A multiscale model of terrain dynamics for real-time earthmoving simulation

Martin Servin1,2*, Tomas Berglund2 and Samuel Nystedt1,2

Abstract

A multiscale model for real-time simulation of terrain dynamics is explored. To represent the dynamics on different scales the model combines the description of soil as a continuous solid, as distinct particles and as rigid multibodies. The models are dynamically coupled to each other and to the earthmoving equipment. Agitated soil is represented by a hybrid of contacting particles and continuum solid, with the moving equipment and resting soil as geometric boundaries. Each zone of active soil is aggregated into distinct bodies, with the proper mass, momentum and frictional-cohesive properties, which constrain the equipment’s multibody dynamics. The particle model parameters are pre-calibrated to the bulk mechanical parameters for a wide range of different soils. The result is a computationally efficient model for earthmoving operations that resolve the motion of the soil, using a fast iterative solver, and provide realistic forces and dynamic for the equipment, using a direct solver for high numerical precision. Numerical simulations of excavation and bulldozing operations are performed to test the model and measure the computational performance. Reference data is produced using coupled discrete element and multibody dynamic simulations at relatively high resolution. The digging resistance and soil displacements with the real-time multiscale model agree with the reference model up to 10–25%, and run more than three orders of magnitude faster.

Keywords: Deformable terrain, Discrete element method, Multibody dynamics, Multiscale, Real-time simulation, Soil mechanics

Introduction

Physics-based simulation of earthmoving equipment and soil is an important tool for developing smarter systems that meet the increasing demands for energy efficiency, productivity and safety in the agriculture, construction, and mining industries. Simulation with real-time performance is essential when developing new control systems, human-machine interfaces or training operators using interactive simulators with hardware or human in the loop. Fast simulation is vital also for applying artificial intelligence to motion planning and control of earthmoving equipment. These techniques are data-hungry, requiring many repeated simulations of a wide range of normal operating conditions as well as many, potentially hazardous, edge cases. Although the simulations can be run...
mostly in parallel, real-time performance, or faster, is necessary for covering a sufficiently large parameter space within reasonable time. Simulating heavy equipment alone at real-time with good accuracy is challenging but feasible using rigid multibody dynamics and a well-optimized solver [1,2]. It may seem out of reach to include also the environment, involving vast amount of soil with complex dynamics on scales that cross six orders of magnitude [3].

The first physics-based models for real-time simulation of heavy equipment and deformable terrain appeared in mid 1990, starting with Li and Moshell assuming the Mohr-Coulomb theory and volume preserving deformations [4]. Park [5] developed an elaborate model of the digging resistance in soil for the purpose of construction excavator simulators. The model starts from the fundamental earthmoving equation (FEE) [6] for a blade cutting a horizontal soil bed, assuming an active zone in the shape of a wedge [7]. The interface between the wedge and the passive soil represent the failure surface, stretching from the cutting edge of the blade to the free surface of the terrain. Park extends the model to a 3D bucket digging in sloped terrain, including the formation of a secondary separation plate by the deadload of material in the bucket and penetration resistance from the tool’s teeth. The FEE-based models take into account the strength of the soil, expressed in terms of its internal friction and cohesion, but are limited to stationary conditions and do not describe the motion of the soil. Holz et al. [8–10] combined the fundamental earthmoving equation (FEE) with contacting particle dynamics to model the cutting resistance and motion of the soil. In this approach, static soil is adaptively converted into particles as the tool cut the terrain. Digging resistance, based on the FEE, is applied as a six-degree of freedom kinematic constraint with force limits on the digging tool. A wedge-shaped active zone is assumed. The portion of terrain in the active zone that has not yet undergone failure and particle conversion provide mass to the FEE. The particle contacts on top of the wedge contribute as surcharge mass, and the contacts with the tool provide additional digging resistance. Later, a similar approach is taken in [11].

Unfortunately, relying on the FEE for the digging resistance has serious drawbacks. It is valid only at steady state and it suffers from unphysical singularities, as pointed out in [10]. Furthermore, the active zone possesses no inertia or momentum. When accelerated motion is involved, the FEE can clearly not provide correct dynamics and proper reaction forces. On the other hand, resolving the active zone with contacting particles challenges the computational performance. If the particles are not too many, real-time simulation is possible using nonsmooth contact dynamics [12] or position-based dynamics [13], with a stationary iterative solver like the Gauss-Seidel algorithm. This is associated with numerical errors that manifest as artificial elasticity or insufficient friction [14,15]. If the errors become too large the soil will behave more like a compressive fluid rather than a stiff soil that yield only if the shear stress reach critical values. The errors grow with the number of particles and applied stress and decrease with the time-step and increased number of solver iterations. In the other limit, of a few coarse particles, the solver error decrease but the spatial discretization errors grow large.

To cope efficiently with the disparate length and time scales, we explore a multiscale model for the soil dynamics. The model targets real-time simulation of earthmoving equipment and terrain with realistic reaction forces and soil deformations. A macroscale model, describing the rigid multibody dynamics, is simulated using a direct solver for high numeric precision. The active soil, resolved as particles in a mesoscale model, is simu-
lated using an iterative solver at high-speed with large error tolerance. The key idea is that the soil dynamics is represented in both models, with a coupling that filters out the discretization and solver errors from the mesoscale model but capture the bulk dynamics. When an earthmoving object come in contact with the terrain, the potential failure surfaces and zones of active soil are predicted. The soil inside the active zones is represented using a hybrid model of contacting particles and continuum solid, which support smooth transitioning between the resting solid, liquid and gas phases. The coupling back to the macroscale model is mediated through aggregate bodies. They have the momentary shape, mass, and velocity of the resolved soil in each active zone. The motion of the aggregates are constrained relative to the equipment and the terrain failure surface, in accordance with the Mohr-Coulomb model. For blunt objects contacting the terrain, the subsoil stress distribution is estimated and the soil compacts if the stress reach critical values. A microscale model, simulated using relatively small particles and high numerical precision, is used for pre-calibration of the particle parameters to match the bulk mechanical properties of a wide range of soils. The primary model parameters of the multiscale model are the bulk mechanical properties of the soil at a given bank state: the internal friction, cohesion, dilatancy, elasticity, and mass density; as well as the equipment’s geometry, surface friction and cohesion. To test the model, simulations of excavation and bulldozing operations in various soils are performed. The digging resistance and soil displacements from the real-time multiscale model are compared with those from the microscale reference model.

Modelling and simulation of soil and earthmoving equipment

Length- and timescales

Soil and granular media are strongly dissipative multiphase materials with multiple length- and timescales [3,16]. They consist of contacting grains with size ranging from clay at $10^{-6}$ m to cobble and boulders at $10^0$ m. Natural soil has a certain moisture content, that may significantly increase the strength of the soil by cohesive forces. At sufficient moisture levels, however, a pore pressure develops that lower the inter-particle normal forces and, consequently, the internal friction. Large deformations are commonly localized in shear bands, sometimes as narrow as a few particle diameters. The soil outside the shear bands is displaced rigidly. Soil is strongly dissipative. Therefore, the solid phase is the natural state. If it is agitated, for instance by an earthmoving equipment, it may transition to the liquid or gaseous phases. For sand and gravel in typical earthmoving operating conditions, the grain collision timescale is in the microsecond regime while the liquid timescale is around a millisecond\(^1\). Earthmoving equipment, on the other hand, has characteristic size of 10 m, operating range of around 100 m and may displace several cubic meters of soil per second. An earthmoving tool can be controlled with a spatial precision of about 10 mm and its geometric features is also on this scale. A typical loading cycle, with a bucket excavator or a wheel loader, has time duration in the range between 15 and 30 s [18,19]. The natural timescales of the rigid body motion is about 1–10 Hz and the control systems typically operate at a frequency below 100 Hz [20]. These multiple and separated length- and

\(^1\)The liquid phase time scale is the characteristic time scale of particle rearrangements, $d/\sqrt{\eta/\rho}$, with particle diameter $d$, mean stress $\eta$, and specific mass density $\rho$. The time scale in the gaseous state is the characteristic collision time $4(\nu^{-1}(m/4k)^{2/5})$, assuming a strongly damped Hertzian contact model [17], for impact velocity $\nu$, particle mass $m$, and contact elasticity $k = \sqrt{\bar{E}/2(1 - \nu^2)}$, with Young’s modulus $E$ and Poisson’s ratio $\nu$. 
timescales are important to consider when modelling soil and granular media interacting with earthmoving equipment.

Distinct particles
The discrete element method (DEM) [21] describe granular media and soil as consisting of distinct contacting particles of finite size. It is a versatile model that automatically describe soil in the solid, liquid, and gaseous phases and the transitions between them. DEM is clearly applicable for simulating the soil dynamics in earthmoving operations [22–24]. It is, however, very computationally intensive. A common solution is to not represent the true grains with their actual distribution of size, shape, and mechanical properties. Instead, the soil is represented by a collection of large, often spherical, pseudo particles with contact parameters—elasticity, friction, cohesion and rolling resistance—that are calibrated such that the bulk mechanical properties match the ones observed in the soil that the model is meant to represent. The mapping between the particle parameters and macroscopic soil parameters can be carried out using the triaxial test or cone penetration test, as described in [25]. But even with pseudo-particles of size 10 to 100 mm, the number of particles for representing a terrain the size of earthmoving equipment exceeds what is currently possible to simulate in real-time using DEM.

Continuum
On length scales larger than the particle size, soil may be modelled using continuum mechanics. The state of the soil is represented with scalar, vector and tensor fields for mass density \( \rho(x, t) \), displacement \( u(x, t) \), velocity \( v(x, t) \), stress \( \sigma(x, t) \) and strain \( \epsilon(x, t) \). The fields obey the equations of mass continuity

\[
\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = 0,
\]

and momentum balance

\[
[\frac{\partial}{\partial t} + v \cdot \nabla] v = \rho^{-1} \nabla \cdot \sigma + f_{\text{ext}},
\]

where we use the short notation for partial time derivative \( \frac{\partial}{\partial t} = \frac{\partial}{\partial t} \), and \( f_{\text{ext}} \) denote any external force, like gravity. In the dense regime, and when the stress is below a certain yield strength condition, the soil behaves as an elastic solid with some constitutive law relating stress to strain, e.g., Hooke’s law for small deformations \( \sigma(x, t) = C : \epsilon(x, t) \). For isotropic and homogenous materials the stiffness tensor \( C \) has only two independent parameters, often represented by the Young’s modulus \( E \) and the Poisson’s ratio \( \nu \). When the stresses reach the yield condition the solid fails and deforms plastically, rupture or flows rapidly. The simplest yield condition for soil is the Mohr-Coulomb criteria. It predicts that the material will fail along any plane with normal \( n \) where the shear stress, \( \tau_n = \sqrt{(\sigma \cdot n)^2 - \sigma_n^2} \), reaches the critical value of

\[
\tau_n = \mu \sigma_n + c,
\]

where the normal stress is \( \sigma_n = \sigma_{\alpha\beta} n_\alpha n_\beta \). The model has two parameters for the strength of the material: the internal friction, \( \mu \), and the cohesion, \( c \). The internal friction is often represented by the angle of internal friction \( \phi = \arctan(\mu) \). Analogously with Coulomb friction, the critical shear strength grows linearly with the normal stress (pressure). The shear stress must also overcome any cohesive strength of the material, which is independent of the stress and strain in the Mohr-Coulomb model. If the material yields quasistatically it may be modelled as an elastoplastic solid, with a plastic flow rule that accounts
for strain-hardening (or softening) and the volume expansion during shear. The latter is known as dilatancy and is defined as \( \text{tr}(\dot{\varepsilon})/3\|\dot{\varepsilon}\| \), where \( \dot{\varepsilon} \) is the deviatoric strain. The hardening can be incorporated in the Mohr-Coulomb law by [16, 26] by an effective internal friction
\[
\mu = \tan(\phi + \psi)
\]
where \( \psi = \arcsin\left[\text{tr}(\dot{\varepsilon})/3\|\dot{\varepsilon}\|\right] \) is the dilatancy angle. Soil may compact plastically under uniform compression if the applied stress is large enough to cause particle rearrangements. This is captured by a soil’s compression index
\[
C_c \equiv \frac{\Delta e}{\Delta \ln(\sigma)}
\]
which measure the change in void ratio, \( e \), by a change in confining stress.

