Sensitivity analysis of the MAGFLOW Cellular Automaton model for lava flow simulation

Giuseppe Bilotta\textsuperscript{a,c,*}, Annalisa Cappello\textsuperscript{a,c}, Alexis Hérault\textsuperscript{b,c}, Annamaria Vicari\textsuperscript{c}, Giovanni Russo\textsuperscript{a}, Ciro Del Negro\textsuperscript{c}

\textsuperscript{a} Dipartimento di Matematica e Informatica, Università di Catania, 95125 Catania, Italy
\textsuperscript{b} Conservatoire National des Arts et Métiers, Paris, France
\textsuperscript{c} Istituto Nazionale di Geoﬁsica e Vulcanologia, Sezione di Catania, Catania, Italy

\textbf{A R T I C L E   I N F O}

\textbf{Article history:}
Received 29 July 2011
Received in revised form 17 February 2012
Accepted 19 February 2012
Available online 13 March 2012

\textbf{Keywords:}
Sensitivity analysis
MAGFLOW model
Forecasting
Hazard
Lava flow simulation
Polynomial chaos
Sobol' indices
ANOVA

2010 MSC:
49Q12
90C31
93B35

\textbf{A B S T R A C T}

MAGFLOW is a physics-based numerical model for lava flow simulations based on the Cellular Automaton approach that has been successfully used to predict the lava flow paths during the recent eruptions on Mt Etna. We carried out an extensive sensitivity analysis of the physical and rheological parameters that control the evolution function of the automaton and which are measured during eruptive events, in an effort to verify the reliability of the model and improve its applicability to scenario forecasting. The results obtained, which include Sobol' sensitivity indices computed using polynomial chaos expansion, confirm the consistency of MAGFLOW with the underlying physical model and identify water content and solidus temperature as critical parameters for the automaton. Additional tests also indicate that flux rates can have a strong influence on the emplacement of lava flows, and that to obtain more accurate simulations it is better to have continuous monitoring of the effusion rates, even if with moderate errors, rather than sparse accurate measurements.

© 2012 Elsevier Ltd. All rights reserved.

\section{Introduction}

Lava flows are complex systems whose evolution is controlled by a variety of properties such as temperature, chemical composition, degree of crystallization and driven by features such as the rheology of the fluid and the topography over which the fluid flows. These properties and features interact with complex and typically nonlinear laws, the knowledge of which has seen a continuous progress.

The advances in the understanding of the processes underlying this phenomenon provide physical models that can be included in scenario forecasting tools for the assessment of lava flow hazard. The use of accurate physical models, as well as the possibility to employ better models as they become available, increases the accuracy and reliability of the scenarios predicted by such forecasting tools. However, the uncertainty in the values and measures of the parameters and data that must be made available to these tools to predict the lava flow paths during eruptive events reduces the confidence of their application, presenting the need for theoretical as well as empirical constraints on the parameter values as well as on the effect of these uncertainties on the employed models.

One of the most successful approaches to lava flow modeling is the Cellular Automaton (CA) (Crisci et al., 1986; Ishihara et al., 1990; Miyamoto and Sasaki, 1997b; Avolio et al., 2006), in which the computational domain is represented by a (usually regular) grid of 2D or 3D cells, characterized by some properties such as lava height and temperature, and where the modeling of the phenomenon is described through an evolution function for the properties of the cells.

Developed at INGV-Catania, MAGFLOW is one of the Cellular Automaton (CA) models for lava flow path simulation based on...
physical modeling of lava flows. Specifically, the evolution function in MAGFLOW is directly derived from a steady-state solution of the Navier–Stokes equation for Bingham fluids, coupled with a simplified heat transfer model (Vicari et al., 2007).

MAGFLOW has been successfully used both to reproduce past events with well-known characteristics, and to predict the paths of lava flows in real time during the Mt Etna eruptions of 2004 (Del Negro et al., 2008), 2006 (Vicari et al., 2009; Herault et al., 2009), 2008 (Bonaccorso et al., 2011; Ganci et al., 2012) and 2011 (Vicari et al., 2011). The aim of this work is to integrate this literature on the validation of the model with a detailed sensitivity analysis on the physical and rheological parameters which are made available to the MAGFLOW model for the simulation of eruptive events.

