Connecting micro dynamics and population distributions in system dynamics models

Saeideh Fallah-Fini [1] (sfallah@jhsph.edu), **Hazhir Rahmandad** [2] (hazhir@vt.edu), **Hsin-Jen Chen**[1] (hschen@jhsph.edu), **Youfa Wang**[1] (yawn@jhsph.edu)

1 Johns Hopkins Global Center on Childhood Obesity, Center for Human Nutrition, Department of International Health, Bloomberg School of Public Health, Baltimore, MD

2 Grado Department of Industrial and Systems Engineering, Virginia Tech, Northern Virginia Center, Falls Church, VA

Abstract

Common system dynamics models capture the mean behavior of groups of indistinguishable population elements (e.g. people, tasks, widgets) aggregated in stock variables. However, many modeling problems require capturing heterogeneity across these elements with respect to some attribute(s) (e.g. weight, errors, price). The representation of heterogeneity could be important for correct characterization of behavior of a system as well as evaluation of policy options. In this paper we develop a method to connect micro-level dynamics (associated with elements in the population) with macro-level population distribution along an attribute of interest. The method enables modelers to efficiently characterize the distribution of attribute of interest without explicitly modeling all the elements in the population. We apply our method for modeling distribution of Body Mass Index and its changes over time in a sample population of 3074 female adults obtained from the National Health and Nutrition Examination Survey (NHANES) data. Comparing our results with the ones obtained from an agent-based model that captures the same phenomena shows that our method offers good precision with computational costs that are significantly less than agent-based models.

1. Introduction

When modeling a population of elements (e.g. people, tasks, widgets, etc.) that are heterogeneous with respect to some attributes (e.g. weight, skill level, productivity, etc.), system dynamics modelers usually focus on capturing the mean of attributes of the population aggregated in stock variables. However, many modeling problems require capturing the heterogeneity and variation of attributes across population elements (Osgood, 2004). For example, in a study that aims to evaluate the effect of a weight loss intervention on a population, we may be interested to see how the intervention impacts people with different "weights or Body Mass Index (BMI)". Similarly, we may be interested in how software developers with different "skill levels" have different outputs as opposed to looking at the output associated with average skill level of developers in a company; how students with different "intelligence levels" have different performance; how social networks with different "number of network connections" have different power/effects ; how athletes with different "age" have different performance; and also how different levels of "health indicators" (e.g., cholesterol, blood pressure, weight, BMI, etc.) can lead to different health outcomes associated with people in a population.

In all these scenarios the heterogeneity of population with respect to its attributes/characteristics is likely to have a strong impact on system behavior. Thus, modelers are interested in capturing the distribution of the characteristics of interest in a population of elements and not only the mean of those characteristics. This is because either the impact of average of the attribute on system evolution is not obvious (due to non-linearity), or we care about extreme cases in population (tails of distribution). For example software developers with very high skills show much higher productivity or people with high BMI (i.e. obese people) are more prone to severe health risks.

In those problem domains that have the above characteristics, modelers are more interested in capturing variations across the population elements. To do so, aging chains are commonly used in the system dynamics literature. In this approach, the elements are broken down to different population groups based on their attribute value (range) and are represented by different stocks. These fundamental stocks are then strung together to form a chain. Aging chains can be used for tracking characteristics of a population (Sterman, 2000), for example tracking experience and promotion chain of employees, or tracking aging of equipment in a company.

Aging chains have a very simple conceptual structure and can easily be implemented. However, aging chains are generally used for capturing characteristics which do not influence the propensity of an element to move across different groups. Thus the movement of an element from one stock to another is a fraction independent of the attributes based on which we have divided the stocks. A common example is the use of aging chains to capture aging of a population. Here each element in each stock has a constant rate of movement to the next stock which can be directly determined based on the average duration of stay in that stock.

However in many interesting real-world problems the movement of population elements across different percentiles on the distribution of attributes of interest is a function of where an element is located on the distribution. For example, current skill of employees influences their skill acquisition rate; current number of links on a social network influences rate of link acquisition; and current weight of an individual influence the rate of weight gain and loss (due to the effect of current weight on energy expenditure of individuals).

The current practice in such settings (where the propensity of each element to move from one population group to another is a function of the characteristic based on which the population groups are specified) is to formulate an aggregate function that controls the transition rate from one group to another without considering the exact micro (i.e. element-level) dynamics that generate those movements. System Dynamic models in obesity domain (for example see the model by Homer et al. (2006)) usually use this approach for capturing movement of an individual from a population group that represents a specific BMI range to another population group that represent a different BMI range. The challenge with the current practice is that we don't have any solid way to formulate the aggregate flow based on the physics of the problem. While the literature can provide some fundamental mechanisms at the micro level which could be captured in an individual level model, macro rates remain rather ad hoc and hard to establish without calibration.

In response to this challenge, many modelers prefer to switch to using agent-based model architectures which allow for capturing the micro mechanisms directly. However, agent-based approach faces computational challenges for very large populations (for example population of a country). This computational expense may reduce the ability of the modeler to conduct adequate calibration and sensitivity analysis and may bog down the project in details not central to the dynamics of interest.

The goal of this paper is to provide a consistent method for connecting micro-dynamics and macro distributional outcomes without the need to use the detailed agent-based models. More specifically, this research targets the problems with the following characteristics:

- We are modeling a population of elements, where we care about the dynamics of their distribution along dimension (attribute) ***d*** (e.g., weight, skill, social capital, etc.).
- We have a mechanism based model at the individual element level, e.g., a model of weight gain and loss for individuals, a model of individual skill acquisition, a model of propensity to gain social capital for individuals.
- The individual level mechanisms that control the movement of the population elements on dimension ***d*** are a function of ***d*** itself. For example, the weight gain/loss of individuals depends on their current weight, since energy expenditure of individuals depends on their weight in the forms of fat mass and fat free mass; skill acquisition rate in employees depends on their current skill level, since less and less skills are retained as the skill level of employees is increased; current number of links on a social network increases the rate by which one can meet other people, thus increases the rate of link acquisition.

After describing our method, we provide a concrete empirical example regarding application of our method to modeling distribution of BMI associated with a sample of female adult population and will present how the distribution of BMI of this population will change due to a shock in the energy intake of the whole population. To capture the micro-level mechanism associated with weight gain and loss in individuals (elements of our population), we use the model for human metabolism and body weight change developed by Hall (2010). To show that this mechanism works correctly, we compare and contrast our method to precise results coming from an agent-based model of the same phenomena.

The rest of this paper is as follows. Section 2 describes the method. Section 3 describes the experimental set up regarding application of the developed method to our empirical dataset. Section 4 discusses the results. Section 5 concludes and discusses areas for future research.

