Groundwater flow and heat transport for systems undergoing freeze-thaw: Intercomparison of numerical simulators for 2D test cases

Christophe Grenier[1], Hauke Anbergen[2], Victor Bense[3], Quentin Chanzy[1&4], Ethan Coon[5,6], Nathaniel Collier[6], François Costard[7], Michel Ferry[8], Andrew Frampton[9], Jennifer Frederick[10], Julio Gonçalvès[11], Johann Holmén[12], Anne Jost[13], Samuel Kokh[14], Barret Kurylyk[15], Jeffrey McKenzie[16], John Molson[17], Emmanuel Mouche[1], Laurent Orgogozo[18], Romain Pannetier[9], Agnès Rivière[19], Nicolas Roux[1&7], Wolfram Rühaak[20], Johanna Scheidegger[21], Jan-Olof Selroos[22], René Therrien[17], Patrik Vidstrand[22], Clifford Voss[23]

[1]{Laboratoire des Sciences du Climat et de l'Environnement, IPSL/LSCE, CEA-CNRS-UVSQ, Université Paris-Saclay, Gif-sur-Yvette, France}
[2]{APS Antriebs-, Prüf- und Steuertechnik GmbH, Rosdorf, Germany}
[3]{Department of Environmental Sciences. Wageningen University, Netherlands}
[4]{ENS Cachan, Département Génie Mécanique, Université Paris-Saclay, Cachan, France}
[5]{Los Alamos National Laboratories, Los Alamos, USA}
[6]{ORNL Climate Change Science Institute, Oak Ridge, TN, USA}
[7]{Geosciences Paris Sud, GEOPS/IPSL, Université Paris-Saclay, Orsay, France}
[8]{MFRDC, Orvault, France}
[9]{Department of Physical Geography, Stockholm University, Stockholm, Sweden}
[10]{Sandia National Laboratories, Albuquerque, NM, USA}
[11]{Aix-Marseille University, CNRS, IRD, CEREGE, Aix-en-Provence, France}
[12]{Golder Associates, Stockholm, Sweden}
[13]{Sorbonne Universités, UPMC Univ. Paris 6, CNRS, EPHE, Metis, Paris, France}
[14]{Maison de la Simulation USR 3441, Digiteo, CEA Saclay, Gif-sur-Yvette, France}
[15]{Centre for Water Resources Studies and Department of Civil and Resource Engineering, Dalhousie University, Halifax, NS, Canada}
[16]{Department of Earth and Planetary Sciences, McGill University, Montreal, PQ, Canada}
[17]{Département de géologie et de génie géologique, Université Laval, Québec, Canada}
[18]{GET, UMR 5563 CNRS-IRD-UPS, Université Toulouse, Toulouse, France}
[19]{PSL Research University, MINES ParisTech, Centre de Géosciences, Fontainebleau, France}
[20]{Federal Institute for Geosciences and Natural Resources (BGR), Hannover, Germany}
[21]{British Geological Survey. Nottingham, UK}
[22]{Swedish Nuclear Fuel and Waste Management Company, Stockholm, Sweden}
[23]{United States Geological Survey, Menlo Park, CA, USA}

Correspondence to: C. Grenier (christophe.grenier@lsce.ipsl.fr)

1 **Abstract**

2 In high-elevation, boreal and arctic regions, hydrological processes and associated water bodies

3 can be strongly influenced by the distribution of permafrost. Recent field and modelling studies

4 indicate that a fully-coupled multidimensional thermo-hydraulic approach is required to

5 accurately model the evolution of these permafrost-impacted landscapes and groundwater

6 systems. However, the relatively new and complex numerical codes being developed for

7 coupled non-linear freeze-thaw systems require verification.

8 This issue is addressed by means of an intercomparison of thirteen numerical codes for two-

9 dimensional test cases with several performance metrics (PMs). These codes comprise a wide

10 range of numerical approaches, spatial and temporal discretization strategies, and

11 computational efficiencies. Results suggest that the codes provide robust results for the test

12 cases considered and that minor discrepancies are explained by computational precision.

13 However, larger discrepancies are observed for some PMs resulting from differences in the

14 governing equations, discretization issues, or in the freezing curve used by some codes.

15 **Keywords**: numerical simulation, code benchmarking, thermo-hydrological coupling,

16 permafrost, sharp interface problems

17

18 **1   Introduction**

19 Climate change has been most pronounced at high latitudes (McBean et al., 2005; IPCC, 2013)

20 and high elevations (Pepin et al., 2015), and these trends are expected to continue in the coming

21 decades (IPCC, 2013). In these cold regions, hydrological processes are influenced by changing

22 precipitation regimes (e.g., Serreze et al., 2000; Hinzman et al., 2005) but are also very sensitive

23 to increasing air and ground temperatures because frozen ground stores and transmits water

24 very differently than unfrozen ground (Rowland et al., 2010; Walvoord and Kurylyk, 2016 and

25 references therein). Frozen ground retards or inhibits horizontal and vertical groundwater flow

26 due to the influence of pore ice on hydraulic conductivity. Thus, permafrost thaw or changing

27 patterns of seasonal subsurface ice can result in new or enhanced surface water distributions

28 (Connon et al., 2014) and can modify subsurface water pathways (Kurylyk et al., 2014a;

29 Frampton and Destouni, 2015) which can thereby affect the hydrologic and hydrogeologic

30 connectivity of a landscape. A number of studies note, for example, long-term increasing trends

31 in groundwater discharge (i.e. river baseflow) (e.g., Duan et al., 2017; Rennermalm et al., 2010;

St. Jacques and Sauchyn, 2009; Walvoord and Striegl, 2007) and decreases in seasonal variability (Frampton et al., 2011, 2013) due to permafrost thaw. Such behavior has a corresponding influence on ecosystem evolution and carbon storage and release (Schädel et al., 2016; Schuur et al., 2015; Vonk & Gustafsson, 2013). This context provides the impetus for developing an improved understanding of the interrelationships between climate change, permafrost thaw, and groundwater flow systems.

Over the past decade, a new class of hydrogeologic computer simulation codes has emerged with specific functionality for cold regions. These codes enable predicting the impacts of climate change on hydrological and hydrogeological systems (cf. special issue of *Hydrogeology Journal*, Cold Regions Hydrogeology, in 2013). These new 'cryo-hydrogeology' codes couple the groundwater flow equation to a heat transfer equation with dynamic freeze-thaw processes (e.g. Coon et al., 2016; Nagare et al., 2015; Karra et al., 2014; Wellman et al., 2013; Grenier et al., 2013; Rowland et al., 2011; Dall'Amico et al., 2011; Frampton et al., 2011; Painter et al., 2011; Bense et al., 2009; McKenzie et al., 2007; Mottaghy and Rath, 2006; Hansson et al. 2004 ; Ippisch, 2001; Hwang et al. 1972).

The numerical solution to the set of coupled, non-linear thermo-hydrologic (TH) equations in these emerging models can be quite challenging because the pore water phase change terms (i.e. which describe freeze-thaw) lead to very narrow transition zones between the frozen and unfrozen regions. These high temperature gradients induce strong non-linear effects due to temperature-dependent thermal properties and hydraulic conductivities that depend on pressure, water content and temperature. The physical processes and solution strategies behind the various cryo-hydrogeology models vary among codes (Kurylyk and Watanabe, 2013), which provides the motivation for further code testing, comparison and validation. Furthermore, while the results presented in the above-mentioned studies are qualitatively intuitive, their accuracy and physical realism is unknown. A rigorous and widely accepted approach for properly validating these codes is a prerequisite for applying these codes to study impacts of climate change on hydrological and hydrogeological systems.

McKenzie et al. (2007) identified the need for such a code validation effort. They first considered an existing 1D analytical solution (the Stefan problem: a homogeneous system with conduction, phase change, and a partially frozen zone) and developed 2D test cases. More recently, Kurylyk et al. (2014b) revisited several 1D analytical solutions for ground freezing and thawing, and proposed that a solution for conduction and phase change with the addition

3

of pore-water advection could be applied for cryo-hydrogeological code benchmarking. In an effort to optimize the type of 2D test cases introduced by McKenzie et al. (2007) for intercomparison purposes, two new complementary 2D benchmark cases with full TH coupling, and with more explicit performance measures than considered by McKenzie et al. (2007), are proposed in the present paper.

The InterFrost project (wiki.lsce.ipsl.fr/interfrost) was launched at the end of 2014 with the intention of validating TH codes. The validation strategy proposes a progression from simple test cases with analytical solutions to more complex laboratory and field-scale systems. The first validation step is the Lunardini analytical solution used by McKenzie et al. (2007) (named the "T1 Case") and the second is the Kurylyk et al. (2014b) case with constant advection (named "TH1 Case"), where T refers to purely thermal conduction cases while TH refers to cases with thermal and hydrodynamic (fluid flow) processes. Both cases consist of analytical solutions to the posed problems. The TH1 case is not a fully coupled system (i.e. water flow is independent of the temperature field), and both analytical solutions are limited to 1D heat transfer.

The solutions to the new set of intercomparison cases presented here are obtained with numerical, rather than analytical, approaches. While numerical solutions do not inherently guarantee exact solutions to the stated mathematical problem as do analytical solutions, agreement among several numerical approaches (represented by the different codes used here) may build confidence that the solutions are correct.

The increased flexibility in the numerical approaches used here also allows for full TH coupling with 2D geometries. The next two intercomparison cases, herein named the "TH2 Case" and "TH3 Case", consider the evolution of more complex 2D systems with changing frozen and unfrozen regions under various hydraulic gradients. Evaluation of the numerical codes against these benchmarks is based on the intercomparison of simulation results through a set of performance metrics (PMs). Conclusions are drawn regarding similarities and differences among model responses over a range of parameters.

Thirteen codes developed and/or applied by research institutes from USA, Canada, Sweden, Germany, UK, the Netherlands, and France were compared in the InterFrost project. A brief presentation of the participating codes and laboratories is provided in Appendix A. All codes were successful in matching the T1 and TH1 analytical-solution test cases before being applied to the TH2 and TH3 test cases. This paper presents results of the TH2 and TH3 test cases as

4

well as a detailed intercomparison of results from the thirteen codes. The initial work conducted for the T1 (McKenzie et al., 2007) and TH1 (Kurylyk et al., 2014b) test cases is not presented, as these tests were intended as a first validation step within InterFrost. However, these preliminary analyses helped to resolve a number of previously unanswered questions such as the appropriate spatio-temporal resolution for such simulations. The TH2 and TH3 results presented herein represent more complex scenarios than the seminal T1 and TH1 cases and should provide additional standards by which to verify future codes.

This verification effort will provide a basis for addressing more general challenges related to numerical simulation of cold-region hydrogeological processes, including coupling with the geo-mechanical processes of soil consolidation and frost heave (Painter et al., 2013). The results are directly relevant to many fields of research, including cold regions hydrology and hydrogeology (see e.g. Lemieux et al, 2016; Nagare et al., 2015; Wellman et al., 2013; Rowland et al., 2011; Frampton et al., 2011; Bense et al., 2009; McKenzie et al., 2007), subsea permafrost and evolution of methane hydrates (e.g. Frederick and Buffett, 2014), nuclear waste storage (e.g. Vidstrand et al., 2013; Grenier et al., 2013; Holmen et al., 2011), fluvial geomorphology (Costard et al., 2014; Dupeyrat et al., 2011; Randriamazaoro et al., 2007), Mars studies (e.g. Painter, 2011; Grimm and Painter, 2009; Aguirre-Puente et al., 1994; Clifford, 1993), climate modelling (1D models, see e.g. Gouttevin et al., 2012), terrestrial ecosystem changes (Kurylyk et al., 2016; Briggs et al., 2011), artificial ground freezing in geotechnical engineering (e.g. Zhou and Meschke, 2013; Pimentel et al., 2012), and acid mine drainage in cold regions (Elberling, 2005).

International benchmarking exercises have proved very fruitful in the past to build confidence in numerical models and propel the development of a new generation of codes in closely related fields (e.g. Kolett et al. 2017; Pruess et al., 2014; Gustafson et al., 2009; Larsson, 1992).

2 Theory, equations and numerical issues

The previously cited literature provides a wide range of expressions for water flow and heat transport equations in porous media. For the purpose of intercomparisons within the InterFrost group, a set of ad hoc TH equations was agreed upon with the intention of avoiding high levels of complexity that could preclude some developers from participating. The adopted option was to look for the simplest combination of equations and parameters while still maintaining the

important physical processes and interrelationships (conduction, advection, latent heat, transient Darcy flow and first-order coupling).

Consequently, the set of equations used in this study is not intended as a reference system for all code developers. It should, however, provide a good platform for the seminal intercomparison of TH code simulations.

This evaluation step first addresses the challenges of simulating the complex behavior of a coupled non-linear set of equations with moving sharp interfaces (phase change regions). A more complete code validation project (as intended later with InterFrost) will include comparisons with laboratory and field conditions. Although the set of equations required in such a second phase may differ from those used here, the overall validation procedure would be similar and would rely on the confidence gained from these simple 2D test cases. Consequently, parameter simplicity is maintained in all comparison cases. For example, we apply a simple empirical freezing curve with an impedance factor, although other options may be preferred (e.g. based on a deduction from the unfrozen soil moisture characteristic curve or heterogeneities, see discussions by Kurylyk and Watanabe, 2013; Painter and Karra, 2014, Amiri et al., 2018).

The Darcy flow and energy transport equations considered for the benchmark are shown in Eqs. (1) and (2), respectively, as:

$$(S_w \epsilon \rho_w g \beta)\frac{\partial p}{\partial t} = \vec{\nabla}.\left[K_w \vec{\nabla}p\right] + \vec{\nabla}.\left[K_w \vec{\nabla}z\right] - \epsilon\left(\frac{\rho_w - \rho_i}{\rho_w}\right)\frac{\partial S_w}{\partial t} \quad (1)$$

$$\left(\rho_w S_w c_w \epsilon + \rho_i S_i c_i \epsilon + (1-\epsilon)\rho_s c_s - \epsilon \rho_i L \frac{\partial S_i}{\partial T}\right)\frac{\partial T}{\partial t}$$

$$= \vec{\nabla}.\left(\lambda_t \vec{\nabla}T\right) + \vec{\nabla}.\left[\rho_w c_w K_w T \, \vec{\nabla}p + \rho_w c_w K_w T \, \vec{\nabla}z\right] \quad (2)$$

All fixed parameters or functions corresponding to Eq. 1 and Eq. 2 are presented in Table 1 (or in the development below). The porous medium is assumed fully saturated with water (liquid or ice), and no source or sink terms are considered.

The heat transfer equation (Eq. 2) includes conduction and advection (with the Darcy velocity split into two terms to solve with pressure head as the dependent hydraulic variable) as well as phase change (expressed with latent heat in the heat capacity term). Local (grain-fluid) thermal equilibrium is assumed. Considering the small range of temperature variations, no density-driven convection is included (Table 2 & 3).