The finite element method is the most widely used and versatile technique for simulating deformable solids. The popular integration schemes for elastoplastic solids [27] first compute trial strain and stress fields, and then use Newton’s method for searching the plastic strain increment that fulfill the constitutive model and plastic flow rule at each time-step. In the regime of loose soil and large shear rate, the material is better described as a viscoplastic solid or a non-Newtonian fluid with some constitutive law between the stress and strain-rate, like the \( \mu(I) \)-rheology for granular media [28]. Cemented soil, with clay or silt particles, may fracture by brittle failure. Mesh-based numerical methods for solid dynamics have various difficulties with large deformations and topological changes associated with soil that undergo plastic or brittle failure, and transitions between the dense solid, dilute liquid and gaseous phases. Meshfree methods, such as the material point method, are showing promising results on 3D problems, coupled with multibody dynamics, but real-time simulation of earthmoving operations remain out of reach [29, 30].

Analytical and semi-analytical solutions, derived from the continuum theory, are valuable for both insight and for creating fast simulation models. This include the Boussinesq type-of-equations for the stress and deformation fields underneath a load applied on the surface of a semi-infinite domain. The reaction forces on a thin blade cutting a soil bed has been thoroughly analysed [7]. A blade, or separating plate, has two basic modes of operation, penetration and separation. Penetration is the motion straight into the soil with relative velocity in tangential direction of the plate only. Soil cavities at the tip of the blade, often equipped with teeth, are thus forced to expand [5, 31]. The penetration resistance may be considerable although the deformations are relatively small. Separation corresponds to movement normal to the plate and is the main cause of soil failure and large displacements. The edge where the blade meet the material is referred to as the cutting edge. See Fig. 1 The shape of the failure surface can be computed analytically in the two dimensional case, applicable for a wide blade, using the method of stress characteristics and assuming the Mohr-Coulomb criteria. The failure surface is often approximated with a plane. This defines an active zone with the shape of a wedge. Rankine’s theory for a flat soil with a blade pushing on it in the horizontal direction predicts that the soil fails at an angle \( \theta = \pi/4 - \phi/2 \) against the horizontal. In three dimensions, the failure surface extends sideways also, which cause long berms along the sides of a pushing blade. The separation force acting on a blade moving horizontally at a constant speed is well described by the Fundamental Earthmoving Equation (FEE) [6]. The FEE is motivated by wedge model of the soil failure. The force resistance per tool width \( L \), in the FEE is composed of four terms.
Fig. 1 Illustration of a blade interacting with a soil bed. There are two modes of operation, penetration, and separation (left). The failure surface forms a wedge-shaped active zone (middle). When a blade cuts the soil there are many forces contributing to the soil resistance (right).

\[
F = \frac{\rho gd^2 N_f}{L} + QN_Q + cdN_c + caN_a,
\]

with specific soil mass density \(\rho\), tool penetration depth \(d\), soil cohesion \(c\), surcharge force \(Q\) (per tool width) and soil-tool adhesion \(ca\). The first term is due to the weight of the wedge, the second term is the additional (vertical) surcharge, the third term the cohesive force in the failure surface, and the fourth term is the resistance due the adhesion between the blade and the soil, see Fig. 1. The four \(N\)-factors (found in [6,7]) depend on the geometry of the tool and the failure zone, the internal friction and cohesion, as well as of the surface friction and adhesion. The quadratic dependency on the cutting depth \(d\) reflect that the weight depends linearly on the cross-section area of the failure zone. Note that the cohesive and adhesive force terms are proportional to the area of the failure surface and blade contact surface, respectively. One key limitation of the FEE is that it assumes steady state and low speed of the blade and soil flow. Also, the \(N\)-factors suffer from singularities at certain geometrical configurations.

Contacting rigid multibodies

The earthmoving equipment can be simulated efficiently using rigid multibody dynamics [1,2]. Articulated and actuated mechanisms are thus modelled as rigid bodies with kinematic constraints that represent the mechanical joints, motors, and driveline. Dynamic contacts between the bodies are best modelled as kinematic constraints and complementarity conditions to express unilaterality, impacts and Coulomb friction. In the current paper, we apply rigid multibody dynamics with contact dynamics for both the equipment and the soil, and in all the levels of the multiscale model. Therefore, we describe the computational framework at greater level of detail.

The state of a rigid multibody system with \(N_b\) bodies, \(N_j\) joints and actuators and \(N_c\) contacts, is represented on descriptor form in terms of the position, \(x(t) \in \mathbb{R}^{6N_b}\), velocity, \(v(t) \in \mathbb{R}^{6N_b}\), and Lagrange multipliers, \(\lambda_j(t) \in \mathbb{R}^{6N_j}\) and \(\lambda_c(t) \in \mathbb{R}^{6N_c}\), that are responsible for the constraint forces in joints and contacts. The system position variable is a concatenation of the spatial and rotational coordinates of the \(N_b\) bodies, \(x = [x, e]\), and the velocity variable holds the linear and angular velocities, \(v = [v, \omega]\). The time evolution of the system state variables \([x, v, \lambda]\) is given by the following set of equations

\[
M \ddot{v} = f_{ext} + G^T \lambda_j + G^T \lambda_c
\]
\[ \epsilon_j \lambda_j + \eta_j g_j + \tau_j G_j v = u_j, \quad (8) \]

contact\_law(\nu, \lambda_c, g_c, G_c), \quad (9)

where \( M \in \mathbb{R}^{6N_c \times 6N_c} \) is the system mass matrix, \( f_{exi} \) is the external force, and \( G_j^T \lambda_j \) and \( \lambda_c \) are constraint forces for joints (and motors) and for contacts, respectively. The forces have dimension \( \mathbb{R}^{6N_c} \) and is composed of linear force and torque. Equation (8) is a generic constraint equation, with constraint function \( g_j(\nu) \), Jacobian \( G = \partial g / \partial \nu \), compliance \( \epsilon_j \) and viscous damping rate \( \tau_j \). An ideal joint is represented with \( \epsilon_j = \eta_j = 0 \) and \( \tau_j = 0 \), in which case Eq. (8) express a holonomic constraint, \( G_j^T \lambda_j = 0 \). A linear or angular motor may be represented by a velocity constraint (setting \( \epsilon_j = \eta_j = 0 \) and \( \tau_j = 1 \)), \( G_j^T \lambda = u_j(t) \), with set speed \( u_j(t) \). The holonomic and nonholonomic constraints can be seen as the limit of a stiff potential, \( U_c = \frac{1}{2} g^T g \), or a Rayleigh dissipation function, \( R_v = \frac{1}{2} (Gv)^T Gv \), respectively [32]. This offer the possibility of mapping known models of viscoelasticity to the compliant constraints. Descriptor form means that no coordinate reduction is made. The system is represented explicitly with its full degrees of freedom, although with the presence of constraints.

We consider the system to have nonsmooth dynamics [33]. That means that the velocity and Lagrange multipliers are allowed to be time-discontinuous, reflecting instantaneous changes from impacts, frictional stick-slip transitions or joints and actuators reaching their limits. This is unavoidable when using an implicit integration scheme\(^2\) because of the coupling between the state variables trough unilateral and frictional contacts.

As contact law between particles we use a model that include cohesive-viscoelastic normal contacts (n), tangential Coulomb friction (t) and rolling resistance (r). These are formulated in terms of inequality and complementarity conditions for the velocities, Lagrange multipliers and constraint functions. The resulting model can be seen as a time-implicit version of conventional DEM and is therefore referred to as nonsmooth DEM [14,34]. We use the following conditions on \( \nu \) and \( \lambda_c = [\lambda_n, \lambda_t, \lambda_r] \) as contact\_law:

\[
0 \leq \varepsilon_n \lambda_n + g_n + \tau_n G_n \nu \perp (\lambda_n + f_c) \geq 0, \quad f_c = c_p A p / \| G_n^T \| \quad (10)
\]

\[
\gamma_n \lambda_t + G_t \nu = 0, \quad \| \lambda_t \| \leq \mu_t \| G_n^T \lambda_n \| \quad (11)
\]

\[
\gamma_t \lambda_r + G_r \nu = 0, \quad \| \lambda_r \| \leq r \mu_r \| G_n^T \lambda_n \|, \quad (12)
\]

where \( g_n \) is a function of the contact overlap and the Jacobians, \( G_n, G_t \) and \( G_r \), are the normal, tangent and rotational directions of the contact forces [14]. The parameters \( \varepsilon_n, \tau_n, \gamma_t \) in Eq. (10) control the contact stiffness and damping, and \( f_c \) the cohesion. Setting these parameters to zero means that no penetration should occur, \( g_n(\alpha) \geq 0 \), and if so the normal force should be repulsive, \( \lambda_n \geq 0 \). The inclusion of \( f_c \) enables cohesive normal force with maximum value \( c_p A_p \), where \( c_p \) is the particle cohesion and \( A_p \) is the particle cross section area. The cohesion is active when the contact overlap is smaller than a certain cohesive overlap, set to a fraction of the particle size, e.g., \( \delta_c = 0.025d \). Eq. (11) state that contacts should have zero slide velocity, \( G_t \nu = 0 \), giving rise to a friction force that is bounded by the Coulomb friction law with friction coefficient \( \mu_t \). Similarly, Eq. (12) states that, as long as the constraint torque is no greater than the rolling resistance law, the

\(^2\)The alternative is to resolve the contact events using smooth trajectories, stiff potentials and small time-step explicit time integration.
contacting bodies are constrained to zero relative rotational motion, \( \mathbf{G} \mathbf{v} = 0 \). Here, \( \mu_r \) is the rolling resistance coefficient and \( r \) is the particle radius. It is a well-known fact that the effect of particle angularity, on internal friction and angle of repose, can be captured using spherical particles with rolling resistance. As explained in [35], a \( n \)-sided polygon can be assigned a rolling resistance coefficient \( \mu_r = (1/4) \tan(\pi/2n) \), which gives \( \mu_r = 0.05 \) for an eight-sided polygon, \( \mu_r = 0.1 \) for a square and \( \mu_r = 0 \) for a sphere (\( n = \infty \)). We map the normal contact law, Eq. (10), to the Hertz-Mindlin model for contacting viscoelastic spheres, \( f_n = k_n \delta^{3/2} n + k_n c_d \delta^{1/2} n \), where \( \delta(x) \) is the contact overlap, \( k_n = \frac{1}{3} E^a \sqrt{d^a} \) is the contact stiffness, \( c_d \) is a damping coefficient, \( E^a = [(1 - v^a_n)/(1 + v^a_n)]/E_a + (1 - v^b_n)/(E_b)]^{-1} \) is the effective Young’s modulus, \( d^a = (d_a^{-1} + d_b^{-1})^{-1} \) is the effective diameter for two contacting spheres, \( a \) and \( b \), with Young’s modulus \( E_a \), diameter \( d_a \) and Poisson’s ratio \( v_a \) etc. The mapping to Eq. (10) is accomplished by \( g_n = \delta^{5/4}, \varsigma_n = 5/4 k_n \) and \( \tau_n = 5c_d/4 \).