2. Method

Consider a system that includes a population of n elements (e.g., people, tasks, etc.), each of which exhibits some values/levels along dimension (attribute) d (e.g. weight, skill, etc.). We choose to disaggregate this population group within several stocks based on their attribute values. For continuous attributes such as Body Mass Index (BMI), each stock contains members of the population whose attribute values fall within a specific range of the attribute of interest (e.g., we may have a stock containing people whose BMI is between 25 and 30). Those population members who change their attribute values (e.g., lose or gain weight, acquire some skills, etc.) will contribute in flows between stocks of sub-populations (macro population groups).

To capture the transition rates between macro population groups, we first allocate a representative agent to each group. The representative agents present average of the individuals in each sub-population group. We then find the transition rates between population groups based on the individual-level dynamics that regulate the representative agents of those groups. To do so, we think of the representative agent in each group to be moving on dimension d based on the micro dynamics, but rather than letting that simulated representative agent to move accordingly, we use the resulting rate of change for that agent to approximate the resulting rate of movement of population elements across different stocks. The detailed steps of our method are as below:

Step 1: Decide on the divisions needed along the dimension (attribute) *d* for disaggregating the population into sub-groups (macro population groups)

At this step, the domain of attribute of interest (dimension d) is divided into several distinct ranges (not necessarily equal) represented by distinct stocks. Each one of these stocks holds a sub-group of population whose attribute values fall within the attribute range associated with that stock. For example, Figure 1 shows the distribution of an attribute in a population. The population has been disaggregated into M distinct sub-group. P_k represents frequency of population elements in group k; and Xi_k and Xf_k represent the initial and final values of the range of attribute associated with population group k, respectively.

We use “subscripting” feature of Vensim as a convenient way for disaggregating the population into distinct stocks according to their attribute range. Let subscript G (g1, g2, …, gM) represents the vector of M stocks, each representing a sub-population. Note that the subscribing technique is conceptually equivalent to using M explicit stocks in the model.

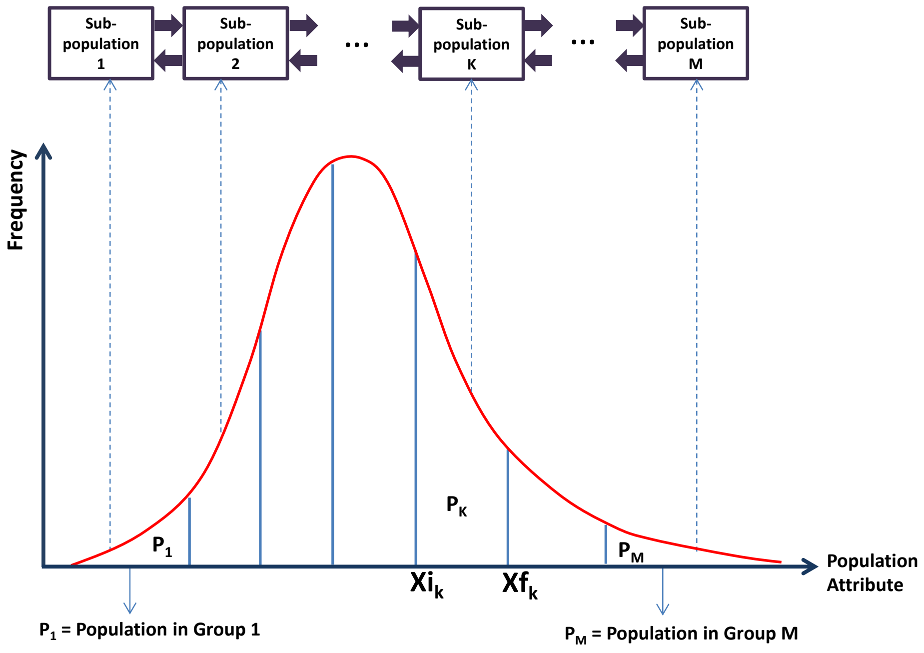

Figure 1: Disaggregation of population into macro population groups

Step 2: Estimate distribution of population elements along dimension (attribute) *d*

To estimate the distribution of population along dimension ***d***, we assume that individuals in each interval on dimension ***d*** are uniformly distributed along the interval range. Thus, as it is shown in Figure 2, the distribution of a population along dimension ***d*** can be estimated using the vertical bars. Knowing the range of attribute values corresponding to each sub-population group (stock) *k* (i.e., Xi_k and Xf_k), as well as the population associated with that group (P_k), we can calculate the height of the vertical bar associated with sub-population group *k* as below:

$$Y_k = \frac{P_k}{Xf_k - Xi_k} \tag{1}$$

By choosing narrower ranges (widths) for attribute values associated with each sub-population, we can get a better estimate of the true distribution of population along that attribute. However, one should note that as number of distinct ranges that is defined for an attribute increases, the number of distinct stocks increases rapidly and this may imposes some computational cost. Moreover, availability of data and the number of elements that are expected to be in each group also affect the size of the intervals.

Next, we allocate a representative agent to each sub-population group. The representative agent of each group represents average of individuals in that group with respect to attribute under analysis. Defining representative agents would enable us to connect the individual-level (micro) dynamics to the population distribution.

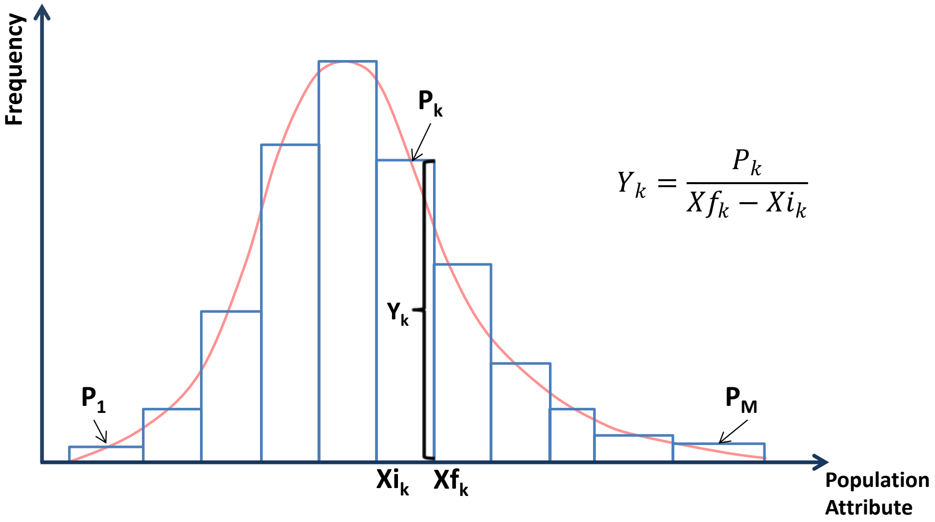

Figure 2: Estimating distribution of population elements along dimension d

Step 3: Develop the individual-level mechanism based model (Micro dynamics associated with representative agents)

At this step a micro-level dynamic model for an individual element (representative individual) is developed. For example, if one is modeling distribution of population with respect to attribute "Body Mass Index", the individual-level model represents dynamics of body weight gain/loss, and consequently changes in BMI of an individual. Similarly, if one is interested in modeling distribution of employees of an organization with respect to their skill level, the individual-level model would capture the dynamics associated with skill acquisition and decay in an employee; and also if one is interested in modeling distribution of social networks in a country with respect to the number of network connections they have, the individual-level model could capture the dynamics associated with social networking and link acquisition.