Eq. 1 corresponding to the water mass conservation was obtained after simplifying the water mass conservation equation (Eq. 3):

$$\frac{\partial(\epsilon \rho_i S_i + \epsilon \rho_w S_w)}{\partial t} = -div(\rho_w \vec{U}) \qquad (3)$$

where $\vec{U}$ is the Darcy velocity. All other variables and parameters are defined in Table 1. No source term is included. Assuming that porosity does not evolve with time, since soil consolidation, thaw settlement or frost heave is not considered (Bear and Bachmat, 1991), and that $S_w + S_i = 1$, the system simplifies to:

$$\epsilon S_w \frac{\partial \rho_w}{\partial t} + \epsilon(1 - S_w)\frac{\partial \rho_i}{\partial t} + \epsilon(\rho_w - \rho_i)\frac{\partial S_w}{\partial t} = -div(\rho_w \vec{U}) \qquad (4)$$

In the absence of ice compressibility (negligible compared with water compressibility), the governing equation reduces to:

$$\epsilon S_w \frac{\partial \rho_w}{\partial t} + \epsilon(\rho_w - \rho_i)\frac{\partial S_w}{\partial t} = -div(\rho_w \vec{U}) \qquad (5)$$

Considering that the water density depends primarily on pressure (and to a relatively negligible extent on temperature in the ranges considered), and introducing the water compressibility coefficient $\beta = \frac{1}{\rho_w}\frac{\partial \rho_w}{\partial P}$, the pressure head $p = \frac{P}{\rho_w g}$, and dividing both sides by ρ_w, Eq. 5 finally yields Eq. 1 considered for the benchmark.

The initial development of the TH equations and solutions of the proposed InterFrost test cases were first completed with the Cast3M code (www-cast3m.cea.fr). Cast3M had already been applied to various 1D and 2D cases (Régnier, 2012; Grenier et al., 2012; Roux et al., 2017) based on a somewhat different system of equations.

3   Presentation of test cases

The TH2 and TH3 test cases were inspired by real-world conditions and from examples presented in the literature. These cases were specifically defined to enable testing the coupled system of equations, while remaining as simple as possible to allow running with the largest range of codes, without major code modifications. In particular, simple 2D geometries were considered to minimize meshing issues and to remove the need for specific mesh generators.

1   Furthermore, all properties remained isotropic, and both cases could be run on a symmetric

2   half-domain to save computational time.

3   These complementary cases address two levels of coupled TH complexity. The first case (Case

4   TH2) deals with the thawing of an initially frozen inclusion subject to a simple water flow

5   regime with constant positive input temperature. The second (Case TH3) represents a simplified

6   talik (an unfrozen zone within permafrost) with competition between permafrost formation

7   (with imposed negative boundary temperatures) and permafrost thaw due to the inflow of warm

8   water. Inflow water temperature and flow rates are the controlling parameters for the evolution

9   of the talik which may expand or shrink depending on whether the head gradients and associated

10   water flow rates exceed a threshold. The threshold depends on the thermal properties, boundary

11   conditions, and system geometry.

12   Test cases TH2 and TH3 are presented in the following sections in more detail, with the purpose

13   of providing a basic understanding of the physics of the system evolution. Simulations obtained

14   with the Cast3M code are provided to support the presentation and analyses. The discretization

15   parameters associated with these reference simulations (e.g. meshing and time steps) were

16   selected after running convergence studies from which a compromise was found between "full"

17   convergence and computational efficiency. The spatial and temporal sensitivity analyses

18   conducted with the Cast3M code are presented in Appendix B. The analysis was used to check

19   the numerical convergence of the preliminary Cast3M simulations as well as to estimate the

20   spread in the performance metrics associated with various levels of space and time

21   discretization.

3.1   Case TH2: "Frozen Inclusion Thaw"

23   The conditions for Case TH2 ("Frozen Inclusion Thaw") are provided in Fig. 1 and Table 2.

24   The parameter values applied in the benchmark are provided in Table 1. This case is inspired

25   by the spherical inclusion problem in the heat transfer literature (e.g. Carslaw and Jaeger, 1959)

26   and it simplifies the geometry of the frozen wall case of McKenzie et al. (2007), which

27   considers similar effects but for freezing rather than thawing.

28   The initial frozen zone inclusion is square (Fig. 1a), with constant head flow boundaries along

29   the vertical sides (right and left) and no-flow conditions on the horizontal boundaries above and

30   below (Fig. 1b). The initial condition heads are assumed uniform, equal to the downgradient

boundary head. The thermal transport boundary conditions are an imposed temperature of 5 °C at the upgradient (left) inflow boundary, with zero temperature-gradient conditions (zero conductive fluxes) along all other boundaries (Fig. 1c). Outside the inclusion, the water temperature is initially uniform at 5 °C. The system is simulated with and without thermal advection. Evolution of the initial frozen zone depends on heat conduction from the transient temperature field surrounding it and on thermal advection from upgradient warm water driven by an imposed hydraulic gradient controlled by the boundary heads.

The head field (Fig. 2a) and the temperature field (Fig. 2b) simulated with Cast3M at time 22,860 s (6.4 hours), assuming a head gradient of 3%, are provided in Fig. 2 (as a color contour plot (upper) and a line-contour plot (lower)). These fields are symmetrical relative to the longitudinal x-axis through the center of the inclusion. The associated time is shortly before the threshold time when the minimum temperature rises above 0°C. The initial frozen inclusion is now rounder in shape and warmer compared to the initial -5°C conditions, while a cooler temperature plume has been transported downgradient (see Fig. 2b). The head and flow fields show that water flows around the frozen inclusion and that hydraulic head gradients ($p+z$ in Eq. 1) are higher (head contours are more densely spaced) within the frozen region due to its lower hydraulic conductivity.

Figure 3 presents the time evolution of the temperature plume is for two different head gradients. For a zero head gradient (without thermal advection), the -5°C inclusion gradually thaws due to conductive heat transfer from the initially warmer 5°C temperature field around it (Fig. 3a). The temperature profiles are symmetrical along the central longitudinal axis at early times but become asymmetrical over time due to the greater distance to the right side boundary condition (zero conductive heat flux) relative to the closer left (inflow) boundary where a 5 °C condition was imposed. The minimum domain temperature increases as the initial inclusion warms. For the simulation in which advection is included with an imposed head gradient of 9% (Fig. 3b), thermal degradation due to the inflowing 5°C water accelerates the thaw process. Because of its lower effective hydraulic conductivity, water initially flows around the inclusion leading to a relatively colder downstream thermal plume and strongly asymmetrical temperature profiles. At later times, when the inclusion has completely thawed, the remnant cold water region is transported downstream by advection and thermal dispersion, forming the lower amplitude temperature profiles such as the profile at 16.6 hours, shown in red (Fig. 3b).

3.2 Case TH3: "Talik Opening / Closure"

The conditions for Case TH3 "Talik Opening / Closure" are provided in Fig. 4. The geometry of the system and the initial conditions are defined by two frozen (-5°C) semi-circular zones within a positive background temperature field of 5°C (Fig. 4a). Flow conditions (Fig. 4b) are the same as for Case TH2 with fixed heads along the upgradient (left) and downgradient (right) faces, and zero-flux (no-flow) conditions elsewhere (compare Fig. 1b & 4b). Thermal boundary conditions (Fig. 4c) are imposed temperatures of -5°C on the upper and lower boundaries, 5°C on the (left) upstream boundary, and zero conductive flux on the (right) downstream boundary. Heat thus exits the system from this right boundary through advection alone. The parameter values are identical to Case TH2 (refer to Table 1).

This case represents the evolution of a talik - an unfrozen zone within permafrost. Such conditions of heat exchange are found in the real world where water flows between two permafrost zones. Horizontal internal taliks exist permanently within some permafrost regions (e.g. Gagarine, 2012; Anisimova et al., 1973), while vertical flow-through taliks can often be found below a lake or a large river and provide a direct connection with a sub-permafrost aquifer (see for instance the cases of Wellman et al., 2013, Rowland et al., 2011, and Bense et al., 2009). Case TH3 (Fig. 4) can apply to both real-world cases as the simulations do not include density effects, and thus the results are independent of orientation. TH3 is simulated under hydrostatic (no-flow) conditions as well as under different background hydraulic gradients controlled by the imposed boundary heads (similar to TH2).

The simulated head field (Fig. 5a) and temperature field (Fig. 5b) at time 19,860s (5.5 hours), assuming a head gradient of 3%, are provided in Fig. 5 (as a color contour plot (lower) and a line-contour plot (upper)). These fields are symmetrical relative to the central longitudinal x-axis. At the indicated time, and with this low head gradient of 3%, the talik is approaching closure as shown by the 0°C isotherm which already reaches the central axis (Figure 5).

The time evolution of the temperature plume is next shown for two different head gradients (Fig. 6). With a zero head-gradient (pure conduction, Fig. 6a), the two initially frozen zones merge across the central initially unfrozen part of the domain. The flow stops rapidly as the talik closes. The case including thermal advection (9% head gradient) is shown in Fig. 6b. In this case, cooling in the central part is progressively retarded due to the inflow of warmer water by advection, and the cooler plume is shifted downgradient (Fig. 5b). After this initial phase,

further evolution of the system depends on the amount of heat advected into the talik zone. For low rates of advection, the system eventually closes because conductive heat loss towards the frozen boundaries in the upper and lower zones dominates, whereas for cases with more rapid thermal advection, such as in the case with a 9% imposed head gradient corresponding to Fig. 6b, heat exchange widens the unfrozen central pathway until steady-state conditions are established and the talik remains open.

4   Performance measures

The performance measures (or performance metrics; PMs) are introduced to facilitate code intercomparisons. With the aim of studying the importance of advection, all performance measures are computed for each test case and for a range of imposed head gradients (see Table 2 and Table 3).

Table 4 provides an overview of the PMs, which include bulk (integrated) system parameters as well as point values. The time evolution of these PMs is computed and displayed for all simulation results and codes. The entire series of PMs provides an efficient coverage of all physical thermal and hydrological conditions at various scales. The chosen PMs, which include total heat, total liquid water volume and the minimum domain temperature, correspond to several physical conditions of interest for applications in the real world (ex. computing total fluxes across a boundary, talik evolution considering equivalent hydraulic conductivity) or are necessary for numerical purposes to check heat and water budgets. The PMs reflect unique system behaviors. It should be noted that the advective heat fluxes and the total sensible heat are evaluated with 0 K as the thermal datum (e.g., Lee, 1998), although the choice of the thermal datum will not influence the calculation of the net advective flux. A more detailed description of these PMs follows.

4.1   Case TH2: frozen inclusion thaw

For Case TH2, three performance measures are introduced, combining thermal and hydrodynamic information:

- The first PM (TH2_PM1) is the minimum field temperature. Approaching complete thaw of the inclusion, the point of minimum temperature will remain within the inclusion, then will start to migrate downgradient with the cold water plume.

- The second PM (TH2_PM2) is the net heat flux (Joules) leaving the system (J_{net}). This heat flux is calculated as the flux density integral at the upstream (*u*) and downstream (*d*) vertical boundaries (note that no energy is transferred across the top and bottom no-flow, perfectly insulating boundaries):

$$J_{out} = \int_d \left(\rho_w c_w T U_x - \lambda_T \frac{dT}{dx} \right) dy - \int_u \left(\rho_w c_w T U_x - \lambda_T \frac{dT}{dx} \right) dy$$

  where U_x is the horizontal component of the Darcy flux. Heat flux leaving the system is defined here as positive. Temperatures are expressed in kelvins taking 0K as the reference. Note 0°C could be introduced as the reference. Such a 2D flux depends on a third dimension, arbitrarily chosen here as unity (1 m).

- The third measure (TH2_PM3) is the total volume of liquid water in the model domain. This measure corresponds to an integral over the total volume of the domain and is obtained by summing the product of the liquid water saturation, porosity, and representative volume for each node across the domain. As was assumed for PM2, the transverse width is set to 1m.

The PMs are here illustrated and analysed with the Cast3M simulations (Fig. 7). Special attention is first placed on TH2_PM1, the evolution of the minimum domain temperature (cf. Fig. 5a), focusing on the case with a head gradient of 3% (purple line) for discussion of evolution Phases 1 to 4. Within the first few minutes, the frozen inclusion rapidly warms due to conductive heat loss to the surrounding flowing water, leading to a rapid rise in the temperature of the center of the inclusion (refer to Phase 1 in Fig. 7a). The temperature curve then reaches a slowly evolving plateau between -1 and 0 °C (Phase 2), corresponding to the phase change region of the simulated porous medium, where adsorption of latent heat slows the rate of temperature increase, known as the 'zero curtain effect' (Romanovsky & Osterkamp, 2000) (see Table 2). After reaching 0°C, phase change is complete, and the temperature increase accelerates again (Phase 3). The warming rate then decreases as the cooler zone corresponding to the initial inclusion becomes mobile and moves downstream through what has become a uniform flow field (Phase 4). The evolution of the minimum field temperature is provided in Fig. 7a for other rates of advection (head gradients from 0 to 20%). This broad range of simulations is provided here for the sake of illustration, while the requirements for the task participants were limited to a subset of 4 head gradient values (Table 4). The shape of the purely

conductive case (0% gradient) also displays a phase change domain of between -1°C and 0°C and is typical of a diffusion (or equivalently conduction) process. Advection modifies the profile by accelerating the early warming and providing a final rapid rise in the temperature once the center of the advected thermal plume has exited the domain.

The times required for the minimum temperature to reach 0°C can be plotted as a function of head gradient (Fig. 7b) integrating the TH2_PM1 results over the entire range of head gradients. Results show intuitively that the larger the head gradient, the narrower the peak and the quicker the low-temperature plume exits the system (Fig. 7b).

The evolution of net heat flux leaving the system (TH2_PM2) for Case TH2 (Figure 7c) exhibits a rapid increase as the low-temperature plume resulting from the thawing of the initial frozen zone crosses the downstream boundary. Net heat fluxes are negative as a result of the colder inclusion and the convention of assuming positive heat fluxes when thermal energy (relative to background temperature) is leaving the system. This performance measure is equivalent to the total heat flux exiting the downstream (right) boundary, removing the base heat flux associated with advection of the background 5°C temperature field.

The total unfrozen (liquid) water volume in the system (TH2_PM3) increases as time proceeds until it stabilizes at a level of 1.11 m^3 when the minimum temperature of the system is above zero (Fig 7d). This performance measure provides information about the initial phase of thawing of the frozen inclusion. As expected, results show that the stronger the advection intensity, the quicker the thermal degradation of the frozen inclusion (Fig. 7d). (Showing the total heat content throughout the domain would be an alternative performance measure but would contain similar or redundant information).

4.2   Case TH3 "Talik Opening / Closure"

Three performance measures are considered here,

- The evolution of the bulk system equivalent hydraulic conductivity (TH3_PM1). This is computed for each time step as the integrated Darcy flux ($Q = \int q \, dx$) at the right boundary divided by the imposed head gradient: $K_{eq} = \frac{Q}{\Delta H / L_X}$

- The evolution of lateral heat flux (TH3_PM2). This is calculated by integrating the conductive heat fluxes across the upper and lower boundaries and considering an arbitrary 1 m length for the transverse dimension.

- The evolution of total sensible heat J_T (TH3_PM3), which is computed as: $J_T = \iint (\rho_w C_w \epsilon S_w + \rho_i C_i \epsilon (1 - S_w) + \rho_S C_S (1 - \epsilon)) T\, dxdy$. The transverse dimension is again assumed to be 1 m.

- The point temperature evolution at two fixed dedicated locations of the domain (TH3_PM4).

The parameters associated with the reference simulation shown in the present section were chosen after convergence studies as a compromise between "full" convergence and computational efficiency (see Appendix B. The following text refers to Fig. 8 and simulations with Cast3M code.