We separate collisions into resting contacts and impacts using an impact threshold velocity \( v_{\text{imp}} \). If the relative contact velocity is smaller than this value the contacts are stabilized terms are \( \dot{\mathbf{v}} \). For numerical integration we employ the §stepper [32]. It is a first order accurate discrete variational integrator, developed particularly for fixed time-step real-time simulation of multibody systems with non-ideal constraints and non-smooth dynamics. It has been proven to be linearly stable and found stable on contact problems in empirical studies [37]. The numerical time integration scheme for advancing the system’s position and velocity from \([x_n, v_n] \) at time \( t_n \) to \([x_{n+1}, v_{n+1}] \) at time \( t_{n+1} = t_n + \Delta t \) consist of a position update \( x_{n+1} = x_n + \Delta t v_{n+1} \) after having computed the new velocity \( x_{n+1} \) and corresponding Lagrange multiplier \( \lambda \). This is done by solving the following mixed complementarity problem (MCP) [38]

\[
\begin{bmatrix}
M - \mathbf{G}_n & -\bar{\mathbf{G}}_n \\
\mathbf{G}_n & \Sigma & 0 \\
\bar{\mathbf{G}}_n & 0 & \Sigma
\end{bmatrix}
\begin{bmatrix}
v_{n+1} \\
\lambda \\
\bar{\lambda}
\end{bmatrix}
- 
\begin{bmatrix}
p_n \\
q_n \\
u_n
\end{bmatrix}
= w_l - w_u,
\]

\[0 \leq z - l \perp w_l \geq 0 \]

\[0 \leq u - z \perp w_u \geq 0\]

where the constraints have been grouped into position constraints (no bar) and velocity constraints (with bar), \( p_n = M \mathbf{v}_n + \Delta t f_n \), \( q_n = -\frac{4}{\Delta t} \mathbf{Y} g + \mathbf{Y} \mathbf{v}_n \), and \( u_n \) is the target speeds of the velocity constraints (zero for frictional contacts). The regularization and stabilization terms are \( \Sigma = \frac{4}{\Delta t} \text{diag}[\varsigma/(1+4\tau/\Delta t)] \), \( \bar{\Sigma} = \frac{1}{\Delta t^2} \text{diag}([\gamma]) \) and \( \mathbf{Y} = \text{diag}[1/(1+4\tau/\Delta t)] \). The slack variables \( w_l \) and \( w_u \) are only used internally by the MCP solver. For details about the solver, see Sec. Implementation.
Coarse-graining

The process of averaging particle kinematics and contacts forces into continuous and differentiable fields is referred to as homogenization. **Coarse-graining** is one particular way of doing homogenization. This is useful when combining particle and continuum models of granular media. The fields are computed by sampling the particle variables using a smoothing kernel, $\zeta(x)$, that is normalized $\int \zeta(x)dx^3 = 1$ and approaches zero on a smoothing length $R$. The fields of mass and momentum density are computed as

$$\rho(x, t) = \sum_a m_a \zeta(x - x^a(t))$$

and

$$p(x, t) = \sum_a m_a \nu^a \zeta(x - x^a(t))$$

respectively, and the velocity field is simply

$$u(x, t) = p(x, t)/\rho(x, t).$$

(16)

The rate of strain tensor can thus be computed as

$$\dot{\varepsilon}_{\alpha\beta} = \frac{1}{2} \sum_a m_a \nu^a \frac{\partial \nu^a}{\partial x^\beta} \zeta(x - x^a(t)).$$

for a small change in density $\rho/\rho_0 = 1 + \Delta V/V$ relative to a reference state with mass density $\rho_0$ and Young’s modulus $E_0$. The sign $\pm$ is positive for compaction and negative for expansion. The dilatancy angle also increase with the level of compaction, and with the internal friction by Eq. (4). Based on numerical simulations and coarse-graining of dense granular media, Roux and Radjai [39] proposed

$$\psi = c_\psi (\varphi - \varphi_c),$$

where $\varphi$ and $\varphi_c$ are the current and critical porosity, at which the soil switch between positive dilation (volume expansion) and negative dilation (volume shrinkage) upon shearing, and $c_\psi$ is a constant that depends on the particle shape.

A multiscale model of terrain dynamics

The multiscale model has three levels of abstraction, illustrated in Fig. 2, that we refer to as **micro-, meso-** and **macroscale**. In the microscopic model the terrain is fully resolved by relatively fine-grained particles with contact properties pre-calibrated to represent various type of soil, e.g., dirt, gravel, or sand. It serves as a ground-truth reference model, simulated off-line in advance with relatively small time-step and high numerical accuracy. The output is used for validation of the coarser scale models and, if necessary, for calibration of parameters not known from theory or experiments. The meso- and macroscale models are simulated in real-time and synchronously coupled to each other. The terrain deformations and soil flow is integrated using the mesoscale model, with a hybrid representation of the soil that combine coarse particles and fields discretised with a regular grid of voxels.
Input to the mesoscale model is the motion and contact forces of the equipment at the interfaces to the terrain. This is provided by the macroscale model, which focus on the rigid multibody dynamics of the equipment, and any other objects interacting with the terrain. The equipment experience the resting terrain as a quasistatic surface and each region of agitated soil as distinct dynamic bodies, whose shape, mass velocity and mechanical strength is aggregated from the mesoscale model.

It is a significant feature that each of the three models use the same parameters to characterize the soil properties. These are mapped to the model parameters for the particle, voxel, and aggregate body dynamics. The parameter mapping relies in part on the theory of continuum soil mechanics and in part on parameter calibration using the microscale model. The primary soil parameters include mass density \( \rho_b \), internal friction angle \( \phi_b \), cohesion \( c_b \), dilatancy angle \( \psi_b \), bulk elasticity modulus \( E_b \), that characterise the mechanics of the soil as found in a natural \textit{bank state}. These are collected in a bulk parameter vector \( \mathbf{p}_b = [\rho_b, \phi_b, c_b, \psi_b, E_b] \). The bank state refers to the state at which the true soil is found in nature. It may be the result of geological, meteorological and machined processes, that leave the soil at a particular packing density and moisture content. Two soils with identical (real) particles may thus have different bank state properties and are considered as two distinct soil with different bulk parameters \( \mathbf{p}_b \).

**Microscopic model**

At highest resolution we represent the terrain as a set of relatively fine-grain frictional-cohesive particles, \( \mathcal{N}_p \), and the equipment as a set of rigid multibodies, \( \mathcal{N}_{RB} \), with some set of joints and actuators. Let \( \mathcal{N}_{BP}^B \) denote the set of particles in direct contact with a body \( B \in \mathcal{N}_{RB} \) of the equipment. The equations of motion for the earthmoving body and the terrain particles, labelled \( a \), are

\[
\begin{align*}
\mathbf{m}_a \ddot{\mathbf{v}}_a &= \mathbf{f}_{ext}^a + \mathbf{G}_a^T \lambda_{as} + \sum_{a' \in \mathcal{N}_B^a} \mathbf{G}_{aa'}^T \lambda_{aa'}' \\
\mathbf{m}_a \ddot{\mathbf{v}}_b &= \mathbf{f}_{ext}^b + \mathbf{G}_b^T \lambda_b + \sum_{a' \in \mathcal{N}_B^b} \mathbf{G}_{ab}^T \lambda_{ab'}
\end{align*}
\]

(19)

(20)

where \( \mathbf{G}_a^T \lambda_{as} \) is the constraint force coupling the earthmoving body to the rest of the equipment, \( \mathbf{G}_a^T \lambda_{as} \) is the contact force on the body from particle \( a \), and \( \mathbf{G}_{ab}^T \lambda_{ab'} \) is the inter-particle contact force between particle \( a \) and \( a' \). The body-particle interfacial force is the source for the digging resistance perceived by the equipment. The interfacial stress alters also the internal stress in the soil, causing shear failure or brittle rupture if it reach the critical stress.

The particles can roughly be divided into two domains. Either they are part of the active zone, \( \mathcal{N}_A^p \), which is sheared or rigidly displaced by the earthmoving body, or they remain part of the resting bed of particles, \( \mathcal{N}_C^p \). The domains are separated by a failure surface \( AC \). If the earthmoving body is a tool with penetrating teeth or a sharp edge, the particles in direct contact with that constitute a third domain, \( \mathcal{N}_D^p \). The particle domains are illustrated in Fig. 2.

We use a \textit{pseudo-particle} representation of the soil. That means that the particles do not represent the true grains with their actual distribution of size, shape, and mechanical properties. Instead, the soil is represented by a collection of relatively large spherical par-
Fig. 2 Overview of the multiscale model. The mesoscale and macroscale models are simulated in real-time coupled to each other while the microscale model is simulated off-line for the purpose of model testing and calibration. The mesoscale model combines a coarse particle and continuum representation of the soil inside the active zone. The active soil is aggregated into a single body, present in the macroscale model, with frictional-cohesive couplings with the resting terrain and the rigid multibody equipment.

particles with specific mass density and contact parameters - elasticity, friction, cohesion and rolling resistance - \( \mathbf{p}_p = [\rho_p, \mu_\tau, \mu_r, c_p, E_p] \). These parameters are calibrated to numerical values that give the particle soil the desired bulk mechanical properties \( \mathbf{p}_b \). The size of the pseudo-particles is chosen small enough to resolve the important features of the equipment and precision at which it can be controlled, i.e., around \( 10 - 50 \) mm in earthmoving applications.

**Mesoscale model**
The mesoscale model is a medium-resolution model of the soil dynamics using a hybrid particle-continuum approach. The soil in the active zone is primarily represented by coarse particles. The continuous soil model has two phases, resting *solid mass* and *fluidized mass*. The former represents resting soil outside the active zone and is considered an elastoplastic solid. The latter complement the use of particles for representing soil displacement in the active zone. The fluidized mass is convected with the coarse-grained velocity field of the particles, and subject to gravity. The macroscale model supplies the equipment’s motion and contact forces at the terrain interface as input to mesoscale model. This is the basis
for predicting the active zones and provide boundary conditions that drives the mesoscale soil dynamics.

