# Dynamic cohorts: a new approach to managing detail

Tom Fiddaman & Larry Yeager
Ventana Systems, Inc., Harvard MA, USA
Contact: tom@ventanasystems.com
6/7/2016

## Abstract

While it is always desirable to minimize the complexity of a model, some problems require the detailed representation of heterogeneous subgroups, where nonlinearities prevent aggregation or explicit chronological aging is needed. It is desirable to have a representation that avoids burdening the modeler or user computationally or cognitively, with excessive detail or arcane formulations.

Eberlein & Thompson (2013) propose *continuous cohorting*, a novel solution to the cohort blending problem in population modeling, and test it against existing aging chain and cohort-shifting approaches. Continuous cohorting prevents blending of ages and other properties. However, this comes at some cost in complexity, because it creates internal states equal in number to the duration of the aging chain divided by the time step.

We propose another new solution, *dynamic cohorts*, that prevents blending with a comparatively low computational burden. More importantly, the approach simplifies the representation of distinct age, period and cohort effects and representation of dynamics other than the aging process, like migration and attribute coflows. By encapsulating the lifecycle of a representative cohort in a single entity, rather than dispersing it across many states over time, it makes it easier to develop and explain the model structure.

These innovations are not limited to population models; they also apply to other quantities, like vehicles, perishable goods or financial instruments. It makes it easy to change aggregation parametrically for empirical determination of the appropriate level of detail, and to mix detailed and aggregate elements in the same model.

## Background

Eberlein and Thompson (2013) provide a thorough treatment of the cohort blending problem. Cohort blending arises in aging chains due to the assumption that the contents of a stock are well mixed. Surprisingly, avoiding blending in a continuous aging chain requires a very high order delay structure (Figure 1). This has long been known, and is often advantageous for representing phenomena in which events are broadly dispersed (Forrester, 1961).

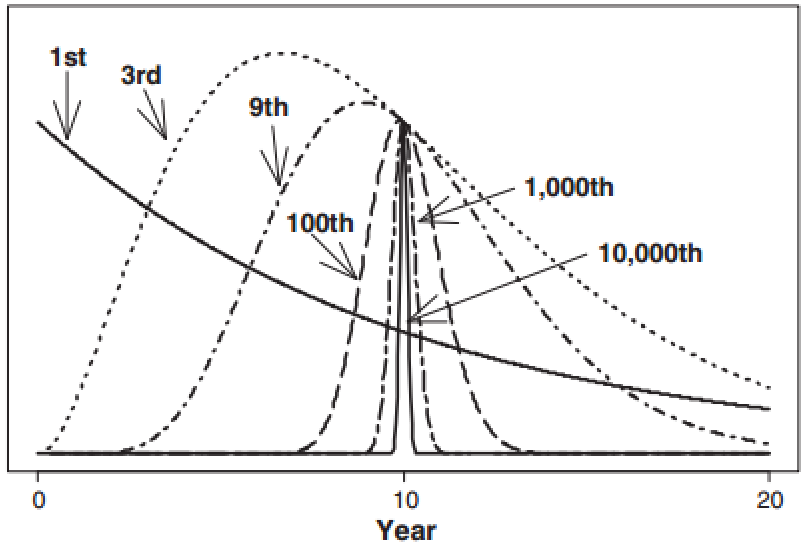


**Figure 1 Eberlein & Thompson (2013) Fig. 1 Output from an impulse for different order delays.**

Blending does not occur when the model is discrete, with a cohort interval equal to the time step. This is the solution commonly employed by demographers. However, this is rather limiting for models that include other dynamics, which may require a time step less than the minimum practical cohort dimension of 1 to 10 years.

Some dispersion may be acceptable in certain contexts. For example, if the purpose of the model is to predict health care demand, the onset of disease is not strictly chronologic, because people age at different rates and therefore are intrinsically mixed. In that case, it may be possible to model incidence without fine age cohorts. However, this eschewing cohorts does not work when one wishes to calibrate the model to data that has age or cohort detail, or to represent a phenomenon that is explicitly chronologic, such as retirement at a particular age of eligibility. In such cases, accurate maintenance of age structure is required.

Continuous cohorting achieves accuracy by adapting the demographer's approach to (potentially) finer time steps. This is done by creating an internal representation of an aging chain that is more detailed than the array dimension of the population cohorts, with one stock per time step. This prevents mixing by discretizing the age distribution. However, it comes at a price in computational complexity: the number of stocks required in the internal representation is (age range)/(time step), so that an aging chain with 100-year duration requires 1600 stocks if time step is .0625 (1/16) year.

## Additional Challenges

Besides the blending problem, there are several additional challenges in the representation of aging chains.

First, the representation of the flow of aging from a younger stock to an older stock is not straightforward. A typical pair of levels in an aging chain has the following configuration (Figure 2).

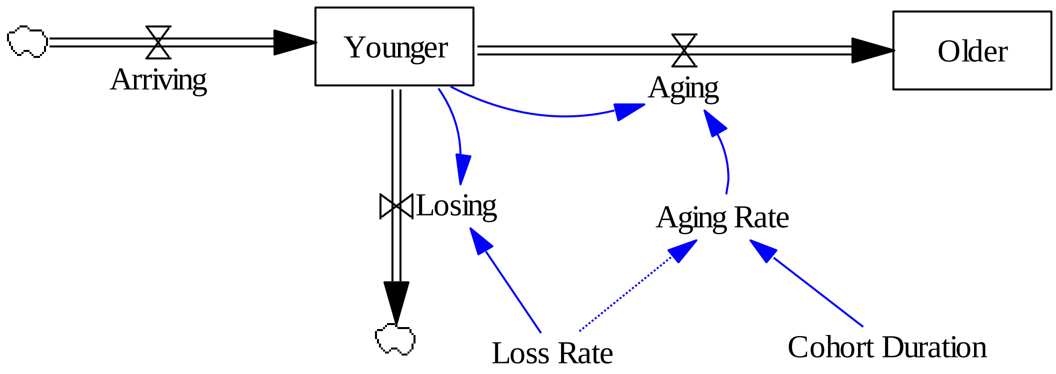


Figure 2 A pair of simplified aging chain levels.