After developing the micro-level dynamic model associated with representative agents, we also need to identify the initial conditions associated with ***M*** parallel micro-level models each of which representing a sub-population. The attribute values of representative individuals can be used to set the initial condition of the micro-level models of representative individuals. The attribute value associated with each representative individual is calculated as the average of the initial (Xi_k) and final (Xf_f) values of the attribute range of each population group. For example, in our empirical example, we calculate Body Mass Index of representative individual in each group as the average of the initial and final BMI of that group. The Body Mass Index of representative individual is then used along with average height of that group to calculate body weight of representative individual that is needed as the initial condition for the micro-level model of body weight dynamics.

Step 4: Perform the following tasks at each time step:

Task 1: Calculate the rate of change for each representative individual using micro-level dynamics

Given the exogenous factors associated with micro-level dynamics of individuals in a population, at any time step, the rate of change of representative individual of each sub-population along dimension ***d*** can be calculated. Let $\boldsymbol{dAtt_k}$ shows the rate of change associated with representative individual of sub-population group *k* at any time step. This rate is used in the next step to calculate the number of individuals that go from one sub-population to another. In our empirical example, this step is equivalent to calculating how much BMI associated with each representative individual changes in each time step due to the imbalance between energy intake and energy expenditure of individuals calculated inside the micro-level model of body weight change.

Task 2: Calculate the rate of transition of individuals between macro population groups based on individual's rate of change

The rate of change of representative individuals (***dAtt***) in each sub-population calculated in previous step, gives us the speed by which population elements move from that sub-population to the neighbor sub-populations. Positive values for the rate of change of representative individual of group *k*, as an example, imply that some of the population elements in group *k* will move to group *k+1*. However, if representative individual of group *k* shows a negative rate of change, this implies that some of the individuals in group *k* will move to group *k-1*. In any case, the rate of population leaving population group *k* is a function of rate of change of representative individual of group *k* along dimension **d** (i.e., $\boldsymbol{dAtt_k}$) and vertical dimension of rectangle associated with group *k* (i.e., $\boldsymbol{Y_k}$) and is calculated as shown in Equation (2). $\boldsymbol{Y_k}$ in this equation represents the frequency (density) of individuals in group *k* that have the potential to move to group *k+1* (if $\boldsymbol{dAtt_k}$ is positive) or group *k-1* (if $\boldsymbol{dAtt_k}$ is negative).

$$\begin{aligned}&Rate\ of\ population\ leaving\ group\ k\ to\ group\ k+1 = Max\ (dAtt_k * Y_k\ , 0)\\&Rate\ of\ population\ leaving\ group\ k\ to\ group\ k-1 = ABS\ (Min\ (dAtt_k * Y_k\ , 0))\end{aligned} \tag{2}$$

where "ABS" means absolute value of the term inside parenthesis. Finally, to calculate the final rate of change in population of group *k* in each time step, we not only need to consider how many individuals leave sub-population *k* (to group *k* or *k+1*), but also how many individuals join sub-population *k* (from groups *k-1* and/or *k+1*).

Figure 3 shows a schematic representation of transition of population elements between adjacent population groups *k-1*, *k*, and *k+1*. According to Equation (2), total number of population elements that leave group *k* to group *k+1* or *k-1* (represented by solid arrows in Figure 3) can be calculated as: $Max\ (dAtt_k * Y_k\ , 0) + ABS\ (Min\ (dAtt_k * Y_k\ , 0))$. On the other hand, population elements that join group *k* are coming from group *k+1* and/or group *k-1*. Using a similar logic, total number of population elements that may leave group *k+1* to group *k* (represented by patterned arrow going from population group *k+1* to group *k*) is $ABS\ (Min\ (dAtt_{k+1} * Y_{k+1}\ , 0))$. Moreover, total number of population elements that may leave group *k-1* to group *k* (represented by patterned arrow going from population group *k-1* to group *k*) is $Max\ (dAtt_{k-1} * Y_{k-1}\ , 0)$. In sum, rate of change in population of group *k* at any time step is calculated as below:

$$\begin{aligned}&Rate\ of\ change\ in\ population\ of\ group\ k = Inflow - Outflow =\\&(Max\ (dAtt_{k-1} * Y_{k-1}\ , 0) + ABS\ (Min\ (dAtt_{k+1} * Y_{k+1}\ , 0)) - (Max\ (dAtt_k * Y_k\ , 0) +\\&ABS\ (Min\ (dAtt_k * Y_k\ , 0))\end{aligned} \tag{3}$$

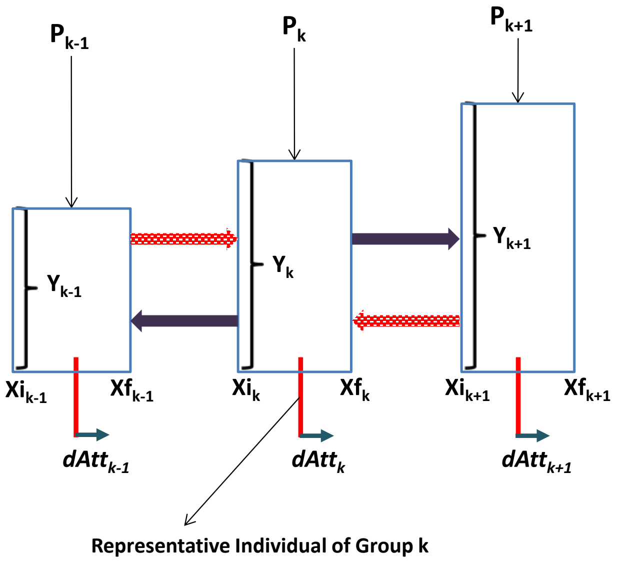

Figure 3: Schematic representation of transition of population elements among adjacent groups[1]

One important note here is that the rate of change of representative person in group *k* is only used for calculating how many individuals go from group *k* to group *k+1* or *k-1*. This rate does not actually lead to a change in the position (state) of representative individual *k* along dimension ***d***. This is due to the fact that representative individual of group *k* at any point of time represents population in group *k*. Those individuals that leave group *k* or join group *k* (from other groups) at any time step only change the percentage of population in group *k* in the next time step. They do not affect the position of representative individual of this group since the initial (Xi_k) and final (Xf_f) values associated with attribute range of group *k* is not changing.

