ORIGINAL ARTICLE

Computational modelling in sport: a hybrid simulation of the runner as a complex adaptive system

E. Vermeulen[a], S. S. Grobbelaar[a,b], A. Botha[a,c], and K. Nolte[d]

[a] Department of Industrial Engineering, Stellenbosch University, Banghoek road, Stellenbosch, South Africa; [b] DST-NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy (SciSTIP), Stellenbosch University [c] Next Generation Enterprises and Institutions, Council for Scientific and Industrial Research, Meiring Naude road, Pretoria, South Africa [d] Division of Biokinetics and Sport Science, Department of Physiology, University of Pretoria, South Street, Pretoria, South Africa

ARTICLE HISTORY
Compiled May 16, 2021

Abstract
Running-related injuries, specifically overuse injuries, are complex and multifaceted. A different way of thinking is required to fully comprehend why, and how, these injuries occur. The systems thinking perspective offers tools and techniques to capture system-wide interactions in causal, closed-loop structures which may be studied and quantified in a practical way. The value of systems thinking in sport may only realise when pragmatic models follow descriptive, qualitative models. This article instantiates a qualitative, systems thinking perspective of the runner as a complex adaptive system with a hybrid simulation model. The approach is substantiated with principles from physics and physiology and is partially driven by data from a running watch. Results demonstrate that the scientific, reductionist method may be augmented with dynamic, closed-loop thinking and simulation modelling.

KEYWORDS
systems thinking, simulation modelling, running related overuse injury, complexity, running wearables

1. Introduction

The chronicity of overuse[1] injuries in long-distance running remain prevalent, despite extensive research into prevention and management thereof (Barton, 2018). Although reductionist thinking have made great inroads to understand cause-effect relationships between variables (Bittencourt et al., 2016), it fails to provide comprehensive explanations for the multi-scale, system wide, complex interactions behind sport injuries (Bittencourt et al., 2016; Hulme and Finch, 2015; Hulme et al., 2019). The reductionist view analyses complex performance outcomes linked with causal factors in a linear pathway, neglecting varying spatio-temporal scales and isolating risk factors to study its direct effect on outcomes (Bittencourt et al., 2016). However, in living sys-

CONTACT E. Vermeulen. Email: euodiav@gmail.com

[1]Recently, the term *gradual onset* injuries have been used instead of overuse, however in order to build on the momentum regarding systems thinking and running-related injuries in Hulme et al. (2017a, 2019) and Hulme et al. (2018), this article remains with *overuse*

tems, properties found at higher levels are not present in the lower levels (Solomon, Berg, and Martin, 2002). Living systems have subtle interconnectedness that becomes sensible only through systems thinking (Senge, 1990). In sports, for instance, a team structure constitute of positions, tactics, set pieces, leadership, collective experience, amongst others, traits not found in the individual athlete who forms part of the team.

The increasing complexity of and technology integration in sport and recreational systems have spurred new approaches to study and solve problems in these domains. The contributions of ergonomics to the sport and recreation domains were published in a thematic issue in the Theoretical Issues in Ergonomics Science journal (volume 18, issue 4) in 2017, recognising that ergonomics research and practice have roles in the design of systems, optimisation of performance, and prevention of injuries. In that issue, Salmon (2017) decomposed the contributions to physical, cognitive, and systems ergonomics research. Ergonomics is defined as a systems discipline by Wilson (2014), applying a systems philosophy and taking systems approaches to address and solve problems related to human interactions with the elements of their environment.

The systems thinking perspective to complexity in sport domain problems has been endorsed by various authors. The conceptual, complex sports model from Bittencourt et al. (2016) moves away from finding causes, to finding relations between risk factors (the interacting web of determinants) that support emerging injuries. Hulme et al. (2017b) validated a prototype systems-theoretic accident mapping and processes model of the Australian distance running system. The purpose of the prototype was to demonstrate how a systems-theoretic approach to accident analysis may apply to sports injury research. The prototype was presented in Hulme et al. (2017a), as part of the sport and recreation thematic issue of the Theoretical Issues in Ergonomics Science journal. Vermeulen, Grobbelaar, and Botha (2020) presented two tools situated in a systems thinking perspective to address running-related overuse injuries, namely the ice-berg model with an embedded web of determinants, and facilitation of a problem (and intervention) definition by re-articulating the generic system archetypes in a sport context. McLean et al. (2019) showed how to effectively use system archetypes to identify systemic factors that contribute to issues in football coaching, in turn also identifying leverage points for long-term change to improve coaching practices.

In the last few years wearable tracking technologies that focus on physiological and physical activity metrics to monitor the performance during running in the real world have assimilated into the running community. For the purpose of this study, the wearable technologies used in running will be referred to as the running wearable to include smart watches, activity trackers and running watches. These quantified-self athletes now present researchers with the opportunity to expand the investigation of the runner's performance in *their* (real) environment, outside of limited and unrealistic laboratory-like conditions or controlled experimental settings (Napier, Esculier, and Hunt, 2017; Passfield and Hopker, 2016). The athlete interacts with the environment, other runners on the road, and via the feedback from their running wearable, even with themselves when they make decisions towards their own fitness goals. They modify their behaviour for optimal performance based on the outcomes from these interactions.

After the thematic publication considering ergonomics in sport and recreation by the Theoretical Issues in Ergonomics Science journal in 2017, Hulme et al. (2017a) collaborated with others and followed through on their theoretical work on complex systems thinking in studying running-related overuse injuries with practical, computational applications. Hulme et al. (2019) constructed a proof-of-concept system dynamics model for running injury development on a population level. Hulme et al. (2018) presented the first agent-based model to study emerging running-related overuse injuries from the interactions between runners and their coaches' advice. The data-driven computational model is also promoted by Hulme et al. (2019) to complement traditional statistical regression analysis.

This article contributes a quantitative development from the qualitative systems thinking perspective presented earlier in Vermeulen, Grobbelaar, and Botha (2020), by describing and testing a closed-loop, dynamic hypothesis of injury development through a causal loop diagram and a hybrid simulation model of the runner as a complex adaptive system (henceforth referred to as the runner) in a proof-of-concept format. Essentially, the earlier work in Vermeulen, Grobbelaar, and Botha (2020) may be considered as part of the qualitative groundwork for the simulation model presented in this article. The objective of the article is to provide a theoretical and practical foundation from which closed-loop thinking may be developed to augment scientific studies. The closed-loop thinking process requires three considerations:

(1) A discourse from a strong scientific base; in this case, the Second Law of Thermodynamics and the spring-mass system.
(2) Input data, representing the runner in their environment. Data from a running wearable and a weather data bank (*MeteoBlue*, `https://www.meteoblue.com`) were mined for this purpose.
(3) Mechanistic mathematical expressions and frameworks to initiate and drive the decision rules and flows in the hybrid simulation model. A training load function is developed to link running with a physiological response that can be captured in the system's computational steps. An interaction framework is presented to sample for heart rate, based on the cadence level (stride frequency) of the runner.

These three considerations contribute to physical ergonomics in sport by applying systems thinking tools to study the dynamic relationships between runners and their immediate environment. The novelty lies in their combination in a computational model to simulate an individual runner as a complex, adaptive system. In doing so, sport scientists and practitioners may learn from this article how to perform closed-loop thinking that may support injury prevention strategies or management of training loads. The article illustrates how computational modelling may be an adjunct tool to make systems thinking practical in the sport science domain.

The article is arranged as follows: § 2 provides the theoretical background for systems thinking tools and the principles from physics as they apply to running; § 3 presents the systems thinking method and the processes contained in the models (§ 4); results and discussion follow in § 5.

2. Literature review

This section briefly discuss the causal loop diagram and two simulation methods, as well as the running science considerations to substantiate the simulation.

2.1. *Addressing complexity through systems thinking tools*

As a scientific field of knowledge, systems thinking deals with understanding complexity and change by studying dynamic cause and effect over time. The rationale is to study the whole as a synergy of its parts; to see the big picture. Systems thinking is a way to conceptualise the world and relationships in a closed loop, that cause and

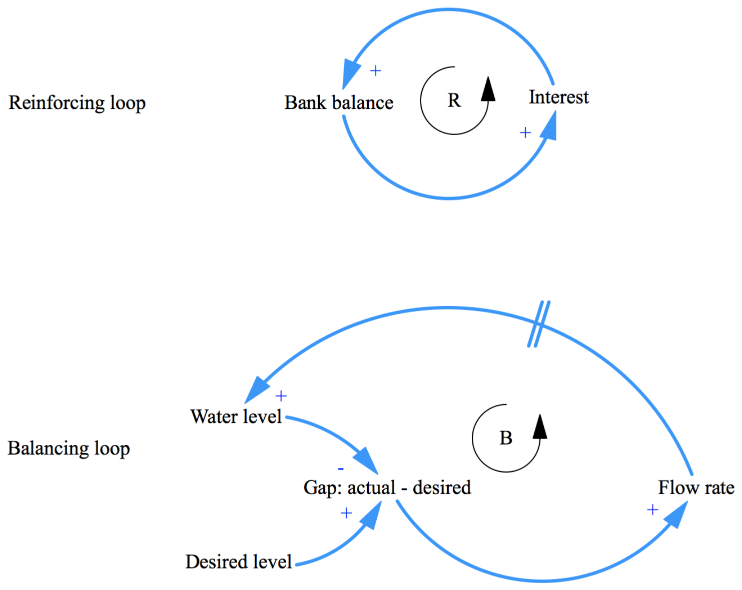


Figure 1.: Causal feedback loops: reinforcing and balancing CLDs

effect are non-linear; often the end (the effect) influences the means (the cause).(Maani and Cavana, 2007). A qualitative method to study systems is causal loop modelling, communicated visually as a causal loop diagram. Quantitative methods include data mining and simulation modelling, specifically agent-based modelling and system dynamics.

2.1.1. Causal loop diagrams

A causal loop diagram is used to qualitatively define the system under study, in terms of its constituent elements (variables) and feedback loops (causal interactions). Polar links indicate the direction of influence of one variable on the other. A positive link (designated by a + sign or the letter *s*) implies the variables move in the same direction, whereas a negative link indicates reciprocal movement (designated by a − sign or the letter *o*). Elements are connected through the links and the feedback loop closes when the initial element is reached again. Delays are the temporal distance between the actions in a system and the (eventual) consequence, indicated by double lines on a connecting link. The feedback loop may reinforce the original action (a reinforcing loop such as seen in a growing bank balance), or counter the initial action (a balancing loop such as seen in a tank filling up with water) (Maani and Cavana, 2007), see Figure 1 for illustration.

Every system has inherent delays (Griffin et al., 2016), although some consequences are further removed in time than others: it takes days or weeks for a planted seed to produce a flower, but there are only seconds between opening the faucet and a full glass of water. To recognise and optimally leverage a gap as a result of a delay is key to a system's survival, for growth and stability (Griffin et al., 2016). Overshooting and undershooting goals, unbounded growth and sudden decline and uncontrolled oscillations (when the system cannot reach stability) are the characteristics of system behaviour when delays and gaps are misunderstood or simply (perhaps unintentionally) ignored (Senge, 1990; Kirkwood, 2013; Griffin et al., 2016).