For a population model, *Losing* and the *Loss Rate* represent mortality, for example. Naively, the *Aging Rate* can be modeled as 1/*Cohort Duration*. However, this is not accurate when losses are significant, in which case the *Aging Rate* must be corrected to account for the fact that not all members of the cohort survive to age out. This is not a problem for fine cohorts (e.g., 1 year duration), where losses integrated over the cohort duration are negligible, but adds complexity when a more aggregate representation is desirable.

Second, the *Loss Rate* and other features of an aging chain are typically not constant. Treating them as constant (e.g., assuming that mortality or fertility rates will not change in the future) is likely to lead to errors at least as large as the dispersion problem. Demographers recognize distinct Age, Period and Cohort effects in the APC model (Hobcroft, Menken & Preston, 1982). These refer to the Age of individuals in a group, the Period or time at which an event occurs, and the Cohort or year in which individuals in a group were born or otherwise created. For a vehicle fleet, this means that there might be three effects on the scrap rate of a given cohort: the chronological age of vehicles (Age), the prevailing driving hazards and maintenance practices (Period), and the cumulative driving and maintenance history of the model year (Cohort). Since vehicles have no particular dependence on chronology, the Period and Cohort effects are the most natural to represent, but vehicle registration data is often provided by Age, so it is important to be able to represent all three.

A natural extension of fixed Cohort effects is to recognize that the health of an individual, or operability of a vehicle, depends on states that may evolve independent of chronology. For an individual, cumulative exposure to war, disease, nutrition and other factors matters, as for a vehicle cumulative mileage driven and maintenance history matters. To represent this thinking operationally, one would like to model additional states associated with the group (Figure 3).

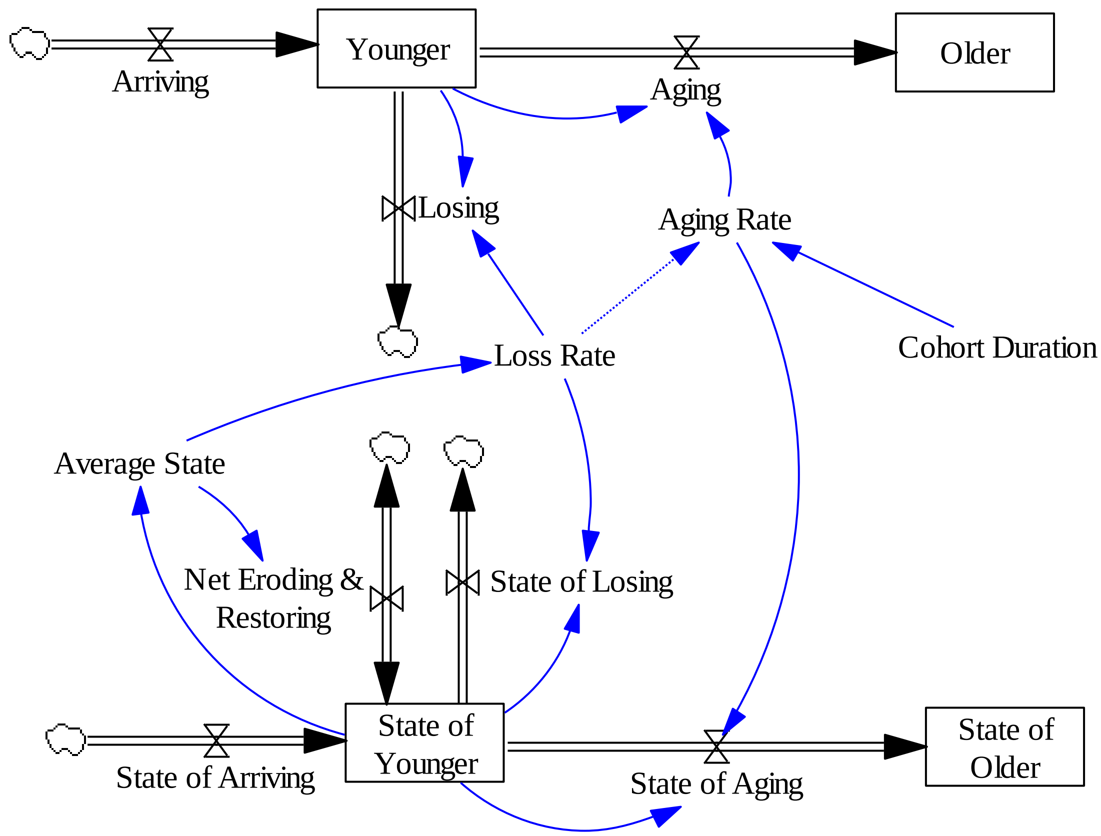


Figure 3 A pair of aging chain levels with co-flowing state.

In addition to states that co-flow with the cohort population, there may be other flows like migration or transshipment of goods that contribute to the complexity of the aging chain.

Regardless of whether variation in rates across a chain is due to chronologic aging or cohort state effects, the variation in de facto rates implies that a certain level of cohort detail is necessary to achieve a given level of resolution. For example, as mortality rates rise steeply with age, an aggregate cohort of 5 or 10 years duration would see substantial variation across its age structure, as in Figure 4. If precision simulation of older ages were required, a finer age structure would be desirable.