The equivalent hydraulic conductivity K_{eq} (TH3_PM1) measures the hydraulic impedance associated with the evolution of the talik. This conductivity is computed for each time step based on the net water flux and the hydraulic gradient. For no or low levels of advection, the system closes and the equivalent hydraulic conductivity decreases to near-zero as the system becomes completely frozen. Sensitivity to the head gradients between 0 and about 6.3% shows that the higher the advection, the slower the decrease in bulk K_{eq} (Fig. 8a). At higher flow rates (head gradients $\geq$ 6.4%), the heat transferred by advection is sufficient to thaw increasing portions of the initially frozen domain thus increasing the bulk hydraulic conductivity. The head gradient threshold between both regimes (i.e. when the competing effects of conduction and advection were equivalent) was found to be between 6.3 and 6.4 % (grey curves in Fig. 8a, TH3_PM1, and dotted line around 6.35% in Fig. 8b providing talik closure times for low head gradients).

The evolution of the total conductive heat flux (TH3_PM2) is also computed (Fig. 8c). As in TH2, the heat fluxes are assumed positive when exiting the system. Results show the greater the advection, the higher the absolute value of the heat flux entering the system (across the imposed fixed-temperature boundary). The total sensible heat (TH3_PM3) exhibits the same dual behaviour as previous performance measures (Fig. 8d): before a threshold gradient is reached, the initial talik closes, reflecting the state where advection cannot transport enough energy in from the boundary and the total heat of the system subsequently declines. For greater

14

levels of advection, the talik opens, leading to increasing levels of total sensible heat and to subsequent thawing.

Some locations in the domain record rapid temperature changes. The case presented in Fig. 8e is the temperature evolution at the center of the system at Pt1 (x,y = 0.5, 0.5 m, TH3_PM4_Pt1). For low imposed head gradients, the temperature drops below zero, reflecting the closure of the talik, while for higher gradients, the talik opens. In the latter case, the temperature initially drops due to high heat conduction from the nearby frozen zones before increasing again under the influence of heat advection. Pt2 is located in the initially frozen zone (Table 4), the divide between the initially frozen and unfrozen zones falling in the middle location between Pt1 and Pt2 along the vertical axis.

The system evolves to reach a permanent equilibrium state where the 0°C isotherm becomes closer to the upper and lower boundaries for increasing head gradients (Fig. 8f) leading to wider central thawed zones. Such steady-state profiles could be considered as complementary performance measures. However, the 0°C interface is not sharp under steady-state conditions, thus its exact location is somewhat uncertain. Moreover, the computational effort becomes high when true steady-state conditions are desired.

5 Presentation of codes and participants

Participation in the InterFrost benchmark was promoted at scientific conferences and accepted on a voluntary basis without dedicated funding. Thirteen codes have joined InterFrost. Several codes were recently developed, but most were developed within existing simulation platforms (i.e., software packages to simulate flow and transport in porous media) that were further enhanced to account for phase change phenomena. The codes span a spectrum of numerical approaches: finite difference, finite element, finite volume, and hybrid approaches. Non-linearities and coupling of equations are dealt with by means of iterative approaches (Picard or Newton schemes) or direct inversion. The meshing strategy is either fixed (structured and unstructured meshes) or dynamically modified to follow steep temperature gradients. The time step strategy is defined in advance or is inherently adaptive. Several of these codes are massively parallel, while others only run on a single core. Information relative to each code is provided in Table 5. More detailed information and a list of references is provided in Appendix A

The simulation conditions considered for the reference calculations and the 13 participating codes are provided in Table 6 (TH2 Case) and Table 7 (TH3 Case). Most codes used convergence studies for refinement of the spatial discretization and time step size. The influence of convergence criteria was investigated as well. As a result, the reference simulations provided are expected to be spatially and temporally converged, and finer discretizations would give the same results. Tables 6 & 7 show that a vast majority of codes simulated domains that contained on the order of tens of thousands of elements or cells, with some taking advantage of spatial symmetry.

6 Results of inter-code comparison

In the absence of any available reference simulation (compared to the analytical solution benchmarks that could be solved directly), solutions for TH2 and TH3 were obtained via an inter-code comparison of the simulation results as well as by using the convergence studies noted above. The inter-code comparison is first developed by comparing the results obtained by all codes for each PM (Section 6.1) and then enhanced by a detailed analysis of specific PMs in the cases where thresholds appear (Section 6.2). A more detailed analysis of the results, as well as a study of the sources of discrepancies, is treated in detail in Appendix C.

The complete series of PMs accounts for 32 comparable metrics (Table 4), composed of 3 PMs associated with TH2 under 4 head gradients (0, 3, 9, 15%) and 5 PMs for TH3 under 4 head gradients (3, 6, 9, 15%). Furthermore, additional plots representing threshold values (see Table 4) and times to reach steady state are provided for the complete series of head gradients. Although the analyses are built upon the full set of PMs, for the sake of saving space, only a representative subset of the PMs is provided in the present paper. Interested readers will find all of the raw plots and data for each PM on the InterFrost web site (wiki.lsce.ipsl.fr/interfrost/). A detailed analysis of the results produced by each code is not included because most codes behaved similarly. Overall conclusions are drawn about the intercomparison project, which provide a general and robust evaluation of the code performance.

It should, however, be noted that some codes taking part in the intercomparison could not fully comply with all the test case requirements within the short time of the project. For instance, simulations with PFLOTRAN-ICE remained bounded to a power law freezing function and thus could only apply an approximation of the exponential expression required in the

16

specifications (see Table 2). However, their simulations were kept in the intercomparison results to illustrate the sensitivity of the simulations to the freezing function. This sensitivity was high as demonstrated in the sensitivity analysis carried out with the Cast3M code, considering an exponential freezing curve and a range of parameters (see Appendix B). Other codes also had somewhat different terms in the equations or expressions for bulk properties but which did not have a significant influence on the results.

6.1 Visual illustration of PM results

Fig. 9 provides two PM compilations including all 13 codes for the TH2 case: the evolution of the minimum temperature within the simulated domain (TH2_PM1, Table 4) and the associated total unfrozen water content (TH2_PM3, Table 4). Both plots in Fig. 9 are for the case of a head gradient of 3%, a realistic value for a hydrogeological setting. The evolution of the PMs is qualitatively intuitive and follows the results proposed using Cast3M and analyzed in Section 4.1.

Intercomparison results show that all curves in Fig. 9 have the same qualitative pattern. From a quantitative perspective, some of the code results differ slightly from the group that achieved essentially identical results. This is especially visible for the minimum temperature curve where some codes show a discrepancy in the negative temperature region, especially in the transition range from -1°C to 0°C corresponding to the phase change. With PFLOTRAN-ICE (light blue curve), for example, this discrepancy in thawing dynamics (Fig. 9a) originates from the different freezing curve that is implemented in the code (cf. Table 1). However, the influence of the different freezing curves is not very apparent in the plot of the total liquid water volume (see Fig. 9b). This is due to the integrative nature of TH2_PM3 (integral of liquid water over the whole domain), a measure which is less discriminating than TH2_PM1. The same general resilience to slight differences in the freezing curve is also evident for TH2_PM2 (total flux exiting the system, which is not presented here). The influence of the freezing curve range on TH2_PM1 and TH3_PM1 is further discussed by means of a sensitivity analysis in Appendix B.

Although the spatially averaged PMs do not provide discriminating power for the system dynamics, they do provide another level of information. For example, the spread of the simulation curves in TH2_PM3 (Fig. 9b) for the initial conditions shows that all codes do not represent exactly the same initial liquid water volume. This point will be further examined

1   below (refer to discussions regarding Fig. C.2 in Appendix C), and highlights differences in the

2   meshing strategy which creates slightly different initial conditions. Fig. 9 further indicates that

3   the same level of discrepancy is maintained throughout the simulation time until the total

4   disappearance of the frozen inclusion (at threshold times around 7×10^4 s).

5   Fig. 10 provides compilations for TH3_PM4 at points Pt1 and Pt2 and for the 3% head gradient

6   case. Since both points are located close to the initial boundary between the frozen and the

7   unfrozen zone (see Table 4), the point temperatures rapidly evolve (4 to 5 °C difference over

8   about 10^4 s). Intercomparison results show that the difference in the freezing function changes

9   the temperature evolution for PFLOTRAN-ICE, but for all other codes the results are very

10   consistent, even at early times (before 10^4 s,). Temperatures in this central region (where Pt1

11   and Pt2 are located) are especially sensitive to the competing effects of advective vs. conductive

12   heat transfer, and this competition leads to either opening or closing of the talik. In coupled

13   thermo-hydraulic (advective-conductive) cases (hydraulic gradients >0), non-linearities could

14   lead to instabilities and oscillations between the frozen and thawed regimes. The similarity of

15   results in Fig. 10 suggests that the codes are sufficiently robust to deal with such difficult

16   conditions and confirms that they have converged to accurate solutions.

17   While Figure 11 (right axis) provides the total sensible heat of the system (TH3_PM3), it only

18   represents a 2D averaged measure where, except for two cases (PFLOTRAN-ICE and Ginette),

19   all heat evolution curves are effectively identical. Fig. 11 (left axis) provides another overview

20   of the performance of the 13 codes, specifically, it presents the equivalent hydraulic

21   conductivity (TH3_PM1) which is more discriminating. This PM is a key parameter which

22   clearly reflects the closure versus the opening of the talik. All codes predict system closure for

23   times ranging from 6.5 to 7.5×10^4 s except in one case (see Fig. 12 as well). Once again, the

24   PFLOTRAN-ICE results are slightly different due to the different applied freezing curve. Three

25   other curves (codes FEFLOW, and to a lesser extent GEOAN and SMOKER) display somewhat

26   different evolution while the remaining nine are nearly identical. Initial conditions for both PMs

27   (Fig. 11, initial time and Fig. C.2d) appear similar and do not explain the discrepancies. The

28   differences for these three curves are partly attributed to coarse meshing and insufficient

29   accuracy in the convergence of the non-linear coupling and partly to the method of computing

30   the equivalent hydraulic conductivity. The same reasons, especially coarse meshing (refer to

31   Table 7) would explain the results by Ginette for the total heat (Fig. 11). The sensitivity tests

32   done by some participants have shown that the discrepancies between equivalent hydraulic

conductivity computed with steady-state flow fields and transient flow fields are negligible. Similarly, other sources of discrepancies involving somewhat different equation sets, for instance specifically incorporating pressure terms associated with the water to ice volume change, proved negligible through additional sensitivity tests.

The general conclusion from visual analysis of these TH2 and TH3 results is that all curves provide qualitatively similar results or trends, except for PFLOTRAN-ICE which uses a different freezing curve. Such results are indeed intuitive and follow from the preliminary study with Cast3M analyzed in Section 4. From a quantitative perspective, the results among the other codes differ to only a limited extent.

A more quantitative measure of the spread among the curves associated with each PM was introduced in the form of a standard deviation curve. This provided insights into the level of discrepancies among the 13 codes which likely resulted from different initial conditions (as a consequence of meshing strategies), or from incomplete spatial convergence of the simulation results. Further, results showed that there was no significant difference in the spread of the results among codes with increasing head gradients for both TH2 and TH3. The reader is referred to Appendix C for a detailed presentation of these issues.

6.2   Effects of thresholds associated with the PMs

The conclusions drawn on the convergence of results do not apply for the case of TH3 at the 6% head gradient value, thus the special case of PMs associated with TH2 and TH3 thresholds is further discussed here. For TH2_PM1, the threshold time was defined as the time for the minimum temperature to reach 0°C. For TH3_PM1, the threshold time is associated with the time to reach 0.1 % of the initial equivalent hydraulic conductivity.

Fig. 12a shows the threshold times computed for TH2. The standard deviation of TH2_PM1 was at a maximum in this time interval for all head gradients (Fig. C.1, see Appendix C), indicating that obtaining accurate simulations for this time period is challenging. The time for the initially frozen zone to reach 0°C and the time to reach steady-state were computed by each code. All requested head gradients are represented (vertical dots: 0, 3, 9, and 15 %) as well as other head gradients when participants simulated intermediate cases. The threshold time decreases as head gradients increase due to the thermal degradation from advection. The range of results also decreases with increasing head gradients, with a moderate spread along the

average general trend. Fig. 12b displays the time required to reach steady-state conditions. The general trend is identical (reduction in difference between codes with increasing head gradients), although the overall spread along this general trend is much lower.

The threshold times appear to be sensitive to the individual simulation set-ups (e.g. mesh size) and individual codes, although results are consistent in the case of the steady-state times. For TH3, the important threshold does not correspond to a distinct time during the thermal degradation process, but is associated with a change in the system's hydrologic regime that is controlled by the head gradient. Fig. 13b displays the equivalent hydraulic conductivity (TH3_PM1) for a head gradient of 6 %, for which four codes predict talik opening while nine codes predict talik closure. A refined study conducted with the Cast3M code provided a threshold head gradient between 6.3 % and 6.4 %. Based on this information, the 6 % head gradient appears critical for this particular test problem. The large spread of observed results is primarily a result of simulating this threshold correctly. Even for the 9 codes predicting talik closure, the spread in the closure times is quite wide as seen from Fig. 13a (6 %). A similarly large dispersion of talik closure times is observed for 3 % (in the same plot). This was already discussed with respect to Fig. C.1 plots (see Appendix C), where closure time differences were attributed to differences in the computation of this PM, in addition to a divergence in the initial conditions. The somewhat different evolution obtained with GEOAN was attributed to specific difficulties met in this case (6%) to converge in the non-linear coupling associated with the latent heat term.

The threshold case therefore appears to be the most challenging problem for accurate quantitative simulations. The simulated characteristic times (time to reach 0°C or steady state) generally had very small standard deviations. However, the talik closure times are apparently difficult to accurately simulate, making predictions challenging.

7   Discussion and perspectives

The main conclusion focuses on the similarity of results, showing that the participating codes performed well in solving the TH benchmark problems. However, several important points are worth discussing and putting in perspective.

The conclusions of the paper are drawn from intercomparison of results from 13 simulation codes. This large number of codes, in addition to the diversity of numerical approaches, is a

valuable basis for drawing conclusions. Nevertheless, increasing both the number of codes and their diversity is important, and other modelers are invited to run these same cases to both test their codes and to complement the present dataset.

The approach taken here relies on an intercomparison of codes. A more reliable approach that actually verifies that a code is working correctly is to compare simulated results with an analytical solution. No analytical solution have been published that simultaneously considers multidimensional spatial groundwater flow and heat transfer with the freeze/thaw process. Thus, intercomparison of a variety of numerical simulation codes is the second best approach to verifying such codes for these types of cryo-hydrogeological problems. In a group process, where there is no pre-existing criterion for assessing which code is 'correct' (i.e. providing a numerical solution that is equivalent to the analytical solution, should one exist), arbitrarily choosing one model output as the correct reference solution is not advisable. No code can provide such a non-challengeable numerical solution. It remains, however, an interesting challenge for modelers and mathematicians to produce such an improved or reference solution in the future.

We adopted a code intercomparison strategy based on performance measures. The strategy combines real physical parameters, sometimes integrated to handle practical issues, and allows simple intercomparisons. Another strategy for intercomparing could have been considered based on direct comparison of the full 2D fields of the simulation variables (e.g. T, p). Rühaak et al. (2015) attempted to compare 2D field results which showed similar results to those demonstrated here. However, their comparison proved to be highly time consuming even with only 4 participating codes, and would be impractical for 13 codes. An extension of this approach to study the transient evolution of specific patterns or behavior associated with specific key processes would require adapted simulation outcomes, including a common file exchange standard, the development of associated interface tools, and a powerful tool for projecting results (given the range of meshing methods). Although promising and complementary, this is beyond the reach of our current project.