Sensitivity analysis is a powerful verification tool whose main purpose is to assess the influence that uncertainties and errors in the input parameters have on the output results of a given model (Nossent et al., 2011; Shahsavani and Grimvall, 2011; Annoni et al., 2011; Makler-Pick et al., 2011; Jing and Yang, 2011). In this sense, it is complementary but distinct from the validation process that assesses the ability of the model to properly represent the phenomenon it describes. For example, validation is typically performed by comparison of simulation results against well-known test cases, while sensitivity analysis focuses more on the more abstract mathematical relationship between fluctuations in the input and the corresponding variations in the output of the model.

The results we will present indicate a strong adherence of the numerical and physical model to the physical phenomenon, within the known limits of application (for example, the MAGFLOW model does not describe tunneling and ephemeral vent opening). The analysis also allows the identification of the critical input parameters, i.e. parameters for which variations in value have a stronger impact on the results.

The additional knowledge gained from this sensitivity analysis can be used to improve on the use of the model for scenario forecasting, an essential tool which has seen recent applications to sophisticated issues including evaluating the impact of protective barrier placement to divert lava flows (Scifoni et al., 2010) and the creation of a hazard map for lava flow invasion on Mt Etna (Cappello et al., 2011a, b). The new insight on the relative importance of the parameters also raises new questions about possible ways to improve the model itself.

The paper has the following structure. The MAGFLOW model is briefly presented, highlighting the essential parameters, in Section 2. This is followed by a brief presentation of the numerical stability of the model (Section 3). The methodology for the sensitivity analysis is introduced in Section 4.1, which is followed by a brief description of the Sobol' sensitivity indices and how they can be computed using polynomial chaos expansion. The actual results of the sensitivity analysis are then presented in Section 5, and some preliminary tests and considerations on the influence of effusion rates are presented in Section 6.

2. The MAGFLOW model

In this section we present the key aspects of the MAGFLOW model that are relevant to our sensitivity analysis. A more detailed description can be found in Vicari et al. (2007).

The computational domain of a simulation is chosen large enough to include the prospected maximum extent of the lava flow emplacement. The domain is then decomposed into square cells whose width matches the resolution of the Digital Elevation Model (DEM) available for the area. The ground elevation provided by the DEM, the lava thickness, the amount of solidified lava, and the thermal energy are the four properties that define the state of each cell at each iteration.

The evolution of the system is purely local, in the sense that each cell evolves according to its present status and the status of its Moore neighborhood (i.e. its eight immediate neighbors).

Lava thickness varies according to lava influx from the vent for cells corresponding to a vent location, plus any lava flux between neighboring cells. Cross-cell lava flux is derived from a steady-state solution for the Navier–Stokes equations for a fluid flowing on an inclined plane with Bingham rheology (Dragoni et al., 1986). The approximation coming from the use of a steady-state solution is justified by the high viscosity of lava flows.

The rheological parameters are the yield strength $S_y$ and the plastic viscosity $\eta$. The Bingham flow condition, requiring that shear stress is higher than the yield strength, is modeled by introducing a critical height $h_{cr}$ and having flux between two adjacent cells only when $h > h_{cr}$, with $h$ being the lava height in the cell with higher total height. The volumetric flux $q$ is therefore zero if 800 K, and is otherwise given by

$$ q = \frac{S_y h_{cr}^2 \Delta x}{3 \eta} \left( a^2 - \frac{3}{2} a^2 + \frac{1}{2} \right). $$

where $\Delta x$ is the distance between adjacent cells, and $a = h/h_{cr} > 1$.

The critical thickness $h_{cr}$ is computed from the yield strength and slope angle to account for both pressure and gravity-driven flows:

$$ h_{cr} = \frac{S_y}{\rho g (\sin \alpha - \partial h/\partial x \cos \alpha)} = \frac{S_y}{\rho g (\Delta z^2 + \Delta x^2)} $$

where $\rho$ is the lava density, $\Delta$ is the slope angle of the inclined plane, $g$ the gravity acceleration, $\Delta$ the overall solid height difference (considering ground elevation and solid lava height), and $\Delta h$ the difference in lava thickness, as computed at the cell with highest total height (Fig. 1).

Following Ishihara et al. (1990), the yield strength is computed according to the formula

$$ \log_{10} S_y = 13.00997 - 0.00897T $$

(the original Ishihara formula was in the CGS system, which we converted into SI units) and the viscosity follows Giordano and Dingwell (2003)

$$ \log_{10} \eta = -4.643 + \frac{5812.44 - 427.04 \times H_2O}{T - 499.31 + 28.74 \ln(H_2O)} $$

with $T$ being the temperature in Kelvin and $H_2O$ the water content in weight percent (wt%). MAGFLOW assumes a constant (average) water content across the flow, and its value is a user-controlled parameter typically in the range 0.02–0.2 wt %.