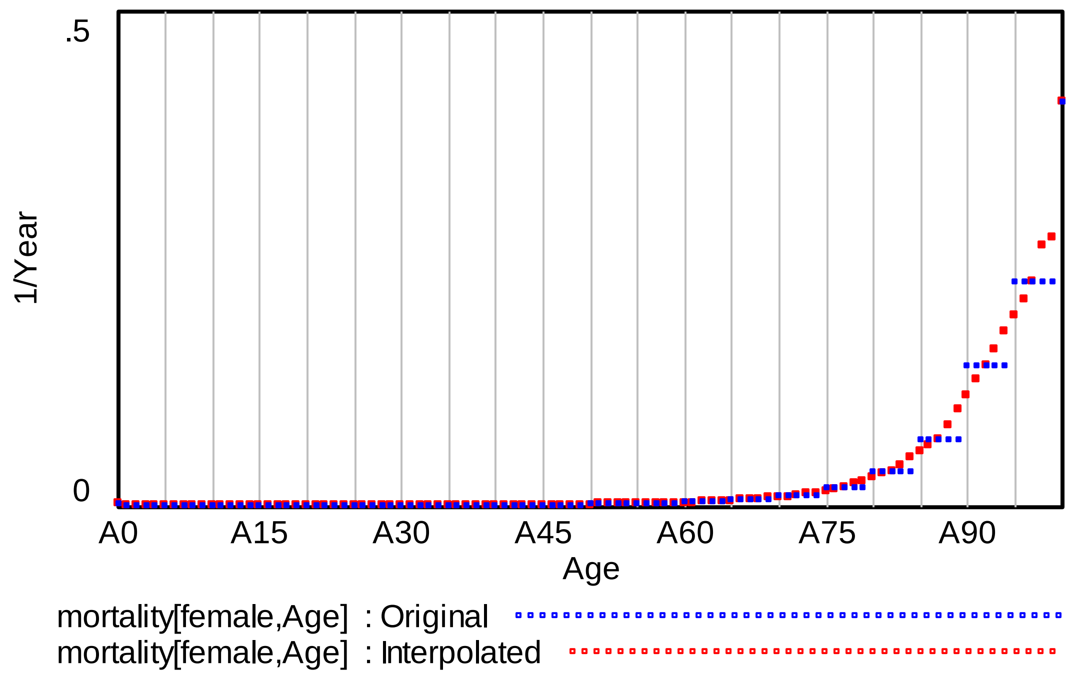


**Figure 4 Age-specific Mortality Rate for Japan in Eberlein & Thompson (2013). Stepwise trajectory is the original, presumably limited in resolution by data availability. The actual mortality rate must resemble something more like the interpolated trajectory shown.**

While demographic processes are generally smooth, one could easily construct pathological cases for other kinds of cohorts, where variations in the inflow rate would propagate (assuming no dispersion) to create large deviations from realistic outflow rates later. In some cases, one could construct rules for determining whether disaggregation is sufficient for accuracy, but in general this is complex and most practical to achieve empirically, by varying the level of detail.

## Aging Chain Behavior

To illustrate the efficiency of dynamic cohorts, we reconsider the demographic model of Japan from Eberlein & Thompson (2013). That model shows that a population with annual cohorts and a fine time step (1/16 year) results in substantial blending, and that continuous cohorting, by avoiding blending, reveals “true” behavior that significantly differs (Figure 5).

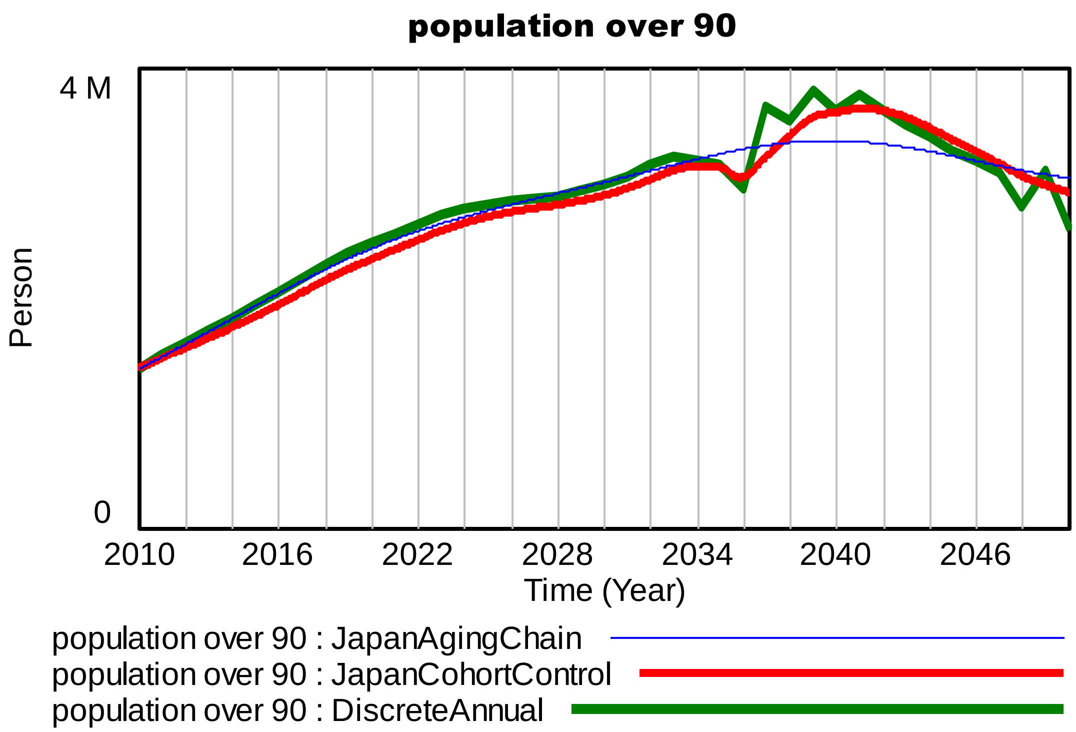


**Figure 5 Population over age 90, comparing annual aging chain in continuous and discrete time to continuous cohorting**

The improvement comes at a computational price: continuous cohorting requires (101 annual cohorts) / (1/16 year time step) x (2 genders) = 3232 stocks and their associated flows. The user does not need to see or manage all of these; just the 202 explicit stocks are visible. However, this detail is not useful, in that it does not improve accuracy over the demographer's solution of discretizing with time step equal to the cohort duration, in part because the input (tables of fertility and mortality rates) has only 1 to 5 year resolution.