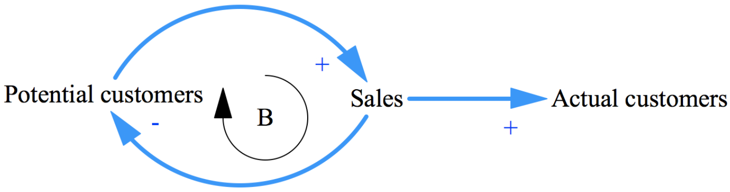
2.1.2. Simulation modelling

A hybrid simulation approach is taken to utilise the strongest features of the various simulation methods to model complex problems. The hybrid simulation model in this article constitute an agent-based component and a system dynamics component. In hybrid simulation, both micro- and macro level of analysis is possible. The individual behaviour of humans, animals or other agents that make autonomous decisions in a bottom-up fashion constitute analysis on a micro-level through agent-based modelling, but their aggregated behaviour and the feedback within the system on a macro level is more easily studied through system dynamics (Martin and Schlüter, 2015; Alvi et al., 2019). The different levels of abstraction allows for the models to be flexible and enable handling of evolving dynamics in a system (Kunc, 2019).

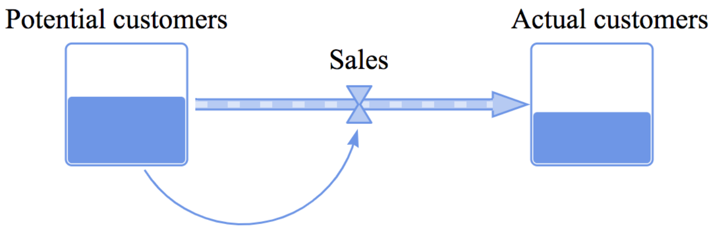
Macal and North (2010) explain that agent-based modelling is an effective simulation method to model complex systems, as it speaks directly to the interconnectedness and their adaptation characteristics. The agent-based model consists of autonomous, heterogeneous agents that interact with each other on a low level of granularity (that is, with a high level of detail). The agent acts and makes decisions based on rules (i.e. algorithms) that have have been programmed in the agent's make-up. The interactions are the self-organising process between multiple, connected agents and influence their behaviour, which in turn aggregates into the changing behaviour of the system (Macal and North, 2010). In complex sport systems, agents can range from large entities such as sporting bodies, facilities, teams to individual athletes and down to the molecular level of muscle fibres. Agents learn over time (based on their experiences), adapt to environmental changes (or changes in other, connected agents), and update internal states (Hulme et al., 2019). Facilities or other non-living assets may 'learn' or adapt through the actions of individuals who manage such assets, for example the groundsmen and women may change their treatment of the grass on pitches, resulting in different surface properties. There are two main purposes to agent-based modelling: generating hypotheses (similarly, theoretical experimentation or experimental thinking) and prognostic studies (to describe the course of behaviour or outcome over time) (Auchincloss and Roux, 2008) in a safe, virtual environment. These capabilities allow researchers to push the boundaries of research that might otherwise be considered as unethical (pushing an athlete into an injured state) or impractical and costly (instantly replace the stadium's track surface with another material).

On the other spectrum of granularity, system dynamics is useful to study the system's behavioural emergence over time at an aggregated level, or a high level of granularity (Hulme et al., 2019). The system dynamics component is constructed as stocks, flows (rates) and parameters that influence stocks and flows (see Figure 12). A system dynamics model constitute a system of differential equations to describe the system. The system process is modelled as a plumbing system, with stocks representing tanks filled with liquid and the flows as the valves (or pumps) that control inter-tank flow. Stocks signify accumulation of a variable (integration) i.e., its level, at any given point in time. Flows are the rates at which the stock level changes (differentiation). Parameters are used to alter the rates. Information on stock levels feeds back into the flows to influence the flow, and in this way the reinforcing and balancing feedback loops are generated (Kirkwood, 2013).

A practical example of a system dynamics model concerning the advertising pro-



(a) Causal loop diagram: advertising, adapted from Kirkwood (2013)



(b) Stock-and-flow diagram: advertising, adapted from Kirkwood (2013)

Figure 2.: Causal loop diagram transformed into a stock-and-flow diagram for the advertising mechanism. The value for `sales` may be represented by a mechanistic model in the form of $\alpha \times$ `potential customers`, where α is a constant proportional rate or percentage.

cess is taken from Kirkwood (2013) (Figure 2). The number of `potential customers` increases the number of possible `sales`, and the number of `sales` adds to the number of `actual customers`. This growth is limited by the negative feedback from `sales` to `potential customers`: as `sales` increase, they eventually erode the `potential customers` base, in turn leading to fewer `sales` over time and slower growth of the `actual customers` base, until (if possible) all `potential customers` have been converted to `actual customers`.

2.1.3. Systems thinking in sport

Complex systems in sport have multiple, scalable levels, namely: micro (on cellular and molecular level), meso (the individual) and macro (the socioeconomic and political). Modulations only at the individual level (for example, through altering the biomechanics) do not suffice to prevent running-related overuse injuries (Hulme et al., 2019). From a biological standing, the individual athlete may be considered as a living organism, consisting of multiple subsystems (the organ systems) that work together to maintain life as a whole organism (Solomon, Berg, and Martin, 2002), implying injuries should be understood at all of these levels.

Multiple agents (or entities), which are fundamentally different to each other, interact within and across system levels and over time. Running-related overuse injuries may be attributed to the interactions between the heterogeneous factors of road infrastructure, equipment and technologies (shoes, tracking devices), training pathways, training workloads, biological predispositions, to name a few (Hulme et al., 2019). Systems thinking offers an alternative perspective to augment the traditional, scientific reductionism to study overuse injuries.

2.2. *Entropy and the athlete*

The concept of entropy was used in both the causal loop diagram and the simulation model to describe the degree of order in an athlete's body. Entropy may be thought of a measure of energy degradation (or devaluation) during metabolism, disorder and randomness of parts (fibres, cells and molecules) that make up the living system and its sub-systems (musculoskeletal, cardiovascular, neuromuscular and so forth) and the number of (infinitely) possible states that the particles might find themselves in (Kattman, 2018; Navrátil, 2011; Solomon, Berg, and Martin, 2002). The nature of human life processes is irreversible and thermodynamic, therefore the laws of thermodynamics may apply to human physiology. The concept of entropy may therefore be used to describe life processes in human physiology (Boregowda et al., 2016; Navrátil, 2011; Batato, 1990).

The key to an athlete's body's ability to repair and maintain its structures, is to dispose of the excess entropy (heat from metabolic processes and degraded particles from structures that have been broken down during activity) while taking in matter that consists of usable energy, or has low entropy (Kattman, 2018). The term *negentropy* is the reciprocal variable of entropy: it stands for order, difference or structure (Kattman, 2018). The athlete must take in negentropy in the form of structured energy (nutritional input from carbohydrates, proteins and fats) and engage in physical training, whereby the body's sub-systems develop resilience to countervail the build-up of internal entropy and disorder.

It must be noted that the athlete's body as an entropically open system does not violate the dynamical laws of thermodynamics. When the athlete's entropy is lowered through structure formation to strengthen muscles, improve cardiovascular function or repair injuries with the intake of negentropy, entropy in the universe (his/her immediate environment) must increase when the athlete radiates heat via the skin or excretes other, lower order materials such as water (from sweat) and carbon dioxide during exhalation (Kattman, 2018). Re-drawing the boundary to include the universe and the athlete (the athlete-environment system can now be considered as a closed system), it is clear that the Second law of Thermodynamics holds: entropy will always increase.

2.3. *Cyclical impact loading*

A mathematical expression is required to provide the computational steps in the simulation model with a quantified value for the training load. The training load is quantified in a mathematical expression that connects the elements of the runner in a composite value to represent the perturbation of homeostasis through cyclical impact loading. The derivation of the impact loading takes inspiration from the established concept of the runner as a spring-mass system (Farley and González, 1996; Ferris and Farley, 1997; Morin, 2018; Girard et al., 2013; Moore, 2016; Kulmala et al., 2018). The runner may be represented as a linear spring-mass system consisting of a single mass (the body) that oscillates around an equilibrium position (defined as where the vertical ground reaction force equals body weight, (Cavagna, Legramandi, and Peyré-Tartaruga, 2008)) with the legs acting as springs. The repetitive up-and-down movement of the center of mass of a runner represents a wave-form function over time (Clark, Ryan, and Weyand, 2017), and can be modelled as simple harmonic motion. In simple harmonic motion the displacement, velocity, and acceleration of an oscillating object is modelled as sinusoidal functions of time.

During running, there is contact (or a vertical collision) with the ground when the foot strikes the ground, through stance phase and ends when the toe is lifted. The ground reaction forces vary with time during the stance phase. The body is encountering cyclical impulse loading from these vertical collisions as it moves over the ground, alternating between being airborne and in contact with the ground (Kulmala et al., 2018; Clark, Ryan, and Weyand, 2017). Capturing this cyclical impulse is essential to establish the cumulative effect of the training load on the athlete's body in the simulation model.

3. Methods

The systems thinking and modelling methodology phases set out by Maani and Cavana (2007) guided the construction of the models. Their method consists of 1) problem structuring, 2) conceptual qualitative modelling (primarily the causal loop diagram), 3) dynamic modelling (primarily the system dynamics simulation model), 4) scenario planning, and 5) implementation and organisational learning. However, not all phases nor sub-steps are necessarily required in a systems thinking and modelling methodology intervention. The degree of effort to solve the problem will determine whether all steps are taken. The detailed steps for each phase may be found in Maani and Cavana (2007).

3.1. *The systems thinking and modelling methodology*

The objective of the article is to establish a foundation from which to study the runner (and potentially other athletes) using systems thinking and computational modelling. The scenario planning and implementation and organisational learning phases were therefore not included. During problem structuring the magnitude and scope of the problem is described. In this case, the problem to be simulated is described as: Running-related overuse injuries develop gradually and as a result of complex interactions between the runner and his/her environment. How can we simulate this process?

The conceptual modelling is done with a causal loop diagram of the runner, incorporating into the storyline the concepts of training loads, damage, entropy, adaptations, and recovery. The intervention policies to prevent or reduce the onset of running-related overuse injuries are identified during this phase. The dynamic modelling instantiates the causal loop diagram and is completed with a hybrid simulation model, consisting of an agent-based component and a system dynamics component. The unit of analysis is the individual runner. The simulation model is partly driven by data mined from running wearables. The simulation model quantitatively tests the intervention policies applied to prevent or reduce the onset of running-related overuse injuries. The main purpose of the computational framework is to generate learning in structural leverage in the runner as a system, and not necessarily to predict an injury from a big data set consisting of data points from a large sample size of athletes. Therefore, the idea of the quantified-self as a $n = 1$ sample size from Sands et al. (2017) and Swan (2013) is used, that is the overall, population level sample size consists of a few archetype runners, but the dataset per individual athlete is large.

Validation of simulation models, specifically system dynamics models, is not entirely possible since they are endogenous abstractions of reality (Sterman, 2000). For this reason, models must be subjected to testing for how well they map to the real world to build confidence in the model as the most applicable version of the abstracted reality. The model tests from Sterman (2000) were applied to the simulation model, of which two are demonstrated in this article, namely the sensitivity analysis and the family member test. Models can and should be tested beyond their design limits to ascertain their useful domain (Wakeland and Hoarfrost, 2005).

3.2. *Processes within the model*

This section explains how the three considerations for a starting foundation upon which to build practical, closed-loop models are applied to the runner. The entropic line represents the process by which damage gradually builds up to the point of injury; the training load that drives the build-up is quantified by means of a mathematical expression; finally, the data mining process to extract input parameter values to the simulation model is explained.

In this proof-of-concept model, the data from one athlete's running wearable were incorporated to quantify the parameters. Data from two different runners were then used to test the simulation model for varying or similar behaviour (the family member test). Ethical clearance was obtained from the Research Ethics Committee from Stellenbosch University.