The actual amount of lava gained or lost by a cell at each iteration is given by the total flux $Q$ of the cell multiplied by the timestep $\Delta t$ for that iteration. To prevent non-physical solutions, the timestep is controlled by ensuring that for each cell we have

$$ Q \Delta t < chA $$

where $h$ is the lava height in the cell and $A$ the cell area. The constant $0 < c < 1$ ensures that only a fraction of the total fluid lava volume is lost at each iteration, and should be selected small enough to ensure that the stationary solution of the Navier–Stokes equation used to compute the flux remains approximately valid during the next step; the influence of the value of $c$ is discussed in Section 3. The timestep used by the CA is then the minimum of the $\Delta t$ computed by each cell according to the inequality (5).
Cells are considered at constant temperature (no vertical temperature gradient), and heat flux is computed from mass flux considering the temperature of the cell losing mass. Heat loss is only considered with radiation from the surface, ignoring the effects of conduction to the ground and convection in the atmosphere.

Solidification follows a simplified model that does not take into consideration phenomena such as plug or crust formation. When the temperature of a cell drops below a given solidus temperature $T_s$, defined as the temperature below which lava stops flowing, a corresponding fraction of the lava in it is converted to solid lava, which is assumed to lie at the bottom of the cell, thereby contributing to the total height of the cell but not to the amount of fluid that can move.

### 3. Numerical errors and stability

Preliminary tests were done to evaluate the numerical robustness of the model as well as of its implementation. Indeed, the evolution function of the CA heavily depends on a discrete condition for the mass flow between neighboring cells: as mentioned in the description of the model, a cell with higher total height cedes mass to a lower-height neighbor only if the ratio $a = h_2/h_1$ between the lava thickness and the critical thickness satisfies $a > 1$. During a simulation, numerical errors due to the non-exact nature of computer arithmetics can accumulate, in such a way that for a given cell at a given time the condition may or may not be satisfied depending on the sequence of operations that preceded the computation of the condition.

For example, traversal of the neighborhood in a different sequence (clockwise rather than counter-clockwise) can lead to slight differences in the total flux computed for a cell, resulting in comparable slight differences in the thickness after one evolution step, differences that can accumulate to the point of triggering the condition when it would otherwise not be triggered, or conversely.

To evaluate the impact of the numerical roundoff error, multiple sets of simulations with identical input data and parameters were run under different numerical conditions: single versus double precision floating point operations, and different neighborhood traversal patterns including different choices for the first neighbor. As expected, differences were observed in the results, but in all cases, both during the evolution and for the final emplacements, these accounted for only a few cells over several thousands. The fitness index (described later in Section 4.4) was always over 99.8%. Due to the very small error introduced by limiting the precision, we opted for single-precision computation to favor speed of execution, given the large number of simulations required by our approach to sensitivity analysis.

The truncation error in time was also tested: since the physical model underlying the evolution function of the automaton relies on the assumption of a stationary flow, the timestep controlling constant $c$ should in general be used to ensure that the timestep is small enough to guarantee that the assumption remains approximately correct during a single step. By running simulations with different values for the constant $c$, it was observed that the differences in emplacements remained within the fitness limits of the numerical error. In our calculations we used $c = 1/16$ as a reasonable compromise between accuracy and efficiency.

### 4. Sensitivity analysis

Our sensitivity analysis (SA) is focused on the rheological and physical parameters which are currently measured during eruptive events, namely water contents and eruptive and solidus temperature. The purpose is to assess the impact that measurement errors have on the results of the simulation and therefore to identify the critical parameters, an information which is important in applications to scenario forecasting during eruptive events, and to provide suggestions for more research in modeling.

Although the effect of the other input data to the model was considered beyond the scope of the work presented here, we have also run some preliminary tests regarding the dependency on the effusion rates and its relation to the parameters analyzed. The results of these additional tests will be discussed in Section 6.1.2.

There are a number of other possible parameters that could be taken into consideration for the SA. For example, the coefficients that appear in the yield strength Equation (3), or the ones controlling the viscosity law (4), are derived from experimental data, and it would be useful to determine to what extent they influence the simulations. However, the extreme computational complexity of evaluating the impact of all these parameters led us to restrict our selection to the ones on which data is more likely to be available on a short notice during eruptive events. A future work might include a detailed analysis of the sensitivity for these parameters.