Task 3: Estimate the new distribution of population elements along dimension (attribute) *d*

Using the rate of change in population in each group (calculated in previous task), the new values associated with population in each group at the start of the next time step is calculated. These new values for population in each group (i.e., P_k) at the start of the next time step would allow us to develop a new estimation for the distribution of population along dimension ***d*** using the method described in Step 2 (i.e., calculating Y_k).

3. Experimental Setup: Modeling distribution of a sample population with respect to their Body Mass Index

We apply the developed method for capturing the heterogeneity of a human population with respect to their Body Mass Index (BMI) and how population individuals move from one weight group to another based on the imbalance associated with their energy intake and expenditure. We use an empirical distribution associated with 3074 female adults aged 18-74 over years 1999-2003 obtained from the National Health and Nutrition Examination Survey (NHANES) data to set the initial condition of the simulation.

---

[1] Note that in any two adjacent groups *k* and *k+1*, we have $Xf_k = Xi_{k+1}$.

Step 1: As the first step, we need to disaggregate the population into several sub-population groups based on their BMI. To do so, we looked at the histogram of the BMI of the female adults and finally defined 17 distinct ranges for BMI, namely, [14, 16), [16, 18), [18, 20), [20,22), [22,25), [25,28), [28,30), [30,32), [32,34), [34,36), [36,38), [38,40), [40,42), [42,44), [44,46), [46,50), [50,58), where [x, y) means the interval contains those individuals whose BMI is greater than or equal to x and less than y (i.e., x <= BMI < y). While defining the intervals, we made sure to allow for some cutting points for BMI such as 20, 25, or 30 so that we can later calculate some important statistics, such as percentage of population who are overweight (i.e., their BMI is >=25 and <30). This can be obtained by summing up the values (levels) of the two sub-population stocks associated with ranges [25, 28) and [28, 30).

Step 2: Using the data associated with 3074 female adults, we then set the initial condition associated with frequency of population in all 17 distinct BMI intervals. The population in each group *k* (i.e., $\boldsymbol{P_k}$) along with the range of BMI intervals associated with that group (as defined in Step 1) would allow us to estimate the empirical distribution of BMI at the start of the simulation (i.e., calculating $\boldsymbol{Y_k}$ associated with all groups).

Step 3: In this step we need to construct a micro-level model that can capture the dynamics of weight gain and loss in representative individuals. To do so, we use the individual-level energy model for adults developed by Hall (2010) to capture the energy balance and weight change of individuals over time. Body weight (BW) in Hall's model is represented by two stocks capturing Fat Mass (FM) and Fat Free Mass (FFM) associated with individuals. The change in body weight is then modeled as the result of an imbalance between energy intake (EI) and energy expenditure (EE) of individuals. Energy expenditure is composed of several components. First, the Resting Metabolic Rate (RMR) which is the energy required to perform valid body functions while body is at rest. Second, the energy needs for physical activity. Third, the energy required for digesting food and nutrients consumed (thermic effect of food). Fourth, the energy required for developing new mass (or digesting existing mass). RMR mainly depends on fat mass and fat free mass. Energy expenditure attributed to physical activity depends on body weight and intensity of physical activity. Equation (4) shows the formula associated with energy expenditure in adults adopted from Hall (2010).

$$EE = K + \gamma_L FFM + \gamma_F FM + \delta BW + \beta \Delta EI + \eta_F \frac{dFM}{dt} + \eta_L \frac{dFFM}{dt} \tag{4}$$

where γ_L= 22 kcal/(kg.day), γ_F = 3.2 kcal/(kg.day), $\delta = 7$ kcal/(kg.day) corresponding to an average sedentary person, β = 0.24, η_F = 180 kcal/kg, η_L = 230 kcal/kg. *K* is a constant determined by the initial energy balance condition. Its value was set to 370.21 kcal/day based on the numerical example provided by Chow and Hall (2008).

Change in FM and FFM depend on the difference (imbalance) between EI and EE (EI-EE). The energy imbalance is partitioned to be added (if positive) or deducted (if negative) from FM and FFM stocks. The partitioning function φ for adults is defined as below (Hall, 2010):

$$\varphi = \frac{C}{C+FM} \ , \quad C = 10.4\, kg * \frac{\rho_L}{\rho_F} \tag{5}$$

where ρ_F = 9400 kcal/kg, ρ_F = 1800 kcal/kg. Thus, the time course of weight loss and gain can be obtained by solving the following differential equations related to change in fat mass (FM) and fat free mass (FFM) (Hall, 2010):

$$\begin{aligned}\rho_F \frac{dFM}{dt} &= (1-\varphi)(EI-EE) \\ \rho_L \frac{dFFM}{dt} &= \varphi(EI-EE)\end{aligned} \tag{6}$$

To set the initial condition for micro-level SD model of representative individuals, we need to define the weight (in form of fat mass and fat free mass) associated with representative individuals at the start of the simulation. To do so, first we calculate BMI of representative individual in each group as the average of the initial and final values of BMI range of that group. The BMI of representative individual in each group is then multiplied by the square of height of representative individual in that group to give us the weight of representative individuals. The height of representative individual in each group was estimated as the average of height of individuals in each group obtained from the empirical dataset (this value was approximately equal to 1.62 meters for all groups). After calculating the body weight associated with each representative individual, this value is multiplied by the fat mass fraction of representative individual to give us fat mass and fat free mass of representative individual in each group. Fat mass fraction associated with representative individual in each group was estimated as the average of fat mass fraction of individuals in each group obtained from empirical dataset.

Simulation: We start the simulation from equilibrium (in terms of weight), where energy expenditure of the population is the same as its energy intake. Thus, no change in the distribution of BMI of population is expected over time. Next, we impose a shock in energy intake of the population at a specific time and estimate the change in distribution of population along BMI attribute (i.e., percentage of people in different BMI intervals) over time.

To validate our method, we compare our results with the ones obtained from an agent-based model that simulates the change in BMI of 3074 explicit agents whose initial condition (in terms of body weight and body mass fraction) has been obtained from the empirical dataset. Thus, the population in agent-based model has the same initial condition (in terms of distribution of BMI) as presented in our model. To set the initial fat mass and fat free mass for each individual in agent-based model, we use body weight and fat mass fraction of 3074 adult females obtained from our empirical dataset (NHANES data). We run the agent-based model under similar scenario, i.e., we start the model from equilibrium where energy intake of each individual (agent) is equal to its energy expenditure. Then we impose the same shock in energy intake that we used in our model for the population.