3.2.1. The entropic line: gauging damage and structural integrity

The terms damage, disorder, disarray, and imbalances are often used to describe the aetiology behind running-related overuse injuries. Injuries are viewed as the runner or a subsystem being in a poor or non-functioning state due to the cumulative disorder brought on by repetitive loading of unprepared structures. This section introduces the concept of the entropic line as a measure to gauge the degree of additive damage to structural integrity:

$$E_n = \frac{E}{I} \tag{1}$$

where E_n represents the net entropy as a unitless *entropy:structure* ratio; E is the amount of internal disorder (entropy) from cumulative micro-damage and I represents integrity of structures. Damage develops to tissues once E_n reaches some physiological limit, that is when the tissue has reached its strain threshold and starts to fail.

Both E and I increase as a result of training, however:

- Entropy, E, is removed through thermoregulation during exercise (to take away excess, unusable heat energy), cellular repairs to disordered structures, and adaptations of these structures through physiological actions during recovery (that is, to once again form structure that has lower entropy).
- Integrity, I, is only established when the cellular and tissue adaptations have reached maturity.

An injured runner showing slower performance time might now be considered as a living system of which the internal entropy is increasing. The body is becoming less capable to dispose of the surplus entropy and make use of negentropy (perhaps due to a short supply or a decrease in resilience in structures), thereby E_n increases. I may remain constant at first, but will start to decrease as tissue is broken down and not repaired in time. The athlete's body struggles to regain order in the system to sustain faster paces due to the imbalance between entropy and negentropy and degrading structures.

Environmental stressors may influence the entropic line gauge. In a hotter environment, E (heat component) cannot be dissipated at the same rate as in a cold environment, therefore recovery of order in the systems slows down. At higher altitude, with less negentropy available in the form of oxygen to burn fuel, internal entropy increases (waste components) and structural integrity decreases.

3.2.2. The training load pulsed input function

To derive the impact force-time waveform for the runner as a spring-mass system undergoing simple harmonic motion (§ 2.3) the linear displacement x of the spring force F_s in Hooke's law is replaced with the time dependent function for displacement, $x(t)$ as a cosine function:

$$\begin{aligned} F_s &= -kx \\ F_s(t) &= -k_v x_m \cos(2\pi f t) \end{aligned} \tag{2}$$

Equation 2 is presented as the runner's cyclical load function, where k_v represents the vertical stiffness of the runner's leg springs. Equation 2 results in a smooth, flipped cosine wave (a negative form of the standard wave), without the distinction between initial peak and maximum force characteristic of impact force-time waveform graphs (Figure 3).

The biomechanical parameters for the cyclical load function in Equation 2 are usually calculated from ground reaction forces and spatio-temporal variables during force plate and 3-D motion capture studies. The reader is referred to the work of others, such as: Cavagna, Legramandi, and Peyré-Tartaruga (2008); Girard et al. (2013); Lieberman et al. (2015) for more detail on how the data are collected and measured. Two of the biomechanical parameters for the impact force-time waveforms are available from running wearables. Cadence (to approximate f) is provided directly, whereas the vertical oscillation may serve as a proxy for amplitude. Leg stiffness is analogue to the stiffness coefficient of a spring. It represents the amount of vertical displacement of a reference point on the runner's leg to the ground reaction force encountered. Vertical oscillation is the cyclic maximum vertical displacement of the runner's centre of mass as they move over the ground. Exact formulas and mechanisms by which to calculate these biomechanical variables are beyond the scope of this article, but the reader is referred to Butler, Crowell, and Davis (2003) for more on leg stiffness, and Moore (2016) for running dynamics.

The integral over time of the positive values from Equation 2 (that is, the values generated during ground contact) will provide the impulse the runner's body has experienced from its collisions with the ground. Mathematically, the impulse may be integrated over the absolute value of $F_s(t)$, then taking a proportion α associated with ground contact. If a symmetrical split between aerial and ground contact time for lower running speeds is followed (Cavagna, Legramandi, and Peyré-Tartaruga, 2008), then $\alpha = 0.5$:

$$
\begin{aligned}
I &= \alpha \int_{t_i}^{t_f} F_s(t)\; dt \\
&= \alpha \int_{t_i}^{t_f} |-k_v x_m \cos(2\pi f t)|\; dt
\end{aligned}
\tag{3}
$$

The total impulse is therefore a function of the load-release cycle that the musculoskeletal system, viewed as a spring-mass system, encounter during running. This impulse in Equation 3 is regarded as the training load absorbed by the runner, to which the body's structures must respond. This training load is used as the mechanistic pulsed input function for the simulation model.

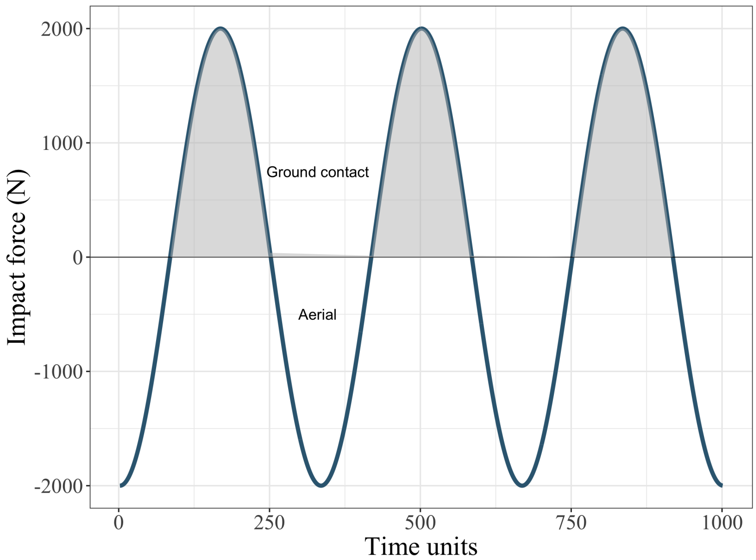


Figure 3.: Simulated force-time waveform of the cyclical load function, where $k_v = 20\ kN/m$, $x_m = 0.1m$, $f = 3\ steps.s^{-1}$. In its current form, the cyclical load function considers the symmetry between aerial and contact time at lower running speeds from Cavagna, Legramandi, and Peyré-Tartaruga (2008), therefore the function is evenly split between positive (ground contact) and negative (aerial time) force values.

3.2.3. Data mining for parameter quantification

The most important features of the analysis are the interaction between the surface type and cadence and between heart rate and cadence. Three years (2017–2019) of running data were extracted directly from the participating athletes' *GarminConnect* profiles. The data collection involved a visual inspection for each run activity's container file on *Google Earth* to ascertain the surface on which the run took place. The container files include the following data points: epoch timestamps, latitude, longitude, altitude, speed, cadence, distance, and heart rate. The interactivity time was calculated as the time elapsed between the end of one activity and the start of the next recorded activity. The different run types represent the dominating surface encoun-

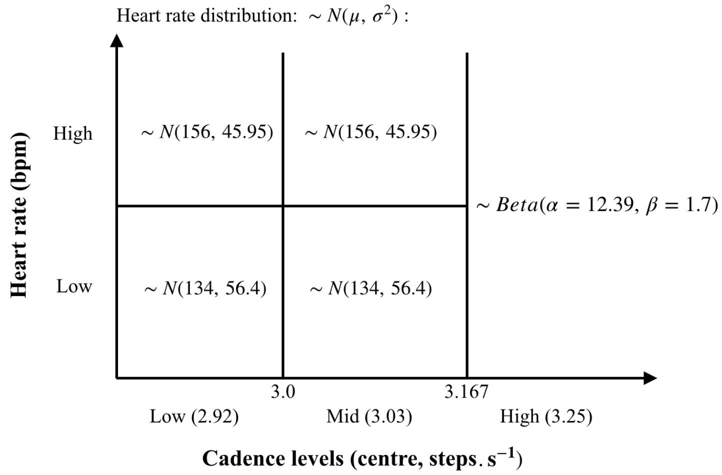


Figure 4.: Parameters for heart rate from cadence classification

tered during the run, namely road running and racing (asphalt and concrete), trail running and racing (grass and gravel) and track running (grass). Historic weather data from the *MeteoBlue* website were extracted bi-monthly over the course of one year (June 2019 to May 2020). The data were aggregated to a monthly level, per location and only for the early morning hours (05:00 to 08:00) and late afternoon to evening hours (16:00 to 20:00) to stabilise the temperature range. The averages, standard deviations, minimums and maximums of temperature were extracted as parameters to inform the simulation model.

The simulation is performed only for the geographic region surrounding Pretoria, South Africa. Cadence is sampled based on the type of surface encountered during the run. The *k-means* clustering algorithm was utilised to better understand general behaviour and distribution of cadence and heart rate for the location-specific data. The clustering served as a guideline to confirm reasonable sampling parameters. A classification framework was developed to sample the heart rate based on a cadence level. The final classification framework is presented in Figure 4. The upper and lower limits between the categories are derived from the minimum and maximum values associated with the cadence and heart rate clusters. The heart rate zones were calculated with cut-off points $K = \{0.65, 0.75, 0.8, 0.85, 0.9\}$ of the maximum heart rate, (Khalil and Sornanathan, 2010), with $\text{HR}_{max} = 208 - 0.7 \times Age$.

4. Models

The dynamic hypothesis of the runner and overuse injury development is first presented in a causal loop diagram, followed by the simulation model which was constructed in the *Anylogic* software's personal learning edition.

4.1. *The runner's causal loop diagram*

The runner's causal loop diagram is presented in Figure 10. Main accumulation points in the system are the training pulse (§ 3.2.2), the integrity of structures, and damaged structures (§ 3.2.1). The training pulse may be articulated as: 'to run for t hours with a leg stiffness k at cadence f with a vertical oscillation of x_m on surface r'. Maturity of structures is a product of the internal elements of the runner's body aggregated as muscle strength, joint alignment and flexibility. Damaged structure accumulates as a result of higher entropy when the restoration process is neglected whilst the cyclical load on structures continues.

Mainly three balancing and two reinforcing loops characterise the runner's causal loop diagram in Figure 10. The first balancing loop, B1, closes the gap between the goal structure and the actual structure of the athlete through the changes in the micro-damage rate. The second balancing loop, B2, also closes the gap between actual and goal structure but through exercise therapy (or conditioning) to build new structure. The third balancing loop, B3, removes damage from the athlete's body by some damage control actions taken by the athlete. The first reinforcing loop, R1, amplifies the training pulse of the runner through adjustments in intensity as mature structure closes the gap. The second reinforcing loop, R2, amplifies an unwanted consequence when damage has occurred and is accumulating, but training is not suspended or lowered. The complete discourse for the runner's causal loop diagram may be found in the supplementary material to this article, § 7.

4.2. *The hybrid simulation model*

The simulation is a hybrid model containing the heterogeneous decision making elements of agent-based modelling in a continuous system dynamics model. A fuller explanation on the model may be found in the supplementary material, § 7.

4.2.1. The agent-based component: training-rest cycle

The athlete makes individual training decisions through a state chart (the training cycle, Figure 11) and trigger functions in the agent-based component. Interactions between the athlete's biomechanical and physiological elements with the environment (leg stiffness, vertical oscillation, cadence, heart rate, temperature, and surface type) are modelled on the micro level through functions within the state chart. The behaviour over time and cluster analyses of the tracking data obtained from the running wearable pointed to micro level interactions between the surfaces encountered, cadence, and heart rate. Interactions between the surface, leg stiffness, and vertical oscillation are based on literature (Ferris and Farley, 1997), specifically the runner as a spring-mass model (§ 2.3). Interaction between temperature and recovery were found in the literature (Kenny and McGinn, 2017).