The demographer's solution is also unattractive, because the aging rate must be adjusted to account for mortality, as discussed above. The oldest cohorts, which have high mortality rates, become unstable and oscillate. While a correction is not hard to implement, it makes the model more complex and therefore harder to interpret.

Moving in the opposite direction, toward greater aggregation, is much more attractive from the perspective of computation and cognitive burden on the modeler. 5 year cohorts increase dispersion over annual cohorts, but might still be usable for some purposes. However, 10 year cohorts or a 4$^{th}$ order population structure (very long cohorts, as in Limits to Growth (Meadows et al., 1972)) are substantially wrong compared to continuous cohorting (Figure 6).

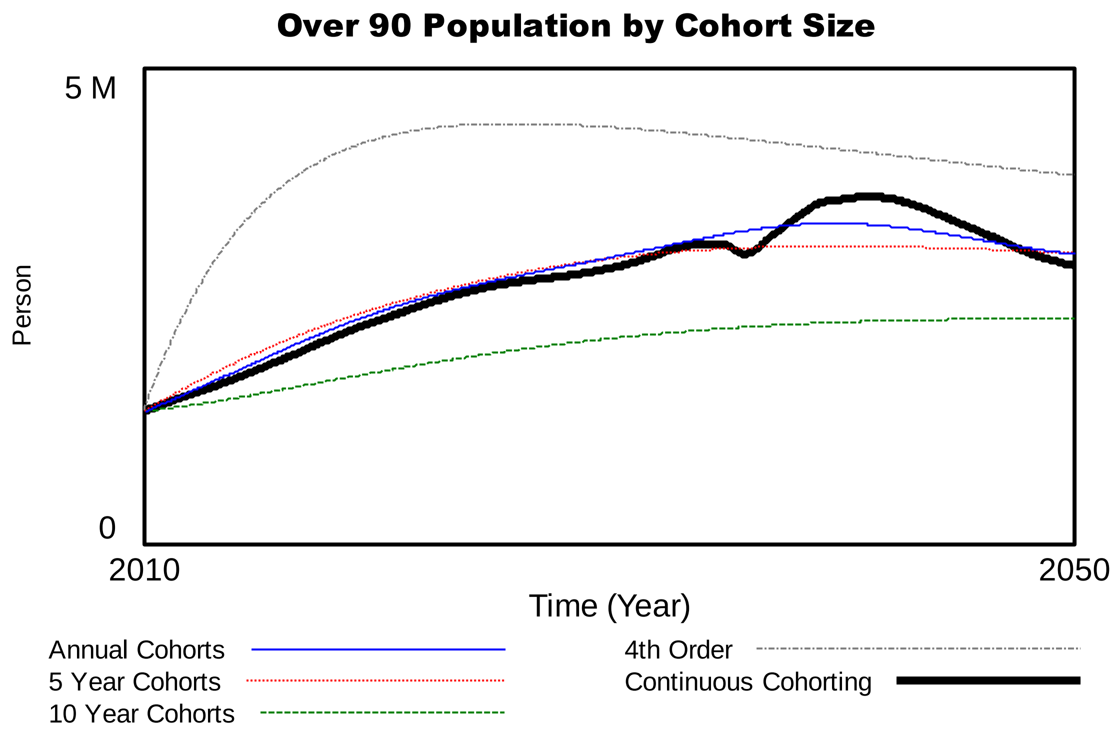


Figure 6 Population variation by cohort duration

Part of the reason the 5 year cohorts perform reasonably well is that the input parameters (mortality and fertility rates) in the original model are quantized in 5 year increments. This data limitation is common when adopting demographic models or projections for reuse in another context.

## Dynamic Cohort Behavior

A desirable cohort formulation:

- Avoids blending (dispersion, mixing)
- Makes it easy to represent Age, Period and Cohort effects, or subsets thereof
- Provides sufficient resolution of effects that vary with cohort age
- Simplifies the representation of multiple inflows, outflows and co-flows
- Permits experimentation with the granularity of the chain, for verification

Dynamic cohorts lend themselves to meeting these criteria. Here, we implement them using the Ventity simulation software (Yeager et al. 2014, Ventana, 2017).

To understand the approach, a few basic Ventity concepts should be introduced:

- *Entities* are modular objects that encapsulate related equations that share the same level of detail. A particular *entity* is an instance of an *entity type*.
- *Attributes* are text tags that provide unique identifiers (*keys*) for entities, and describe other categorical data. Attributes are much like the elements of array dimensions, except that they are

stored as ad hoc lists, and therefore it is not necessary to specify all elements when a multidimensional cube would be sparsely populated.

- *References* provide a way to access variables in other entities via their key attributes.
- Entities belong to a *collection* listing all entities of a given type; users may also create *subcollections* that segment the list of entities according to certain attributes.
- *Aggregate functions* compute sums, averages and other metrics on the list of entities belonging to a collection.
- *Actions* are discrete events that can be executed to change system structure by adding or deleting entities, iterating over collections, or changing states in the model. This may change the effective structure by altering loop gains or change the actual logical structure of the model.