Another point concerns our conclusions about the general convergence of the results over the 13 participating codes while discrepancies within the results do exist. Their sources are discussed now. Discrepancies found in the intercomparison results among codes are likely in part associated with differences in the simulation set-up when implementing individual problems. Over the course of the study, some implementation errors were identified and

corrected by participants. Other differences from the specified conditions agreed on by all participants (e.g. related to equation formulation) could not be easily explained. The test cases were designed to be as simple and reproducible as possible. However, some codes could not be fully adapted to the different implementations of freezing characteristic curves and/or equivalent properties. Therefore discrepancies occur where differences in the basic relationships differed to some extent (e.g. methods used for calculating average bulk, water and ice properties). Furthermore, a remaining small difference in the simulations likely results from a lack of computing precision. Several sources were identified, especially coarse or non-dedicated spatial representation of the initial system conditions which can lead to subsequent code divergence, in addition to weakly converged simulations requiring more demanding conditions in terms of the number of internal non-linear iterations and spatial and temporal refinement. The latter more specifically concerns the computation of the non-linear latent heat term in Eq. 2 for which large non-linearities occur, leading to oscillations and/or difficulties obtaining high precision at low computational costs. Furthermore, the latent heat term requires the derivative of ice saturation as a function of temperature. This can be achieved in various ways (e.g. taking directly the analytical expression into account or a linear or higher order expression of the derivative) producing similar though quantitatively different results. Guidelines for such choices should be found first in the mathematical formulations associated with each specific numerical scheme considered. They should then be tested against 1D basic cases providing analytical expressions (e.g. Stefan problem with Lunardini analytical solution in the Introduction Section). Although the extrapolation of discretization rules obtained from 1D cases is not straightforward to 2D and 3D cases, ad hoc spatial and temporal strategies have been developed from such convergence studies, and are easy to implement on 1D geometries. The same is true for the choice of some key control parameters (e.g. convergence criteria for non-linear loops, parameters considered in under-relaxation algorithms for reducing the amplitude of the oscillations).

All these issues were studied to some extent with the Cast3M or PermaFOAM codes as reported in Appendix B or by other participants (not reported here) by means of sensitivity analyses. The latter issues (examined separately) showed more or less minor influences on the results. However their combined effects are more complex, possibly cumulative, and difficult to address within a reasonable period of time. The cases associated with the code results that differ the most have already been addressed in the results section.

The primary limitations appeared when comparing threshold predictions, especially for TH3_PM1 in the case of the tipping point between the talik opening and closure regimes that occurred for gradients close to 6 %. A better intercomparison strategy could have been associated with the prediction of the head gradient threshold but would have required all participants to sample the domain of head gradients which would have substantially increased the work load. The TH3 Case is more demanding than the TH2 Case, further challenging code capabilities – an important factor for code comparison exercises.

Beyond the code intercomparison presented in this study, other evaluation elements could be included in a more global verification procedure. InterFrost is considering a verification of the numerical solutions to test cases through cold-room experiments. For example, the TH2 experiment has been attempted in the laboratory, however the experimental set-up and measurements have been very challenging. Simulations of experimental laboratory results will require additional modelling effort and adaptations. Some level of calibration will probably be required as well, making the approach not as straightforward as first envisioned, but more realistic with respect to how codes are applied to real field problems.

This effort, however, including real world cases, can be a valuable part of the code verification process considered by InterFrost and is complementary to the existing intercomparison exercises. Comparison of code results with carefully-measured experimental (or field) data can help provide 'validation' of a code (a stronger context than 'verification'). This means that not only does a code solve the governing equations correctly (definition of 'verification'), but also, that the governing equations correctly describe the physical processes and external conditions that drive the system. InterFrost members are further planning the incorporation of field cases into future test cases.

Relatively favorable conditions were used in both test cases for reducing computational loads associated with time and space discretization. This will not necessarily be the case for all real-world applications. Simulating large spatial systems with small-scale heterogeneities, for example, will be challenging. Similarly, the temporal discretization needs to satisfy stability criteria for such cryo-hydrological systems, but also needs to be practical for realistic computation times. These questions may be considered in the future within the InterFrost project.

Applying our modelling approaches to real world cases will also cast the present discussions about numerical simulation precision in a practical context. The representation of the coupled TH processes in the real world will require constraining parameter values associated with properties that are subject to variations in heterogeneous media. The precision in the measurements and the spatial variability added to the difficulty to constrain some parameters (e.g. the freezing curve is often calibrated) will probably result in relatively high levels of uncertainty. These sources of error and uncertainty will have to be compared to the discrepancies resulting from the numerical simulation bounds identified in the present study. Preliminary simulations applied to the experimental case with the frozen inclusion already show that the numerical errors associated with the simulation are low compared to the uncertainties resulting from parameter values.

8   Conclusions

The development of coupled multi-dimensional TH numerical groundwater models with dynamic phase change for cold region applications is relatively recent. The simulation of such systems is challenging due to the coupled and non-linear structure of the equations, creating a complex evolution between frozen and unfrozen zones. The present study attempts to improve the quality and confidence in such codes through intercomparison of results from synthetic 2D test cases.

The present InterFrost intercomparison exercise was valuable for gathering a critical number of participants (13 participating codes) and examining a large variety of numerical approaches and code implementations. Based on the two test cases (TH2 and TH3), most code simulation results generally converged around a single (unique) solution; thus, a consensus was reached among the participants. A group of these models, resolving the same equation sets and identical characteristic curves, behaved similarly in all test cases, for all PMs (performance measures), and over the given range of head gradients. This suggests that the codes are all solving the governing equations equally well, although it is still not guaranteed that this solution is equivalent to the correct analytical solution to the problem, should one exist. A measure of code divergence was introduced in the form of a standard deviation function; normalized ranges of the PMs are generally within a few percent, with a maximum of 17% for TH2_PM1.

Reaching this general consensus with a high level of agreement of the simulation results is a promising outcome, considering that these coupled and non-linear TH systems are difficult to solve. Despite the prevailing differences in the code approaches, meshing strategies, mesh sizes in the reference simulations, and apparent maturity of codes (e.g. some codes had already been extensively tested while other codes were limited to, at most, testing only specific parts of the system of equations), the results exhibited general agreement. The following guidelines are intended for future model developers who decide to employ the present test cases for model validation. 1) Numerical models should be tested against the 1D analytical solutions (e.g. T1, Lunardini solution and TH1, see Kurylyk et al., 2014), before attempting the TH2 and TH3 cases. 2) Implementation of the advective-conductive system without phase change (equivalent to Eulerian mass transport) could be tested in a step by step verification approach. In addition, the simplicity of the 1D systems allows simple convergence studies and identification of stability and accuracy parameters, including time and space discretization requirements. These 1D results will be beneficial to understand the more complex contexts. 3) When running TH2 and TH3, the averaged PMs should be tested first to allow quick error checks (e.g. TH2_PM3, TH3_PM3). 4) Convergence studies on the main PMs (spatial and temporal refinements) should be used to provide the final reference simulations, considering a subset of the head gradient values. An estimation of the accuracy of the results from the convergence tests will provide added value when finally compared with the 2D temperature fields at given times and when compared with the PMs in this study and with additional PMs published on the InterFrost web site in electronic format (wiki.lsce.ipsl.fr/interfrost).

Acknowledgements

References

Aguirre-Puente, J., Costard, F., Posado-Cano, R.: Contribution to the study of thermal erosion on Mars. J. Geophys. Res.-Planets. 99 N0. E3 pp. 5657-5667, 1994.

Amiri, E.A., J.R. Craig, B.L. Kurylyk.: A theoretical extension of the soil freezing curve paradigm. Adv. Wat. Resour. 111, 319-328. doi: 10.1016/j.advwatres.2017.11.021, 2018.

Anbergen H, Rühaak W, Frank J, Sass I.: Numerical simulation of a freeze–thaw testing procedure for borehole heat exchanger grouts. Can. Geotech. J. 2014;14 pp 1-14. doi:10.1139/cgj-2014-0177, 2014.

Anisimova, N.P., Nikitina, N.M., Piguzova, V.M., Shepelyev, V.V.: Water sources of central Yakutia. Guidebook of the II International Conference on Permafrost, 1973.

Atchley, A.L., S.L. Painter, D.R. Harp, E.T. Coon, C.J. Wilson, A.K. Liljedahl, V.E. Romanovsky: Using field observations to inform thermal hydrology models of permafrost dynamics with ATS (v0.83). Geosci. Model Dev. Discuss., Volume 8, 3235-3292. 2015.

Bear, J., and Bachmat, Y.: Introduction to modeling of transport phenomena in porous media. Kluwer Academic Publishers, 1991.

Bense V., Koi H., Ferguson, G., Read, T.: Permafrost degradation as a control on hydrogeological regime shifts in a warming climate. JGR, Earth Surface, Volume 117, Issue F3, September 2012

Bense, V.F., Ferguson, G., Kooi, H.: Evolution of shallow groundwater flow systems in areas of degrading permafrost. Geophys Res Lett 36:L22401, 2009.

Bernard-Michel, G., Le Potier, C., Beccantini, A., Gounand, S., Chraibi M.: The Andra Couplex 1 test case: comparisons between finite-element, mixed hybrid finite element and finite volume element discretizations. Comput Geosci 8:187–201, 2004.

Briggs, M.A., Walvoord, M.A., McKenzie, J.M., Voss, C.I., Day-Lewis F.D., Lane J.W.: New permafrost is forming around shrinking Arctic lakes, but will it last? Geophysical Research Letters, 41, 5, 1585-1592.

Carslaw, H.S., Jaeger, J.C.: Conduction of heat in solids. University Press, Oxford, 1959.

Clifford, S.M.: A model for the hydrologic and climatic behaviour of water on Mars. Journal of Geophysical Research: Planets (1991–2012). Volume 98, Issue E6, pages 10973–11016, 1993

Connon, R.F., Quinton, W.L., Craig, J.R., Hayashi M.: Changing hydrologic connectivity due to permafrost thaw in the lower Liard River valley, NWT, Canada. Hydrological Processes 28, 4163-4178, 2014.

Coon, E.T., Moulton J.D., and Painter S.L.: Managing Complexity in Simulations of Land Surface and Near-surface Processes. Env. Modell. Softw. Volume 78, 134-49. 2016a. doi:10.1016/j.envsoft.2015.12.017

Coon E.T., Sannel A.B.K, Pannetier R., Harp D., Frampton A., Painter S.L., Lyon S.W.: Thermal effects of groundwater flow through subarctic fens – a case study based on field observations and numerical modeling. Water Resour. Res. Volume 52, Issue 3, 1591-606. 2016b.

Costard, F., Gautier, E., Fedorov, A., Konstantinov, P., Dupeyrat, L.: An assessment of the erosional potential of fluvial thermal process during ice breakups of the Lena River (Siberia). Permafrost and Periglacial Processes, 25, 3, 162-171, 2014.

Dall'Amico, M., Endrizzi, S., Gruber, S., Rigon, R.: A robust and energy-conserving model of freezing variably-saturated soil. Cryosphere 5:469–484, 2011.

Diersch H-JG.: FEFLOW: Finite Element Modeling of Flow, Mass and Heat Transport in Porous and Fractured Media. 1st ed. Springer Heidelberg New York Dordrecht London, 2014. doi:10.1007/978-3-642-38739-5.

Duan, L., Man, X., Kurylyk, B. L., Cai T.: Increasing winter baseflow in response to permafrost thaw and precipitation regime shifts in northeastern China, Water, 9, 25 2017.

Dupeyrat, L., Costard, F., Randriamazaoro, R., Gailhardis, E., Gautier E., Fedorov A.: Effects of ice content on thermal erosion of permafrost - Implications for coastal and fluvial erosion, Permafrost and Periglacial Processes, 22, 2, 179-187, 2011.

Elberling, B.: Temperature and oxygen control on pyrite oxidation in frozen mine tailings. Cold. Reg. Sci. Technol., 41(2): 121-133, 2005.

Frampton A., Painter, S., Lyon S.W., Destouni G.: Non-isothermal, three-phase simulations of near-surface flows in a model permafrost system under seasonal variability and climate change. Journal of Hydrology 01/2011; 403:352–359, 2011.

Frampton, A., Painter, S.L., Destouni, G., 2013. Permafrost degradation and subsurface-flow changes caused by surface warming trends. Hydrogeology Journal, Hydrogeology of Cold Regions 21, 271–280. doi:10.1007/s10040-012-0938-z

Frampton, A., Destouni, G., 2015. Impact of degrading permafrost on subsurface solute transport pathways and travel times. Water Resources Research 51, 7680–7701. doi:10.1002/2014WR016689

Frederick, J. M. and Buffett, B. A.: Taliks in relict submarine permafrost and methane hydrate deposits: Pathways for gas escape under present and future conditions. Journal of Geophysical Research: Earth Surface, Volume 119, Issue 2, pages 106–122, 2014.

Gagarine, L.A.: The peculiarities of the mechanism of suffusion processes in the permafrost zone (the case of Central Yakutia). Proceedings: Tenth International Conference on Permafrost (TICOP), Salekhard, 2012.

Ge S., McKenzie J.M., Voss C.I., Wu Q.: Exchange of groundwater and surface-water mediated by permafrost response to seasonal and long term air temperature variation. Geophysical Research Letters, Volume 38, Issue L14402, 2011

Grenier, C., Roux, N., Fedorov, A., Konstantinov, P., Séjourné, A., Costard, F.: Ground thermal impact of a small alas-valley river in Syrdakh (Central Yakutia) in a continuous permafrost area - comparative study between monitoring and 1D numerical analysis. Accepted PPP.

Grenier, C., Régnier, D., Mouche, E., Benabderrahmane, H., Costard, F., Davy, P.: Impact of permafrost development on underground flow patterns: a numerical study considering freezing cycles on a two dimensional vertical cut through a generic river-plain system. Hydrogeology Journal, Volume 21, Issue 1, 257-270, 2013.

Grenier, C., Bernard-Michel, G., Benabderrahmane, H.: Evaluation of retention properties of a semi-synthetic fractured block from modelling at performance assessment time scales (Äspö Hard Rock Laboratory, Sweden). Hydrogeology Journal 17: 1051–1066, 2009.

Grimm, R. and Painter, S.: On the secular evolution of groundwater on Mars. Geophysical Research Letters, Volume 36, Issue 24, 2009.

Gouttevin, I., Krinner, G., Ciais, P., Polcher, J., Legout C.: Multi-scale validation of a new soil freezing scheme for a land-surface model with physically-based hydrology. The Cryosphere, 6, 407–430, 2012.

Gustafson, G., Gylling, B., Selroos, J.O.: The Äspö Task Force on groundwater flow and transport of solutes: bridging the gap between site characterization and performance assessment for radioactive waste disposal in fractured rocks. Hydrogeol J 17: 1031, 2009.

Hansson K, Simunek J, Mizoguchi M, Lundin L-C, Genuchten MT. Water flow and heat transport in frozen soil. Nume Sol Freeze-Thaw Appl Vadose Zone J 2004;3:693–704

Hinzman, L.D. et al.: Evidence and implications of recent climate change in northern Alaska and other Arctic regions. Climatic Change 72, 251-298, 2005.