One important part of our model in our empirical example is related to using co-flows to capture the stock of total energy supply (intake) associated with each population group. Co-flows allow us to consistently keep track of the energy supply for each population group and the individuals within it, in parallel to what would happen at the individual-level (agent-based) model. When population is in equilibrium in terms of weight, energy supply (energy intake) is equal to energy demand (energy expenditure) for the whole population and similarly, for each representative individual. Thus, the initial value associated with energy stock of each population group is calculated by multiplying the energy demand of representative individual in each group by the number of individuals in that group. Whenever we change the energy intake of a population group, that change is accumulated in the energy stock associated with that group. Thus, at any point of time we have total energy supply associated with each group in co-flows. By dividing the total energy stock of each group by the number of people in that group, average energy supply for representative individual for each group is obtained, and that energy supply will be used (as energy intake of representative individual) to calculate the rate of change in the BMI of that representative individual, allowing us to find out the rate of change in the respective population group.

As individuals leave or are added to population group stocks, we also need to capture how much energy is removed from or is added to the co-flow energy stocks corresponding to population groups. To do so, we simply try to mimic what happens in terms of change in energy of individuals in an individual-level (agent-based) model. In an individual-level model, when an individual is going out of one group, he is taking his energy, which is reflected in the total energy of that group, with himself. Thus, in our model we need to calculate how much energy intake is taken out from each group's energy stock when an individual from that population group leaves to a neighboring group. Below are the steps needed to be taken at any time step to calculate the energy that is taken out from a population group (say group *k*) to be added to the neighbor group *k+1*, when an individual leaves group *k* to group *k+1* (the same logic is applied to all groups):

1. Calculate the energy required (energy demand) of an individual that is in the border of groups *k* and *k+1.* We call this individual a "marginal individual" associated with groups *k* and *k+1*. The energy demand associated with that marginal individual is calculated as the weighted average of energy demand (energy expenditure) of representative individuals associated with groups *k* and *k+1*. The weights are based on the distance of marginal person from representative individuals in each group, such that representative individual that is closer to marginal person get a higher weight.
   As an example, assume we have two adjacent groups, corresponding to BMI ranges [20, 22) and [22, 26), respectively. Also, assume that energy expenditures of representative individuals associated with these two groups are 1600 kcal/day and 1750 kcal/day, respectively. The distance between marginal individual (that lies in the border of these two groups) and representative individual of first group (that lies in the middle of first group) is one unit, while the distance between marginal individual and representative individual of second group is two units. Thus, the energy expenditure (energy demand) associated with marginal individual is calculated as $\frac{2}{3} * 1600 + \frac{1}{3} * 1750 = 1650\ kcal$.
2. Calculate the ratio of energy supply to energy demand for the representative person of group *k*. Energy supply of each representative individual is obtained by dividing the energy supply stock of that group to the population stock of the group. Energy demand associated with each representative individual is equal to the energy expenditure calculated by the micro-level model.

3. Multiply the energy demand of marginal individual between groups *k* and *k+1* by the ratio of energy supply to demand calculated for representative individual in group *k*. The obtained result is the energy that marginal individual between group *k* and *k+1* takes out to group *k+1* as that individual leaves group *k* to group *k+1*. This energy is multiplied by the number of people that leave group *k* to group *k+1* to calculate the total energy that leaves group *k* to group *k+1* at any time step.

We also need to discuss the steps needed to be taken at any time step to calculate the energy that is taken out from population group *k+1* to be added to the neighbor group *k*, when an individual leaves group *k+1* to group *k* (i.e., losing weight):

1. Calculate the energy required (energy demand) of an individual that is in the border of groups *k* and *k+1.* We call this individual as "marginal individual" associated with groups k and k+1.
2. Calculate the ratio of energy supply to energy demand for the representative individual of group k+1.
3. Multiply that energy demand of marginal individual between groups *k* and *k+1* by the ratio of energy supply to demand calculated for representative individual of group *k+1*. The obtained result is the energy that marginal individual between group *k* and *k+1* takes out to group *k* as that individual leaves group *k+1* to group *k*. This energy is multiplied by the number of people that leave group *k+1* to group *k* to calculate the total energy that leaves group *k+1* to group *k* at any time step.

As it was mentioned before, the steps mentioned above are used to regulate the energy stocks associated with population groups as individuals move between groups. The energy intake that is used by representative individual of each group is obtained by dividing the total energy stock of each group by the number of people in that group at each time step. Appendix A contains complete documentation of the model.

4. Results and discussion

In this section we run the developed system dynamics model under various scenarios with respect to change in energy intake of the population and will discuss the results.

Scenario 1: In the first scenario we start the simulation from equilibrium (in terms of weight), where energy intake of population is equal to its energy expenditure. Next, we impose a 200 kcal/day shock to the base (equilibrium) energy intake of the population at year 2. We let both our model and the agent-based model run under the same condition till the population reaches a new equilibrium (which happens at year 15), given the new energy intake. Table 1 shows the distribution of population along dimension BMI both at the start of simulation (initial condition) and also at year 15 obtained from both our SD model and the agent-based model. Starting from the same initial population distribution (shown in the third column of Table 1), we observe that the new distribution of population along dimension BMI obtained using both our method and agent-based model are very similar after the models reach their new equilibrium (at year 15). In fact, using the Kolmogorov–Smirnov test (K–S test), it was shown that the difference between the two distributions (the last two columns of Table 1) is not statistically significant at level α=0.01. Moreover, Figure 4 shows the graph representation of the initial distribution as well as the shift in distribution of BMI after population reaches a new equilibrium (at year 15) in both SD model and agent-based model.