## Implementation

The basic idea is to recognize that the root cause of blending, and a substantial source of complexity, is the necessity of moving people or goods from one age stock to another. The solution is simple: *don't move them.*

Instead, we maintain a dynamic collection of cohorts. Each cohort is created at a particular time, i.e. it contains a particular range of birth dates, or model years in the case of vehicles. A cohort and its contents exist *over* time, but there are no aging flows *among* cohorts. A cohort acquires members through birth (and possibly other processes, like immigration), but thereafter it only loses members through mortality. This makes the fate of an individual age cohort easy to understand, because one need not observe different age categories at different times in order to track behavior. One can observe the cohort directly.

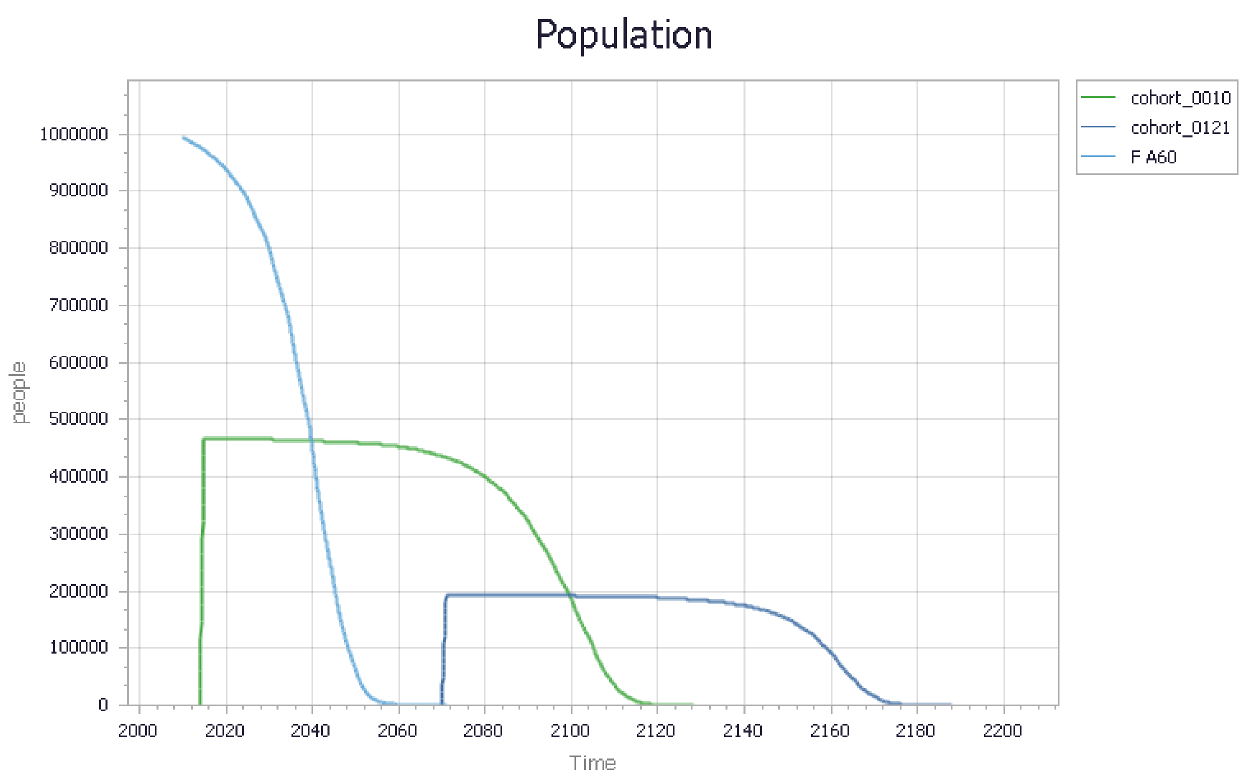

Figure 7 Lifecycles of three representative cohorts, showing females aged 60 existing at the start of the simulation, and females born 10 and 60 years after the start of the simulation. Note that the maximum population of each cohort is declining as births decline.

The populations of existing cohorts are initialized by a table of data (Table 1), extracted from the original Vensim version of the model. Each cohort is identified by a unique attribute (its CohortID in this case), and may contain other attributes, such as the "group gender" and "group region," which identify a Group to which the cohort belongs. Conceptually, the Group of females in Japan is related to a collection of roughly 120 annual age cohorts. Age-specific fertility and mortality rates are provided by lookup tables. Each cohort could have unique fertility and mortality rates, but since the original model lacks this feature, we omit it here for comparability.

Table 1 Cohort Initialization Data

| Enabled | Time | CohortID | group gender | group region | Creation Time | Life Years | Population |
|---|---|---|---|---|---|---|---|
| TRUE | | F A0 | F | Japan 1yr | 2009 | | 501613 |
| TRUE | | F A1 | F | Japan 1yr | 2008 | | 512203 |
| TRUE | | F A2 | F | Japan 1yr | 2007 | | 522909 |
| TRUE | | F A3 | F | Japan 1yr | 2006 | | 530882 |
| TRUE | | F A4 | F | Japan 1yr | 2005 | | 536693 |
| … | | | | | | | |

Each Cohort in existence at the start of the simulation has a creation time. Here, one must be a little cautious about terminology. The simulation starts in 2010. In the original aging chain version of the model, the initial population of array element [female,A0] represents the initial population between ages 0 and 1. There is an analogous Cohort of females aged 0 to 1, but births into that cohort began in 2009.