A regular grid is used for the discrete representation of the continuum model. The grid cells, or voxels, are labelled \( i = (i, j, k) \) by the triplet of integer positions in the grid, aligned with the global coordinate axes. Each voxel has a centre point \( x_i = [x_i, y_i, z_i] \) and constant volume \( V_0 = l_0^3 \). Each voxel has a variable mass \( m_i \), velocity \( v_i \), compaction \( w_i \) and occupancy \( \phi_i \). These are collected in a state vector \( s_i = [m_i, v_i, w_i, \phi_i] \). The evolution of \( s_i \) is treated by a cellular automata, with transition rules that implement convection of fluidized mass, plastic compaction at critical sub-soil stress, conversion between the particle and continuum representation, and surface flow if the local slope exceeds the soil’s angle of repose \( \delta_b \) \cite{4,40}. For cohesion-free materials it coincides well with the internal friction angle, given by Eq. (4). With cohesion it may be larger \cite{41}. Beyond this slope the terrain is not in a stable equilibrium, and will fail by avalanching into a valid state. The transitions are constructed to preserve the total mass to machine precision.

The voxel mass density, \( \rho_i = \rho_{s,i} + \rho_{f,i} \), is composed by the density of resting solid, \( \rho_{s,i} \), and fluidized mass, \( \rho_{f,i} \). All voxels in the terrain are fully occupied, \( \phi_i = 1 \), with solid mass and empty of fluidized mass, except for surface voxels, that may have occupancy less than unity, \( \phi_i = V_i^s / V_0 \in [0, 1] \). These voxels, and the ones above the surface, may also contain fluidized mass. The solid mass density has a natural bank state value \( \rho_b \) but can vary locally within the range \( \rho_s^i \in [\rho_{\text{min}}, \rho_{\text{max}}] \) if subject to compaction or swelling, \( w_i \equiv \rho_{s,i} / \rho_b \). The amount of solid mass in a voxel \( i \) is consequently \( m_i^s = \rho_s^i V_i^s = w_i \phi_i \rho_b V_0 \). The upper and lower limit on the mass density imply that the compaction is bounded by the upper and lower values \( w_{\text{max}} = \rho_{\text{max}} / \rho_b \) and \( w_{\text{min}} = \rho_{\text{min}} / \rho_b \). We identify \( S \equiv w_{\text{min}}^{-1} \) as the swell factor of a soil that is transformed from bank state to its maximally loose packed state.

The surface voxels define a surface heightfield, \( z = h(x, y) \), that is used for contacts with the particles, the equipment, or any other objects in the macroscale model. It has a discrete representation \( h_{ij} = h(x_i, y_j) \). The height value in a column with index \((i, j)\) is the centre position, \( z_{i'} \) of the top-most non-empty voxel, \( i' = (i, j', k) \), plus the local mass fill ratio relative to that voxel centre, i.e.,

\[
    h_{ij} = z_{i'} + (\phi_{i'} - 1/2) l_0,
\]

This make the surface heightfield a continuous function of the solid occupancy, see Fig. 3. Between the grid points the surface height field is interpolated linearly.

The response by the terrain is different for contacts with sharp and blunt geometries. The former lead to shear failure with a localized failure surface while the latter cause soil compaction. If the contacting body has a sharp cutting edge, a co-moving active zone is predicted. The motion of the terrain inside the active zone is resolved by particles and fluidized mass. See Fig. 3 for illustration. The basic shape of the active zone is that of a wedge, defined by the cutting edge, the soil failure surface that extends from the edge to the free surface of the terrain, and the separation plane of the cutting body. The slope of the failure surface, \( \theta \), depends on the soil’s internal friction, \( \phi \), and the orientation of the separation plane relative to the to the terrain surface, \( \beta \). From simulations with the
As a blade and its active zone moves into the terrain a, new solid voxels are resolved into particles or fluidized mass that form the aggregate body. The voxel height value, corresponding to the solid occupancy, is found by projection to the failure plane of the active zone. When the blade is raised b, the gap is filled by solid mass converted from fluidized or particle mass in the vicinity of the edge.

As the cutting body moves through the soil, the active zone overlap with new voxels. Overlapping soil is converted from solid mass, to particles or fluidized mass. The amount of converted mass is computed using Eq. (21), such that the surface voxel’s fill ratio take the exact value where the surface heightfield coincides with the failure surface, see Fig. 3. The liberated mass is converted into new particles or growth of existing particles in the vicinity of the voxel, provided the local void ratio allows it. To avoid ending up with too small particles (poor performance) or too large particles (discretization error) the particle size is restricted to a given range \([d_{\text{min}}, d_{\text{max}}]\). The size range is chosen such that the number of particles remain within the limits of real-time simulation. Any residual mass, that cannot be converted into particle mass, is converted into fluidized mass in the voxel. At the next time-step, the fluidized mass is again candidate for formation and growth of particles. If it reaches a surface voxel outside the active zone, it is converted back to resting solid mass, whereby the surface heightfield is raised accordingly. The inertia of the coarse particles is assumed to dominate over the dilute fluidized mass. Therefore, we simply model the flow of fluidized mass by conserved advection.

\[
\partial_t \rho_f + \nabla \cdot (\rho_f \mathbf{u}) = r_f, \tag{23}
\]

where \(\mathbf{u}(x)\) is the coarse-grained particle velocity field, using Eq. (16), and \(r(x, t)\) is a source (or sink) term for mass converted from (or to) particle mass or solid mass. Any fluidized mass that is found in a voxel with no particles is projected in the direction of gravity towards the surface where it is converted into solid mass.
When a cutting body is raised from the terrain surface a gap may arise between the edge and the surface. In reality, the gap is normally closed by fine-grain soil flowing from the active zone. Mesoscale particles that are coarser than the gap cannot represent this. Instead, this is modelled by particles (and any fluid mass) in the vicinity of the cutting edge losing mass to the surface voxels at the gap. This is illustrated in Fig.3. The amount of mass necessary to fill the gap is computed using Eq. (21). Particles losing mass shrink correspondingly until they reach the lower size limit, where they are converted to fluidized mass. The process of converting soil from solid mass to particle and fluidised mass, and back, enable cutting and grading with high precision, despite the particle and voxel discretization being much coarser. It should also be noted that the operations for mass conversion and transport are by construction guaranteed to preserve the total mass to machine precision.

The time evolution of the coarse particles, $N_{CP}$, is governed by the equation of motion

$$m_b\dot{v}_b = f^\text{ext}_b + G^T_{bs}\lambda_{sb} + G^T_{bc}\lambda_{cb} + \sum_{b'\in N_{CP}^b} G^T_{bb'}\lambda_{bb'}$$

(24)

and mass balance at voxel $i$ containing the particles $N_{CP}^i$

$$\sum_{b\in N_{CP}^i} m_b = -\dot{m}'_i - r'_i \|\dot{m}_b\| \leq \frac{3m_b}{d_b} d'_\text{max}, \ d \in [d_{\text{min}}, d_{\text{max}}].$$

(25)

where $G^T_{bs}\lambda_{sb}$ is the contact force from the cutting body and $G^T_{bc}\lambda_{cb}$ from the terrain surface. The change in solid mass, $\dot{m}'_i = w_i\psi_i\rho_b V_0$, is linked to the change in the surface heightfield imposed by the moving cutting body. The change in particle mass depends on the change in solid mass, but is constrained by a maximum particle growth rate, $d'_\text{max}$, and size range. If the change in solid and particle mass does not match, the residual mass is converted to or from fluidized mass at rate $r'_i$, provided $m'_i \geq 0$. Since the contact model is scale-invariant by construction, the pre-calibrated parameter $p_p$ will give the mesoscale model the desired bulk mechanical properties $p_b$. Particles that come to rest outside the active zone are converted to solid mass, whereby the surface heightfield increases correspondingly. Soil that has undergone rapid flow tend to be more loosely packed than the original bank state. To model this, each soil is assigned a swell factor $S = \rho_b / \rho_\text{min}$. Voxels with mass converted from particles to solid mass are thus given the compaction value $w_i = 1/(1 + s)$ and the local solid mass density $\rho^*_i = \rho_b / (1 + S)$. The effect on the soil’s strength by the change in compaction is addressed below.

We assume that contacting blunt bodies cause soil compaction if the subsoil stress reaches critical values. Since the displacements are small it is not necessary to resolve these deformations using particles. Instead, we operate directly on the local compaction of solid mass $w_i$. The contact forces $f_k = G^T_{bs}\lambda_{sc}$ between the blunt body $B$ and the terrain surface $C$ are obtained from the macroscale model. The forces act at some contact points $x_k$ that enclose a contact patch of area $A_{bsc}$. The subsoil stress $\sigma_i$ in a voxel $i$ at depth $z_i$, and the resulting plastic deformation, can be estimated from analytical solutions and numerical extensions [42,43]. As a first order approximation following [44], we use the model, $\alpha_i = \sigma_{bsc} \left[1 - \left(z_i / \sqrt{A_{bsc} + z_i^2}\right)^3\right]$, for a circular distributed normal load, $\sigma_{bsc} = \sum_k f_k / A_{bsc}$, on a semi-infinite elastic solid. The compaction can be estimated from the compression index in Eq. (5) and by noting the relation to the void ratio $w \approx 1 + e_b - e$. 
Noting that the void ratio and compaction are related by \( w \approx 1 + e_b - e \) we obtain the following model for the compaction in a voxel \( i \) due to subsoil stress at a depth \( z_i \) beneath the contact

\[
w_i = 1 + C_c \ln(\sigma_i/\sigma_b),
\]

with clamping to the set bounds \( w_i \in [w_{\text{min}}, w_{\text{max}}] \). Typical values for the compression index range between 0.001 (dense sand) and 1 (soft clay) and we use \( \sigma_b = 1 \text{kPa} \) as the consolidation stress for the bank state. A change in compaction alters the mass density in the voxel to \( \rho_i = \rho_b w_i \) and mass is shifted downwards column wise to preserve the total mass. This reduce the fill ratio of the surface voxels at the contact and the surface heightfield is reduced correspondingly. The increased level of compaction also makes the material stiffer and stronger. From Eqs. (17) and (18) we obtain the following models for the bulk elasticity, dilatancy angle and effective angle of internal friction

\[
E_{\text{bulk}}^I = E_{\text{bulk}}^0 (1 \pm k_E \| w_i - 1 \| n_E),
\]

\[
\psi_i = c\psi (w_i - w_c),
\]

\[
\phi_i = \phi_b + \psi_i,
\]

where the stiffening parameters \( k_E, n_E \) and have default values 1 and 0.5, respectively, but can be tuned to represent more complex soil. The parameter \( c\psi = \partial\psi/\partial w \) control the rate of hardening, with 0.1 \(-\) 1.0 radians being a typical range. The critical compaction, \( w_c \), determines whether the soil at bank state is expanding or compressing under shear. It is related to the bank state dilatancy angle \( \psi_b \) by \( w_c = 1 - \psi_b / c\psi \) The compacted state is permanent until the soil is disturbed, e.g., by earthmoving equipment interacting with it.