Table 1: Comparing distribution of population along BMI dimension obtained from our method as well as agent-based model under scenario 1 in which we impose a 200 kcal/day shock to the equilibrium energy intake associated with population

Population Group	BMI Range	Initial Distribution (at Year 0)	Distribution at Year 15 (Agent-based Model)	Distribution at Year 15 (SD Model)
1	14-16	0.002	0	0.0001
2	16-18	0.0208	0.0003	0.0013
3	18-20	0.0677	0.0042	0.0081
4	20-22	0.1217	0.0329	0.0240
5	22-25	0.2053	0.1345	0.1030
6	25-28	0.1737	0.199	0.1857
7	28-30	0.0865	0.1205	0.1271
8	30-32	0.0768	0.1019	0.1167
9	32-34	0.0693	0.0821	0.0940
10	34-36	0.0452	0.072	0.0888
11	36-38	0.0403	0.0713	0.0593
12	38-40	0.029	0.0459	0.0483
13	40-42	0.0192	0.0397	0.0455
14	42-44	0.015	0.0296	0.0336
15	44-46	0.0114	0.0205	0.0201
16	46-50	0.012	0.0287	0.0280
17	50-58	0.0042	0.0169	0.0163

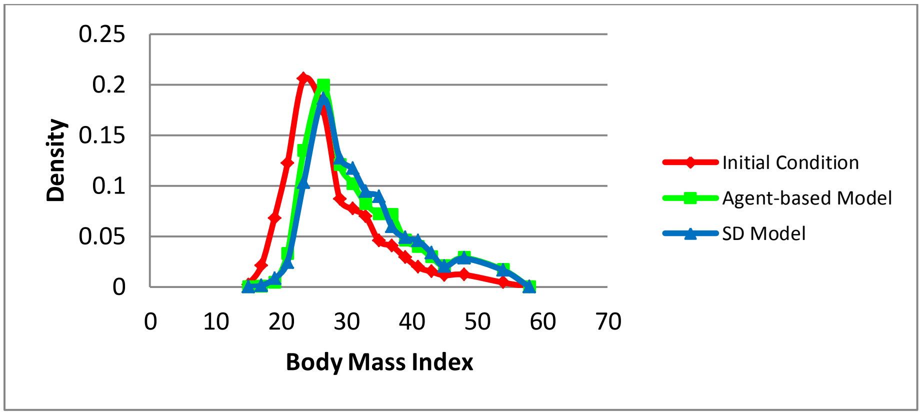

Figure 4: Shift in distribution of BMI in population after 15 years under scenario 1 in which we impose a 200 kcal/day shock to the equilibrium energy intake associated with population

To obtain the percentages of population that are underweight, normal weight, overweight, and obese, we can add up percentage of population in corresponding BMI intervals presented in Table 1. Table 2 shows the percentage of population in four BMI categories estimated by both individual model and our method after imposing 200 kcal/day shock to the base (equilibrium) energy intake. Starting with prevalence of obesity as 32% at the start of simulation, our models estimated that prevalence of obesity will reach to 55% after 15 years if we impose a 200 kcal/day shock to energy intake of population over time. This is close to 51% estimation obtained by agent-based model.

Table 2: Comparing weight status distribution obtained from our SD method as well as agent-based model under scenario 1 in which we impose a 200 kcal/day shock to the equilibrium energy intake associated with population

BMI Categories	Initial Condition	Agent-based Model (Year 15)	SD Model (Year 15)
Underweight Population (BMI < 20)	0.0905	0.0045	0.0095
Normal Weight Population (20 <=BMI < 25)	0.327	0.1674	0.1271
Overweight Population (25 <= BMI <30)	0.2602	0.3195	0.3128
Obese population (BMI >= 30)	0.3224	0.5086	0.5507

Scenario 2: In the second scenario we start the simulation from equilibrium (in terms of weight). Starting at year 2, every individual always is given 30 kcal/day more energy intake than what is needed. This means that energy intakes of individuals over time are always 30 kcal/day more than their energy expenditures. This is equivalent to the calorie that you get from having one hard candy per day in addition to the energy that your body needs to be maintained. Note that implementing this scenario does not require us to use the co-flow structure for distributing energy intake among population groups. In this scenario, we need to simply calculate energy expenditure of each representative individual and then increase that energy intake by 30 kcal/day over time (using a step function) starting at year 2. Table 3 shows the distribution of population along dimension BMI obtained from both our SD model and the agent-based model at year 20. Moreover, Figure 5 shows the graph representation of the initial distribution as well as the shift in distribution of BMI at year 20 in both SD model and agent-based model. Kolmogorov–Smirnov test (K–S test) showed that the difference between these two distributions at year 15 is not statistically significant. Performance and robustness of our method can be further evaluated using other scenarios for changing the energy intake imposed to the population.

Table 3: Comparing distribution of population along BMI dimension obtained from our method as well as agent-based model under scenario 2 in which every individual is always given 30 kcal/day more energy intake than what is needed

Population Group	BMI Range	Initial Distribution (at Year 0)	Distribution at Year 20 (Agent-based Model)	Distribution at Year 20 (Our Model)
1	14-16	0.002	0	0.0000
2	16-18	0.0208	0	0.0003
3	18-20	0.0677	0	0.0023
4	20-22	0.1217	0	0.0092
5	22-25	0.2053	0.0101	0.0556
6	25-28	0.1737	0.0979	0.1101
7	28-30	0.0865	0.1263	0.1056
8	30-32	0.0768	0.1504	0.1083
9	32-34	0.0693	0.1331	0.1101
10	34-36	0.0452	0.1047	0.1030
11	36-38	0.0403	0.0861	0.0904
12	38-40	0.029	0.0786	0.0763
13	40-42	0.0192	0.0626	0.0619
14	42-44	0.015	0.0427	0.0487
15	44-46	0.0114	0.0359	0.0373
16	46-50	0.012	0.0427	0.0442
17	50-58	0.0042	0.0287	0.0465

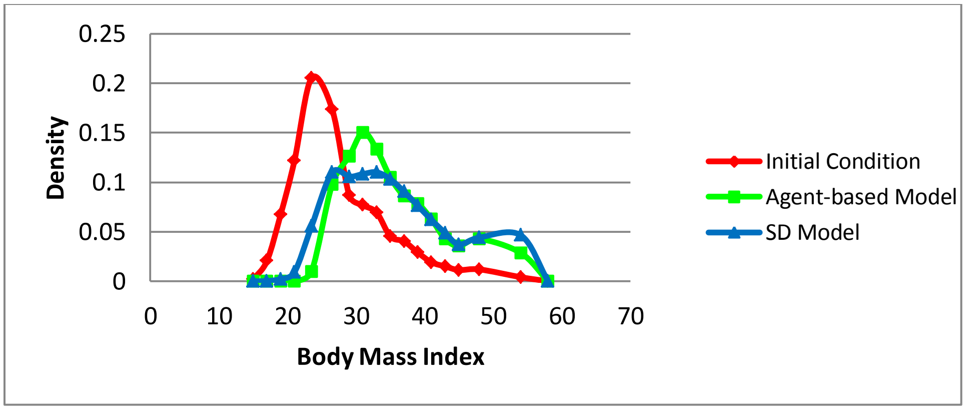

Figure 5: Shift in distribution of BMI in population after 20 years under scenario 2 in which we impose a 30 kcal/day shock to the equilibrium energy intake associated with population

Table 4 shows the percentage of population in four BMI categories estimated by both individual model and our SD model in scenario 2 in which energy intakes of individuals over time are always 30 kcal/day more than their energy expenditures. Starting with prevalence of obesity as 32% at the start of simulation, our SD model estimated that prevalence of obesity will reach to approximately 73% after 20 years if each individual always take 30 kcal/day more energy intake than what is needed. This is close to 77% estimation obtained by the agent-based model.