Future cohorts could also be instantiated from data, but it is computationally wasteful to create them before they are needed. Instead, a timer generates a train of pulses at regular intervals (annually, i.e. every 16$^{th}$ time step, Figure 8). When a pulse occurs, a Create Cohort action executes. The Action creates a new Cohort and establishes its connection to the Group. Similarly, old Cohorts are deleted when their population falls below some value of interest (e.g. 1 person).

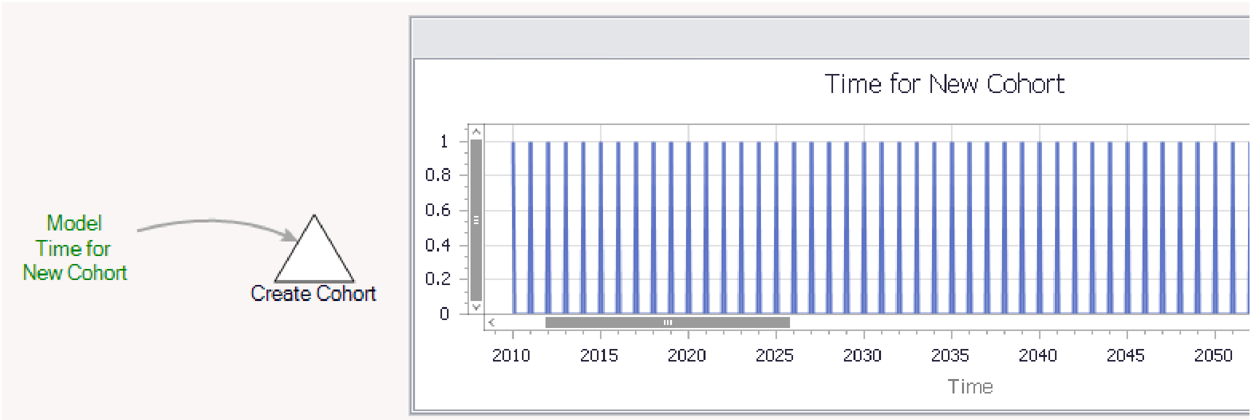


**Figure 8 Cohort Creating Timing**

As a result, the number of active cohorts rises from 202 (the 101 provided by data for each gender) to a near-equilibrium value of about 238, then slightly decreases as declining births mean that populations fall below the specified cutoff of interest sooner (Figure 9). The complexity of the model, as measured by counting stocks, therefore is less than 8% of the aging chain version's, though the time step remains 1/16 year.

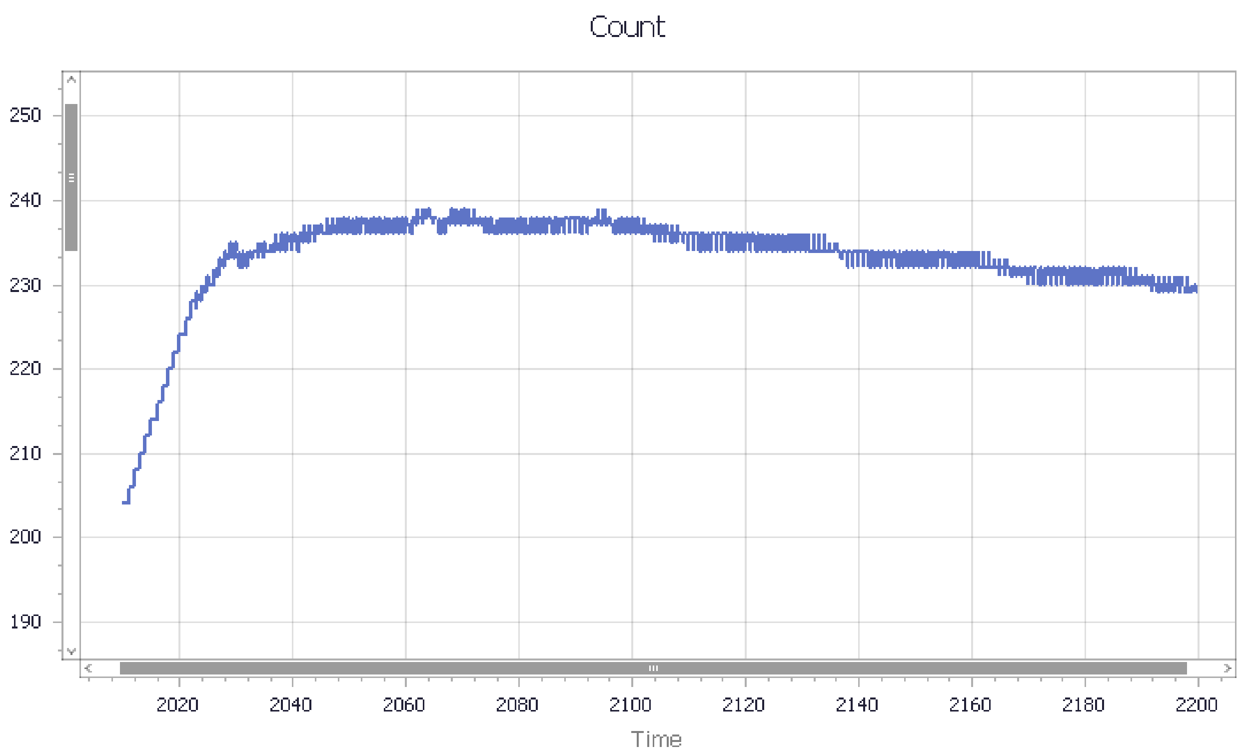


**Figure 9 Count of active cohorts. Note nonzero minimum Y scale.**

In spite of the dramatic reduction in complexity, the model yields results that are nearly identical to the Continuous Cohorting version (Figure 10).