The state chart `stcTrainingStatus` models the athlete's transition between resting and training (Figure 11). Only while in the healthy status may an athlete train, once injured the training stops. The transition to `stInjured` is modelled as a condition, which is based on the time spent in the damage zones and managed by a return function, `fDriveInjury`. The entire simulation allows for 720 days of transitioning between training and resting, before transitioning to the final state and the simulation ends.

4.2.2. The system dynamics component: stock-and-flow diagram of the runner

The stock-and-flow model in Figure 12 is the quantified, dynamic abstraction from the qualitative runner's causal loop diagram in § 4.1. Pressure points develop as a result of unwanted accumulation of damage. Actions taken by the athlete (governed through

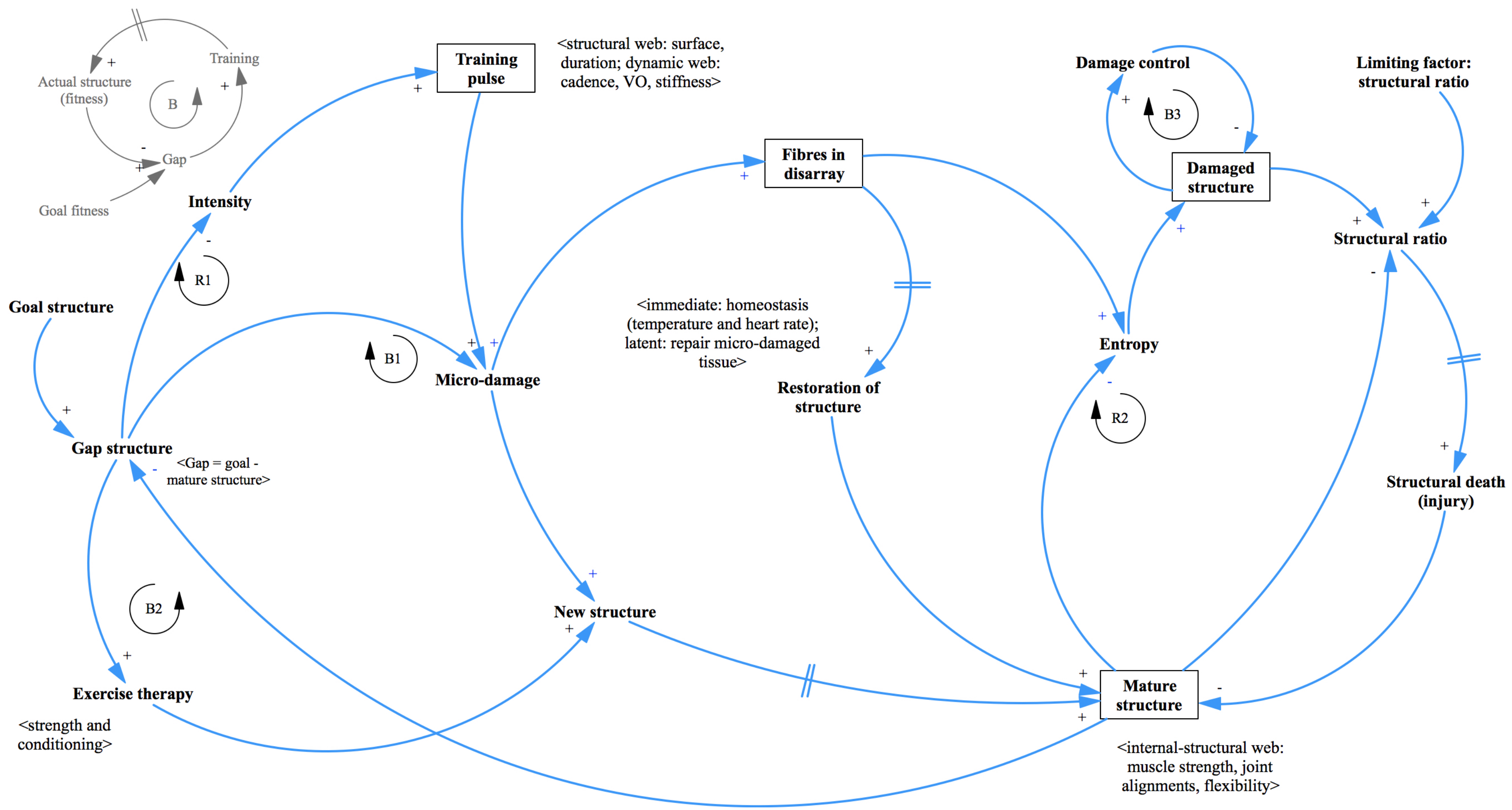


Figure 5.: The causal loop diagram for the runner

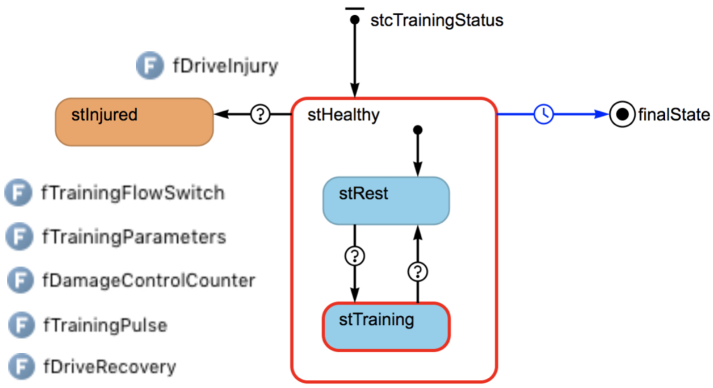


Figure 6.: The training status statechart

the decision rules in the agent-based component) opens and closes the valves (flows) into and out of the accumulation tanks to relieve pressure or divert content. The athlete's system works toward closing the fitness gap, that is, the difference between actual structure and the goal structure.

The path to injury is a two-step process. First, net entropy (in `dvEntropy`) is a fractional relational measure between temporary disarrayed fibres (internal entropy), as a result of micro-damage from training, and existing structure. This calculation is the first pressure point, derived from the entropic line gauge in § 3.2.1). The `dvEntropy` updates continuously to keep track of the relationship between the number of fibres in disarray and structural integrity. In the second step towards injury, when `dvEntropy` reaches a certain threshold, the `flowDamage` is activated and takes as value a fraction of the `stkFibresDisarray`. The outflow from `stkDamage`, `flowDamageOut`, is additive, accounting for natural removal of damaged tissue and damage control measures taken by the athlete.

The $^d/_s$-ratio considers the amount of damaged fibres to the amount of structural integrity, and serves as a measure to gauge the cumulative damage from overload. This is the second pressure point in the plumbing system. Two damage zones exist, with the upper and lower limits for each zone demarcated using the $^d/_s$-ratio. An athlete may spend some time in these damage zones, however it is the cumulative time in the damage zone that will ultimately incite the injury. The damage zones are defined as lower and higher risk.

The `dvRatioDamageStructure` controls the flow of the damage zone drivers through associated dynamic variables: either flow, `flowDamageZone1` or `flowDamageZone2`, depending on the magnitude of `dvRatioDamageStructure`, is activated and flows at a rate one unit-day per simulation-day to add to the stock for the damage zones (`stkDamageZone1` or `stkDamageZone2`). An injury is incited when a certain number of days has accumulated in the damage zone stocks.

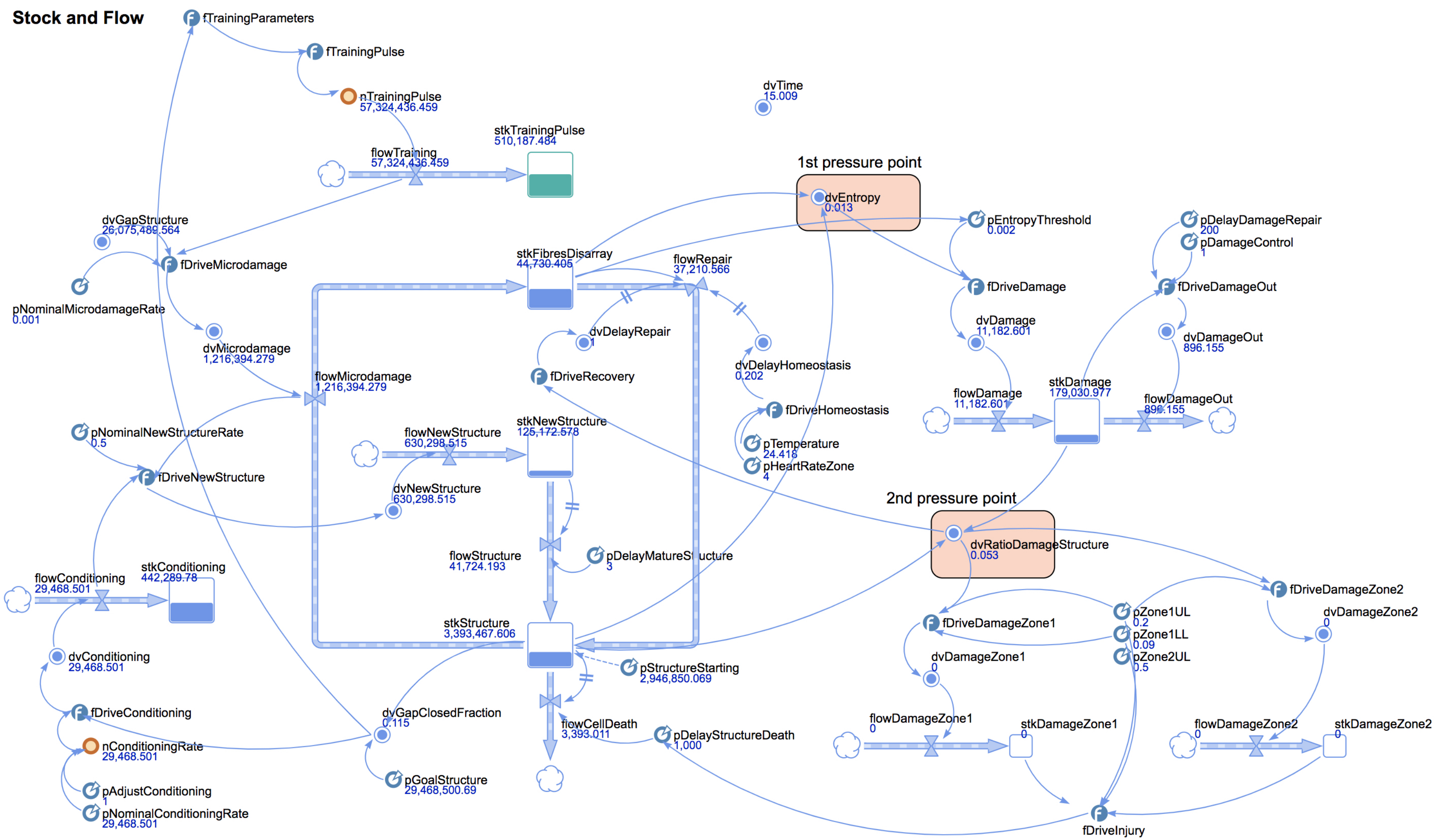


Figure 7.: The detailed stock-and-flow model of the runner. (Not all the arrowed lines in the figure are functional as per traditional stock-and-flow configurations. In the *Anylogic* computational configuration, arrows between driving functions, and parameters, flows, stocks, and variables are not necessary. However, they are shown in the figure to provide the reader with a visual of the closed-loop, ‘pressurised plumbing system’.)

