlander B (1898) Some notes on the Volcanoes of the Taupo District. *Trans. New Zealand Institute*, 31:498–510
- GeoNet Home. <https://www.geonet.org.nz/>. Accessed 2022-02-01
- Historic Eruptive Activity Mt. Ruapehu. (2022) GeoNet. <https://github.com/GeoNet/data/tree/main/historic-volcanic-activity>. Accessed 09 Mar 2022
- Huang DZ, Schneider T, Stuart AM (2022) Iterated Kalman methodology for inverse problems. *J Comput Phys* 463:111262. <https://doi.org/10.1016/j.jcp.2022.111262>
- Hurst AW, Bibby HM, Scott BJ, McGuinness MJ (1991) The Heat Source of Ruapehu Crater Lake; Deductions from the Energy and Mass Balances. *J Volcanol Geotherm Res* 46(1–2):1–20. [https://doi.org/10.1016/0377-0273\(91\)90072-8](https://doi.org/10.1016/0377-0273(91)90072-8)
- Hurst AW, Dibble RR (1981) Bathymetry, Heat Output and Convection in Ruapehu Crater Lake. *New Zealand. J Volcanol Geotherm Res* 9(2–3):215–236. [https://doi.org/10.1016/0377-0273\(81\)90005-6](https://doi.org/10.1016/0377-0273(81)90005-6)
- Hurst T, Christenson B, Cole-Baker J (2012) Use of a Weather Buoy to Derive Improved Heat and Mass Balance Parameters for Ruapehu Crater Lake. *J Volcanol Geotherm Res* 235–236:23–28. <https://doi.org/10.1016/j.jvolgeores.2012.05.004>
- Hurst AW, Hashimoto T, Terada A (2015) Crater Lake Energy and Mass Balance. In: *Lakes Volcanic* (ed) Rouwet D, Christenson B, Tassi F, Vandemeulebrouck J. Springer, Berlin, Heidelberg, pp 307–321
- Kalman RE (1960) A New Approach to Linear Filtering and Prediction Problems. *J Basic Eng* 82(1):35. <https://doi.org/10.1115/1.3662552>
- Mayhew YR, Rogers GFC (1978) *Thermodynamic and Transport Properties of Fluids: SI Units*, 2nd edn. Basil Blackwell, Oxford
- Rauch HE, Striebel CT, Tung F (1965) Maximum Likelihood Estimates of Linear Dynamic Systems. *AIAA J* 3(8):1445–1450
- Särkkä S (2013) *Bayesian Filtering and Smoothing*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9781139344203>
- Sartori E (2000) A Critical Review on Equations Employed for the Calculation of the Evaporation Rate from Free Water Surfaces. *Sol Energy* 68(1):77–89. [https://doi.org/10.1016/S0038-092X\(99\)00054-7](https://doi.org/10.1016/S0038-092X(99)00054-7)
- Scott BJ (1994) Cyclic Activity in the Crater Lakes of Waimangu Hydrothermal System. *New Zealand. Geothermics* 23(5–6):555–572. [https://doi.org/10.1016/0375-6505\(94\)90019-1](https://doi.org/10.1016/0375-6505(94)90019-1)
- Scott BJ (2013) A Revised Catalogue of Ruapehu Volcano Eruptive Activity: 1830–2012. Technical Report 45. GNS Science, New Zealand
- Stevenson DS (1992) Heat Transfer in Active Volcanoes: Models of Crater Lake Systems. PhD thesis, The Open University
- Torrence C, Compo GP (1998) A Practical Guide to Wavelet Analysis. *Bull Am Meteorol Soc* 79(1):61–78. [https://doi.org/10.1175/1520-0477\(1998\)079<0061:APGTWA>2.0.CO;2](https://doi.org/10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2)
- Van der Merwe R (2004) Sigma-Point Kalman Filters for Probabilistic Inference in Dynamic State-Space Models. PhD thesis, Oregon Health & Science University
- Vandemeulebrouck J, Stemmelen D, Hurst T, Grangeon J (2005) Analogue Modeling of Instabilities in Crater Lake Hydrothermal Systems. *J Geophys Res Solid Earth* 110(B2). <https://doi.org/10.1029/2003JB002794>
- White JT (2018) A Model-Independent Iterative Ensemble Smoother for Efficient History-Matching and Uncertainty Quantification in Very High Dimensions. *Environ Model Softw* 109(May):191–201. <https://doi.org/10.1016/j.envsoft.2018.06.009>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