Holmen, J., Benabderrahmane, H., Buoro, A., Brulhet, J.: Modelling of permafrost freezing and melting and the impact of a climatic cycle on groundwater flow at the Meuse/Haute-Marne site. Phys. Chem. Earth 36, 1531-1538, 2011, doi:10.1016/j.pce.2011.10.021.

Hwang CT, Murray DW, Brooker EW. A thermal analysis for structures on permafrost. Canad Geotech J 1972;9:33–46

IPCC Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker et al.]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

Ippisch O. Coupled transport in natural porous media. Combined faculties for the natural sciences and for mathematics. Heidelberg, Germany: Rupertus Carola University; 2001. p. 123.

Karra, S., Painter, S.L., Lichtner, P.C., 2014. Three-phase numerical model for subsurface hydrology in permafrost-affected regions (PFLOTRAN-ICE v1.0). The Cryosphere 8, 1935–1950. doi:10.5194/tc-8-1935-2014

Kollet S., Sulis M., Maxwell R., Paniconi C., Putti M., Bertoldi G., Coon E. T., Cordano E., Endrizzi S., Kikinzon E., Mouche E., Mügler C., Park Y.-J., Stisen S., Sudicky E.: The integrated hydrologic model intercomparison project, IH-MIP2: A second set of benchmark results to diagnose integrated hydrology and feedbacks. Water Resources Research 53, 867-890, 2017.

Kurylyk, B.L., M. Hayashi, W.L. Quinton, J.M. McKenzie, C.I. Voss: Influence of vertical and lateral heat transfer on permafrost thaw, peatland landscape transition, and groundwater flow, Water Resources Research, Volume 52, Issue 2, 1286-1305, 2016.

Kurylyk, B.L., MacQuarrie, K.T.B., McKenzie, J.M.: Climate change impacts on groundwater and soil temperature in cold and temperate regions: Implications, mathematical theory, and emerging simulation tools. Earth-Science Reviews 138: 313-334, 2014a.

Kurylyk, B.L., McKenzie, J.M, MacQuarrie, K.T.B., Voss, C.I.: Analytical solutions for benchmarking cold regions subsurface water flow and energy transport models: One-dimensional soil thaw with conduction and advection, Advances in Water Resources 70, 172–184, 2014b.

Kurylyk B.L., MacQuarrie K.T.B., Voss C.I.: Climate change impacts on the temperature and magnitude of groundwater discharge from shallow, unconfined aquifers. Water Resources Research, Volume 50, Issue 4, 3253-3274, 2014c.

Kurylyk, B. L. & Watanabe, K.: The mathematical representation of freezing and thawing processes in variably-saturated, non-deformable soils. Advances in Water Resources 60 160–177, 2013.

Larsson, A. (1992), The International Projects INTRACOIN, HYDROCOIN and INTRAVAL, Advances in Water Resources 15, 85-87.

Lee, T-C. 1998. Applied Mathematics in Hydrogeology, CRC Press, Boca Raton, Florida.

Leismann, H.M., and E.O. Frind, A symmetric-matrix time integration scheme for the efficient solution of advection-dispersion problems, Water Resour. Res., 25(6), 1133-1139, 1989.

Lemieux J-M., Fortier, R., Talbot-Poulin, M-C., Molson, J., Therrien, R., Ouellet, M., Banville, D., Cochand, M., Murray, R., Groundwater occurrence in cold environments: Examples from Nunavik, Canada. Hydrogeology Journal, http://dx.doi.org/10.1007/s10040-016-1411-1, 2016.

Lunardini, V.J.: Freezing of soil with phase change occurring over a finite temperature difference. Proceedings of the 4th International Offshore Mechanics and Arctic Engineering Symposium. ASM, 1985.

McBean, G., G. Alekseev, D. Chen, E. Forland, J. Fyfe, P.Y. Groisman, R. King, H. Melling, R. Vose, and P.H. Whitfield,. 2005. 'Arctic climate: past and present', in C. Symon, L. Arris, and B. Heal (eds.) Arctic Climate Impact Assessment, Cambridge, Cambridge University Press.

McKenzie, J.M., Voss, C.I.: Permafrost thaw in a nested groundwater-flow system. Hydrogeology Journal, DOI: 10.1007/s10040-012-0942-3, 2013.

McKenzie, J.M., Voss, C.I., Siegel, D.I.: Groundwater flow with energy transport and water–ice phase change: numerical simulations, benchmarks, and application to freezing in peat bogs. Advances in Water Resources 30(4):966–983, 2007.

McKenzie, J.M., Siegel, D.I., Rosenberry, D., Glaser, P., Voss, C.: Heat transport in the Red Lake Bog, Glacial Lake Agassiz Peatlands, *Hydrological Processes*, DOI: 10.1002/hyp.6239, 2006.

Mottaghy D, Rath V.: Latent heat effects in subsurface heat transport modelling and their impact on palaeotemperature reconstructions, Geophysical Journal, Volume 164, Issue 1, pp. 236-245, 2006

Nagare, R. M., Bhattacharya, P., Khanna, J., and Schincariol, R. A.: Coupled cellular automata for frozen soil processes, SOIL, 1, 103-116, doi:10.5194/soil-1-103-2015, 2015.

Orgogozo L. 2015. RichardsFOAM2: a new version of RichardsFOAM devoted to the modelling of the vadose zone. Computer Physics Communications 196: 619-620. DOI: 10.1016/j.cpc.2015.07.009.

Orgogozo L., Renon N., Soulaine C., Hénon F., Tomer S.K., Labat D., Pokrovsky O.S., Sekhar M., Ababou R., Quintard M. 2014. An open source massively parallel solver for Richards equation: Mechanistic modelling of water fluxes at the watershed scale. Computer Physics Communications 185: 3358-3371. DOI: 10.1016/j.cpc.2014.08.004

Painter, S.L., Coon E.T., Atchley A.L., Berndt M., Garimella R., Moulton J.D., Svyatsky D.: Integrated surface/subsurface permafrost thermal hydrology: Model formulation and proof-of-concept simulations. Water Resour. Res. Volume 52, Issue 8, 6062-77, 2016.

Painter, S. and Karra, S.: Constitutive Model for Unfrozen Water Content in Subfreezing Unsaturated Soils. Vadose Zone Journal, Vol. 13 No. 4, 2014.

Painter, S., Moulton, J., Wilson, C.: Modelling challenges for predicting hydrologic response to degrading permafrost. Hydrogeology Journal 21: 221–224, 2013.

Painter, S.: Three-phase numerical model of water migration in partially frozen geological media: model formulation, validation, and applications. Computational Geosciences, 15, 1, 69-85, 2011.

Pepin, N. et al. [Mountain Research Initiative EDW Working Group]: Elevation-dependent warming in mountain regions of the world. Nature Climate Change 5, 424-430, 2015.

Pimentel, E., Papakonstantinou, S., Anagnostou, G.: Numerical interpretation of temperature distributions from three ground freezing applications in urban tunnelling. Tunnelling and Underground Space Technology. Volume: 28, Pages: 57-69, 2012

Randriamazaoro, R., Dupeyrat, L., Costard, F., Carey- Gailhardis, E.: Fluvial thermal erosion: heat balance integral method. Earth Surface Processes and Landforms, 32 (12): 1828-1840, 2007.

Régnier, D.: Modélisation physique et numérique de la dynamique d'un pergélisol au cours d'un cycle climatique. Implications pour le site Meuse / Haute-Marne. Mémoire de thèse, Université Rennes I, 2012.

Rennermalm, A.K., Wood, E.F., Troy, T.J.: Observed changes in pan-arctic cold-season minimum monthly river discharge, Climate Dynamics. 35, 923-939, 2010.

Rivière, A., Gonçalvès, J., Jost, A., Font, M.: Experimental and numerical assessment of transient stream-aquifer exchange during disconnection. Journal of Hydrology, 517, 574–583, 2014. DOI: 10.1016/j.jhydrol.2014.05.040.

Rivière, A., Jost, A., Gonçalvès, J., Font, M.: How confined is the subpermafrost aquifer? Experimental and numerical investigation. Cold Regions Science and Technology, submitted.

Romanovsky V.E., Osterkamp, T.E.: Effects of unfrozen water on heat and mass transport processes in the active layer and permafrost. Permafrost Periglac. Process. 11: 219-239 (2000)

Roux, N., Costard, F., Grenier, C.: Laboratory and numerical simulation of river talik evolution. Permafrost and Periglac. Process., 2017, doi: 10.1002/ppp.1929

Rowland, J.C., Travis, B.J., Wilson, C.J.: The role of advective heat transport in talik development beneath lakes and ponds in discontinuous permafrost. Geophys Res Lett 38:L17504, 2011.

Rowland, J.C., Jones, C.E., Altmann, G., Bryan, R., Crosby, B.T., Geernaert, G.L., Hinzman, L.D., Kane, D.L., Lawrence, D.M., Mancino, A., Marsh, P., McNamara, J.P., Romanovsky, V.E., Toniolo, H., Travis, B.J., Trochim, E., Wilson, C.J.: Arctic Landscapes in Transition: Responses to Thawing Permafrost. Eos, Vol. 91, No. 26, 229-230, 2010.

Rühaak W, Anbergen H, Grenier C, McKenzie J, Kurylyk BL, Molson J, Roux N, Sass I. Benchmarking Numerical Freeze/Thaw Models. Energy Procedia 76 (2015) 301 – 310

Rühaak W., Bense V., Sass I.: 3D hydro-mechanically coupled groundwater flow modelling of Pleistocene glaciation effects. Computers & Geosciences. Vol. 67, June 2014, Pages 89-99.

Schädel C., et al.: Potential carbon emissions dominated by carbon dioxide from thawed permafrost soils. Nature Climate Change 6, 950–953 (2016).

Scheidegger, J., Busby, J., Jackson, C., McEvoy, F., Shaw, R.: Coupled modelling of permafrost and groundwater: A case study approach. British Geological Survey Commissioned Report, CR/16/053, under review.

Schuur, E. A. G., McGuire, A. D., Schadel, C., Grosse, G., Harden, J. W., Hayes, D. J., Hugelius, G., Koven, C. D., Kuhry, P., Lawrence, D. M., Natali, S. M., Olefeldt, D., Romanovsky, V. E., Schaefer, K., Turetsky, M. R., Treat, C. C., Vonk, J. E.: Climate change and the permafrost carbon feedback. Nature, 520, 171–179, doi:10.1038/nature14338

Shojae-Ghias, M., Therrien, R., Molson J., and Lemieux, J-M.: Controls on permafrost thaw in a coupled groundwater flow and heat transport system: Iqaluit Airport, Nunavut, Canada, Hydrogeology Journal, http://dx.doi.org/10.1007/s10040-016-1515-7, 2016.

Serreze, M.C., Walsh, J.E., Chapin, F.S., Osterkamp, T., Dyurgerov, M., Romanovsky, V., Oechel, W.C., Morison, J., Zhang, T., Barry R.G.: Observational evidence of recent change in the northern high-latitude environment, Climatic Change, 46, 159-207, 2000.

Stefan, J.: Über die Theorie der Eisbildung, insbesondere über die Eisbildung im Polarmeere, Sitzungs-Berichte der Österreichischen Akademie der Wissenschaften Mathematisch-Naturwissenschaftliche Klasse, Abteilung 2, Mathematik, Astronomie, Physik, Meteorologie und Technik, 98, 965–983, 1889.

St. Jacques, J-M., Sauchyn, D.J.: Increasing winter baseflow and mean annual streamflow from possible permafrost thawing in the Northwest Territories, Canada, Geophysical Research Letters, 26, L04104, 2009.

Svensson, U., Ferry, M., Kuylenstierna. H.-O.: DarcyTools, Version 3.4. Concepts, Methods and Equation. SKB R-10-70, Svensk Kärnbränslehantering AB. 2010.

Svensson, U., Ferry, M.: DarcyTools: A Computer Code for Hydrogeological Analysis of Nuclear Waste Repositories in Fractured Rock. Journal of Applied Mathematics and Physics, 2, 365-383, 2014.

Vidstrand, P., Follin, S., Selroos, J.-O., Näslund, J.-O., Rhén, I.: Modelling of groundwater flow at depth in crystalline rock beneath a moving ice-sheet margin, exemplified by the Fennoscandian Shield, Sweden. Hydrogeology Journal 21: 239–255, 2013.

Vonk, J., Gustafsson, Ö.: Permafrost-carbon complexities. Nature Geoscience, 6, 2013.

Voss, C.I., Provost, A.M.: SUTRA, A model for saturated-unsaturated variable-density ground-water flow with solute or energy transport, U.S. Geological Survey Water-Resources Investigations Report 02-4231, 291 p, 2002.

Voss, C.I., Provost, A.M., McKenzie J.M., Kurylyk B.L.: SUTRA, A model for saturated-unsaturated, variable-density groundwater flow with solute or energy transport – Documentation of the freeze-thaw capability, saturation and relative-permeability relations, spatially varying properties, and enhanced budget and velocity output, U.S. Geological Survey Techniques and Methods Report, 150 p, in review.

Walvoord, M.A., Kurylyk, B.L.: Hydrologic impacts of thawing permafrost – A review. Vadose Zone Journal, doi: 10.2136/vzj2016.01.0010, 2016.

Walvoord, M.A., Striegl, R.G.: Increased groundwater to stream discharge from permafrost thawing in the Yukon River basin: Potential impacts on lateral export of carbon and nitrogen. Geophysical Research Letters, 34, L12402, 2007.

Weill, S., Mouche, E., Patin, J.: A generalized Richards equation for surface/subsurface flow modelling. Journal of Hydrology 366, 9-20, 2009.

Wellman, T.P., Voss C.I., Walvoord M.A.: Impacts of climate, lake size, and supra- and sub-permafrost groundwater flow on lake-talik evolution, Yukon Flats, Alaska (USA). Hydrogeology Journal 21: 281–298, 2013.

Zhou, M. M., Meschke, G.: A three-phase thermo-hydro-mechanical finite element model for freezing soils. International Journal for Numerical and Analytical Methods in Geomechanics. Volume: 37 Issue: 18 Pages: 3173-3193, 2013.

Tables

Physical properties	Parameter values and expressions
Porosity, ε	0.37
Porous medium compressibility, β (m s² kg^{-1})	10^{-8}
Gravitational acceleration, g (m s^{-2})	9.81
Thermal conductivity of water, λ_w (W m^{-1} K^{-1})	0.6
Thermal conductivity of ice, λ_i (W m^{-1} K^{-1})	2.14
Thermal conductivity of solid matrix, λ_s (W m^{-1} K^{-1})	9.
Total or bulk thermal conductivity, λ_T (W m^2)	$\lambda_T = \varepsilon S_w \lambda_w + \varepsilon(1-S_w)\lambda_i + (1-\varepsilon)\lambda_s$
Specific heat of water, c_w (J kg^{-1} K^{-1})	4182
Specific heat of ice c_i (J kg^{-1} K^{-1})	2060
Specific heat of solid grains, c_s (J kg^{-1} K^{-1})	835
Water density, ρ_w (kg · m^{-3})	1000
Ice density, ρ_i (kg m^{-3})	920
Solid grain density, ρ_s (kg m^{-3})	2650
Dynamic viscosity of liquid water, μ (kg m^{-1} s^{-1})	1.793×10^{-3}
Total volumetric heat capacity $(\rho c)_T$ (J m^3 K^{-1})	$(\rho c)_T = \varepsilon S_w \rho_w c_w + \varepsilon(1-S_w)\rho_i c_i + (1-\varepsilon)\rho_s c_s$
Latent heat of fusion, L (J kg^{-1})	334,000
Water saturation curve, $S_w(T)$ For T > 273.15 K, $S_w(T) = 1$ For T < 273.15 K, $S_w(T)$, exponential fct of T Residual saturation, S_{Wres} W (K)	[McKenzie et al. 2007], exponential expression $Sw(T) = (1 - S_{Wres}) \exp[-((T - 273.15)/W)^2] + S_{Wres}$ 0.05 0.5
Hydraulic conductivity, K_w (m s^{-1})	$K_w = k_r \, k_{int} \, \rho_w \, g / \mu$
Relative permeability curve, $k_r(T)$ Intrinsic permeability, k_{int} (m^2) Ω	[McKenzie et al. 2007], impedance factor expression $k_r(T) = \max(10^{-6}, 10^{-\varepsilon\Omega(1-Sw(T))})$ 1.3×10^{-10} 50.