### 4.1. Methodology

An initial attempt to the SA of MAGFLOW's parameters was based on Affine Arithmetic (AA) (de Figueiredo and Stolfi, 2004), to exploit its capability of correlating the final results with the uncertainty in the input parameter, a significant advantage during the analysis of the influence of multiple parameters.

However, the high non-linearity of the numerical operations quickly cancel the benefits of the use of AA: most of the data in the result loses most of its correlation to...
the prescribed uncertainty in the parameters after a few iterations of the automaton, and since most simulations run for thousands of iterations, the final results contain overestimated, uncorrelated uncertainties which give little insight on the parameter dependency. Moreover, the discrete nature of the evolutionary processes of the CA substantially increases the computational complexity, which make the application of AA, or any other self-verified numerical method, too complicated for practical use in this case.

In the case of cellular automata, commonly used approaches to sensitivity analysis are focused on the specific nature of cellular automata, and analyze the topological influence of properties such as cell size and shape or neighborhood structure and width (Kocabas and Dragicevic, 2006; Moreno et al., 2008). However, our focus is instead on the rheological and physical parameters driving the cellular automaton evolution function.

The actual SA has therefore been completed with a more specific and effective approach based on the recent development of the use of polynomial chaos expansion (PCE) for uncertainty quantification (Sudret, 2008; Crestaux et al., 2009). In this context, polynomial chaos expansion can be used to determine the Sobol’ sensitivity indices, and therefore quantify the impact that each parameter has on the variance of a given function.

4.2. Sobol’ functional decomposition and sensitivity indices

In what follows we give a brief description of the Sobol’ functional decomposition, and define the sensitivity indices; the next section will show instead how polynomial chaos expansions can be used to compute them. We will closely follow Crestaux et al. (2009), which the reader is referred to for further details and proofs.

Let \( \xi_k, \ldots, \xi_d \) be independent identically distributed random variables with range \( \Omega \), and probability density function (pdf) \( p(\xi) \).

Since the variables are independent, the random input vector \( \xi = (\xi_1, \ldots, \xi_d) \) with range \( \Omega = \prod_j \Omega_j \), has pdf \( p(\xi) = \prod_j p(\xi_j) \).

Consider the function space \( \mathcal{X} = L^2(\Omega, p(\xi)) \) equipped with the usual inner product

\[
\langle \phi, \psi \rangle = \int_{\Omega} \psi(\xi)\phi(\xi) p(\xi) d\xi.
\]

A stochastic output \( y \) is a function \( y = f(\xi) \) of our random variables, with \( f \in \mathcal{X} \). We can decompose \( f \) in the form

\[
f(\xi) = \sum_{u \in \{1, \ldots, d\}} f_u(\xi_u)
\]

where the summation is extended over all possible subsets \( u = \{i_1, \ldots, i_s\} \) of \( \{1, \ldots, d\} \) and for each \( u \) we have \( \xi_u = (\xi_{i_1}, \ldots, \xi_{i_s}) \), where \( s \) is the cardinality of \( u \). If \( f_{\emptyset} \) is some constant \( f_0 \) and \( f_u \) with \( u \neq \emptyset \) satisfies

\[
\int_{\Omega} f_u(\xi_u) p(\xi_u) d\xi = 0 \quad \forall \xi \in \Omega,
\]

then the decomposition is called the ANOVA representation of \( f \) (with ANOVA being a contraction of ANalysis Of VAriance).

It can be shown that the ANOVA representation of a function is unique (Sobol, 2001), with the \( f_u \) constructed inductively as:

\[
\begin{align*}
\mathcal{O} & = \int_{\Omega} f(\xi) p(\xi) d\xi, \\
\mathcal{O}(i) & = \int_{\prod_i \Omega_i} f(\xi) \prod_{k \neq i} p(\xi_k) d\xi_k - \mathcal{O}, \\
\mathcal{O}(i_1 i_2) & = \int_{\prod_{i} \Omega_i} f(\xi) \prod_{k \neq i_1, i_2} p(\xi_k) d\xi_k - \sum_{l \in \{i_1, i_2\}} \mathcal{O}(i_l),
\end{align*}
\]

and so on. It can also be shown that the components are orthogonal, i.e.

\[
\int_{\Omega} f_u(\xi_u) f_v(\xi_v) p(\xi_u) d\xi = 0 \quad \forall u \neq v.
\]

The total variance \( D \) of \( y \) can then be computed

\[
D = \int_{\Omega} f(\xi)^2 p(\xi) d\xi - f_0^2
\]

and the conditional variances \( D_u \) are given by

\[
D_u = \int_{\prod_{i \in u} \Omega_i} f_u(\xi_u)^2 p(\xi_u) d\xi_u.
\]

The \( 2^d - 1 \) Sobol’ sensitivity indices are then defined as

\[
S_u = D_u/D \quad \text{for } u \neq \emptyset \quad \text{and are such that } \sum_u S_u = 1.
\]

Essentially, the Sobol’ index \( S_u \) with \( u = \{i_1, \ldots, i_s\} \) measures the influence that the parameters \( \xi_{i_1}, \ldots, \xi_{i_s} \) have on \( y \) when they are considered together, not counting the influence each of them has separately or in conjunction with only a subset of the others.