4.3. *The interventions to augment fitness*

The athlete makes decisions regarding training load and damage control in order to augment their fitness. Interventions are characterised as fundamental or symptomatic. Fundamental solutions have a longer term focus and is subject to delays, whereas symptomatic solutions immediately address the problem (they act as quick-fixes) but loose efficacy over time. Interventions that were considered are:

- Fundamental: increased conditioning, slower ramp-up of training load, reduce training load once in damage zone 1 or 2, and respecting the repair delays by increasing the intervals between training sessions (recovery days).
- Symptomatic (i.e. quick-fix): increasing damage control, increasing the training load.

Damage control is a collective term for any action that an athlete takes to address shortcomings or training induced damage, in the real world this would include orthrotic braces, compression wear, new shoes and so forth. These interventions vary in successful outcomes, but were grouped together since the basic behaviour of the solution remains the same, irrespective of the mechanism: damage will be removed starting immediately. In the model, damage control is therefore dimensionless, and articulated as: activities that remove x units of damaged fibres per day.

4.4. *Simulation input data*

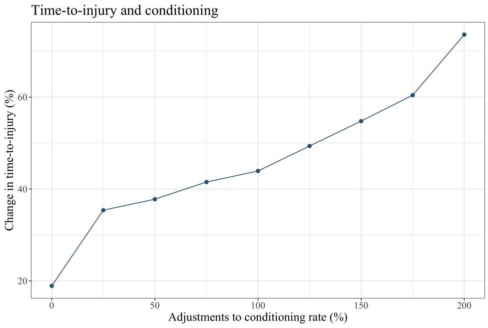
Table 1 contains the input values for the running dynamics that drive the pulsed training load function. Cadence and surface types were mined from the running wearable. With guidance from the cluster analysis, the track training subset is split into two values for cadence, namely high and low to represent the bi-modality in the data according to a probability mass function that distinguishes between the clusters. Runners adjust their leg stiffness in reaction to the stiffness of the surface encountered, thereby maintaining similar biomechanics between different surface types (Ferris and Farley, 1997; Wang et al., 2012). Data ranges from Ferris and Farley (1997) were used as sensible estimates for adjustments to vertical oscillation and vertical stiffness, k_v. Values for other input parameters may be found in the supplementary material.

Table 1.: Input values for parameters per surface type. Cadence and vertical oscillation are both mean ± st.dev.

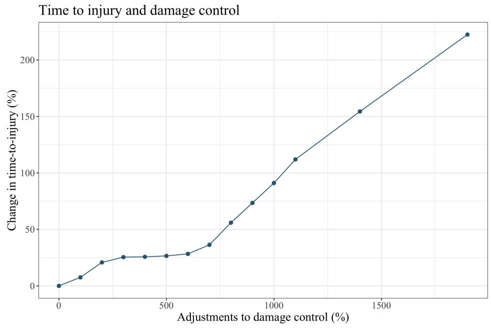
Run (surface) type j	**Probability mass function**	**Cadence** (steps.s^{-1})	**Vertical oscillation** ($meter$)	**Leg stiffness** k_v (kN/m)
Road race (rc)	0.003	3.038 ± 0.080	0.12 ± 0.012	20
Road running (rr)	0.399	2.981 ± 0.077	0.12 ± 0.012	20
Trail race (tc)	0.006	3.075 ± 0.059	0.08 ± 0.008	30
Trail run (tr)	0.154	2.953 ± 0.095	0.08 ± 0.008	30
Track training (tt)	0.438	low: 2.926 ± 0.0605	0.06 ± 0.006	40
		high: 3.244 ± 0.0712	0.06 ± 0.006	40

5. Results and discussion

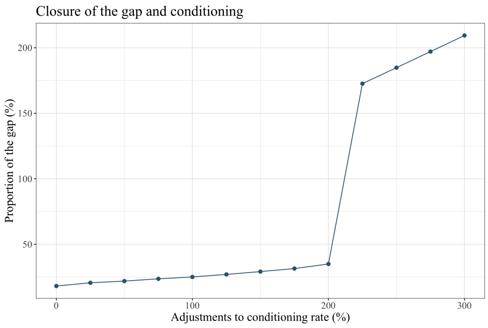
The simulation model is tested and demonstrated with a sensitivity analysis and the family member test. Experimentation serves to show how the model may be programmed differently to observe the causal behavioural changes. Experimentation with



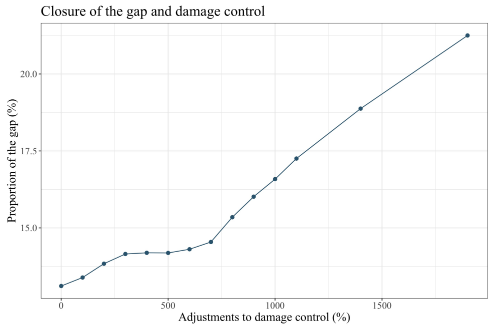
(a) Time-to-injury and conditioning



(b) Time-to-injury and damage control



(c) Closure of the fitness gap and conditioning



(d) Closure of the fitness gap and damage control

Figure 8.: Sensitivity analysis

the model is categorised along the type of intervention that was followed: fundamental interventions consist of conditioning through exercises, whilst the symptomatic intervention consist of damage control. Conditioning and damage control are increased by multiples of the nominal rates. Two outcome variables are reported. The time-to-injury is the number of days that an athlete spent training before incurring an injury. Closure of the fitness gap measure the extent of the structural integrity that was reached at the time the simulation ended, either at the time of injury or after 720 days.

5.1. *Sensitivity analysis*

There is a visible difference in the trends between the effects of conditioning and damage control on time-to-injury and closure of the fitness gap (Figure 8). Although both conditioning and damage control extend the time-to-injury, it takes considerable higher upward adjustments of damage control to extend the time-to-injury. There is a step-wise change in the effect of conditioning on closing the fitness gap. At 3.25x the nominal conditioning rate, the athlete did not enter the damage zone at all and closed the fitness gap on day 328 without incurring an injury. However, for damage control, injury is inevitable despite increases in the parameter. Damage control did not have a significant effect on closing the fitness gap.

5.2. *The family member test*

The simulation model was tested for behaviour anomalies by substituting values of the input parameters with values from other athletes' running wearables, henceforth referred to as 'athlete 1' and 'athlete 3'. The simulation was run for the base case (1x damage control, 0 conditioning) and for increasing conditioning rates on a 8x damage control level for the two additional athletes. The different athletes are essentially representing family members of the same system, or different versions of the runner as a complex adaptive system. Dissimilar outcomes were observed in these tests, which were expected. The magnitude of some events differed, although the general behaviour of the system remained similar.

Some disparities are visible in the time-to-injury for the athletes (Figure 9). This is indicative that the systems' behaviour is different in magnitude of outcome metrics, but the patterns are similar. Athlete 2 consistently takes longer than athlete 1 to reach an injury, however athlete 2 does not get injured for a conditioning rate of 1.75x whereas athlete 1 does get injured. Athlete 3 is consistently injured earlier than the other two.

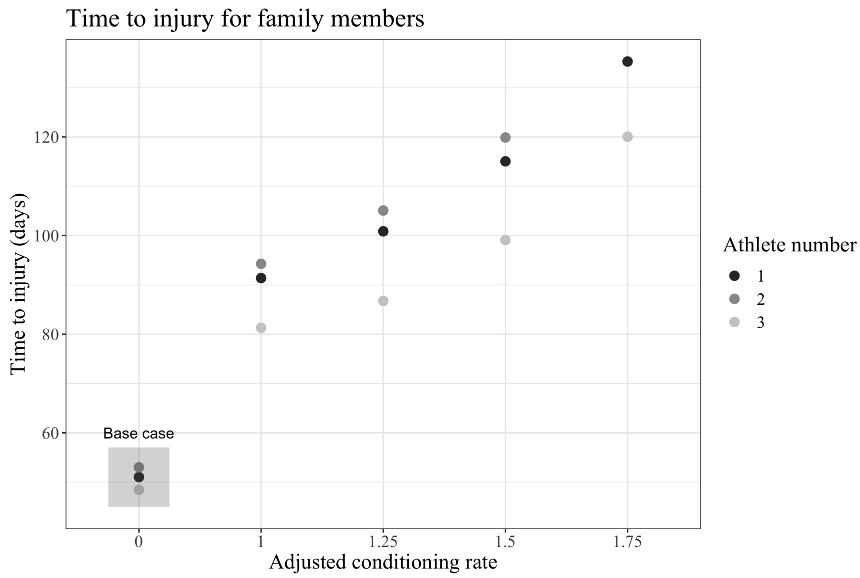


Figure 9.: Time to injury for athletes 1 to 3

5.3. *Results in context*

Other recent systems thinking frameworks and simulation models in sport focus on outcomes on higher (macro) complex system levels (Hulme et al., 2019, 2018), whereas the simulation model demonstrated here is subject-specific on the individual (meso) and subsystem (organ) level with incorporation of the musculoskeletal system and cardiovascular system, to dynamically and quantitatively augment understanding of the runner in their environment. The subject-specific systems thinking approach was also qualitatively illustrated in the complex sport model from Bittencourt et al. (2016), showing different injury development patterns (or risk profiles) for athletes from the sports of basketball and ballet.

The system dynamics model in Hulme et al. (2019) accounts for active runners becoming injured, then recover and re-join the running cohort or leave the system. The authors acknowledge that the model may be expanded to include the influence of larger systems on running (social-, healthcare-, economic-, technological, and sports marketing systems), yet the basic form of the model is sufficient to investigate leverage points and policies to alter the running system. Hulme et al. (2018) introduced the first practical application of an agent-based modelling in sports injury science research. From a systems thinker perspective this model demonstrates the dynamic relationships between an athlete's training load adherence, their acute to chronic workload, and the running-related overuse injuries incidence on population level. The model presented in this article differs from the aforementioned computational models. It is subject-specific, by focussing on the individual runner within their immediate environment, leveraging the $n = 1$ experimentation from the quantified-self movement. The model is driven by primary data collected from running wearables, whereas Hulme et al. (2019) and Hulme et al. (2018) used illustrative and/or secondary data. Hulme et al. (2019) and Hulme et al. (2018) were concerned with aggregating individual level outcomes of injuries to population level, whereas this article showed what processes are needed to simulate the emergence of injury (or protection against) at the individual level.

A subject-specific, yet systemic approach, aligns with the paradigm shift in medicine and healthcare in the 21st century, often referred to as the P's of healthcare (Golubnitschaja et al., 2016): first, *personalised* activity as each athlete has unique abilities and weaknesses; second, *participation* from the individual by the honing of skills and subsequent physical performance; third, focus is placed on *prevention* of injuries; fourth, the individual must be *precise* in their training and execution to attain their specific sporting goals and be an effective athlete. Hulme et al. (2019) also encouraged data driven computational modelling applications. The utilisation of personal data from running wearables to drive the simulation model adds to the personalisation and preciseness of the model, similar to how Kosmidis and Passfield (2015) and Ahmad, Jamaludin, and Hafidz Omar (2018) found subject-specific metrics in their respective analyses of data from running wearables.