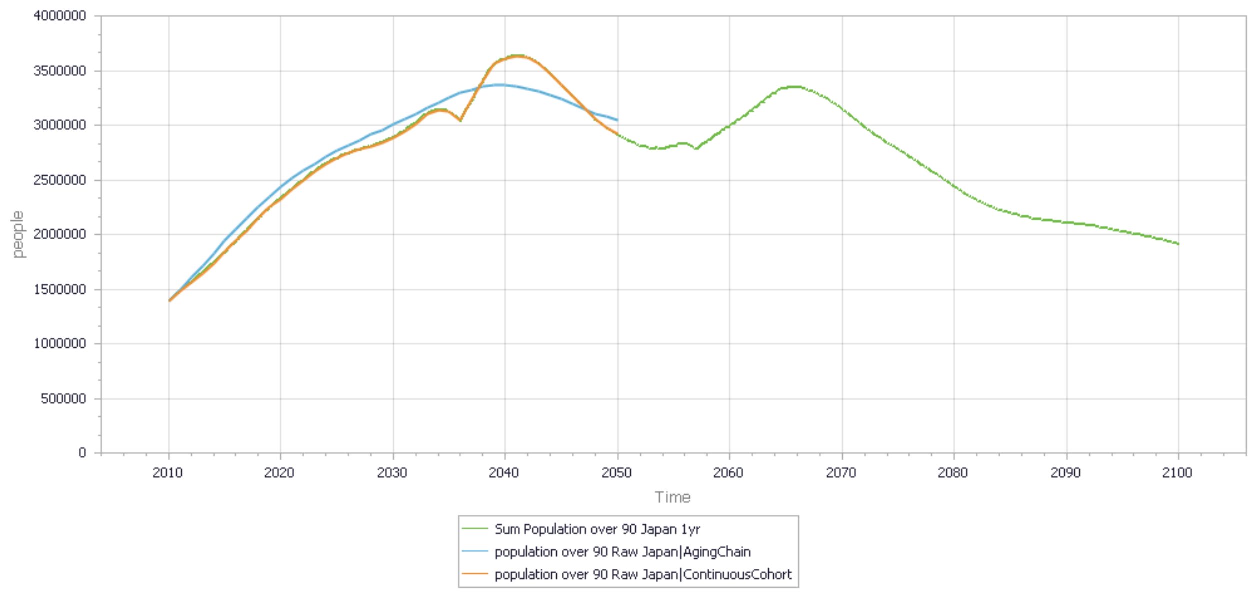


**Figure 10 Comparison of dynamic cohorts (green) to continuous cohorting (orange) and aging chain (blue).**

## A Vehicle Fleet Example

In order to illustrate the behavior of a more general model, taking full advantage of the capabilities of dynamic cohorting, we abstract from data and illustrate with a purely synthetic vehicle fleet model. Here, we emphasize the organization of the model of the model rather than the details of its structure and behavior. In other words, it is the infrastructure supporting cohort modeling that is of interest; given this platform, one could surely create a more realistic fleet model for a particular problem.

The primary model structure resides in two Entities, a Fleet and a FleetCohort. The FleetCohort represents a single cohort of vehicles sharing a common model year or range of years (Figure 11). The FleetCohort contains a stock of vehicles, which is depleted over time by scrapping of inoperable vehicles. It also accumulates mileage driven, for determination of the average mileage of a vehicle in the cohort. Increasing mileage, relative to a standard for the durability of the vehicle, degrades its condition, which in turn diminishes utilization. The mileage standard for vehicle durability is a state that is locked in at the time of vehicle production, owning to the vehicle design (Figure 12), therefore this is a *cohort* effect in the demographer's APC framework. Condition also has a pure *age* influence, driven by the nominal age of the vehicle, which is determined by comparing the current time to the cohort's midpoint year of creation.

The loss rate from scrap depends vehicle condition, as well as a loss rate from accidents. Here, the accident rate is independent of condition, but declines over time, owing to improvements in driver behavior and highway safety. This is a *period* effect in the APC framework, because it depends on time but not the age or state of the cohort. Because the degradation of vehicle condition is nonlinear, increasing rapidly with age and mileage (Figure 13), the model shares an important feature with demographic models: the practical aggregation may be limited by the nonlinearity of losses.

A hybrid cohort-period effect is also present: vehicle utilization depends on fuel cost per mile, which depends on the fuel intensity designed into the vehicles (a cohort effect) and the current price of fuel (a period effect). The fuel intensity might, in a more elaborate model, have its own dynamics. For example, fuel efficiency can decline due to component wear and poor maintenance of older vehicles.

Once they contain too few vehicles to be "interesting," FleetCohorts self-delete. This keeps the computational complexity manageable. In simulations over 100 years with a time step of ¼ year, about 20 annual cohorts typically exist concurrently, or five 5-year cohorts. A similar aging chain would suffer from considerable dispersion, and a similar continuous cohorting approach would require 4 to 20 times as many (mostly internal) levels.