The total influence that a parameter has on the input can be computed using the total indices introduced by Homma and Saltelli (1996):

\[
S_T = \sum_{u \ni i} S_u
\]

which measure the influence a parameter has on the input when taken alone as well as when considered in combination with any and all other possible parameters.

4.3. Polynomial Chaos Expansion and Sobol’ indices

The PCE of a stochastic function \( y \) is written as

\[
y = f(\xi) = \sum_{k=0}^{w} \hat{\beta}_k \Psi_k(\xi)
\]

where \( \hat{\beta}_k \) are the expansion coefficients and \( \{\Psi_k\} \) is an orthogonal polynomial basis of \( L^2(\Omega, p(\xi)) \). The choice of the basis functions depends on the assumed random distribution. For example, the original theory of polynomial chaos (Wiener, 1938) used Hermite polynomials for variables following a Gaussian distribution with domain \( \Omega \). In our case, a uniform distribution with compact support will be assumed, and we will use Legendre polynomials (Xiu and Karniadakis, 2002).

Although the polynomial chaos expansion of a stochastic function assumes an infinite summation, this is generally limited to a finite polynomial order \( p \):

\[
y = f(\xi) = \sum_{k=0}^{p} \hat{\beta}_k \Psi_k(\xi)
\]

where the last index \( P \) is such that \( P + 1 = (p + d)!/(p!d!) \), i.e. the dimension of the space of polynomials of degree at most \( p \) in \( d \) variables.

For multivariate functions, the basis \( \Psi_k \) is built as the tensor product of the univariate basis \( \phi_i \):
the univariate polynomial basis. Then for \( p = 2 \) (which implies \( P = 5 \)) the \( \Psi_k \) family can be built as follows:

\[
\begin{align*}
\alpha_0 &= \{0, 0\}, & \Psi_0 &= \phi_0 \phi_0; \\
\alpha_1 &= \{1, 0\}, & \Psi_1(x_1, x_2) &= \phi_1(x_1) \phi_0, \\
\alpha_2 &= \{0, 1\}, & \Psi_2(x_1, x_2) &= \phi_0 \phi_1(x_2), \\
\alpha_3 &= \{1, 1\}, & \Psi_3(x_1, x_2) &= \phi_1(x_1) \phi_1(x_2), \\
\alpha_4 &= \{2, 0\}, & \Psi_4(x_1, x_2) &= \phi_2(x_1) \phi_0, \\
\alpha_5 &= \{0, 2\}, & \Psi_5(x_1, x_2) &= \phi_0 \phi_2(x_2).
\end{align*}
\]

Assuming the PCE coefficients \( \beta_k \) are known, the total and conditional variances of \( y \) can be computed as

\[
D = \sum_{k=1}^{P} \beta_k^2 \langle \Psi_k, \Psi_k \rangle \tag{6}
\]

and

\[
D_u = \sum_{k \in k_u} \beta_k^2 \langle \Psi_k, \Psi_k \rangle \tag{7}
\]

where

\[ k_u = \{ k \in \{1, \ldots, P\} | \beta_k \neq 0 \forall i \in u \} \]

i.e. \( k_u \) is the set of all polynomial basis functions which have an explicit dependency on the variables \( z_i \) for \( i \in u \).

Of course, since the finite-order polynomial expansion is an approximation of the original function, Equations (6) and (7) only give an approximation of the total and conditional variances. On the upside, they are trivial to compute when the \( \beta_k \) coefficients are known.

Among the many possible ways to compute the PCE coefficients, we used the so-called non-intrusive spectral projection. Exploiting the orthogonality of the polynomial basis, we can write:

\[
\beta_k = \frac{\langle y(z), \Psi_k(z) \rangle}{\langle \Psi_k(z), \Psi_k(z) \rangle}
\]

where, again, \( \langle \cdot, \cdot \rangle \) marks the inner product in the function space. While the denominator can be computed analytically, the numerator must be computed numerically when an analytical form of \( y \) is not known, such as in our case.