Table 4: Comparing weight status distribution obtained from our SD method as well as agent-based model under scenario 1 in which we impose a 200 kcal/day shock to the equilibrium energy intake associated with population

BMI Categories	Initial Condition	Agent-based Model (Year 15)	SD Model (Year 15)
Underweight Population (BMI < 20)	0.0905	0.0000	0.0027
Normal Weight Population (20 <=BMI < 25)	0.327	0.0101	0.0648
Overweight Population (25 <= BMI <30)	0.2602	0.2242	0.2158
Obese population (BMI >= 30)	0.3224	0.7655	0.7268

5. Conclusion

In this paper we develop a modeling extension that can bridge micro dynamics (associated with elements in a population) and macro behavior of the population with significantly less computational costs and fairly good precision. This method would allow us to model dynamics associated with distribution of a population with respect to its attributes/characteristics. Given that many modeling problems require capturing the heterogeneity of population elements with respect to population attributes/characteristics, there are many potential application areas for our developed method. For example, it can be used for modeling distribution of health indicators such as weight, Body Mass Index, blood pressure, cholesterol, etc. in a human population; modeling distribution of skill level or expertise in employees of a company; modeling distribution of number of network connections in a population of social networks, etc.

Our method is mainly applicable in problem domains in which the movement of population elements across different percentiles on the distribution of attributes of interest is a function of where an element is located on the distribution. For example, current weight of an individual influences the rate of weight gain and loss (due to the effect of current weight on energy expenditure of individuals), and consequently the new position of that individual along BMI dimension.

In comparison with agent-based models, our method has significantly less computational costs and is easy to apply. We applied our method to an empirical dataset of 3074 female adults obtained from NHANES data and looked at the change in distribution of BMI in this population over time under different scenarios related to change in energy intake of the population. Our results showed a fairly good precision in comparison with the results obtained from an agent-based model that explicitly modeled 3074 agents (with the same initial condition as our model) and imposed the same shock to the energy intake of individuals.

One of the important limitations of our model is that it can only be used when disaggregating the population along one dimension. If characteristics on multiple dimensions contribute to rate of change across population groups on all those dimensions, this method will not extrapolate easily. In our future work, we aim to evaluate validity of our method under more scenarios of change in energy intake. Also, a sensitivity analysis on the number of intervals that is defined for disaggregating the population into different groups can also provide valuable insight. Such analysis would allow us to see if disaggregating population into more groups (i.e., developing groups of more homogeneous individuals) will lead to a more precise estimation of the distribution of attribute under analysis and its changes over time in the population.

Acknowledgement: The study is supported by National Institute of Health (NIH) research grants from the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD, 1R01HD064685-01A1 and U54HD070725) and the National Institute of Diabetes and Digestive and Kidney (NIDDK, R01DK81335-01A1). The U54 project is co-funded by the NICHD and the Office of Behavioral and Social Sciences Research (OBSSR). This study is also supported by National Collaborative on Childhood Obesity Research (NCCOR) Envision's Comparative Modeling Network (CompMod) program and NIH Office of Behavioral and Social Sciences Research (OBSSR) contract: HHSN276201000004C.

Appendix A: Model Documentation

****

Parameters based on original models/literature

****

GammaF=3.2*365

Unit: kcal/(kg*Year)

GammaL=22*365

Unit: kcal/(kg*Year)

Beta= 0.24

Unit: Dmnl

EtaF= 180

Unit: kcal/kg

EtaL= 230

Unit: kcal/kg

C= 10.4

Unit: kg

Delta=7*365

Unit: kcal/(kg*Year)

RhoF= 9400

Unit: kcal/kg

RhoL= 1800

Unit: kcal/kg

KConstant=370.21*365

Unit: kcal/Year

CConstant=C*RhoL/RhoF

Unit: kg

****

The rest of the model, including adult body weight and BMI distribution modules

****

Energy Supply Representative Individual[WGroup]= Avg Energy of WG[WGroup]

Unit: kcal/People

WGroup2: WGroup

Energy Going Out[W1Mid]= Weighted Energy Demand Marginal Person[W1Mid]*Supply Demand Ratio Representative Individual[W1Mid]

Unit: kcal/People

Energy Coming in[W1Mid]= Weighted Energy Demand Marginal Person[W1Mid]*Vector Elm Map(Supply Demand Ratio Representative Individual[W1], W1Mid)

Unit: kcal/People

Energy Demand Representative Individual[WGroup]=Base EI[WGroup]

Unit: kcal/People

Weighted Energy Demand Marginal Person[W1Mid]=Energy Demand Representative Individual[W1Mid]*(1-ownWeight[W1Mid])+Vector Elm Map(Energy Demand Representative Individual[W1], W1Mid)* ownWeight[W1Mid]