** Macroscale model**

The macroscale model focus on the rigid multibody dynamics of the equipment, \( B \). The resting terrain is perceived as a surface heightfield, \( C \). In each active zone, evolved in time by the mesoscale model, the soil is aggregated into a single body with the inertia, centre of mass and velocity of the particles and fluidized mass. These aggregate bodies, labelled \( A \), have the degrees of freedom of rigid bodies but finite mechanical strength. This is accomplished by compliant frictional-cohesive contact constraints at the interfaces of the aggregate to the terrain surface, which currently coincide with the failure plane, and to the equipment. The equations of motion are

\[
M_A \ddot{\mathbf{v}}_A = \mathbf{f}_{\text{ext}}^A + \mathbf{G}_{\text{AA}}^T \lambda_{\text{AA}} + \mathbf{G}_{\text{AC}}^T \lambda_{\text{AC}},
\]

\[
M_B \ddot{\mathbf{v}}_B = \mathbf{f}_{\text{ext}}^B + \mathbf{G}_{\text{BA}}^T \lambda_{\text{BA}} + \mathbf{G}_{\text{BC}}^T \lambda_{\text{BC}} + \mathbf{G}_{\text{BD}}^T \lambda_{\text{BD}},
\]

where the constraint forces \( \mathbf{G}_{\text{AA}}^T \lambda_{\text{AA}} \) and \( \mathbf{G}_{\text{AC}}^T \lambda_{\text{AC}} \) act on the aggregate from the equipment and the terrain failure surface, respectively. The equipment is held together and actuated by \( \mathbf{G}_{\text{BA}}^T \lambda_{\text{BA}} \). The separation resistance and the inertia of the soil in the active zone is mediated by \( \mathbf{G}_{\text{BC}}^T \lambda_{\text{BC}} \). Through \( \mathbf{G}_{\text{BD}}^T \lambda_{\text{BD}} \), the equipment experience direct contact with the terrain surface, e.g., tyres, tracks or the exterior of the bucket. Soil cutting objects may also be subject to a penetration resistance force \( \mathbf{G}_{\text{BD}}^T \lambda_{\text{BD}} \). See Fig. 2 for illustration of the interaction interfaces.
Each mesoscale active zone, $A$, is aggregated to a macroscale body with the following mass, inertia tensor, centre of mass, linear and angular velocity

$$M_A = \sum_{a \in N_A^p} m_a + \sum_{i \in N_A^v} m_i,$$

$$I^\alpha_{A\beta} = \sum_{a \in N_A^p} m_a \left( \| x_a - x_A \| \delta_{\alpha\beta} + x_a^\alpha x_a^\beta \right) + \sum_{i \in N_A^v} m_i \left( \| x_i - x_A \| \delta_{\alpha\beta} + x_i^\alpha x_i^\beta \right),$$

$$x_A = M_A^{-1} \sum_{a \in N_A^p} m_a x_a + M_A^{-1} \sum_{i \in N_A^v} m_i x_i,$$

$$v_A = M_A^{-1} \sum_{a \in N_A^p} m_a v_a + M_A^{-1} \sum_{i \in N_A^v} m_i v_i,$$

$$\omega_A = I_A^{-1} \sum_{a \in N_A^p} m_a (x_a - x_A) \times v_a + I_A^{-1} \sum_{i \in N_A^v} m_i (x_i - x_A) \times v_i,$$

where $N_A^p$ and $N_A^v$ denote the set of particles and voxels with fluidized mass in active zone $A$.

The contacts aggregated from the $AB$ and $AC$ interfaces are turned into the following velocity constraint

$$0 \leq \gamma_n \lambda_n + G_n v \perp (\lambda_n + f_c) \geq 0, \quad f_c \equiv c_b A,$$

$$\gamma_t \lambda_t + G_t v = 0, \quad \| \lambda_t \| \leq \mu \| \lambda_n \|.$$

Eq. (37) prevent relative motion in the normal direction, i.e., compressive or tensile deformations. Note that, unlike model Eq. (10), this model does not include a contact overlap. This make the aggregate viscoplastic in nature, with no sense of any reference configuration. The viscous damping is controlled by the parameter $\gamma_n = \Delta t / (c_{as} E_b)$, where $c_{as}$ is the aggregate’s damping coefficient (referred to as aggregate stiffness multiplier in the implementation). If the force is tensile and reach the cohesive force limit, given by $c_b A$, separation may occur at the interface. Eq. (38) prevent sliding motion in the failure surface and in the interface to the equipment, unless the forces reach the Coulomb condition. The friction coefficient in the failure surface is set to the soil’s internal friction $\mu = \mu_t$. At the equipment interface the friction coefficient is set by the tool’s surface friction $\mu = \mu_{tool}$.

The contact points between the aggregate body $A$, the equipment $B$ and the failure surface $C$ is the reduced set of contact points from the mesoscale model that best approximate the two contact areas, $A_{AB}$ and $A_{AC}$, with four contact points each. This is illustrated in 2D in Fig. 2.

If the cutting edge has a thickness or is equipped with teeth, there may be a significant penetration force, $G_{TBD}^T \lambda_{BD}$, in addition to the separation force. This is modeled with the following velocity constraint

$$G_{TBD} v_s = 0, \quad G_{TBD}^T \lambda_{BD} \leq f_{pt} \equiv n_t [p_t + (c_t + p_t \mu_t) / \tan \theta_t] A_t.$$
where $G_{Bd} = t_B$ is a unit vector for the direction of penetration, e.g., pointing in the direction of the teeth. For penetration to occur, the tool must overcome the force limit $f_{pt}$ [45], which depend on the tool’s friction coefficient $\mu_t$ and adhesion $c_t$, and on the number of teeth $n_t$, with cross-section area $A_t$, tooth angle $\theta_t$. The tooth pressure, $p_t$, is modelled using the finite cavity expansion theory [31]. In the plastic limit the tooth pressure become $p_t = p_0 + \left[ \Upsilon + (\alpha - 1)p_0 \right] / (2 + \alpha)$ under cylindrical expansion, where $\Upsilon = 2c_t \cos \phi / (1 - \sin \phi)$, $\alpha = (1 + \sin \phi)/(1 - \sin \phi)$. The mean pressure is computed by $p_0 = \frac{1}{2} \rho gz \left[ (1 + K_0) + (1 - K_0) \cos(2\beta) \right]$, taking the lateral earth pressure and tool inclination into account [45], where $K_0$ is the coefficient of lateral earth pressure, $\rho$ is the specific soil mass density, $z$ is the penetration depth and $\beta$ is the insertion angle. A simple model for the coefficient of lateral earth pressure is given by Jaky [46] as $K_0 = 1 - \sin \phi$.

**Complex digging tools**

A bucket used for excavation or wheel loading is more complex than a blade. It can be seen as a bottom plate, with a curved back wall and sidewalls that allow material to accumulate in the bucket. This deadload form a secondary separation plate, at an angle much steeper than the bottom plate. This affect the stress distribution, the shape of the active zone, by Eq. (22), and ultimately also the digging resistance. Furthermore, the forces from the soil that surrounds the bucket act both tangentially (friction) and normally. This cause additional digging resistance as well as strong resistance for motion in the lateral directions. Buckets are also used for other purposes than digging. It is a common operation to do surface leveling and compaction using the bucket exterior. To support these features we treat a bucket as a composite model of an elementary digging tool and a set of soil deformers as described below.
A digging tool has a cutting edge $e_c$, a parallel top edge $e_t$, and a penetration direction $t_c$ orthogonal to these edges and to the normal of the bottom plate $n_c$. This is the primary separation plate of the digging tool. See Fig. 4 for an illustration. A convex digging tool has an inner shape that is the void enclosed by the cutting edge, top edge and the concave tool surface connecting them. The face that connect the cutting edge and the top edge has normal $n_{tc}$. If the bucket is full, this face act as a secondary separation plate. If the digging tool is a simple blade there is no inner shape and $n_c = n_{tc}$. For concave tools we track the amount of material accumulated in the inner shape (deadload) and do linear interpolation of the secondary separation plate $n_{fill}$ between $n_c$ and $n_{ct}$. This changes the angle $\beta$ in Eq. (22) and thus the slope of the failure surface.

A soil deformer is simply a separation plate with no penetration resistance and no inner shape. It is defined by a deformation edge $e_d$ and a parallel top edge $e_t$. They form a face with normal $n_d$. The active zone from a soil deformer is not automatically resolved using particles, unless the velocity of the deformation edge exceed a given threshold. This avoid using particles to simulate the terrain when it remains at steady state.

A bucket can thus be represented with soil deformers for the exterior faces and with a digging tool for the cutting blade and the bucket interior. Figure 4 shows a bucket with two side walls, each assigned a soil deformer. Together with the digging tool, this gives rise to a total of three active zones, which are discretised in the lateral direction. The backside of the bucket can also assigned a deformer, in which case the number of active zones sum to four. If not, the soil cannot be displaced by the backside of the bucket other than by pure compression.

The aggregate normal forces is modeled by Eq. (37). It is a velocity constraint, which are prone to numerical drift. The drift has no significance in soil cutting but prohibit proper resistance of soil deformers resting or being pressed onto a soil bed. A drift will cause them to sink unnaturally into the soil. This is remedied by adding a stabilizing perturbation $\xi g_n$ to the constraint, i.e., $\gamma_n \lambda_n + G_n v \rightarrow \gamma_n \lambda_n + \xi g_n + G_n v$, which turn it into a ordinary contact constraint. Transitioning between the velocity constraint (providing smooth soil cutting) and contact constraint (avoiding artificial sinkage) is achieved by making the stabilization coefficient $\xi$ a function of the orientation of the deformer relative to the terrain surface.

**Algorithm**
The multiscale terrain model can be summarized by the Algorithm 1.
Algorithm 1 Multiscale terrain

1: initialize terrain
2: set surface heightfield \( h(x, y) \) and bank state soil parameters \( p_b = [\phi_b, c_b, \psi_b, E_b] \)
3: set initial voxel states \( s_i = [m_i, v_i, w_i, \phi_i] \)
4: apply cellular automata to ensure the surface heightfield is soil-consistent
5: initialize earthmoving equipment \([x, v, g]\) \( B \)
6: set contact parameters for bodies and terrain surface \([\mu, c, e, E]_{BC}\)
7: set digging tool edges and direction vectors \([e_c, e_t, \theta_c]\) and \([n_t, A_t, \theta_t]\)
8: set deformers \( \rightarrow [e_d, e_t] \)
9: while running simulation do
10: conversion of resting and active soil
11: for each body \( B \) intersecting the terrain \( C \)
12: compute active zones, discretise in wedges with failure angle \( \theta(\phi, \beta) \)
13: convert resting soil in active zones to particle and fluidized mass
14: convert resting particles outside zones to loose solid mass with \( w_i = w_{\text{min}} \)
15: apply cellular automata to re-distribute resting soil violating the angle of repose
16: update the surface heightfield \( h(x, y) \) from the voxel state \( s_i \)
17: do collision detection
18: add contact constraints in the macroscale and mesoscale simulations
19: for each active zone do
20: aggregate body \( A \) from voxels and particles using Eq. (32)-(36)
21: add aggregate velocity constraint (37)-(38) with interfaces \( AB \) and \( AC \)
22: end for
23: penetration resistance
24: estimate pressure on edge/tooth and compute maximum penetration resistance
25: add the penetration constraint Eq. (39)
26: time-step Eq. (7)-(9)
27: solve the macroscale MCP using direct LDLT solver \( \rightarrow [v, \lambda]_B \)
28: solve the mesoscale MCP using iterative PGS solver \( \rightarrow [v, \lambda]_b \)
29: update positions \( x_{i+1} = x_i + \Delta t v_{i+1} \)
30: soil compaction
31: estimate the sub-soil stress from the surface contact forces \( BC \)
32: update the soil compaction \( w_i \) using Eq. (26)
33: displace resting solid mass vertically in compacted voxel columns
34: update the soil strength parameters \( p(w) \)
35: end while