For the numerical integration we opted for the Smolyak formula with Clenshaw–Curtis integration points (Gerstner and Griebel, 1998) (see Fig. 2). Our choice was guided by two factors: the sparseness of the grid and its nesting property, i.e. the fact that the grid for integration with degree \( m \) contains the points of the grids for integration with degree \( n < m \); this important properties allowed us to greatly reduce the number of functional evaluations when increasing the accuracy of the integration.

Section 4.4. Fitness function

To be able to use PCE we need a single scalar value which quantifies the variation in the result of different simulations. The morphological aspect of a lava flow that is most important in the applications of our model (scenario forecasting, hazard evaluation) is the emplacement, i.e. the area subject to invasion by the flow. The difference in emplacement between distinct lava flow fields (real or simulated) can be quantified by a single scalar value with a fitness value, computed as the ratio between the intersection and the union of the affected areas (Fig. 3).

In formulas, for any emplacement \( A \) let \( |A| \) be the measure of its area, and for any two emplacements \( A, B \) let \( A \cup B, A \cap B \) denote as usual the union and intersections of the respective areas. Then the fitness function is given by

\[
\phi = \frac{|A \cap B|}{|A \cup B|}.
\]

Figure 2. Two-dimensional Clenshaw–Curtis integration grid on the unit square. Blue points are the integration points for level 4, red points are the additional points to complete the set of integration points at the next level. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Figure 3. Graphical representation of fitness computation obtained overlaying a test simulation with an actual emplacement (reference). Blue (0.58 km²) denotes common invasion areas, yellow (0.08 km²) denotes underestimated areas (in the reference but not in the tested simulation), red (0.03 km²) denotes overestimated areas (in the tested simulation but not in the reference). Fitness is computed as the ratio between the blue area and the union of all the areas, and in this case would be \( \phi = 0.58 / (0.58 + 0.08 + 0.03) = 0.8406 \) (or \( 84.06 \% \)). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
This fitness function is the same used e.g. by Favalli et al. (2009) in the sensitivity analysis of the DOWFLOW probabilistic model for lava flow simulation. Other choices of fitness functions are also possible; for example, Rongo et al. (2008) define a fitness function $e_1 = \sqrt{\phi}$ to evaluate the application of the SCIARA cellular automata model for lava flows.

The behavior of $\phi$ and $e_1$, and therefore the preference of one over the other, is largely dependent on the shape of the emplacement.

Assume that the reference emplacement is a perfect circle of radius $R$ (ideal case of eruption from circular vent on a horizontal plane), with the test simulation being a concentric perfect circle of radius $r$. We then have $\phi = r^2/R^2$ while $e_1 = r/R$, with $e_1$ seemingly more appropriate than $\phi$. By contrast, for a channel flow of reference length $L$, test simulations of length $l$ have $\phi = l/L$ and $e_1 = \sqrt{l/L}$, and $\phi$ would seem more appropriate than $e_1$ as a fitness index.

In our sensitivity analysis of the rheological parameters, we will provide values for both $\phi$ and $e_1$, to ease comparisons with the other cellular automata models mentioned earlier. While the specific quantitative values of the fitness evaluations in our sensitivity analysis depend on the choice of fitness function, the qualitative behavior of the parameters will be the same for both choices, as illustrated by the overall sensitivity indices (Table 2).

4.5. Sensitivity analysis data

The tests for the SA were built with the following possible ranges of the parameters considered:

- water percentage; range: 0.02 wt%–0.2 wt%;
- solidus temperature; range: 800 K–1100 K;
- eruption temperature; range from 1360 K–1450 K; 

The range of variations were chosen based on the current geophysical knowledge about the properties of Etnean lavas, as well as from the results of the calibration and validation tests run during the development of the MAGFLOW model. Due to the wide amplitude of its range (one order of magnitude), the logarithm of the water percentage parameter was considered as raw parameter in the SA, rather than the water percentage itself.

The topography DEM and the flux rates for all the simulations were taken from the available data for the 2006 eruption (see also Vicari et al. (2009)). Since our aim is not validation but sensitivity analysis, the reference eruption for the fitness function has been chosen as the simulation with the midpoint values $H_2O = 0.06325$ wt%, $T_s = 950$ K, $T_e = 1405$ K, rather than the lava flow extent of the actual 2006 eruption.