Table 1: Physical parameter values and expressions considered for Case TH2 "Frozen inclusion" and for Case TH3 "Talik Opening/Closure". The compressibility β includes water and matrix compressibility.

Simulation domain, longitudinal extent (m), L_x	3.
Simulation domain, lateral extent (m), L_y	1.
Position of inclusion centre	
L_{CX} (m)	1.
L_{CY} (m)	0.5
Size of square shaped inclusion L_{Sq} (m)	0.333
Temperatures	
T_{in} (°C)	+5.
$T^+_{initial}$ (°C)	+5.
$T^-_{initial}$ (°C)	-5.
Applied head gradients	0.
$\Delta H/L_X$	0.03
	0.09
	0.15

1   Table 2: Case TH2 "Frozen Inclusion", geometric parameter values, initial and imposed

2   boundary temperatures, and head gradients

3

Simulation domain, longitudinal extent (m), L_x	1.
Simulation domain, lateral extent (m), L_y	1.
Position of lower circle centre (symmetric for upper circle)	
L_{CX} (m)	0.5
L_{CY} (m)	0.1
Radius, R (m)	0.5099
Temperatures	
T_{in} (°C)	+5.
$T^+_{initial}$ (°C)	+5.
$T^-_{initial}$ (°C)	-5.
$T^-_{imposed}$ (°C)	-5.

4   Table 3: Case TH3 "Talik Opening / Closure", geometric parameter values, initial and imposed

5   boundary temperatures

6
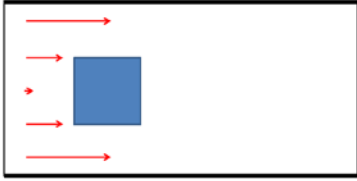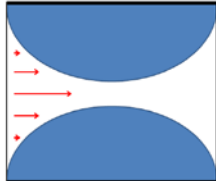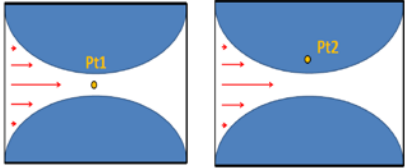
TH2	TH3
TH2	**TH3**
Head gradients (%)	Head gradients (%)
0, 3, 9, 15	3, 6, 9, 15
TH2_PM1	**TH3_PM1**
Minimum domain temperature	Equivalent hydraulic conductivity (m/s) calculated as the total water flow through the system divided by the imposed hydraulic gradient $$K_{eq} = \frac{Q_{Hydro}}{\Delta H / L_X}$$
Thresholds	**Thresholds**
Times for minimum (Temperature) = 0°C (optionally time to reach steady state)	Time to reach $K_{eq} = 0$ (essentially 0.1% of initial hydraulic conductivity, all for lower head gradient values)
TH2_PM2 Net total heat flux: J$_{net}$ = J$_{Out}$ - J$_{In}$ $$J_{net} = \int_{\Sigma_{Out}} \left(\rho_w c_w T U_x - \lambda_T \frac{dT}{dx} \right) dy - \\ - \int_{\Sigma_{In}} \left(\rho_w c_w T U_x - \lambda_T \frac{dT}{dx} \right) dy$$	**TH3_PM2** Conductive heat fluxes at the horizontal boundaries (upper & lower boundary)
TH2_PM3 Total domain liquid water volume	**TH3_PM3** Total sensible heat $$\iint (\rho_w c_w \epsilon S_w + \rho_i c_i \epsilon (1 - S_w) + \rho_S c_S (1 - \epsilon)) T \, dx dy$$
	TH3_PM4_Pt1&2 Temperature evolution at point Pt1 (domain center, initially unfrozen) and Pt2

Table 4: Overview of the TH2 and TH3 test cases and performance measures.

1

Code	Numerical scheme	Non-linearities	TH coupling	Linear solvers	Pre-conditioner	Time step strategy	Automatic mesh refinement	Parallel computing
Cast3M	FV	Picard	Picard & Under-relaxation	BiCGSTAB	ILUO	Prescribed	No	No
PermaFOAM	FV	Picard	Sequential operator splitting	PCG, BiCG	DIC, DILU	Adaptive	No	Mesh Partitioning, ~100-1000 cores
COMSOL	FE	Damped Newton	Damped Newton	MUMPS	ILU	Adaptive	No	shared-memory: 8 cores
DarcyTools	FV	Picard	Picard	GMRES	Multi-grid, ILUO	Prescribed	Yes	PM, 64 cores
MELT	FV	Picard, IMPES	Sequential	GMRES, CG	ASM, ILU	Adaptive	No	No
SMOKER	FE	Picard	Picard & relaxation	PCG (+LF)	ILU factorization	Prescribed	No	No
ATS	MFD	Non-linear Krylov acceleration	Simultaneous inversion	GMRES, CG	Boomer AMG	Adaptive	No	PM, 4 cores
SUTRA	Hybrid FE & IFD	Modified Iterative Method	Modified Iterative Method	CG, GMRES, Orthomin, Direct	ILU, Cholesky	Prescribed	No	No
PFLOTRAN-ICE	FV	Newton	Simultaneous inversion	GMRES	ILU	Adaptive	No	PM, thousands of cores
FEFLOW	FE	Newton	Newton	BICGstab, PARDISO (direct solver)	incomplete factorisation	Adaptive	Yes	Open MP (here 16 cores)
GEOAN	FD	Picard	Picard	Iterative Crank-Nicolson	Non standard	Prescribed Adaptive Mixed	No	Up to 48 cores
FlexPDE	FE	Newton	Newton	CG, GMRES	ILU, Cholesky	Adaptive	Yes	Up to 8 cores
Ginette	FV	Picard	Picard	PCG	BuGC	Adaptive	No	No

1   FV: Finite Volume; MHFE: Mixed Hybrid Finite Element; FE: Finite Elements; FD: Finite Differences, MFD: Mimetic Finite Differences;
2   DD: Domain Decomposition; IFD: Integrated Finite Difference; (Picard = sequential = iterative); PCG: Preconditioned conjugate-gradient;
3   LF = Leismann & Frind (1989) scheme for symmetric transport matrix, MUMPS: Multifrontal massively parallel sparse direct solver, ILU:
4   Incomplete Lower-Upper.
5
6   Table 5: Basic information on all codes participating in the InterFrost benchmark.

1
2
3

Code	Mesh size, number of nodes	Mesh size, Number of elements	Element type	Domain modelled
Cast3M	31,609	31,609	Quadrilateral	Half
PermaFoam	963,202	480,000	Hexahedra (3D)	Full
COMSOL	35,000	35,000	Quadrilateral	Full
DarcyTools	14,810	14,786	Cartesian Hexa (3D)	Half
MELT	30,000	30,000	Squares	Full
SMOKER	60,802	30,000	Hexahedra (3D)	Full
ATS	1,728	1,728	Hexahedra (3D)	Full
SUTRA	20,301	20,000	Quadrilateral	Half
PFLOTRAN-ICE	30,000	30,000	Hexahedra (3D)	Full
FEFLOW	55,213	105,144	Triangles	Full
GEOAN	32,130	32,130	3D	Full
FlexPDE	4,882	2,345	Triangles	Full
Ginette	21,600	21,600	Quadrilateral	Half

4   Table 6: TH2 simulation conditions

5

1

Code	Mesh size, number of nodes	Mesh size, Number of elements	Element type	Domain modelled
Cast3M	8,848	17,272	Triangles	Half
PermaFoam	1,283,202	640,000	Hexahedra (3D)	Full
COMSOL	40,000	40,000	Quadrilateral	Full
DarcyTools	18,076	16,050	Cartesian Hexa (3D)	Half
MELT	10,000	10,000	Squares	Full
SMOKER	20,402	10,000	Hexahedra (3D)	Full
ATS	4,225	4,225	Hexahedra (3D)	Full
SUTRA	45,451	45,000	Quadrilateral	Half
PFLOTRAN-ICE	6,720	6,720	Prisms (3D)	Full
FEFLOW	32,197	63,720	Triangles	Full
GEOAN	40,804	40,804	3D	Full
FlexPDE	90,586	45,127	Triangles	Half
Ginette	5,000	5,000	Quadrilateral	Half

2   Table 7: TH3 simulation conditions

3
4

Figure captions

Figure 1: Case TH2 conditions ("Frozen Inclusion"): (a) geometrical features, (b) hydrological boundary conditions, (c) thermal initial and boundary conditions. The associated parameter values and the applied head gradients are provided in Table 1 & 2.

Figure 2: Surface and contour plots at time 22,860 s (6.4 hours) of (a) the head field (m), and (b) the temperature field (°C).

Figure 3: Evolution of Case TH2 showing simulated temperature profiles along a horizontal line (left to right) through the centre of the system: (a) without advection (pure conduction, upper plot, blue curve for initial time, dotted line, time 1260 s, red line final simulation time 5.9 d), (b) with advection (head gradient of 9%, lower plot, blue line for initial time, dotted line for 930 s, red line for final simulation time 16.6 h). Black lines provide the profile evolution for additional times.

Figure 4: Case TH3 conditions ("Talik closure / opening") (a) geometry, (b) hydrological boundary conditions, (c) thermal initial and boundary conditions. All parameters are provided in Table 2 & 3.

Figure 5: Case TH3: Contour line and contour colour plots at time 19,860 s (5.5 hours) of (a) the head field (m), and (b) the temperature field (°C).

Figure 6: Case TH3: Evolution of simulated temperatures along vertical profiles crossing the centre of the system: (a) without advection (blue curve for initial time, dotted line for 120 s, red line final simulation time 2.71 d) and (b) with a head gradient of 9% (blue line initial time, dotted line 765 s, red line final simulation time 2.07 d). Black lines provide the profile evolution for a constant time interval.

Figure 7: Overview of all PMs associated with the TH2 Case and for head gradients ranging from 0 to 20%. (a) TH2_PM1, minimum temperature, with 4 curve-evolution phases corresponding to the 2% head gradient conditions (blue curve), (b) Time to reach 0°C as a function of head gradient intensity, (c) TH2_PM2, net heat flux, (d) TH2_PM3, total liquid water content; all curves reach the plateau of 1.11 m^3 corresponding to the total system volume composed of liquid water.

Figure 8: Overview of all PMs associated with the TH3 Case and for head gradients ranging from 0 to 15%. (a) TH3_PM1, equivalent hydraulic conductivity, (b) talik closure time as a function of head gradient intensity, (c) TH3_PM3, total sensible heat, (d) TH3_PM2, total lateral conductive flux, (e) TH3_PM4_Pt1, temperature evolution at the centre of the simulation domain, (f) steady state 0°C isotherm profiles for selected head gradients above threshold.

Figure 9: (a) Evolution of the minimum of the temperature field (TH2_PM1) and (b) the total water volume in the domain (TH2_PM3).

Figure 10: (a) Evolution of the temperature at point Pt1 in the middle unfrozen zone of the simulated domain (TH3_PM4_Pt1) and (b) at point Pt2 in the initially frozen zone (TH3_PM4_Pt2). A head gradient of 3% is considered.

Figure 11: (a) Evolution of the equivalent hydraulic conductivity of the domain (TH3_PM1) and (b) the total heat of the domain (TH3_PM3).

Figure 12: Threshold values for TH2. (a) Time for minimum domain temperature to reach 0°C and (b) time to reach steady state. The horizontal axis is head gradient. The test was based on 0, 3, 9, and 15 % head gradients for TH2, and some participants provided extra simulations for intermediate head gradients.

Figure 13: (a) Threshold values for TH3, i.e. time for system closure (time when equivalent hydraulic conductivity reaches zero). The horizontal axis is head gradient. The test was based on 3, 6, 9, and 15 % hydraulic gradients for TH2, while some participants completed extra simulations for intermediate head gradients. (b) Evolution of hydraulic conductivity is plotted with time, showing the divergence of flow regimes between the different codes.

Fig. B.1: Case TH2, sensitivity of minimum temperature evolution to the freezing curve temperature range.

Figure C.1: Standard deviation evolution for two cases issued from TH2: (a) TH2_PM1, minimum temperature, and (b) TH3, TH3_MP4_Pt1, temperature evolution at point Pt2 for hydraulic gradients (GH) of 0, 3, 9 and 15%.

1 Figure C.2: (a) Standard deviation evolution for the total liquid water volume (TH2_PM3) and
2 (c) the total heat (TH3_PM3); (b & c) the corresponding initial conditions where indexes refer
3 to related codes (1: Cast3M, 2: PermaFoam, 3: COMSOL, 4: DarcyTools, 5: MELT, 6:
4 SMOKER, 7: ATS, 8: SUTRA, 9: PFLOTRAN-ICE, 10: FEFLOW, 11: GEOAN, 12: FlexPDE,
5 13: Ginette). Convergence test information is provided in the left column (a & c) in addition to
6 the standard deviation curves for all head gradients (0, 3, 9, 15 % for TH2; 3, 6, 9, 15 % for
7 TH3).

8 Figure C.3: Standard deviation evolution for TH2_PM1 (minimum temperature in solid lines)
9 with standard deviations issued from convergence studies from participants in dashed lines with
10 the colour associated to the specific head gradient.

11 Figure C.4: Integrals of standard deviation for the TH2 performance measures and head
12 gradients of 3, 9, and 15 %. To fit on the same plot, all curves are normalized by the maximum
13 value. For TH2_PM1 & 3, curves decrease from an initial maximum 3 % gradient case. For
14 TH2_PM2, the standard deviation is nearly constant regardless of the head gradient.
15

Appendix A – Participating code information

The Advanced Terrestrial Simulator (ATS, github.com/amanzi/ats): **ATS** is a ecosystem hydrology code focusing on solving problems of integrated surface/subsurface hydrology including vegetation, surface energy balance, snow, and other environmental interactions with hydrology (Painter et al., 2016). ATS was originally developed with permafrost applications in mind, and includes models of flow in partially frozen, variably saturated soils that have been extensively validated against laboratory experiments to capture both cryosuction and density variations between ice and water (Painter 2014). Several groups have used ATS to characterize cold region physics and field sites across the globe (Atchley et al., 2015, Sjoberg et al., 2016). ATS builds on the Arcos multiphysics framework (Coon et al., 2016) and leverages unstructured meshes, second-order, conservative mimetic finite difference discretizations, and an extensive assortment of libraries for linear and non-linear solvers.