Implementation

The multiscale model was implemented in the C++ physics engine AGX Dynamics [47]. It supports real-time simulation of rigid multibody and particle systems with contacts, articulation joints and non-smooth dynamics. Numerical time integration is made with the SPOOK stepper. A block-sparse LDLT solver with pivoting [48] is used as direct solver for the macroscale model, and for the equipment subsystem in the meso- and microscale models. The solutions are exact to machine precision. The dynamics of the contacting particles is solved to lower precision using a projected Gauss-Seidel (PGS), which is accelerated using domain decomposition for parallel processing and warm-starting [14, 15]. The microscale reference simulations are run with 1 ms time-step and 500 PGS iterations. The multiscale model is run with 16.7 ms time-step, for real-time simulation at 60 Hz,
Table 1  A small library of soils with particle parameters pre-calibrated to desired bulk parameters using a triaxial test

| Soil name    | Bulk parameters | Particle parameters |
|--------------|-----------------|---------------------|
|              | $\phi_b$  | $c_b$  | $\psi_b$ | $\mu_t$ | $\mu_r$ | $c_p$ |
| gravel-1     | 44°  | 0  | 11°  | 4.6  | 0.4  | 0.3  | 0  |
| sand-1       | 39°  | 0  | 9°   | 4.5  | 0.3  | 0.1  | 0  |
| sand-2       | 32°  | 0  | 8°   | 3.5  | 0.3  | 0.02 | 0  |
| wet-sand-1   | 33°  | 8.7| 8°   | 4.0  | 0.3  | 0.1  | 30|
| dirt-1       | 40°  | 2.1| 13°  | 6.5  | 0.4  | 0.1  | 2.5|
| dirt-2       | 37°  | 4.9| 10°  | 4.6  | 0.4  | 0.1  | 12.7|
| dirt-3       | 35°  | 21| 10°  | 5.8  | 0.4  | 0.1  | 63.7|
| fs-strong    | 42°  | 0  | 13°  | 3.8  | 0.5  | 0.1  | 0  |
| fs-weak      | 35°  | 0  | 8°   | 2.8  | 0.3  | 0.05 | 0  |
| cfs-strong   | 32°  | 7.1| 7°   | 4.6  | 0.3  | 0.05 | 23|
| cfs-medium   | 32°  | 4.3| 7°   | 4.2  | 0.3  | 0.05 | 12|
| cfs-weak     | 21°  | 4.8| 6°   | 3.0  | 0.15 | 0.025| 23|
| cfs-weakest  | 12°  | 2.4| 1°   | 1.0  | 0.06 | 0.01 | 23|
| cs-weak      | 5°   | 3.1| 2°   | 0.7  | 0.0  | 0.0  | 50|
| frictionless | 6°   | 0  | 1°   | 0.25 | 0.0  | 0.0  | 0  |

Units deg kPa deg MPa kPa

and with 10 PGS iterations for the mesoscale particle model. This corresponds to an error tolerance of about 1% and 10%, respectively, according to the model in [14] and for the test systems in this paper. The voxel data representations and operations are implemented using the OpenVDB library [49]. It is optimized for large, sparse, time-varying volumetric data discretised on a 3D grid and support hierarchical representations.

Soil library

It is important that the bulk and particle representations are dynamically consistent. That means that the particle parameters, for each soil, need to be calibrated to the values that produce the desired bulk properties, e.g., internal friction, cohesion, angle of repose etc. Otherwise they do not describe the same material and there is risk that the conversions between particle and continuum inject energy to the system, which can lead to numerical instability. Therefore, 15 soils were calibrated in advance to the bulk mechanical properties. These are listed in Table 1. The bulk parameters, at bank state, were chosen from tabulated values for different materials, e.g., gravel and sand. To narrow down the particle parameter search space, we were guided by friction measurements of sand grains. The particles are given the Young’s elasticity of $E = 10^9 Pa$, Poisson’s ratio 0.15, specific mass density 2200 kg/m$^3$ and coefficient of restitution 0. Since there may exist many types of sand and gravel (with different distributions of size, shape, microscopic friction, packing density and moisture level), each with distinct bulk properties, an integer is added to the name to distinguish between them. We also created a set of purely frictional soils (fs) and cohesive-frictional soils (cfs) with no particular real soil in mind. For testing purposes a frictionless and a cohesive-frictionless soil (cs) were also created, but they are not expected to be of practical use.

A (virtual) triaxial test was used for parameter calibration, following the procedure and setup in [25], but with a lesser packing density. The model of the triaxial cell consists of frictionless rigid bodies for walls, 250 mm wide, driven by motors that are serv-
controlled to maintain a given normal stress, \( \sigma_i, i = 1, 2, 3 = xx, yy, zz \). The particle samples have uniform size distribution with diameter ranging between 27 and 33 mm, and initial porosity 0.42. The strength of each soil is tested by driving the horizontal walls at a vertical speed of 5 mm/s while controlling the vertical walls to maintain a given consolidation stress \( \sigma_2 = \sigma_3 \). Tests are performed with different consolidation stresses in the range from 1 kPa and 75 kPa. The motion of the walls and the stresses on them are registered during the simulation. Example measurements of the stress deviator, \( \sigma_1 - \sigma_3 \), as a function of the lateral strain are shown in Fig. 5 for the materials gravel-1 and wet-sand-1. The internal friction and bulk cohesion were determined from the Mohr’s circles at failure stress, see Fig. 6. The procedure was repeated, for each soil and consolidation stress, two or three times with different particle initial states. The results are found in Table 1. Due to fluctuations and uncertainty in the Mohr method, the cohesion-free soils show an apparent cohesion of up to 2 kPa at most. In these cases, the bulk cohesion is explicitly set to zero. From the triaxial tests we also measure bulk elasticity as the secant modulus, \( E_b = \Delta \sigma_1 / \Delta \epsilon_1 \) halfway before failure, and the dilatancy angle from \( \psi = \arcsin [\Delta \text{tr} \epsilon / 3 \Delta \| \vec{\epsilon} \|] \). The dilatancy angles are computed from the strain curves in Fig. 7. These bulk parameters are determined for a medium consolidation stress of 15 kPa. The procedure was repeated for all the 15 soils in Table 1.

**Simulation tests**

The multiscale model was tested and evaluated using two primary test systems, which are bulldozing and excavation in a flat soil bed. For reference, the same situations
were performed using the microscale model. The digging resistance, resulting terrain surface heightfield and computational performance of the two models was then compared. All materials in Table 1 were subjected to tests, see the Additional file 1: Video S1 and Additional file 2: Video S2. The results for three frictional soils (gravel-1, sand-1, frictionless) and three cohesive soils (dirt-1, wet-sand, cfs-weak) are reported more detail. In the microscale simulations the soil bed consisted of 50e3 particles, with size uniformly distributed between 55 and 45 mm and specific mass density of 2200 kg/m$^3$. The voxel size in the multiscale model was 0.1 m, the specific mass density also 2200 kg/m$^3$ but the bulk mass density was set to 1474 kg/m$^3$ which correspond to a void ratio of 0.33. Default values for the aggregate stiffness multiplier was set to 0.001 and 1.0 at the terrain and tool interface, respectively.

In the bulldozing test, shown in Fig. 8, a blade cuts a bed of sand-1 soil and pushes the material in front of it. After some distance, the material is dumped in a pile, as the blade is stopped, lifted, and reversed. The cutting depth is 0.05 m and the length is 5 m. The blade has mass 100 kg, sectioned vertically in two plates, 1.6 m wide, 0.37 m tall, 0.02 m thick, and with a relative angle of 55°. It is attached with a lock constraint to a kinematic
body that is driven with horizontal speed of 0.5 m/s. See the Additional file 3: Video S3 for bulldozing in sand-1 and dirt-1. For the penetration resistance, the blade’s edge is discretized by 80 teeth with 10 mm maximum radius and length, and with 2.5 mm minimum radius. The penetration force scaling is then calibrated to 10 for all materials except for gravel, dirt-1 and wet-sand which are given the value 20.

Figures 9 and 10 show the force from the terrain on the blade as a function of the horizontal position $x$. The force from the multiscale and reference model largely follow each other, especially during the intermediate phase, starting after the blade is lowered into the ground ($x = -2.75$ m) and ending where the blade begins to rise ($x = 1.5$ m). In the intermediate phase, the horizontal force grows gradually as soil accumulates in front of the blade, until material start spilling off at the same rate. The vertical force is nearly constant in this phase. The bulldozing resistance for gravel is larger than for sand and for
The digging resistance from bulldozing a bed of sand-1 soil simulated using the multiscale model. The force is divided in the contributions from the penetration and separation models.

Fig. 11

It can be concluded that the penetration force is overestimated during the phase of raising the blade. It dominates the vertical component of the force. The resting and densely packed soil is forced to expand at the tip of the blade. When the blade is raised, the cutting edge moves through the shear zone, which should be looser and offer lesser resistance to penetration than dense soil. This is automatically captured by the resolved reference model but not in the multiscale model. In all cases, the pile of accumulated material in front of the blade is higher and steeper with the reference model than with the multiscale model. This is consistent with the larger time-step and lower PGS solver iteration count in the multiscale model, which imply larger numerical errors that manifest themselves as excessive slipping and rolling in the particle contacts. However, since the soil’s bulk properties have been calibrated in advance the errors affect the size of the aggregate but not on its fundamental shape or strength. In a sense, these mesoscale errors are filtered out in the aggregation process to the macroscale model and do not propagate into the digging resistance. As expected, the digging resistance is much smaller for frictionless soil. It is smaller for the multiscale model than for the reference model. This can be understood by the fact that particles in the multiscale model slide over a frictionless plane while the motion of the particles in the reference model are damped by the dissipative contacts (zero restitution) with the irregular surface formed by the resting particles.

The resulting surface height profiles, after a completed bulldozing cycle, are shown in Figs. 12 and 13. The multiscale and reference models are in good agreement for gravel-1 and sand-1. The depth of the cut strips agrees within 10 millimetres and the dimensions of the side berms are agree by roughly 10%. The deviations in height profiles for the frictionless soil have the reason that is explained above. For the cohesive soils the
deviation between the models is larger. The reference model with dirt-1 give wider side berms and a wider pile. The relative error is up to 50%. For the strongly cohesive soils (wet-sand-1 and cfs-medium) the difference is even larger, up to 100%. For the reference model almost all soil is accumulated in front of the blade. The particle cohesion appear to be much stronger in the reference model than with the multiscale model.