The polynomial chaos expansion was computed with a maximum polynomial degree $p = 4$, resulting in $P = 34$. The $P + 1 = 35$ coefficients were computed using the Clenshaw–Curtis grids up to level 8, resulting in a total of 2561 simulations.

5. Results

As a preliminary to the results of the sensitivity analysis, we show fitness variation plots for the three parameters when considered independently. The results, shown in Figs. 4–6, can already give us a qualitative idea of the influence that each parameter has on the simulation, independently of the others.

We thus see that the solidus temperature is a significant parameter, as it can produce a worst fit as low as 55%, but the influence of water content is even higher, since in this case the fitness can drop by over 50%; the solidus temperature fitness is also close to being constant for values less than the reference value,
while water content has a significant drop on the whole range. The least significant parameter is the eruption temperature, for which the worst case loss is around 30%.

With three independent parameters \((H_2O, T_s, T_e)\) we have 7 Sobol' indices. The polynomial chaos expansion gives us the approximate values shown in Table 1, and the total index for each parameter, as shown in Table 2.

The values confirm the qualitative analysis of the plot, introducing some additional information, in particular with respect to the combined influence of multiple parameters, indicating that the water content also has a significant influence on the sensitivity of the solidus and eruption temperatures independently, to the point that the combined influence of water and solidus temperature is higher than the influence of solidus temperature on its own.

The obtained results indicate water content as the most important factor: the model has a strong direct dependency on this parameter, and additionally water content has a significant impact on the dependency of the model on the solidus temperature, which is the other most significant parameter.

5.1. A brief analysis of the results

The strong interaction between water content and solidus temperature can be explained by considering the physical model underlying the evolution function of the automaton. Indeed, viscosity is known to be a very significant parameter for the morphology of lava flows, and water content in MAGFLOW is the parameter for the viscosity law (4).

As shown in Fig. 7, according to this law viscosity at a given temperature can change by several orders of magnitude for different water content. The effect is particularly prominent in the range considered for the solidus temperatures, where the variation of water content can cause over five orders of magnitude differences in viscosity.

6. Effusion rate tests

Although the sensitivity analysis presented here was focused on the rheological and physical parameters of the MAGFLOW model, simulations also rely on other input data, among which effusion rates have been indicated as potentially significant by Miyamoto and Sasaki (1997a). Indeed, for a given composition, the discharge rate of lava by a vent is an essential parameter in determining the dimensions of the flow.

For MAGFLOW, effusion rates during an eruption are provided as a sampled curve, with given values of the flux at given times. The data is usually obtained by either field observations or, more recently, also from infrared satellite data (Vicari et al., 2009), and interpolated linearly between the sampled points.

An analysis of the effect of effusion rates on the model should therefore consider at least two separate aspects: as for the other parameters, we must consider the potential error in the estimation of the flux rate at a given time, but also the overall trend of the flux rate during the eruption. This led us to consider two sets of tests for the preliminary work presented here.

Table 1

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Sobol' index ((\phi))</th>
<th>Sobol' index ((e_1))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(H_2O)</td>
<td>0.3834</td>
<td>0.3836</td>
</tr>
<tr>
<td>(T_s)</td>
<td>0.1843</td>
<td>0.1882</td>
</tr>
<tr>
<td>(T_e)</td>
<td>0.1004</td>
<td>0.0916</td>
</tr>
<tr>
<td>(H_2O, T_s)</td>
<td>0.2303</td>
<td>0.2448</td>
</tr>
<tr>
<td>(H_2O, T_e)</td>
<td>0.0698</td>
<td>0.0640</td>
</tr>
<tr>
<td>(T_s, T_e)</td>
<td>0.0281</td>
<td>0.0235</td>
</tr>
<tr>
<td>All</td>
<td>0.0038</td>
<td>0.0043</td>
</tr>
</tbody>
</table>

Table 2

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Total index ((\phi))</th>
<th>Total index ((e_1))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(H_2O)</td>
<td>0.6873</td>
<td>0.6967</td>
</tr>
<tr>
<td>(T_s)</td>
<td>0.4465</td>
<td>0.4608</td>
</tr>
<tr>
<td>(T_e)</td>
<td>0.2020</td>
<td>0.1834</td>
</tr>
</tbody>
</table>

Fig. 6. Fitness variation for the eruption temperature, assuming fixed average values for the water percentage and solidus temperature. The red line marks the lowest value achieved. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 7. Semi-log plot of viscosity over temperature for different values of water content, from 0.02 wt% (upper) to 0.2 wt% (lower), in steps of 0.02 wt%.
In the first set, we assumed a systematic relative error in the flux rates provided by satellite data; by altering the sampled values by a given percentage (up to ±50%), we obtained new emplacements with fitness as shown in Fig. 8. We observe that water content once again affects the dependency on the input data. The observed drop in fitness is however always of the same order of magnitude as the variation in flux rates, and often lower.