Cast3M (www-cast3m.cea.fr/**)**: Cast3M is a multi-physics code dealing with various applications, initially developed with a finite element scheme for nuclear reactor applications. It consists of various elemental bricks called procedures that can be organized together for the resolution of more complex problems or equations. Treated domains are solid and structural mechanics, as well as fluid flow and heat transfer. Resolution of the governing equations for transport in porous media has been under development since the 90s. Cast3M now provides tools to resolve saturated flow, unsaturated flow (Richard's equation and multi-phase flow), Eulerian and Lagrangian transport by means of finite volume and mixed hybrid finite element schemes. The latter has proved accurate and efficient for nuclear waste storage applications (flow and transport) within an intercomparison exercise (Bernard-Michel et al., 2004). Several extensions have been developed for coupled physics, taking advantage of the modular properties of the code. For instance, the approach for coupling between surface and sub-surface transfer was part of a recent intercomparison exercise (Kollet et al., 2016). One may refer to Grenier et al. 2013, Roux et al. 2017 and Grenier et al. (accepted PPP) for applications in cryo-hydrology.

COMSOL Multiphysics (www.comsol.com/comsol-multiphysics) is a general-purpose software platform for numerical modelling of coupled and multiphysics problems. Here, the equation-based modelling for general second-order systems of non-linear partial differential

equations is used. The application for cryo-hydrology has been developed for paleo-permafrost simulations in Great Britain (Scheidegger et al. under review).

DarcyTools is a code for simulation of flow and transport in porous and/or fractured media. The code implements a finite volume approach on adapted Cartesian grids with equivalent continuous porous medium upscaling techniques. It is intended to be applicable to a wide range of groundwater flows although initially developed for the analysis of nuclear waste repositories. DarcyTools is strongly coupled for thermo-hydraulics, hydro-mechanics, hydro-chemistry and unsaturated flow. The coupling is achieved by modified Picard schemes and the sets of linear equations are solved by a parallel multigrid preconditioned block-GMRES solver. For more details one may refer to (Svensson & al. 2010; Svensson & Ferry 2014).

FEFLOW (Finite Element subsurface FLOW system) is a computer program for simulating groundwater flow, mass transfer and heat transfer in porous media and fractured media. The program uses finite element analysis to solve the groundwater flow equation under both saturated and unsaturated conditions as well as mass and heat transport, including fluid density effects and chemical reaction kinetics for multi-component reaction systems. Refer to (Diersch, 2014; Anbergen et al. 2014).

FlexPDE (www.pdesolutions.com) is a general purpose scripted finite element model builder for partial differential equations. Developments and recent applications include the field of coupled thermo-hydrological modelling (Bense et al., 2012; Bense et al., 2009) as well as geo-mechanical couplings (Rühaak et al., 2014).

The **GEOAN** computer code is a numerical model for calculation of groundwater head, flow (saturated/unsaturated), and transport (solute concentration and heat) in three dimensions, as well as surface water flow. The temperature calculations can include creation of ice and permafrost. A simulation may also include density effects, deformation of the medium and hydromechanical effects. The mathematical model is based on the continuum approach, the finite difference method and a block centered flow approach. GEOAN includes solvers for parallel processing and a model may include tens of millions of nodes/cells. Considering permafrost simulations the GEOAN computer code was used in a paper that deals with the impact of climate cycles and permafrost on future groundwater flow in the Paris Basin (Holmén et al., 2011).

Ginette: Ginette was initially developed at Metis (UPMC) to deal with interactions between streams and aquifers, as they fluctuate from a connected to a disconnected status. Numerical simulations of experimental laboratory results reproducing such conditions provided the opportunity to test the coupled 1D surface water - 2D variably saturated groundwater code (Rivière et al., 2014). Ginette was then extended to include coupled heat transfer and water flow in saturated porous media and is now jointly developed by Metis and MINES ParisTech (PSL Research University). The code was compared to experimental data acquired on a complex laboratory system to provide validation on the physical processes and mathematical formulations, in particular for the representation of density change between frozen and liquid water (Rivière et al., submitted). Real-world cryo-hydrogeological paleo-applications, which have been presented in conferences (e.g. Jost, 2011; Jost et al., 2014), were also proposed using Ginette, requiring some additional adaptation to the specific needs of basin-scale calculations.

MELT is a 2D multiphase flow and transport simulator developed for modelling freeze-thaw processes in saturated porous media. It was designed specifically for modelling the interactions of submarine permafrost, gas hydrate, and multi-phase submarine pore fluid flow, with specific attention to density-driven flow of fresh and saline pore fluids at Arctic coastlines (e.g. submarine groundwater discharge). It has been developed by Jennifer M. Frederick (now at Sandia National Laboratories, NM, USA) over the course of her thesis work at U.C. Berkeley, and post-doctoral studies at the Desert Research Institute, Reno, NV, USA. MELT uses the finite volume method to solve for flow and scalar transport based on two-phase Darcy's Law, conservation of mass (hydro), and conservation of energy (thermo). MELT is based on the IMPES algorithm with up-winding (i.e., implicit pressure, explicit saturation) for the pressure/flow solution, with scalar transport (i.e. heat and mass) solved sequentially. Refer to Frederick and Buffet (2014).

PermaFoam: PermaFoam is an OpenFOAM® solver dedicated to cryo-hydrogeology modelling. OpenFOAM® (openfoam.com and openfoam.org) is an open-source tool box for computational fluid dynamics that is broadly used in both industrial and academic applications. One of its main strengths is its capability to use efficient parallel computing techniques. The PermaFoam solver, which deals with coupled heat transfer and water flow in variably saturated and heterogeneous porous media, takes advantage of these parallel computational options to handle the numerical difficulties associated with the strong couplings and non-linearities encountered in cryo-hydrogeology modelling. A paper related to the application of PermaFoam

to the study of the permafrost dynamics in an experimental watershed of central Siberia is currently under review (Orgogozo et al., submitted). Regarding resolution of the water flow equation, since PermaFoam is based on the RichardsFoam2 solver, the reader may refer to the publications associated with this latter solver (Orgogozo et al., 2014, Orgogozo, 2015) for additional numerical details.

PFLOTRAN-ICE refers to the **PFLOTRAN** code (http://www.pflotran.org/) with the multiphase ice-water-vapor flow physics module enabled. PFLOTRAN is an open source, massively parallel subsurface flow and reactive transport code which can solve a system of generally non-linear partial differential equations describing multiphase, multicomponent and multiscale reactive flow and transport in porous materials. The code is designed to run on massively parallel computing architectures as well as workstations and laptops. Parallelization is achieved through domain decomposition using the PETSc (Portable Extensible Toolkit for Scientific Computation) libraries. Capabilities in addition to ice-water-vapor flow physics include modules for thermo-hydro-chemical interactions, supercritical CO_2, surface flow, sorption, precipitation and dissolution processes and reactive transport. Refer to Karra et al. (2014).

SMOKER is a finite element numerical model for solving complex density-dependent groundwater flow, contaminant transport, groundwater age, and thermal energy transport problems. The model can be used to solve one, two, or three-dimensional transport problems within a variety of hydrogeological systems, including discretely-fractured porous media. Originally developed as a research tool to study ATES (Aquifer Thermal Energy Storage) systems, the model can be used for virtually any application involving the storage or transport of thermal energy in the subsurface where temperatures remain < 100 °C. The SMOKER model and related CHeat model (Chemistry & Heat) have been tested and applied to a variety of hydrogeological systems, including applications to hydrothermal systems (e.g. "Black Smokers", Yang et al. 1996), heat storage systems (Molson et al. 1992), groundwater age (Molson & Frind, 2012), and permafrost degradation (Shojae-Ghias et al., 2016). For further information, see: http://www.science.uwaterloo.ca/~molson/

.

SUTRA is an established USGS groundwater flow and coupled transport model (Voss and Provost, 2002) that has been applied since the early 1980s to simulate density-driven groundwater flow problems including saltwater intrusion and thermal convection. More

1 recently, the code has been enhanced to allow for pore water phase change in the governing

2 equations (McKenzie et al., 2007; McKenzie and Voss, 2013) to simulate hydrogeologic

3 systems that experience ground freezing and thawing. Various iterations of this expanded code

4 have been applied to simulate groundwater flow and coupled energy transport in environments

5 with permafrost or seasonally freezing ground (e.g., McKenzie et al., 2006; 2007; Ge et al.,

6 2011; McKenzie and Voss, 2013; Wellman et al., 2013; Kurylyk et al., 2014; 2016; Briggs et

7 al., 2014). Recent advancements beyond the code described by McKenzie et al. (2007) include

8 freeze-thaw capabilities in the unsaturated zone, more complex density functions, and pre-

9 programmed functions for soil freezing, relative permeability, and desaturation.

10

Appendix B – Spatial and temporal sensitivity analyses, sensitivity to the freezing curve

Spatial and temporal sensitivity analyses

In a preliminary convergence test, sensitivity analyses to spatial discretization (while keeping the time step fixed as in the reference simulation) and to temporal discretization (while keeping the mesh size fixed as in the reference simulation) were conducted for both benchmark cases with the Cast3M code. PermaFoam (Orgogozo et al., 2016) was also used in a spatial convergence study (using an automatically calculated time step), covering a large range of elements from a few thousand to millions. In addition, a convergence study of the threshold parameters for the non-linearities and coupling convergence loops was completed. Such tests had been previously provided with the Lunardini case since its analytical solution provides a useful reference.

Here, convergence studies carried out with Cast3M were first completed with Case TH3 since it has a smaller domain allowing the use of refined grids at limited computational cost. A head gradient of 3% was imposed corresponding to realistic values. The mesh size for Case TH2, with a larger simulation domain, was directly adapted from the Case TH3 convergence study.

The variability issued from the convergence tests and for the full set of performance measures is provided in Table B.1 for Cases TH2 & TH3. Variability is expressed in percentage, as the absolute range taken by the performance measure divided by the median value. A large range of time and space increments were considered starting from relatively coarse conditions. Results show variable sensitivities. For the spatial convergence study, the larger range in variability corresponds to the Case TH3 performance measure of conductive fluxes through the upper and lower boundaries (29.2%). This result highlights the need for accurately simulating the interplay between advective heat flux (which requires a sufficiently refined longitudinal discretization) and conductive heat flux from the imposed boundary conditions (which requires a sufficiently refined transverse discretization to accurately capture the local thermal gradient at the boundaries). The time for system closure is less impacted by the discretization (0.7 %). For the time-step convergence study, the greatest variability (7.2 %) is obtained for the first performance measure (time for total thaw). This punctual measure is clearly more sensitive to

1 the discretization strategy than the other performance measures which correspond to integrated

2 averages (of flux or total water volume) over the entire domain.

3 These results present important consequences in terms of simulation intercomparisons. First,

4 these ranges provide a quantitative estimation of the maximum uncertainty associated with any

5 simulation. For practical reasons, due to the high computational costs associated with

6 "converged" simulations, the reference simulations provided during the course of the

7 benchmark exercise will probably not be "fully converged". This evaluation of the associated

8 ranges, however, can provide insight to explain the discrepancies between the results of the

9 various codes running the same cases. Further, some performance measures that would appear

10 too sensitive to the spatial and temporal discretization should probably be discarded because

11 they would not allow sufficient discrimination between simulation results.

Case TH2 – "Frozen Inclusion"	Time for total thaw	Total heat flux after 9.10^4 s	Total water volume after 4.10^4 s
Δx sensitivity range	1.4 %	1.2 %	0.015 %
Δt sensitivity range	7.2 %	0.6 %	0.05 %
Case TH3 – "Talik Opening/Closure"	Closure time	Upper flux after 3.10^4 s	Total heat after 3.10^4 s
Δx sensitivity range	0.7 %	29.2 %	0.03 %
Δt sensitivity range	3.4 %	0.6 %	0.2 %
Case TH3 – Regular mesh	Closure time	Upper flux after 3.10^4 s	Total heat after 3.10^4 s
Δx sensitivity range	7.4 %	63.5 %	6.9 %

12 Table B.1: Variability of the performance measures for the full range of time steps and mesh

13 sizes considered in the sensitivity study (Case TH2, "Frozen Inclusion"): for time convergence,

14 time steps are 480, 240, 120, 60, 30, and 15 s. For spatial convergence, discretization ranges

15 from 1.96×10^{-2} m to 3.98×10^{-3} m.

16

17 The reference case is run with a time step of 60 s and a mesh size of 6.94×10^{-3} m for a head

18 gradient of 3%. For Case TH3, "Talik Opening/Closure", the time steps considered are 1500,

1 600, 300, 120, 60, 30, 12, 6, and 2.4 s. Lateral spatial discretization ranges from 1.96×10^{-2} m to

2 2.49×10^{-3} m (with the circular interface explicitly reproduced using triangular element

3 automatic meshing). The reference simulations are completed with an average mesh size of

4 5.4×10^{-3} m. For Case TH3, a similar spatial convergence study was conducted, while

5 considering a regular mesh onto which the geometry of the initial condition is projected.

6 Another practical point related to the influence of the meshing strategy is now briefly discussed.

7 These simulations were first made on a regular mesh as in the previous test case (Case TH2,

8 "Frozen Inclusion"). The initial conditions were directly projected onto the regular mesh

9 without any special effort to match, for instance, the total initial heat requirements. Results

10 showed that a strict spatial convergence is more computationally demanding with such a

11 meshing strategy due to the variability of the actual simulated initial condition. A dedicated

12 discretization strategy was therefore preferred. Moreover, the total range of variability as

13 previously studied notably increases: results are compared in Table B.1. For the codes taking

14 part in a benchmark for Case TH3 which had no dedicated meshing strategy, the simulations

15 could likely be improved by imposing greater control on the total initial heat and initial shape.

16

17 Following the convergence studies, the parameters associated with the Cast3M reference

18 simulations presented below were chosen as a compromise between "full" convergence and

19 computational efficiency. This allowed running accurate sensitivity cases (presented later)

20 while maintaining manageable computational effort. The associated parameters are provided in

21 Table B.2.

	Nb Elements	Δx	Δt	Nb time steps	Average Nb of iterations
Case TH2	31,014	6.94 mm	7.5 s – 60 s	3500 – 15000	20
Case TH3	17,272	(2.7 – 7.6) mm	7.5 s – 60 s	3500 – 20000	20

22 Table B.2: Parameters associated with the Case TH2 & Case TH3 reference simulations:

23 number of elements, size of the square elements (Δx), range of time step (Δt) varied depending

24 on head gradients, and number of iterations within a time step in the initial phase change period

Sensitivity to the saturation curve parameters

All test cases consider a saturated porous medium with an exponentially shaped freezing curve varying over approximately 1°C between the start of freezing at 0°C and complete freezing by about -1°C (leaving a residual saturation of liquid water) (Table 2). For real world applications, a large range of other freezing functions could also be considered (e.g. linear, exponential, power law) as well as smaller or larger freezing ranges. A recent discussion on this issue is provided by Kurylyk and Watanabe (2013). The 1°C freezing range chosen for the benchmark was intended as a good compromise between very steep freezing fronts, which are numerically hard to handle (with large changes in unfrozen water content over small temperature ranges), and smooth fronts with smaller changes in unfrozen water content over larger temperature ranges.

The influence of the freezing curve for the simulations appears very important from a sensitivity analysis conducted on Case TH2. In these cases, the W parameter in Table 2 controlling the slope of the freezing curve was varied, resulting in freezing ranges extending over temperature intervals between 0.1°C and 4°C (B_t value in Figure B.1, the reference case is in bold and blue color: 1°C), while keeping the residual water saturation (S_{Wres}) fixed at 0.05. The corresponding sensitivity of TH2_PM1 for a head gradient of 3% is presented in Fig. B.1. Results show a dramatic influence of the freezing range: phase change starts earlier for a large freezing range which also modifies the times to reach the 0°C threshold. The relative threshold time variability is 17.3 %. The impact on TH3_PM1 was also significant (but not illustrated here), leading to a variability of the relative threshold times for talik closure of 103.5%.