The results provide support for a systems thinking perspective in sport science, whereby the mental models of athletes and practitioners may be challenged both conceptually and empirically. The quantitative modulations of the causal interdependencies within the hybrid system dynamics model helped identify optimal leverage points (or targets for interventions, as mentioned by Hulme et al. (2019)), as well as the system elements reactions to their implementation. The simulation model showed that respecting delays and opting for fundamental solutions, even though they take longer to materialise, are optimal for structural integrity and athletic longevity. These concepts and interventions are not new to the sporting world; however the method in which they are explored is novel. The implications for prevention of running-related overuse injuries are therefore related to the methodological process behind closed-loop thinking. Thinking about training and overuse injuries in terms of a pressurised plumbing system with accumulating content in tanks and actions as valves to increase or release content by allowing movement between the tanks, departs from a linear cause-effect thinking pattern to arrive at causal, closed-loop thinking. The site of accumulation in the system drives the development of healthy structure or injury. Intervention actions that add to the flow of structure formation (conditioning) and allow for less outflow from structure (longer recovery) result in accumulation of the structure. Actions that open the outflow from structure more than the inflow to structure (short recoveries) result in damage build-up elsewhere in the system, leading to injury. The actions that open the outflow of accumulated damage are not effective to prevent damage build-up.

The relationship between the level of structure and damage (the d/s-ratio) is considered as the main pressure point and an indicator of health or injury. The inciting of the injury was not based on published injury rates, but rather on emergent behaviour and the critical values of the relationship between structure and damage. Inciting an injury as an anticipated future event (through the simulation parameters) adds to the confidence of the system dynamics model to study structural interactions in the runner.

5.4. *Limitations*

Dealing with a system with expressions on various levels of detail, it was necessary to decide upon the level of unit analysis for tractability. This is similar to what Coveney et al. (2013) refer to as the scale upon which system biologists perform analysis, then moving up and down to incorporate higher and lower levels of detail. The goal of this article was to demonstrate a computational framework that may contribute to the pragmatism of systems thinking in sport, founded upon three considerations, namely: a scientific discourse, data mining, and mathematical expressions to drive the actions in the simulation model. To achieve this feat and maintain parsimony of a computational model based on a living system, the level of analysis was limited to the individual runner and its subsystems. The $n = 1$ experimentation from the quantified-self movement limits the external validation of the simulation model. This thinking may well be aggregated to the macro-level, by aggregating the results of multiple individual-level simulations into higher system level behaviour. Societal-level influences may also be imposed on individual runners, such as dissemination of new shoe technology that claim to reduce impact loading, which would directly influence the mechanism by which the simulation model deals with cumulative damage. Both expansions would require a larger sample size of runners. The running wearable data collection process is tedious and time consuming, limiting the amount of data that was extracted from participants. Expanding the sample size must consider the time cost to download and extract the data, or find an alternative that would accelerate data collection.

To remain parsimonious, the simulation is an abstraction of the real world and does not account for all the possible permutations that an athlete may undergo during training. The model assumes the athlete's capability to adapt, but already these adaptation rates differ amongst individual athletes and is subject to change under environmental conditions (Borresen and Lambert, 2009). The training load calculation through the pulsed functions (Equations 2 and 3) only quantifies the interaction between the musculoskeletal system and the surface. Although the heart rate zones are utilised to gauge repair time (the full return to homeostasis), more work is required to quantify the aerobic response. Some physiological and biological variables that are excluded from the model are genetic factors and biometrics. Resource constraints, psychological processes, and socio-economic factors were also not considered.

Future work to improve on these limitations are related to both the data sources and additional variables. An alternate data source for wearable data include the *Strava* platform, which would allow for a large data set at a small time cost. Another alternative data source is an empirical study design to collect validated data of a disparate group of runners at a substantial time cost, or some combined strategy. Energy balances as a variable through nutritional intake may also be considered, which may form the bridge to socio-economic aspects. Other variables to consider for inclusion are psychological and socio-economic factors, environmental factors (specifically, relative humidity, slopes, altitude, air quality), and resource constraints. Either way, augmenting and refining the simulation model as presented here would require additional data and alternative data collection processes.

6. Conclusion

The value of systems thinking and simulation modelling for practitioners lies in the methodological thinking to discern between fundamental and symptomatic interventions in the dynamic, complex solution space. The solution space consists of multiple, interacting levels (complex) and has inherent time components (dynamic). Systems thinking promotes the process to find optimal leverage through changing the elements of structure (through their interactions, how they are connected or altering individual properties) instead of finding solutions to address symptoms of problematic structure. Progressing from the qualitative descriptions to practical, quantifiable applications is crucial to move systems thinking forward in sport science, which was done in Hulme et al. (2019) and Hulme et al. (2018), and now with the hybrid simulation model presented in this article.

References

Ahmad, Zulkifli, Najeb Jamaludin, and Abdul Hafidz Omar. 2018. "Monitoring and Prediction of Exhaustion Threshold during Aerobic Exercise Based on Physiological System using Artificial Neural Network." *Journal of Physical Fitness, Medicine and Treatment in Sports* 3: 1–3.

Alvi, M.S.Q., I. Mahmood, F. Javed, A.W. Malik, and H. Sarjoughian. 2019. "Dynamic behavioural modeling, simulation and analysis of household water consumption in an urban area: A hybrid approach." In *Winter Simulation Conference*, Vol. 2018-December, 2411–2422. Doi: `10.1109/WSC.2018.8632309`.

Auchincloss, Amy H., and Ana Diez Roux. 2008. "A New Tool for Epidemiology: The Usefulness of Dynamic-Agent Models in Understanding Place Effects on Health." *American journal of epidemiology* 168: 1–8. Doi: `10.1093/aje/kwn118`.

Barton, C J. 2018. "Managing RISK when treating the injured runner with running retraining, load management and exercise therapy." *Physical Therapy in Sport* 29: 79 – 83. Doi: `https://doi.org/10.1016/j.ptsp.2017.10.002`.

Batato, M. 1990. "Energetics of the human body." *Schweizerische Zeitschrift fur Sportmedizin* 38 (3): 133 – –141. Original article in French.

Bittencourt, Natalia, W H Meeuwisse, Luciana De Michelis Mendonça, Alberto Nettel-Aguirre, Juliana Ocarino, and Sérgio Fonseca. 2016. "Complex systems approach for sports injuries: Moving from risk factor identification to injury pattern recognition-narrative review and new concept." *British Journal of Sports Medicine* 50: 1–7. Doi: `10.1136/bjsports-2015-095850`.

Boregowda, Satish, Rod Handy, Darrah Sleeth, and Andrew Merryweather. 2016. "Measuring Entropy Change in a Human Physiological System." *Journal of Thermodynamics* 2016: 1–8. Doi: `https://doi.org/10.1155/2016/4932710`.

Borresen, Jill, and Michael Ian Lambert. 2009. "The Quantification of Training Load, the Training Response and the Effect on Performance." *Sports Medicine* 39 (9): 779–795. Doi: `10.2165/11317780-000000000-00000`.

Butler, R J, H P Crowell, and Irene McClay Davis. 2003. "Lower extremity stiffness: implications for performance and injury." *Clinical Biomechanics* 18: 511 – 517. Doi:`10.1016/S0268-0033(03)00071-8`.

Cavagna, G. A., M. A. Legramandi, and L. A. Peyré-Tartaruga. 2008. "Old men running: mechanical work and elastic bounce." *Proceedings of the Royal Society B* 275: 411 – 418. Doi: `10.1098/rspb.2007.1288`.

Clark, Kenneth P., Laurence J. Ryan, and Peter G. Weyand. 2017. "A general relationship links gait mechanics and running ground reaction forces." *Journal of Experimental Biology* 220 (2): 247–258. Doi: `10.1242/jeb.138057`.

Coveney, Peter, Vanessa Díaz-Zuccarini, Norbert Graf, Peter Hunter, Peter Kohl, Jesper Tegner, and Marco Viceconti. 2013. "Integrative approaches to computational biomedicine." *Interface Focus* 3 (20130003): 1–4. Doi: `10.1098/rsfs.2013.0003`.

Farley, Claire T., and Octavio González. 1996. "Leg stiffness and stride frequency in human running." *Journal of Biomechanics* 29: 181 – 186. Doi: `10.1016/0021-9290(95)00029-1`.

Ferris, Daniel P., and Claire T. Farley. 1997. "Interaction of leg stiffness and surface stiffness during human hopping." *Journal of applied phsyiology* 82 (1): 15–22.

Girard, Olivier, Gregoire Millet, Jean Slawinski, Sebastien Racinais, and Micallef Jean-Paul. 2013. "Changes in Running Mechanics and Spring-Mass Behaviour during a 5-Km Time Trial." *International journal of sports medicine* 34.

Golubnitschaja, Olga, Babak Baban, Giovanni Boniolo, Wei Wang, Rostyslav Bubnov, Marko Kapalla, Kurt Krapfenbauer, Mahmood S. Mozaffari, and Vincenzo Costigliola. 2016. “Medicine in the early twenty-first century: paradigm anticipation – EPMA position paper.” *EPMA Journal* 7 (23). Doi: 10.1186/s13167-016-0072-4.

Griffin, Paul M., Christopher J. DeFlitch, Harriet Nembhard David A. Munoz, Hyojung Kang, and Nathaniel D. Bastian. 2016. *Healthcare systems engineering.* New Jersey: John Wiley & Sons, Inc.

Hulme, A., P. M. Salmon, R. O. Nielsen, G. J. M. Read, and C. F. Finch. 2017a. “Closing Pandora’s Box: adapting a systems ergonomics methodology for better understanding the ecological complexity underpinning the development and prevention of running-related injury.” *Theoretical Issues in Ergonomics Science* 18 (4): 338–359. Doi: 10.1080/1463922X.2016.1274455.

Hulme, Adam, and Caroline Finch. 2015. “From monocausality to systems thinking: a complementary and alternative conceptual approach for better understanding the development and prevention of sports injury.” *Injury Epidemiology* 2: 1–12. Doi: 10.1186/s40621-015-0064-1.

Hulme, Adam, Scott Mclean, Paul Salmon, Jason Thompson, Ben R Lane, and Rasmus Oestergaard Nielsen. 2019. “Computational methods to model complex systems in sports injury research: Agent-Based Modelling (ABM) and Systems Dynamics (SD) modelling.” *British Journal of Sports Medicine* 53 (24): 1507–1510. Doi: 10.1136/bjsports-2018-100098.

Hulme, Adam, Paul Salmon, Rasmus Nielsen, Gemma Read, and Caroline Finch. 2017b. “From control to causation: Validating a ‘complex systems model’ of running-related injury development and prevention.” *Applied Ergonomics* 65: 345–354. Doi: 10.1016/j.apergo.2017.07.005.

Hulme, Adam, Jason Thompson, Rasmus Nielsen, Gemma Read, and Paul Salmon. 2018. “Towards a complex systems approach in sports injury research: Simulating running-related injury development with agent-based modelling.” *British Journal of Sports Medicine* Doi: 10.1136/bjsports-2017-098871.

Kattman, U. 2018. “A biologist’s musing on teaching about entropy and energy: towards a better understanding of life processes.” *SSR* 99 (368).

Kenny, G P, and R McGinn. 2017. “Restoration of thermoregulation after exercise.” *Journal of Applied Physiology* 122: 933 – 944. Doi: 10.1152/japplphysiol.00517.2016.

Khalil, I, and L Sornanathan. 2010. “Fitness monitoring system based on heart rate and SpO2 level.” In *Proceedings of the 10th IEEE International Conference on Information Technology and Applications in Biomedicine*, 1–5. IEEE.

Kirkwood, C W. 2013. *System Dynamics Methods: A Quick Introduction.* Arizona State University, United States: Creative Commons.

Kosmidis, I, and L Passfield. 2015. *Linking the performance of endurance runners to training and physiological effects via multi-resolution elastic net.* Technical Report arXiv:1506.01388. Cornel University Library.