In the second set of tests, the effect of the effusion rate curve was evaluated by running simulations with three different trends for the effusion rates, all giving the same total amount of emitted lava. The effusion rates provided by satellite data were thus compared with a constant flux rate and a (discretized) Gaussian effusion rate curve (Fig. 9).

The results are summarized in Table 3. Since the trend of effusion rates cannot be synthesized as a single quantity, a proper relation between trend and fitness is difficult to establish. Qualitative considerations highlight however that the trend of the effusion rates can significantly alter the emplacement of the flow, and that once again the effects are more significant with high water contents.

Fig. 8. Variation in fitness ($f$) assuming a systematic relative error in the flux rates estimation, for different values of the water content. An ideal linear decrease in fitness is also shown for comparison.

Fig. 9. The different effusion rates with the same total volumes used to test the effect of the effusion rate trend on emplacement.
7. Conclusions and future perspectives

The sensitivity analysis of the MAGFLOW model highlights that, among the physical and rheological properties employed in the model, water content and solidus temperatures are the parameters MAGFLOW is most sensible to.

The worst fitness achieved during the tests is accounted for by the difference between extremal values of the solidus temperature, and it’s around 50%, corresponding to about one third over- or underestimation of the emplacement.

The sensitivity analysis, as well as the preliminary tests conducted on the effusion rates, show that water content is important not only in itself, but also due to its influence on the dependency of other parameters as well as of the input data, with lower water contents making the automaton less sensitive to variations in the other parameters.

The results indicate that MAGFLOW is consistent with the physical model underlying its evolution function, with the importance of viscosity as a factor for determining the morphology of a lava flow being reflected in the importance of water content, the tunable parameter to model lava viscosity.

Combined with the extensive validation done against effusive events (Vicari et al., 2007, 2009, 2011; Del Negro et al., 2008) on the Etna and with the assessed numerical robustness of the implementation (Section 3), the sensitivity analysis therefore confirms the overall reliability of MAGFLOW for lava flow simulations, but it also indicates possible work for future research and analysis.

Improvements to the CA could include for example a model for the evolution of water content during the flow, to be used instead of the average value currently employed. More sophisticated laws to describe the viscosity of lava by taking into consideration also the chemical composition, degree of crystallization and amount of gas are also currently available, and they could be integrated into MAGFLOW, with the caveat that the values of the additional parameters required by these models are rarely available at the onset of an eruption.

The preliminary tests on effusion rates variation also confirm the findings of Miyamoto and Sasaki (1997a) about its relevance for flow emplacement. Our tests provide however further information, showing that the overall shape of the effusion rate curve can have a much higher impact than the errors in the instantaneous magnitude of the mass flux itself.

While a more thorough analysis of the effects of the overall flux rate trend on the emplacement should be considered, to understand which aspects of the flux rate curve are more relevant (such as number of peaks, slope, etc.), the information offered by our tests already has very important implications in practical applications: in particular, the results indicate that the accuracy of predicted lava flows can be considerably improved by developing techniques for continuous or high-frequency measurements of the flux rate with a reasonable error rather than by improving the accuracy of less frequent measurements. Remote (Vicari et al., 2009, 2011) and ground-based (James et al., 2009) sensing techniques are a good step in that direction.

Finally, the sensitivity tests provide useful information to improve the reliability of forecast scenarios, highlighting the parameters and data for which accuracy is more important. When such data are nevertheless missing, reliable estimations of minimum and maximum possible extent of the emplacement can be obtained by running multiple simulations considering the extremal possible values of the unknown parameters.

Acknowledgments

This study was undertaken with the financial support from the V3-LAVA project (INGV-DPC 2007–2009 contract).

The authors wish to thank the editor, A.J. Jakeman, and the unknown reviewers for their precious suggestions on how to improve our methodology and the overall quality of the paper.

References


By comparing the results of this pair of tests, the most important result that emerges is that the overall trend of the effusion rate can be more important than the individual magnitude of the flux rates: for example, the simple knowledge of the total emitted mass, even if exact, will not provide sufficient information to correctly reproduce a real event, since the lack of knowledge of the flux rates during the eruption can impact the emplacement reducing the fitness by as much as 40%, a drop that would only be matched by errors in the flux rates estimate of 45% or more.