The freezing range thus has a major control on the temperature evolution curves and any departure from the benchmark specifications clearly induces large discrepancies.

Appendix C – Quantitative assessment of PM results, sources of discrepancies, sensitivity to head gradients

Reassessment of PM results with quantitative measurements: standard deviation evolution

We introduce here a measure of the spread of the curves to obtain a more rigorous quantitative assessment of the PM dispersion. The standard deviations of all curves, or equivalently the square root of the L2 norm, based on the average of all curves, are computed to quantify the divergence in the results. As mentioned earlier, since no reference simulation can be proposed, the results among the participants themselves were employed to compute the mean. The sample mean and sample standard deviation are expressed below (Eq. C.1 & C.2) where each PM time evolution curve is the $S_i(t)$ signal, with i varying from 1 to $N = 13$.

$$Mean(t) = S(t) = \frac{1}{N}\sum_{i=1}^{N} S_i(t) \qquad (C.1)$$

$$Standard\ Deviation\ (t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(S_i(t) - S(t))^2} \qquad (C.2)$$

These statistical parameters are useful to (i) visualize and discuss the time evolution of major discrepancies throughout the various PMs, (ii) obtain a quantitative basis to understand the impact of spatial and temporal convergence on the results and their effect on the general spread of the curves, and (iii) discuss the influence of the advective flux on the different model results.

Figure C.1 provides the evolution of the discrete (i.e. not integrated) PMs for TH2 and TH3 (TH2_PM1 and TH3_PM4_Pt1). For example, the green curve in Fig. C.1a (TH2_PM1) displays the changes in standard deviation with time for a hydraulic gradient (GH) of 3% (refer also to corresponding results shown in Fig. 9a, TH2_PM1). All of the codes begin with the correct minimum temperature of 5°C, and thus the initial standard deviation is zero (Fig. C.1). The discrepancies then increase to a maximum, followed by a local minimum value after 10^4 s and subsequent plateau. The largest spread of the curves is observed close to the threshold time (when the minimum temperatures reach 0°C, at approximately 7×10^4 s). Afterwards, the standard deviation again generally declines, with a slight increase before the disappearance of

the plume. Finally, the standard deviation returns to near-zero as the simulations return to a steady-state uniform temperature of 5°C after 18×10^4 s.

The standard deviation is greatest around the threshold time as a result of the cumulative thermal degradation by conduction, advection and phase change, from the initial inclusion. The first standard deviation maximum is related to how the initial conditions were managed. The initial Boolean temperature conditions are numerically difficult to simulate, leading to oscillations that have to be damped numerically. When reaching the phase change period, the standard deviation plateau is a result of the impact of PFLOTRAN-ICE using a different freezing function while other codes are almost identical. The later-stage standard deviation evolution (after 10^5 s) probably results from accumulated shorter time differences or from differences in advection and dispersion of the colder temperature plume evolving from the initial inclusion.

Similar standard deviation variations are observed for the other imposed head gradients but correspond to shorter time frames (Fig. C.1). The final steady-state condition is reached more quickly for higher head gradients as a result of stronger advection leading to more rapid thermal degradation of the initial inclusion (refer to the blue and rose curves in Fig. C.1a, representing 9 and 15 % gradients). The structure for the 0 % case (red) is different and corresponds to longer simulation times due to purely conductive heat transfer.

Normalized standard deviations were also calculated to better compare the results among the PMs (Table C.1). With TH2_PM1, for example, the standard deviations were divided by 10°C, which represents the range of simulated temperatures (-5 to +5°C). Similar procedures were conducted to normalize all of the PMs listed in Table C.1, with the results indicating that the models were in general agreement with low normalized standard deviations of below 5 % for TH2_PM3, TH3_PM3, and TH3_PM4. Standard deviations are above 10 % for TH2_PM1 where the influence of the freezing curve expression is significant. No normalization was performed for all flux-based PMs (TH2_PM2 and TH3_PM2), while the special sensitivity of the 6% head gradient case, not included here, is further discussed in the main body.

Maximum (Norm. Std. Dev.)	Norm	Gradient = 0 %	Gradient = 3 %	Gradient = 9 %	Gradient = 15 %
TH2_PM1	$\Delta T = 10°C$	9	12	16	17
TH2_PM3	$\Delta V = 4\times10^{-2}\ m^3$	3	3	3	3
TH3_PM1	$\Delta K = 4\times10^{-4}\ m/s$	-	9.5	7	6
TH3_PM3	$\Delta E = 8\times10^{7}\ J$		2.3	4.7	5.2
TH3_PM4_Pt1	$\Delta T = 10°C$		3.1	3.2	2.5
TH3_PM4_Pt2	$\Delta T = 10°C$		1.9	3.9	4.3

Table C.1: Maximum values for the normalized standard deviation evolution (%)

Sources of discrepancies

A critical question for the code comparison exercise is to understand the origins and causes of discrepancies. One source of discrepancy was already mentioned, being related to differences of equation sets and characteristic curves.

Another important source of discrepancy is the codes' sensitivity to different initial conditions. As seen from the total liquid water volumes (TH2_PM3, Fig. C.2a) or total sensible heat initial conditions (TH3_PM3, Fig. C.2c), the standard deviation for this initial time propagates throughout the entire simulation time. Steady-state conditions are a 5 °C uniform domain, which all codes clearly agree on. Fig. C.2b displays the variability of the initial water volumes, and Fig. C.2d shows the initial total heat, among all 13 codes, the dotted line giving the calculated theoretical value. This variability results from different meshing strategies, in some cases mixed with differences in the characteristic curve expressions. This is particularly the case for TH3, in which specific meshing was required to represent the spherical-shaped initial conditions (see Fig. 4). Some codes use a dedicated meshing routine which accurately represents the boundaries of the frozen zone, whereas other codes project the initial temperature fields onto a regular orthogonal mesh.

A third cause of discrepancy may be related to the convergence accuracy. Although participants were asked to provide spatially and temporally converged simulations, all participants were

likely not able to complete a proper convergence study. Briefly, a convergence study is an exercise to demonstrate that the mesh and time step size do not affect the simulation outcomes. A full convergence study is ideally a pre-requisite for both test cases (TH2 & 3), and should include all gradients considered. This exercise is very computationally intensive and was only partially completed, for example it was not carried out for all head gradients. In an effort to study the influence of mesh and time refinement, the convergence studies made available by two participants are integrated into the analysis (see Fig. C.2a & b, dashed lines, Cast3M and PermaFoam codes). Mean and standard deviations were computed, based on Eq. C.1 & C.2, and considering the PM curves obtained from various levels of spatial and temporal discretizations (refer to Appendix B). The general conclusion is that these convergence studies show similar evolution in time and amplitudes as compared with the intercomparison results (compare plain and dotted lines in Fig. C.3). This is especially true for TH2_PM1 (Fig. C.3) where the similar standard deviation curves suggest that the spatial and temporal discretization could be responsible for such spreads in the code results, in particular at the same times as the systems undergo critical changes. The convergence studies conducted by both the Cast3M and PermaFoam codes considered a large range of variability (including coarse meshes that were not used for the final reference simulations) so that no direct conclusion can be drawn in terms of relative amplitude contributions. These results, however, indicate the importance of convergence studies, and how they affect the inter-code comparison results.

A fourth source of discrepancy could be related with the accuracy in the convergence of non-linear loops, especially that associated with the latent heat term. Although not illustrated here, the influence of this accuracy was often referred to among modelers, especially in the case of TH3 where conflicting conditions lead to strong oscillations in the resolution of the temperature field. As suggested in the main text (Discussion and Perspective section), valuable rules to handle this difficulty can be obtained from running the 1D Lunardini analytical solutions.

Effect of thermal advection

Advection is responsible for accelerated thermal degradation of frozen zones by heat exchange and heat transport. Difficulties were expected in representing advection since most of the numerical approaches (FE, FD, FV) are adapted primarily for solving conduction-dominated processes. The benchmark cases including advection were considered for various levels of head gradients (0, 3, 9, and 15 % for TH2; 3, 6, 9, and 15 % for TH3).

In all Figures C.1 to C.3, the maximum standard deviations associated with all gradients lie within a factor of two, indicating that codes achieve a similar convergence independent of head gradient levels. However, the maximum standard deviation may not be the most appropriate measure. Another metric for assessing the impact of gradient on the results was proposed - the integrals of the normalized standard deviation curves, computed over identical time frames. The results for TH2 are plotted in Fig. C.4. The concept behind this metric is to normalize all integrals to have all of the information on one plot. The same could not be done for TH3 because the integration interval would encompass the entire transient period before reaching steady state. The time was too long for low gradients and such long simulation times were not provided by all participants. The same is true for TH2_PM1, therefore the 0 % gradient is also not represented in Fig. C.4. Results show that the spread of the curves is of the same order of magnitude for the large range of flow velocities considered (imposed head gradients of 3, 9, and 15 %) with a weak tendency for the spreading to be reduced as head gradients increase, thus providing no clear and marked trend with increasing head gradients.

These results suggest that the advection term is apparently well handled or at least similarly well between codes regardless of the rate of water flow. Though some codes applied an upwind scheme for high advection levels, thus introducing some numerical dispersion, while others did not, the convergence is not altered by high advection for the test cases considered.

(a)

(b)

Zero flux

Imposed head $H_0+\Delta H$

Imposed head H_0

Zero flux

(c)

Zero conductive flux

Zero conductive flux

$T_{in} = 5°C$

Heat

Advection

$T^-_{initial} = -5°C$

$T^+_{initial} = 5°C$

Zero conductive flux

H (m)

> 9.8E−5
< 8.9E−2

9.0E−2 8.1E−2 7.2E−2 6.3E−2 5.4E−2 4.5E−2 3.6E−2 2.7E−2 1.8E−2 9.0E−3 0.0

8.1E−2   6.3E−2   4.5E−2   2.7E−2   9.0E−3

(a)

T (°C)

> −9.E−1
< 5.0

−5.0 −4.0 −3.0 −2.0 −1.0 0.0 1.0 2.0 3.0 4.0 5.0

0. 1. 2. 3. 4.

1 m

(b)

Temperature (°C)

MAXIMUM : 5.000
MINIMUM : -5.000

(a)

Distance along axis (m)

Temperature (°C)

MAXIMUM : 5.000
MINIMUM : -5.000

(b)

Distance along axis (m)

(a)

L_Y L_X

A L_{CY} R B

L_{CX} C

(b)

Zero flux

Imposed head $H_0 + \Delta H$

Imposed head H_0

Zero flux

(c)

$T_{imposed} = -5°C$

$T^-_{initial} = -5°C$

Zero conductive flux

$T_{in} = 5°C$

Heat Advection

$T^+_{initial} = 5°C$

$T^-_{initial} = -5°C$

$T_{imposed} = -5°C$

4

(a)

H (m)
> 1.5E−6
< 3.0E−2

3.0E−2
2.7E−2
2.4E−2
2.1E−2
1.8E−2
1.5E−2
1.2E−2
9.0E−3
6.0E−3
3.0E−3
0.0

1 m

(b)

T (°C)
> −4.97
< 4.98

5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
−3.0
−3.5
−4.0
−4.5
−5.0

5

(a)

(b)

(c)

(d)

7

(a)

(b)

(c)

(d)

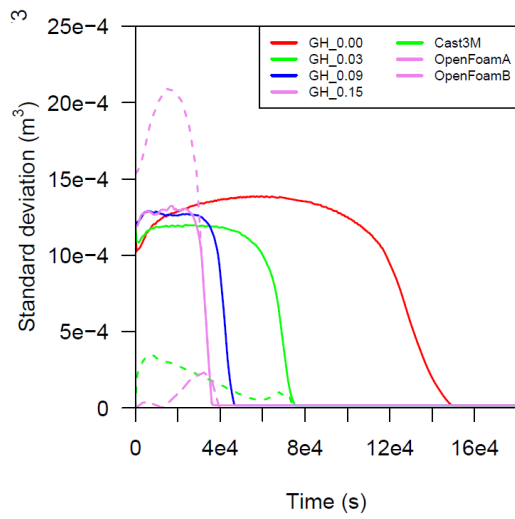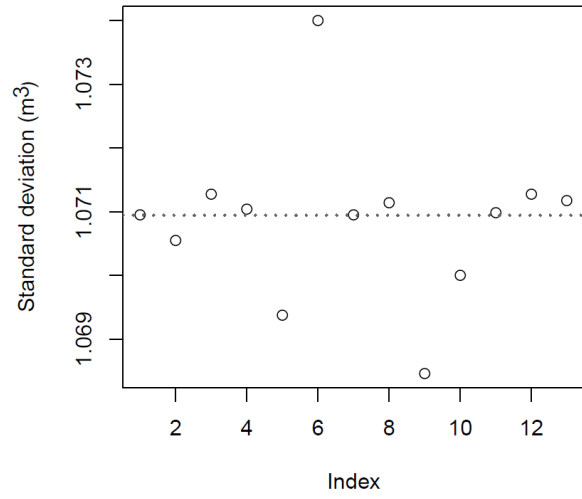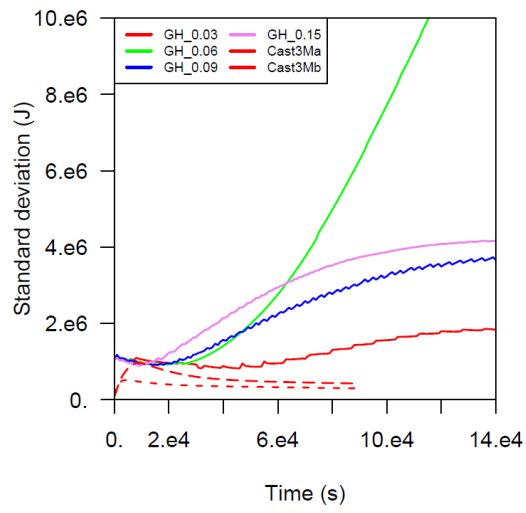
(e)

(f)

(a)

(b)

(a)

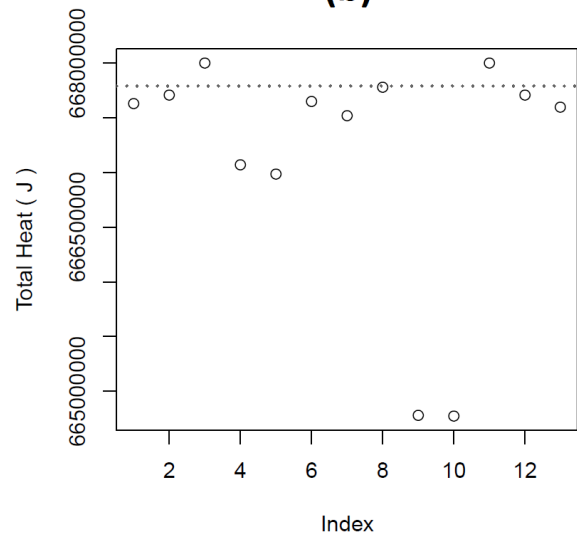**(b)**

12

(a)

(b)

(a)

(b)

15

(a)

(b)

(c)

(d)