The excavation test with dirt-1 can be seen in Fig. 14 and in the Additional file 4: Video S4. The excavator model consists of four dynamic bodies and 4 joints: a bucket on the end of an articulated arm, with an outer and inner boom, attached to a revolving base on a static foundation. The test follows the setup of [24] where only the boom articulation joint is actuated, i.e., the inner arm is held fix. The bucket is 1.2 m wide, 1.2 m long, 0.9 m high. It weighs 100 kg and has a thickness of 0.05 mm. The outer arm is 3.24 m long and weighs 300 kg. A hinge motor drives the arm to rotate at target speed 0.1 rad/s which translates to 0.4 m/s at the bucket cutting edge. For the penetration resistance, the blade’s edge is discretized by 24 teeth with 50 mm maximum radius and length, and 10 mm minimum radius. The penetration force scaling is then calibrated to 2 for all materials except frictionless soil for which it is set to zero. The force exerted on the bucket during the digging can be seen in Fig. 15 and 16. Observe that the bucket moves in the opposite direction to the bulldozing test, i.e., from right to left. The bucket start penetrating the soil and fill the bucket at $x = 2.0$. The horizontal force peak around $x = 0$ m. Up to this point,
the force in the vertical direction has been dominated by the penetration resistance, see Fig. 14. The weight of material inside and in front of the bucket increase gradually, as do the resistance to shearing the soil along the failure plane. From $x = 0.5$ m to $x = 0$ m these are in balance and after this the soil weight and the shear resistance dominate the vertical component of the digging force. At $x = -1.0$ m the bucket breaks loose from the ground. The weight of the material in the bucket can be observed as the residual vertical force $F_z$ at $x = -2.5$. There is good agreement between the models for the peak horizontal force and the residual vertical force. Overall, the time-averaged forces agree within 25%. There is a trend for the multiscale model to overestimate the vertical digging resistance during breakout around $x = -0.5$ m. From Fig. 17 it is not clear whether the penetration force or the separation force is to blame. As in the bulldozing test, there are significant deviations between the models for the strongly cohesive soils, in particular wet-sand-1. With the reference model, the force fluctuates heavily between $x = 1.0$ m and $x = -1.0$ m, see
Fig. 16. Force resistance from excavation using three cohesive soils and the microscale model, with snapshot taken at 5 s.

Fig. 17. The digging resistance from excavation in a bed of dirt–1 soil simulated using the multiscale model. The force is divided in the contributions from the penetration and separation models.

The reason for this is the strong cohesion. It prevents the soil from flowing freely and gradually fill the interior of the bucket. Instead, a strong soil beam is formed. This can be seen in the cross-section image in Fig. 16 for wet–sand. The force fluctuations grow large when the beam hit the interior back wall of the bucket. The beam starts to compress, buckle and fail in an irregular manner. The aggregate in the multiscale model does not represent such modes of deformations and failure and does not produce large force fluctuations. The height profiles after an excavation cycle are shown in Figs. 18 and 19. For gravel–1 and sand–1 the width and height of the trench, side berms and the pile are in relatively good agreement. The depth of the trench agrees within 10 mm and the other features within 10% on average. For dirt–1 the side berms and pile is roughly 50% larger for the reference model, and for wet–sand and cfs–weak the pile is more than twice as tall. The behaviour of the different soil materials under excavation is shown in Additional file 2: Video S2. It is clear that the strongly cohesive soils are much more cohesive in the reference model compared to the multiscale model.

The active zone model, Eq. (22), is validated using the reference model. The induced particle motion when digging in flat and sloping soil of the type dirt–1 with an excavator bucket can be seen in Fig. 20 and in the Additional file 5: Video S5. The shape of the wedge formed by the mobilized soil is indicated. The angle between the dynamic separation plate...
Fig. 18 The resulting surfaces from excavation in three frictional soils using the multiscale model (top row) and the microscale reference model (middle row). The difference is shown at the bottom row.

Fig. 19 The resulting surfaces from excavation in three cohesive soils using the multiscale model (top row) and the microscale reference model (middle row). The difference is shown at the bottom row.

Fig. 20 Analysis of the active zone when digging in a flat (left) and sloped (right) soil bed with a complex bucket. The particles are colour coded by their velocity, with blue to red ranging from 0 to 1 m/s. See Additional file 5: Video S5.

and the soil surface is estimated to \( \beta_{\text{flat}} = 42^\circ \) and \( \beta_{\text{sloped}} = 45^\circ \) in the two configurations. Eq. (22) imply failure planes with inclination \( \theta_{\text{model flat}} = 47^\circ \) and \( \theta_{\text{model sloped}} = 45^\circ \), respectively. From the reference model we estimate the failure angles at \( \theta_{\text{ref flat}} = 53^\circ \) and \( \theta_{\text{ref sloped}} = 40^\circ \), respectively, which correspond to a 12% error.

We analysed also the computational speed of the multiscale model and the reference model, aware that the results depend on implementation, optimization efforts and the
Table 2  Performance analysis of the excavation test involving $N^{rb}$ rigid bodies and $N^{p}$ particles

|        | $N^{rb}$ | $N^{p}$ | $N^{rb}_{eq}$ | $N^{p}_{eq}$ | Time-step [ms] | Real-time |
|--------|----------|---------|---------------|--------------|----------------|-----------|
| Multiscale | 8        | 950     | 85            | 1.8e4        | 16.7           | 1.5       |
| Reference | 4        | 50e3    | 48            | 1.2e6        | 500            | 1         |
| Ratio   | 2        | 0.019   | 1.77          | 0.0015       | 0.02           | 16.7      |
|         |          |         |               |              | 1500           | 0.001     |

The direct MCP solver process $N^{rb}_{eq}$ equations, and the iterative solver $N^{p}_{eq}$ equations. The multiscale model is 1500 times more efficient due to the smaller number of equations for the iterative solver, fewer iterations, and larger time-step.

Fig. 21  A wheel loader digging in a step pile of soil. The resistance is too large for breaking out and instead the rear of the vehicle is raised from the ground. The right image shows the active zones, discretised in soil wedges, and the cohesive-frictional contacts between the aggregates and the terrain. See the Additional file 6: Video S6.

The analysis was performed on the excavation test system shown in Fig. 14 and the result is summarized in Table 2. The number of rigid bodies $N^{rb}$ and particles $N^{p}$ are given for the multiscale model and reference model. The number of equations, for the velocity update and constraint multiplier, are divided by what is treated by the direct solver, $N^{rb}_{eq}$ and by the iterative solver, $N^{p}_{eq}$. The multiscale model has a real-time factor of 1.5, i.e., the computational time is 11 ms per simulation time-step 16.7 ms. That means that there is room for simulating a more complex vehicle at real-time speed. The reference model has a real-time factor of 0.001, meaning that each 1 ms time-step on average take 1 s to compute. In both cases, solving the MCP is what dominate the computational time. The price of introducing the aggregate bodies (the main aggregate and a back deformer) and penetration constraints is an increased number of equations ($N^{rb}_{eq}$) processed by the direct solver, namely 36 additional equations (77% increase). The relative speedup of 1500 is an effect of the 16.7 times larger time-step, 67 times fewer equations for the iterative solver, and 50 times less iterations. To check the sensitivity of the multiscale model to the number of iterations and spatial resolution this was tested, but neither had strong effect on the reaction force.

Finally, the multiscale model is demonstrated in use with full vehicle models in complex earthmoving scenarios. Fig. 21 shows a wheel loader digging in a steep wall of soil. See Additional file 6: Video S6. The active zone in the cutting direction, discretised in five parallel wedges, is visualized in the right image. There are also active zones on each side of the bucket originating from the soil deformers. The different size and inclination of the wedges reflect the nonuniform distribution of mass. The active zone in the digging direction is resolved with particles, but not the side deformers because of low lateral

3The tests were run on a desktop computer with Intel Core i7-7800X CPU at 3.5 GHz and 32 GB RAM.
Fig. 22 Demonstration of variable soil compaction and swelling. The subsoil stress from the tires cause compaction (dark grey). Particles that are converted back to resting solid become more loosely packed (light grey) than nominal bank state.

Fig. 23 Demonstration interaction with other rigid bodies (red rocks) while digging a trench with a tracked excavator. See the Additional file 7: Video S7.

Variable soil compaction is illustrated in Fig. 22 and visualized by the intensity of the grey terrain. Medium grey represents nominal compaction at the bank state, at which the bulk strength parameters are defined. Light grey represent soil that has dilated due to shear deformations, e.g., have been dug or pushed with the bucket. That soil has lower compaction and weaker strength according to the swell factor and dilatancy angle, respectively. Dark grey represent soil that has been compacted, e.g., due to compressive stress from the tires. It has higher compaction and strength than the bank state value. The right images show the wheel loader driving into the left pile of loose soil that is easily compacted.
The involvement of additional rigid bodies interacting both with the earthmoving machine and the terrain is demonstrated in Fig. 23 with a full-sized tracked excavator digging a deep trench. The rocks contact directly with the excavator bucket, via the direct solver, but also with the particles, through the iterative solver. The rocks force the soil to distribute around the rock inside the bucket or pile up around the rocks on the ground. The terrain is initialized with a high state of compaction such that the trench can sustain steep side walls. See the Additional file 7: Video S7.

The capability of representing large terrains, everywhere deformable, is demonstrated with the bulldozing example in Fig. 24 and the Additional file 8: Video S8. Observe that the terrain is cut precisely at the dozer’s cutting blade, with a precision much finer than the coarse particles, and match the vertical motion of the blade. The effect of bulldozer chassis and blade oscillations can be observed as wave pattern in the cut terrain surface.

Discussion

The results show good performance of the multiscale model and mostly fair agreement with the reference model, needing little calibration. There are, however, a number of limitations with the method as well. The most significant deviation occur for strongly cohesive soils, where the particle cohesion appear much stronger in the reference model than in the multiscale model. Furthermore, the shape of the active zone is a crude approximation. It is clear from both the underlying theory and numerical studies that the shear failure surface is not a plane and the active zone is not well approximated by a wedge, or several parallel wedges. The deviation in the digging resistance at breakout indicate that the slope of the failure surface is not correct, or that the assumption of shear band localization is false. The active zone model in the present paper requires some manual work with defining edges, direction vectors, teeth etc. Ideally, the dependency on such geometric features should emerge from a model that merely takes the 3D geometry of a digging tool as input. The damping coefficient of the rigid aggregate has not been derived from the underlying theory or scrutinized using the microscale reference model. An alternative to improving the active zone model by analytical means is to take a data-driven approach. Following Wallin [50], it is possible to train a model, from resolved reference simulations, to rapidly predict the digging resistance and the flow field in a granular bed from a digging tool of particular 3D shape without explicitly defining any cutting edges, direction
vectors or teeth. Potentially this can be used for predicting the shape and mechanical strength of the active zone with higher precision and generality, and possibly also the soil displacements. The drawback of the data-driven approach is the need for running many simulations in advance, covering a wide range of soils, terrain shapes, and tool trajectories to have a useful model.

The computational bottleneck in the tested implementation is the mesoscale particle simulation. It is run synchronously on CPU with the macroscale simulation using fix and identical time-step. This is not necessary but made so for simplicity because the simulation software, AGX Dynamics, is designed for strong coupling between the multibody dynamics and the (nonsmooth) DEM simulation. It might be worth investigating alternative, possibly asynchronous, methods for simulating the active soil, e.g., smooth DEM on GPU.