Unit: kcal/People

Supply Demand Ratio Representative Individual[WGroup]=Energy Supply Representative Individual[WGroup]/ Energy Demand Representative Individual[WGroup]
Unit: Dmnl
OutFlowE[W1Mid]=ownOFP[W1Mid]*Energy Going ut[W1Mid]+ nxtOFP[W1Mid] *Energy Coming in[W1Mid]
Unit: kcal/Year
OutFlowE[WEND]=0
Unit: kcal/Year
Avg Energy of WG[WGroup]=Energy[WGroup]/Population[WGroup]
Unit: kcal/People
Sum of halfs[W1Mid]=lenght of WG[W1Mid]/2+Vector Elm Map(lenght of WG[W1],W1Mid)/2
Unit: kg/(meter*meter)
ownWeight[W1Mid]=lenght of WG[W1Mid]/2/Sum of halfs[W1Mid]
Unit: Dmnl
lenght of WG[WGroup]=Xf[WGroup]-Xi[WGroup]
Unit: kg/(meter*meter)
AvePopBW[WGroup]=Initial BMI[WGroup]*Height[WGroup]*Height[WGroup]
Unit: kg
Base EI[WGroup]= INITIAL(EInumerator[WGroup]/EIdenominator[WGroup])
Unit: kcal/Year
BW[WGroup]=FFM[WGroup]+FM[WGroup]
Unit: kg
BWfactor[WGroup]=BW[WGroup]*Delta[WGroup]
Unit: kcal/Year
DEnergyIntake[WGroup]=EnergyIntake EI[WGroup]-Base EI[WGroup]
Unit: kcal/Year
dIfactor[WGroup]=Beta*DEnergyIntake[WGroup]
Unit: kcal/Year
EIdenominator[WGroup]=1-(IFactor[WGroup])/(1+IFactor[WGroup])
Unit: Dmnl
EInumerator[WGroup]= (KConstant[WGroup]+FFMfactor[WGroup]+FMfactor[WGroup]+BWfactor [WGroup]-Beta*0)/(1+IFactor[WGroup])
Unit: kcal/Year
Energy[W1]= INTEG (-OutFlowE[W1]+Pulse shock[W1],Initial population[W1]*Base EI[W1])
Unit: kcal
Energy[WGMid]= INTEG (-OutFlowE[WGMid]+Vector Elm Map(OutFlowE[W1],WGMid-2)+Pulse shock[WGMid], Initial population[WGMid]*Base EI[WGMid])
Unit: kcal
Energy[WEND]= INTEG (Vector Elm Map(OutFlowE[W1],WEND-2)+Pulse shock[WEND],Initial population[WEND]*Base EI[WEND])
Unit: kcal
Energy Expenditure without K[WGroup]=(FFMfactor[WGroup]+FMfactor[WGroup]+BWfactor [WGroup]+dIfactor[WGroup]+Iterm[WGroup])/(1+IFactor[WGroup])
Unit: kcal/Year
EnergyExpenditure[WGroup]=KConstant[WGroup]/(1+IFactor[WGroup])+Energy Expenditure without K[WGroup]
Unit: kcal/Year
EnergyIntake EI[WGroup]=Avg Energy of WG[WGroup]

```
            Unit:      kcal/Year
EnergyPFunction[WGroup]=CConstant/(CConstant+FM[WGroup])
            Unit:      Dmnl
FFM[WGroup]= INTEG (0,AvePopBW[WGroup]*(1-InitFMFraction[WGroup]))
            Unit:      kg
FFMfactor[WGroup]=FFM[WGroup]*GammaL[WGroup]
            Unit:      kcal/Year
FM[WGroup]= INTEG (0,AvePopBW[WGroup]*InitFMFraction[WGroup])
            Unit:      kg
FMfactor[WGroup]=FM[WGroup]*GammaF[WGroup]
            Unit:      kcal/Year
Height[WGroup]=1.62
IFactor[WGroup]=EtaF*(1-EnergyPFunction[WGroup])/RhoF+EtaL*EnergyPFunction[WGroup]/RhoL
            Unit:      Dmnl
Indicated dFFM[WGroup]=EnergyPFunction[WGroup]*(EnergyIntake EI[WGroup]-
EnergyExpenditure[WGroup])/RhoL
            Unit:      kg/Year
IndicateddBMI[WGroup]=(IndicateddBW[WGroup]/(Height[WGroup]*Height[WGroup]))
            Unit:      kg/(Year*meter*meter)
IndicateddBW[WGroup]=Indicated dFFM[WGroup]+IndicatedFM[WGroup]
            Unit:      kg/Year
IndicatedFM[WGroup]=(1-EnergyPFunction[WGroup])*(EnergyIntake EI[WGroup]-
EnergyExpenditure[WGroup])/RhoF
            Unit:      kg/Year
InitFMFraction[WGroup]:DATA
Initial BMI[WGroup]=(Xi[WGroup]+Xf[WGroup])/2
            Unit:      kg/(meter*meter)
Initial population[WGroup]:DATA
Iterm[WGroup]=EnergyIntake EI[WGroup]*IFactor[WGroup]
            Unit:      kcal/Year
nxtOFP[W1Mid]=Min(0,Vector Elm Map(IndicateddBMI[W1], W1Mid))*Vector Elm Map(Y[W1],W1Mid)
            Unit:      People/Year
nxtOFP[WEND]=0
            Unit:      People/Year
OutFlowP[W1Mid]=ownOFP[W1Mid]+nxtOFP[W1Mid]
            Unit:      People/Year
OutFlowP[WEND]=0
            Unit:      People/Year
ownOFP[W1Mid]=Max(0,IndicateddBMI[W1Mid])*Y[W1Mid]
            Unit:      People/Year
ownOFP[WEND]= 0
            Unit:      People/Year
Physical activity level[WGroup]= 0.5
            Unit:      Dmnl
PopPercentage[WGroup]=Population[WGroup]/TotPop
            Unit:      Dmnl
```

```
Population[W1]= INTEG (-OutFlowP[W1],Initial population[W1])
        Unit:   People
Population[WGMid]= INTEG (-OutFlowP[WGMid]+Vector Elm Map(OutFlowP[W1],WGMid-2),Initial
population[WGMid])
        Unit:   People
Population[WEND]= INTEG (Vector Elm Map(OutFlowP[W1],WEND-2),Initial population[WEND])
        Unit:   People
Pulse shock[WGroup]=PULSE(2,0)*(200*365)*Initial population[WGroup]/TIME STEP
        Unit:   kcal/Year
TotPop=SUM(Population[WGroup!])
        Unit:   People
W1Mid:W1, WGMid
WGMid:(W2-W16)
WGroup:W1, WGMid, WEND
Xf[WGroup]:DATA
        Unit:   kg/(meter*meter)
Xfin[WGMid]=Xf[WGMid]
        Unit:   kg/(meter*meter)
Xfin[W1]=Xf[W1]
        Unit:   kg/(meter*meter)
Xfin[WEND]=Xf[WEND]
        Unit:   kg/(meter*meter)
Xi[WGroup]:DATA
        Unit:   kg/(meter*meter)
Xinit[WGroup]=  Xi[WGroup]
        Unit:   kg/(meter*meter)
Y[WGroup]=Population[WGroup]/(Xfin[WGroup]-Xinit[WGroup])
        Unit:   People/kg/(meter*meter)
```

References:

Chow, C.C., K.D. Hall., 2008, The Dynamics of Human Body Weight Change. *PLoS Computational Biology*, 4(3) e1000045.

Hall, K.D., 2010, Mechanisms of metabolic fuel selection: modeling human metabolism and body-weight change. *IEEE Engineering In Medicine And Biology Magazine* 29(1) 36-41.

Homer, J., Milstein, B., Dietz, W., Buchner, D., Majestic, E., 2006, Obesity Population Dynamics: Exploring Historical Growth and Plausible Futures in the U.S., Proceedings of the 24th International System Dynamics Conference, Nijmegen.

Osgood, N., 2004, *Representing heterogeneity in complex feedback system modeling: Computational resource and error scaling*, Proceedings of the 22nd International Conference of the System Dynamics Society, Oxford.

Sterman, J. D., 2000, *Business Dynamics: Systems Thinking and Modeling for a Complex World*, McGraw-Hill.