Kulmala, Juha-Pekka, Jukka Kosonen, Jussi Nurminen, and Janne Avela. 2018. “Running in highly cushioned shoes increases leg stiffness and amplifies impact loading.” *Nature* 8: 1–7. Doi: 10.1038/s41598-018-35980-6.

Kunc, M. 2019. “Strategic Planning: The Role of Hybrid Modelling.” In *Winter Simulation Conference*, December, 1280–1291. Doi: 10.1109/WSC40007.2019.9004881.

Lieberman, Daniel E., Anna G. Warrener, Justin Wang, and Eric R. Castillo. 2015. “Effects of stride frequency and foot position at landing on braking force, hip torque, impact peak force and the metabolic cost of running in humans.” *Journal of Experimental Biology* 218 (21): 3406–3414. Doi: 10.1242/jeb.125500.

Maani, KE, and RY Cavana. 2007. *Systems thinking, system dynamics: managing change and complexity.* New Zealand: Pearson.

Macal, C, and Michael North. 2010. "Tutorial on agent-based modelling and simulation." *Journal of Simulation* 4: 151–162. Doi: 10.1057/jos.2010.3.

Martin, Romina, and Maja Schlüter. 2015. "Combining system dynamics and agent-based modeling to analyze social-ecological interactions: an example from modeling restoration of a shallow lake." *Frontiers in Environmental Science* 3: 66. Doi: 10.3389/fenvs.2015.00066, https://www.frontiersin.org/article/10.3389/fenvs.2015.00066.

McLean, Scott, Gemma J. M. Read, Adam Hulme, Karl Dodd, Adam D. Gorman, Colin Solomon, and Paul M. Salmon. 2019. "Beyond the Tip of the Iceberg: Using Systems Archetypes to Understand Common and Recurring Issues in Sports Coaching." *Frontiers in Sports and Active Living* 1: 1–12. Doi: 10.3389/fspor.2019.00049.

Moore, Isabel. 2016. "Is There an Economical Running Technique? A Review of Modifiable Biomechanical Factors Affecting Running Economy." *Sports Medicine* 46: 1–15. Doi: 10.1007/s40279-016-0474-4.

Morin, J B. 2018. *Biomechanics of Training and Testing*, book A Simple Method for Measuring Lower Limb Stiffness During Running. Springer, Cham.

Napier, C, J F Esculier, and M A Hunt. 2017. "Gait re-training: out of the lab and into the streets with the benefit of wearables." *British Journal of Sports Medicine* 0.

Navrátil, V. 2011. "Health, ageing and entropy." *School and health* 21: 329 – 335.

Passfield, L, and J G Hopker. 2016. "A Mine of Information: Can Sports Analytics Provide Wisdom From Your Data?" *International journal of sports physiology and performance* 12 (7): 851—855. Doi: http://dx.doi.org/10.1123/ijspp.2016-0644.

Salmon, P.M. 2017. "Ergonomics issues in sport and outdoor recreation." *Theoretical Issues in Ergonomics Science* 18 (4): 299–305. Doi: 10.1080/1463922X.2017.1300355.

Sands, W A, A A Kavanaugh, S R Murray, J R McNeal, and M Jemni. 2017. "Modern Techniques and Technologies Applied to Training and Performance Monitoring." *International Journal of Sports Physiology and Performance* 12: S2 63–72. Doi: 10.1123/ijspp.2016-0405.

Senge, Peter M. 1990. *The Fifth Discipline: The Art And Practice of the Learning Organization.* New York: Doubleday/Currency.

Solomon, E P, L R Berg, and D W Martin. 2002. *Biology.* 6th ed. London: Thomson Learning.

Sterman, John D. 2000. *Business Dynamics Systems Thinking and Modeling for a Complex World.* United States of America: McGraw-Hill Higher Education.

Swan, Melanie. 2013. "The Quantified Self: Fundamental Disruption in Big Data Science and Biological Discovery." *Big Data* 1: 85–99. Doi: 10.1089/big.2012.0002.

Vermeulen, Euodia, Sara S. Grobbelaar, and Adele Botha. 2020. "Conceptualising a systems thinking perspective in sport studies." *Theoretical Issues in Ergonomics Science* 0: 1 – 16. Doi: 10.1080/1463922X.2020.1788662.

Wakeland, Wayne, and Megan Hoarfrost. 2005. *The Case For Thoroughly Testing Complex System Dynamic Models.* Systems Science Faculty Publications and Presentations. https://pdxscholar.library.pdx.edu/sysc_fac/78, [Online]. Accessed 13 April 2020.

Wang, Lin, Youlian Hong, Jing-Xian Li, and Ji-He Zhou. 2012. "Comparison of Plantar Loads During Running on Different Overground Surfaces." *Research in Sports*

Medicine 20 (2): 75–85. Doi: 10.1080/15438627.2012.660816.
Wilson, J.R. 2014. "Fundamentals of systems ergonomics/human factors." *Applied Ergonomics* 45 (1): 5–13. Doi: 10.1016/j.apergo.2013.03.021.

7. Supplementary material

7.1. *Full causal loop model*

The runner's training cycle in the runner's causal loop diagram reads as follows: *Training pulse* inflicts some micro-trauma to structures, thereby adding to *micro-damage. Micro-damage* adds to the internal *entropy* of the runner through increased *number of fibres that are now in some state of disarray.* Internal *entropy* adds to the *damage* accumulated by the body, which increases the *damage:structure ratio* (also influenced by a *limiting factor*). The *damage:structure ratio* (E_n from Equation 1 in § 3.2.1) represents the relationship between damaged tissue and the state of integrity of tissues. Once this ratio has exceeded a limit from accumulated damage, *cell (structural) death* occurs, representing an injury to structure. The injury takes away from the *integrity of structures* and negates the fitness and health of the runner, whereas maturity of structures lowers the *damage:structure ratio.*

Internal entropy also leads to *restorative actions* taken by the body, of which two modes are differentiated as a function of time: an immediate restorative action by the body after the physical activity to return to homeostasis and a latent component, whereby the body's physiological processes set in to repair micro-damage after the activity. The restoration delay increases with damage build-up and is associated with the environmental temperature and the heart rate-zone in which training took place. Restored integrity of structures lowers the entropy, and as the runner becomes fitter, the restorative actions stabilise due to the physiological adaptations in the runner.

The body's response to micro-damage as a result from the training pulse is to form *new structures* to augment structural integrity to deal with the higher loads imposed by the higher levels of physical activity. *Exercise therapy* through strength and conditioning regimens adds to the new structure formation and reinforces the structural integrity. Mature structure closes the gap between the *structural (fitness) goals* of the runner and the current level of structure, and subsequently the intensity of training may increases, thereby increasing the training pulse. As the gap closes, the training load may increase until the gap is closed and the training load will stabilise.

7.2. *Full simulation model explanation*

The athlete makes individual training decisions through a state chart (the training cycle, Figure 11 is reproduced here for easier referral) and trigger functions in the agent-based component. Interactions between the athlete's biomechanical and physiological elements with the environment (leg stiffness, vertical oscillation, cadence, heart rate, temperature, and surface type) are modelled on the micro level through functions within the state chart. The behaviour over time and cluster analyses of the tracking data obtained from the running wearable pointed to micro level interactions between the surfaces encountered, cadence, and heart rate. Interactions between the surface, leg stiffness, and vertical oscillation are based on literature (Ferris and Farley, 1997), specifically the runner as a spring-mass model (§ 2.3). Interaction between temperature and recovery were found in the literature (Kenny and McGinn, 2017).

The state chart `stcTrainingStatus` models the athlete's transition between resting and training (Figure 11, reproduced here for easier referral). The state chart is nested: at the highest levels the athlete is either healthy (in state `stHealthy`) or injured (in state `stInjured`). The lower sub-levels of `stHealthy` consist of active training (`stTraining`) or at rest (`stResting`). Only while in the healthy status may an athlete