The mesoscale model support plastic compaction of the solid terrain but not shear deformations. That is a major limitation for simulating deep ruts and vehicles or other objects sinking or being buried in the terrain.

Finally, the presented method relies on a having a soil library where particle parameters are pre-calibrated to match the bulk mechanical parameters. The current library, involving 15 different soils, can easily be extended to a wider range of more soils, e.g., to include the over 100 virtual soil samples that was mapped in [25]. From such data-sets it is possible to identify mapping functions $p_p \rightarrow p_b$, such that new soils can be introduced on-the-fly. That is important for making machine learning models robust and transferable from one domain to another, e.g., from the simulation domain to the physical domain, using domain randomization [51]. However, the current model does not support inhomogeneous soil or mixing of two or more soils. That extension is left for future development.

A weakness in the presented study is that the validation and calibration is not based on field measurements but rely on reference simulations of relatively high resolution. Field experiments are left for future work, for example, using also the methodology in [52] for determining the soil’s bulk parameters using the earthmoving equipment.

It is interesting to also compare the investigated method with other multiscale methods for vehicle-terrain simulation. A method for hierarchical multiscale vehicle-terrain simulation is investigated in [53]. It combines vehicle multibody dynamics with a finite element terrain model, where parallel DEM simulations of relatively small representative volume elements at the quadrature points in place of an assumed elastoplastic constitutive law. This is an attractive alternative to relying on an idealized model for predicting the failure surface around a digging tool, or the sub-soil deformations under tyres, but real-time performance appear remote at the present time.

**Conclusion**

It has been found possible to simulate earthmoving operations in real-time with a model that captures the rigid multibody dynamics of the equipment, the reaction forces from the terrain, and much of its deformations and flow dynamics. With a multiscale model the terrain’s active zones are represented simultaneously as a rigid body, as particles and as a continuum. A direct solver is applied to the multibody system for high numerical precision and an iterative solver to the particle system for scalability. The models are made dynamically consistent through a soil library that is calibrated in advance using a reference
model simulated at high-resolution. The performance, realism, and capability to represent a wide range of materials and scenarios make the solution suitable for simulation-based development control and automation of earthmoving equipment.

**Supplementary Information**

The online version contains supplementary material available at https://doi.org/10.1186/s40323-021-00196-3.

Additional file 1: Video S1. Simulated bulldozing in soils of different strength using the microscale reference model. Particles are color coded by velocity.

Additional file 2: Video S2. Simulated excavation in soils of different strength using the microscale reference model. Particles are color coded by velocity.

Additional file 3: Video S3. Simulated bulldozing in dirt-1 using the microscale reference model (top) and the microscale reference model (bottom).

Additional file 4: Video S4. Simulated excavation in dirt-1 using the microscale reference model (top) and the microscale reference model (bottom).

Additional file 5: Video S5. Simulated excavation in dirt-1 using the microscale reference model and a full-scale excavator. Particles are color coded by velocity.

Additional file 6: Video S6. Simulation of a wheel loader digging in a step pile of soil using the multiscale model. The terrain heightfield and active zones, discretised in soil wedges, and the cohesive-frictional contacts between the aggregates and the terrain are visualized.

Additional file 7: Video S7. Simulation of an excavator digging a trench using the multiscale model with the presence of dynamic rocks.

Additional file 8: Video S8. Simulation of high-precision bulldozing in a large terrain using the multiscale model.

**Acknowledgements**

Anders Backman, Viktor Knutsson, Fredrik Nordfelth and Martin Nilsson at Algoryx Simulation has made important contributions in the software implementation, testing and optimization.

**Authors' contributions**

The theory and algorithms was developed jointly by the authors. The implementation was made primarily by TB and SN. The simulation analysis was carried out primarily by MS. The authoring of the paper was lead by MS.

**Funding**

Open access funding provided by Umeå University.

**Author details**

1Department of Physics, Umeå University, 90187 Umeå, Sweden, 2Algoryx Simulation AB, Kuratorvägen 28, 90736 Umeå, Sweden.

**Received:** 1 November 2020  **Accepted:** 1 April 2021

**Published online:** 12 May 2021

**References**

1. Bender J, Erleben K, Trinkle J. Interactive simulation of rigid body dynamics in computer graphics. Comput Graph Forum. 2014;33(1):246–70.

2. Haug EJ. Computer aided kinematics and dynamics of mechanical systems, basic methods, vol. 1. Boston: Allyn and Bacon, 1989.

3. Fang Y-G, Li B. Multiscale problems and analysis of soil mechanics. Mech Mater. 2016;103:55–67. https://doi.org/10.1016/j.mechmat.2016.09.003.

4. Li X, Moshell JM. Modeling soil: realtime dynamic models for soil slippage and manipulation. In: In computer graphics proceedings, annual conference series; 1993. p. 361–8.

5. Park B. Development of a virtual reality excavator simulator: a mathematical model of excavator digging and a calculation methodology. PhD thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia; October 2002.

6. Reece AR. The fundamental equation of earth-moving mechanics. In: Proc. of the Institution of Mechanical Engineers; 1964, p. 179.

7. McKyes E. Soil cutting and tillage. Developments in agricultural engineering. Amsterdam: Elsevier; 1985.

8. Holz D, Beer T, Kuhlen T. Soil deformation models for real-time simulation: a hybrid approach. In: Prautzsch H, Schmitt A, Bender J, Teschner M, editors. Workshop in virtual reality interactions and physical simulation "VIPHYS" (2009), 2009. https://doi.org/10.2312/PE/viphys/viphys09/021-030.

9. Holz D, Azimi A, Teichmann M, Mercier S. Real-time simulation of mining and earthmoving operations: a level set-based model for tool-induced terrain deformations. In: Proceedings of the 30th ISARC, Montréal, Canada, 2013.
36. Wang D, Servin M, Berglund T, Mickelsson K-O, Rönnbäck S. Parametrization and validation of an nonsmooth discrete element approach to granular matter. Int J Numer Meth Eng. 2014;97:878–902. https://doi.org/10.1002/nme.4612.

37. Wang D, et al. Warm starting the projected gauss-seidel algorithm for granular matter simulation. Comput Part Mech. 2016;3(1):43–52.

Andreetto B, Forterre Y, Pouliquen O. Granular media. Between fluid and solid. 1st ed. Cambridge: Cambridge University Press; 2013.

Antypov D, Elliott JA. On an analytical solution for the damped hertzian spring. Europhys Lett. 2011;94(5):50004.

Filla R. Quantifying operability of working machines. PhD thesis, Linkoping University. Linkoping University, Fluid and Mechatronic Systems, The Institute of Technology; 2011.

Bouchaud J, Cates M, Prakash JR, Edwards S. A model for the dynamics of sandpile surfaces. Journal De Physique. 1994;I(4):1383–410.

The Eurographics Association; 2014. https://doi.org/10.2312/vrphys.20141232.

Filla R. Quantifying operability of working machines. PhD thesis, Linkoping University. Linkoping University, Fluid and Mechatronic Systems, The Institute of Technology; 2011.

Daviet G, Bertails-Descoubes F. A semi-implicit material point method for the continuum simulation of granular materials. ACM Trans Graph. 2016. https://doi.org/10.1145/2897824.2925877.

Estrada N, Azima E, Radjai F, Taboada A. Identification of rolling resistance as a shape parameter in shared granular media. Phys Rev E. 2011;84:011306. https://doi.org/10.1103/PhysRevE.84.011306.

Wang D, Servin M, Berglund T, Mickelsson K-O, Rönnbäck S. Parametrization and validation of a nonsmooth discrete element method for simulating flows of iron ore green pellets. Powder Technol. 2015;283:475–87. https://doi.org/10.1016/j.powtec.2015.05.040.

Lacoursière C, Linde M. Spook: a variational time-stepping scheme for rigid multibody systems subject to dry frictional contacts. UMINF report from Umeå University. 2011;11(9). https://webapps.cs.umu.se/uminf/index.cgi?year=2011&number=9.

Murty KG. Linear complementarity. Linear and nonlinear programming. Heidelberg: Helderman-Verlag; 1988.

Roux S, Radjai F. Texture-dependent rigid-plastic behavior. In: Luding S, Hovi J-P, Herrmann HJ, editors. Physics of dry granular media. Berlin: Springer; 1998. p. 229–36.

Bouchaud J, Cates M, Piashash JR, Edwards S. A model for the dynamics of sandpile surfaces. Journal De Physique. 1994;I(4):1383–410.

Mitarai N, Nori F. Wet granular materials. Adv Phys. 2006;55:1–45.

Keller T, Défossez P, Weisskopf P, Arvidsson J, Richard G. Soilflex: a model for prediction of soil stresses and soil compaction due to agricultural field traffic. In: Analysis of analytical approaches. Soil Tillage Res. 2007;93(2):391–411. https://doi.org/10.1016/j.still.2006.05.012.

Zhu Y, Chen X, Owen GS. Terramechanics based terrain deformation for real-time off-road vehicle simulation. In: International symposium on visual computing; 2011. Springer. p. 431–40.

Madsen J. Mobility prediction of multi-body vehicle dynamics handling simulations on deformable terrain. Doctoral thesis. University of Wisconsin-Madison; 2014.
45. Bennett N, Walawalkar A, Heck M, Schindler C. Integration of digging forces in a multi-body-system model of an excavator. Proc Inst Mech Eng Part K Multibody Dyn. 2016;230(2):159–77.
46. Jaky J. Pressure in silos. In: 2nd Int. Conf. Soil Mech, vol. 1. Balkema, Rotterdam; 1948. p. 103–7.
47. AGX Dynamics. https://www.algoryx.se/products/agx-dynamics. Accessed: 2020-06-12. https://www.algoryx.se/products/agx-dynamics.
48. Lacoursière C, Linde M, Sabelström O. Direct sparse factorization of blocked saddle point matrices. Para 2010: state of the art in scientific and parallel computing, Reykjavik, June 6–9, 2010.
49. Museth K. Vdb: high-resolution sparse volumes with dynamic topology. ACM Trans Graph. 2013;32(3):27–12722. https://doi.org/10.1145/2487228.2487235.
50. Wallin E, Servin M. Data-driven model order reduction for granular media. Manuscript submitted for publication; 2020.
51. Tobin J, Fong R, Ray A, Schneider J, Zaremba W, Abbeel P. Domain randomization for transferring deep neural networks from simulation to the real world. In: 2017 IEEE/RSJ international conference on intelligent robots and systems (IROS); 2017. p. 23–30.
52. Zhao Y, Wang J, Zhang Y, Luo C. A novel method of soil parameter identification and force prediction for automatic excavation. IEEE Access. 2020;8:11197–207. https://doi.org/10.1109/ACCESS.2020.2965214.
53. Yamashita H, Chen G, Ruang Y, Jayakumar P, Sugiyama H. Hierarchical multiscale modeling of tire-soil interaction for off-road mobility simulation. J Comput Nonlinear Dyn. 2019; https://doi.org/10.1115/1.4042510.

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