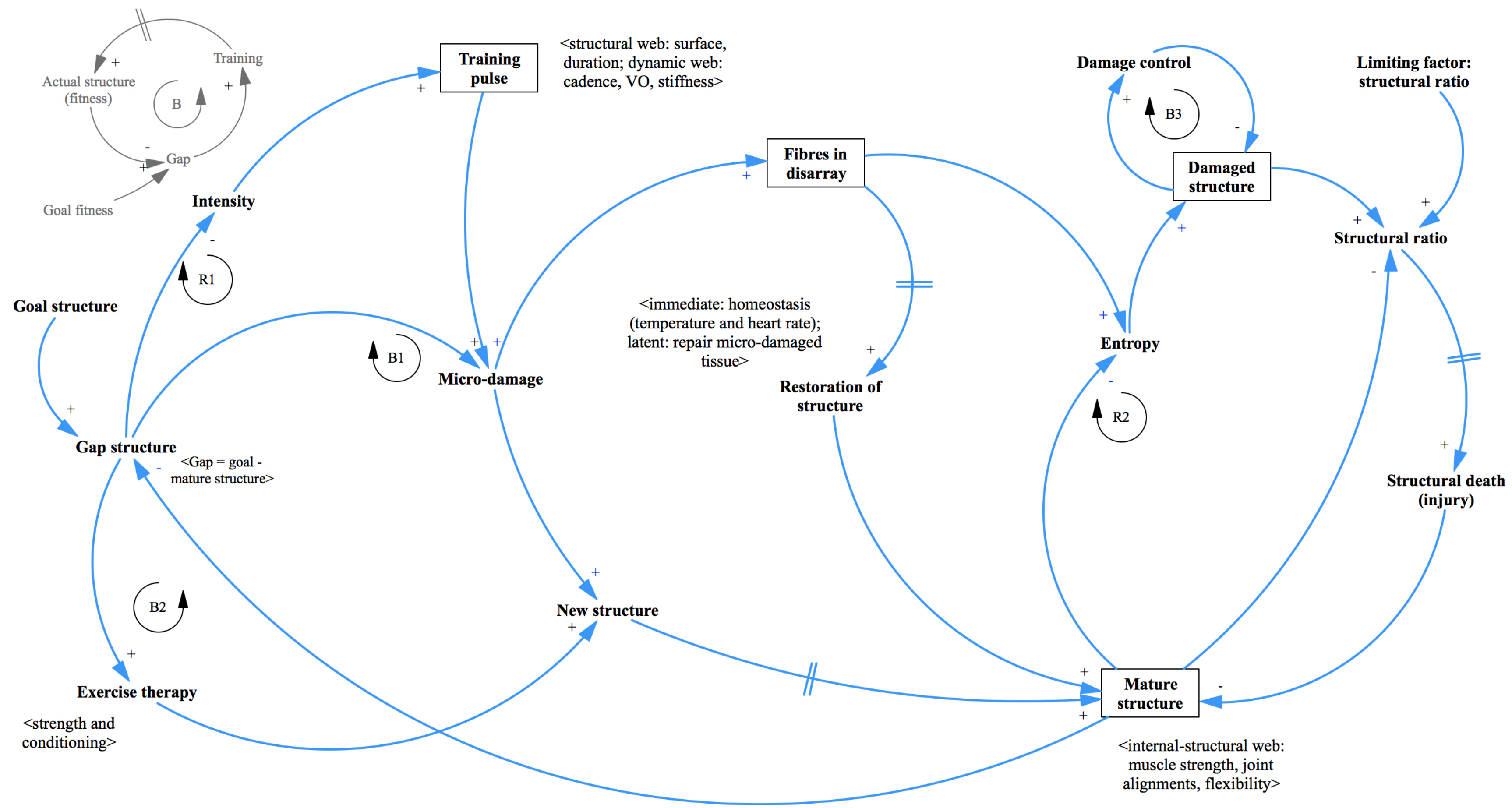


Figure 10.: The causal loop diagram for the runner

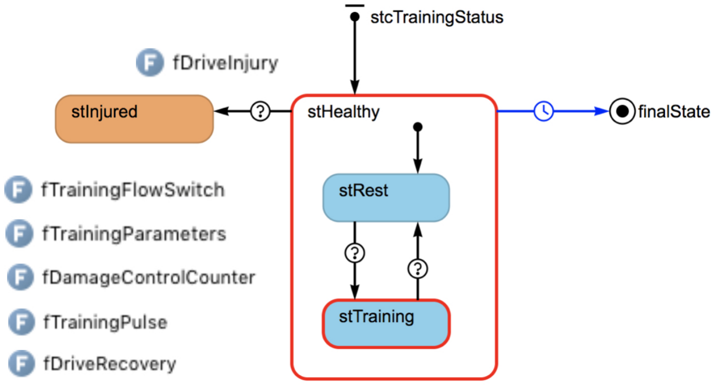


Figure 11.: The training status statechart

train, once injured the training stops. The transition to `stInjured` is modelled as a condition, which is based on the time spent in the damage zones and managed by a return function, `fDriveInjury`. The functions `fTrainingParameters`, `fTrainingPulse`, `fDamageControlCounter` are all called upon in the `stRest` to enforce the runner's decision rules for the upcoming training session and to deal with damage. Recovery rules are coded into a return function, `fDriveRecovery` and adapts staggered parameters for each damage zone to quantify a continuous dynamic variable, `dvDelayRepair` that manages recovery and conditioning over time. The entire simulation allows for 720 days of transitioning between training and resting, before transitioning to the final state and the simulation ends.

The stock-and-flow model in Figure 12 is the quantified, dynamic abstraction from the qualitative runner's causal loop diagram in § 4.1. Pressure points develop as a result of unwanted accumulation of damage. Actions taken by the athlete (governed through the decision rules in the agent-based component) opens and closes the valves (flows) into and out of the accumulation tanks to relieve pressure or divert content.

Once per day, the training flow, `flowTraining` is activated to pulse the wave function that characterises the spring-mass system of a moving runner (Equation 2). The function, `fTrainingPulse` calculates the size of the pulse before the training session starts. The `flowTraining` is activated and deactivated by the function `fTrainingFlowSwitch`, which monitors the position of time in the simulation as to when each training session starts and ends. The total, positive impulse generated during the training session is absorbed as the training load, into stock `stkTrainingLoad`. The gap between `stkStructure` and the goal structure, `pGoalStructure`, is continuously updated in the dynamic variable, `dvGapStructure`, and is calculated as the difference between them: $\texttt{dvGapStructure} = \texttt{pGoalStructure} - \texttt{stkStructure}$.

The path to injury is a two-step process. First, net entropy (in `dvEntropy`) is a fractional relational measure between temporary disarrayed fibres (internal entropy), as a result of micro-damage from training, and existing structure. This calculation is the first pressure point, derived from the entropic line gauge in § 3.2.1). The `dvEntropy` updates continuously to keep track of the relationship between the number of fibres in disarray and structural integrity. At some threshold, e_t, a portion of fibres in disarray is considered damaged and is then added to the damage stock at a rate of $\alpha_d\%$.

As such, `dvEntropy` controls the flow into the stock for damaged structure, stkDamage. In the second step towards injury, when dvEntropy reaches a certain threshold, the flowDamage is activated and takes as value a fraction of the stkFibresDisarray. The outflow from stkDamage, flowDamageOut, is additive, accounting for natural removal of damaged tissue and damage control measures taken by the athlete.

The $^d/_s$-ratio considers the amount of damaged fibres to the amount of structural integrity, and serves as a measure to gauge the cumulative damage from overload. This is the second pressure point in the plumbing system. Two damage zones exist, with the upper and lower limits for each zone demarcated using the $^d/_s$-ratio. An athlete may spend some time in these damage zones, however it is the cumulative time in the damage zone that will ultimately incite the injury. The damage zones are defined as lower and higher risk.

The dvRatioDamageStructure controls the flow of the damage zone drivers through associated dynamic variables: either flow, flowDamageZone1 or flowDamageZone2, depending on the magnitude of dvRatioDamageStructure, is activated and flows at a rate one unit-day per simulation-day to add to the stock for the damage zones (stkDamageZone1 or stkDamageZone2). An injury is incited when a certain number of days has accumulated in the damage zone stocks. An injury is incited by instantaneous emptying the content of stkStructure: accelerating the flow of cell death, flowCellDeath by shortening the delay, pDelayStructureDeath to one day.

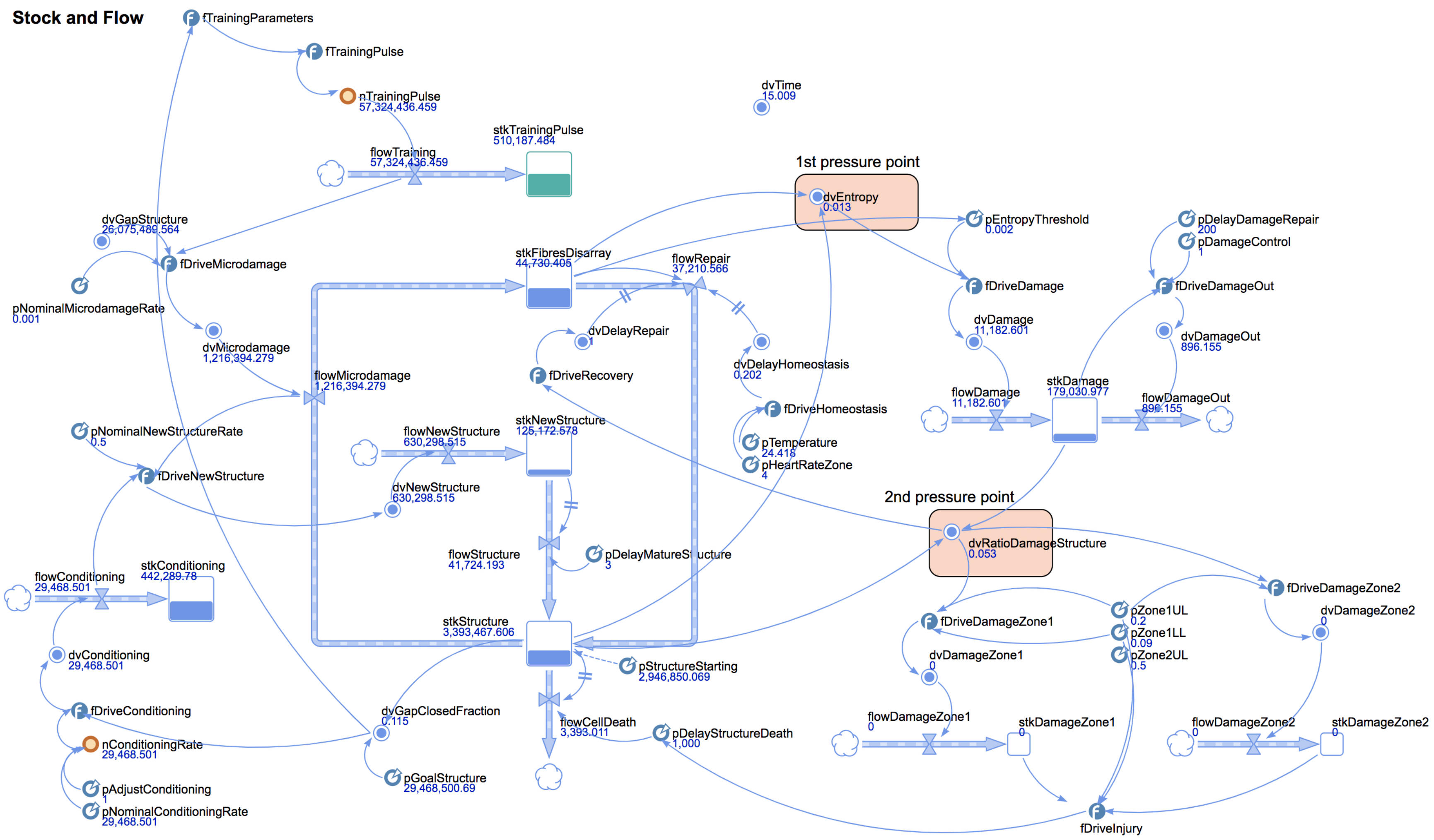


Figure 12.: The detailed stock-and-flow model of the runner. (Not all the arrowed lines in the figure are functional as per traditional stock-and-flow configurations. In the *Anylogic* computational configuration, arrows between driving functions, and parameters, flows, stocks, and variables are not necessary. However, they are shown in the figure to provide the reader with a visual of the closed-loop, 'pressurised plumbing system'.)

7.3. *Simulation input parameters*

Tables 2 and 3 contain the parameters for the simulation. The running type is used as a proxy for the surface. The collection of running types (rc - road racing, rr - road running, - tc - trail racing, tr - trail running, tt - track training on grass) are the starting point from where the other running parameters are set up, based on literature and empirical work. The nominal conditioning rate is set at 1.2% of the initial training impulse.

Table 2.: Input training parameters

Parameter (`name in Anylogic`)	**Expression in *Anylogic* and units**
Primary data	
Surface (`pSurface`)	$j \in \{rc, rr, tc, tr, tt\}$
Cadence (frequency) (`pFrequency`)	$\sim N(\mu_j, \sigma_j)$ steps.s^{-1} for surface j
Heart rate (`pHeartRate`)	Follow classification framework, Figure 4
Interactivity time (`nTrainingIntervalTime`)	0.958 days (23 hours)
Temperature (`pTemperature`)	$\sim N(\mu_m, \sigma_m)$°C for month m
Secondary data	
Vertical oscillation (amplitude, x_m) (`pXm`)	$\sim N(\mu_j, \sigma_j)$ meter for surface j
Leg stiffness (`pStiffness`)	$\in \{20, 20, 30, 30, 40\} kN/m$ for surface j
Illustrative data	
Training duration (`pTrainingDuration`)	1 hour per day: 0.0417 days
Nominal conditioning (`pNominalConditioningRate`)	29469 (unit-less)
Fitness zone limits (upper and lower)	zone 1: $(0.25, 0.5)$ zone 2: $(0.5, 0.75)$

Table 3.: Illustrative input parameters for structure management and delays. These values were calibrated during sensitivity analysis and extreme condition tests.

Parameter (`name in Anylogic`)	**Expression in *Anylogic* and units**
Delay repair per damage zone (`pDelayRepair`)	$\in \{1, 1.5, 2\}$ days
Delay mature structure (`pDelayMatureStructure`)	3 days
Nominal new structure rate (`pNominalNewStructureRate`)	0.5 (unit-less)
Nominal micro-damage rate (`pNominalMicrodamageRate`)	0.001 (unit-less)
Delay: natural cell death per damage zone (`pDelayStructureDeath`)	$\in \{1000, 800, 800\}$ days
Entropy threshold (`pEntropyThreshold`)	0.002 (unit-less)
Damage zone limits (upper and lower)	zone 1: $(0.09, 0.2)$ zone 2: $(0.2, 0.5)$
Nominal damage control per damage zone (`pDamageControl`)	$\in \{1, 4000, 5000\}$ (unit-less)
Increase training duration (`pIncrTrainingDurationSwitch`)	$\in \{0, 1\}$ (unit-less)