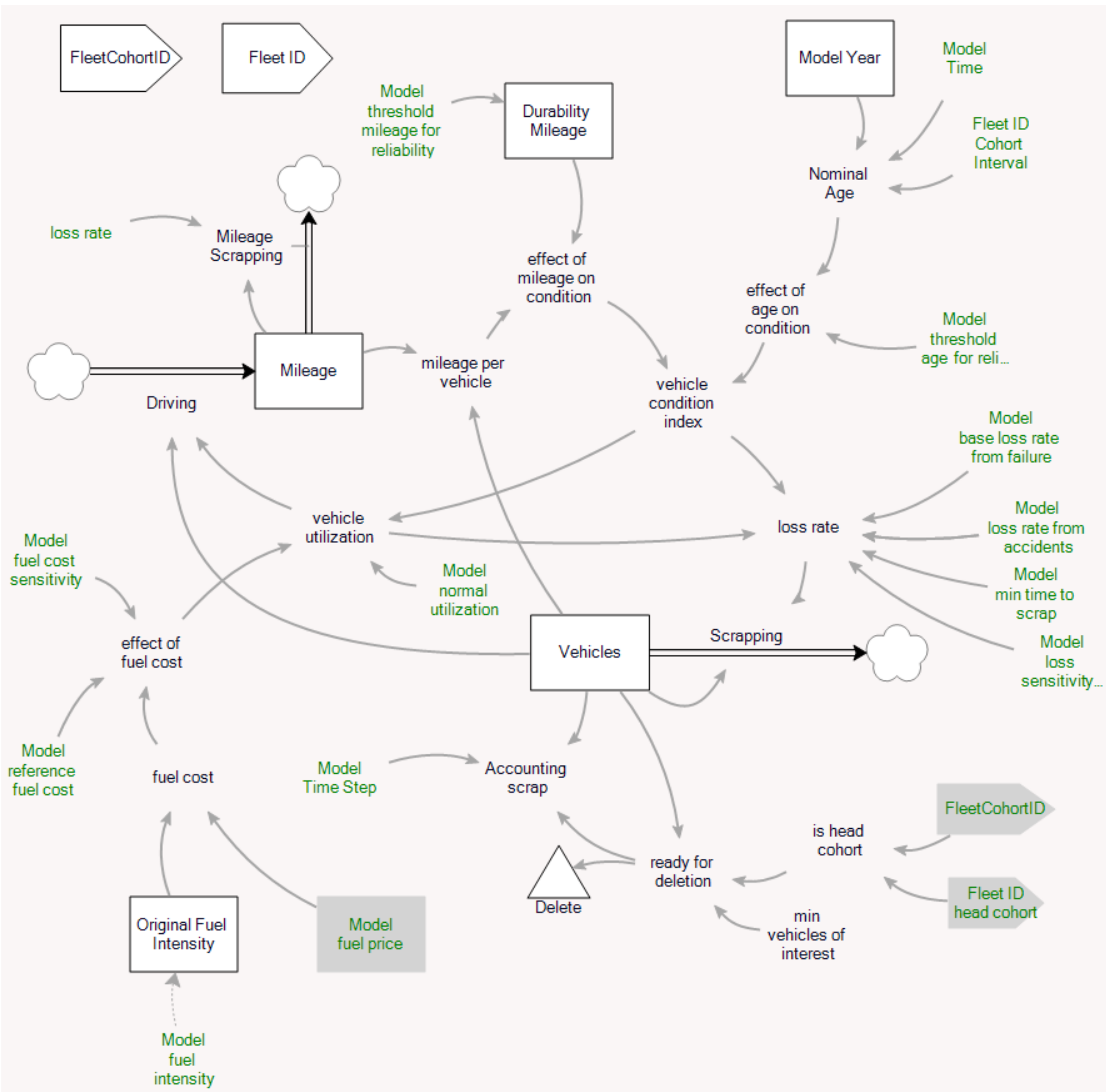


Figure 11 FleetCohort entity structure. Variables with green lettering are references from other entities in the model.

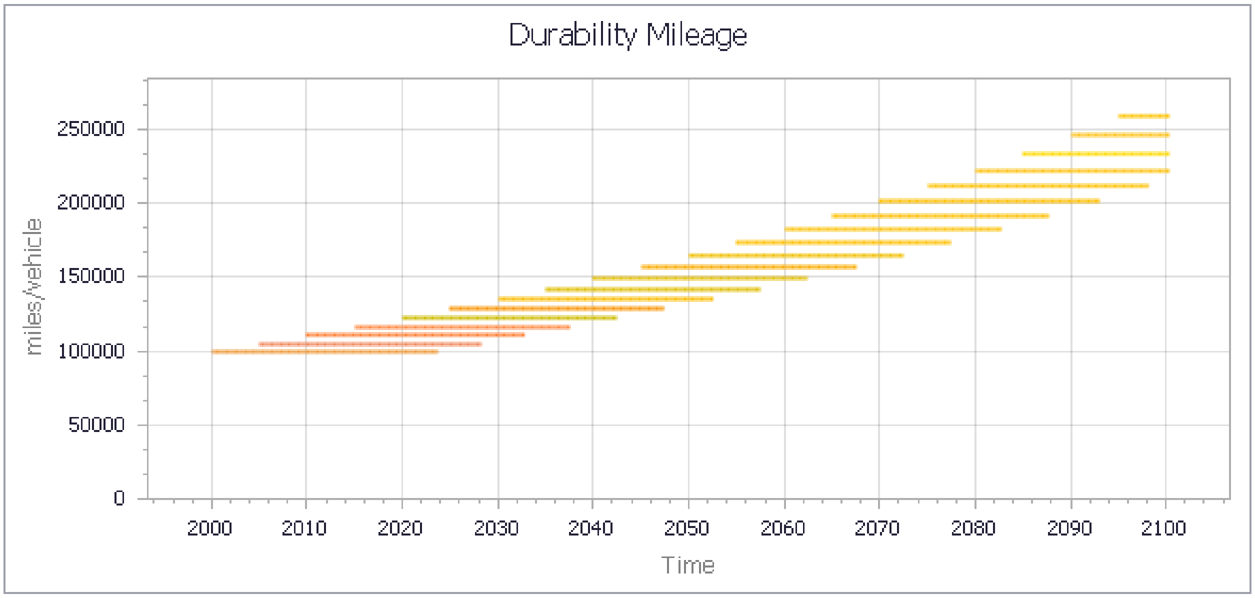


**Figure 12 Cohort durability, locked in at time of creation.**

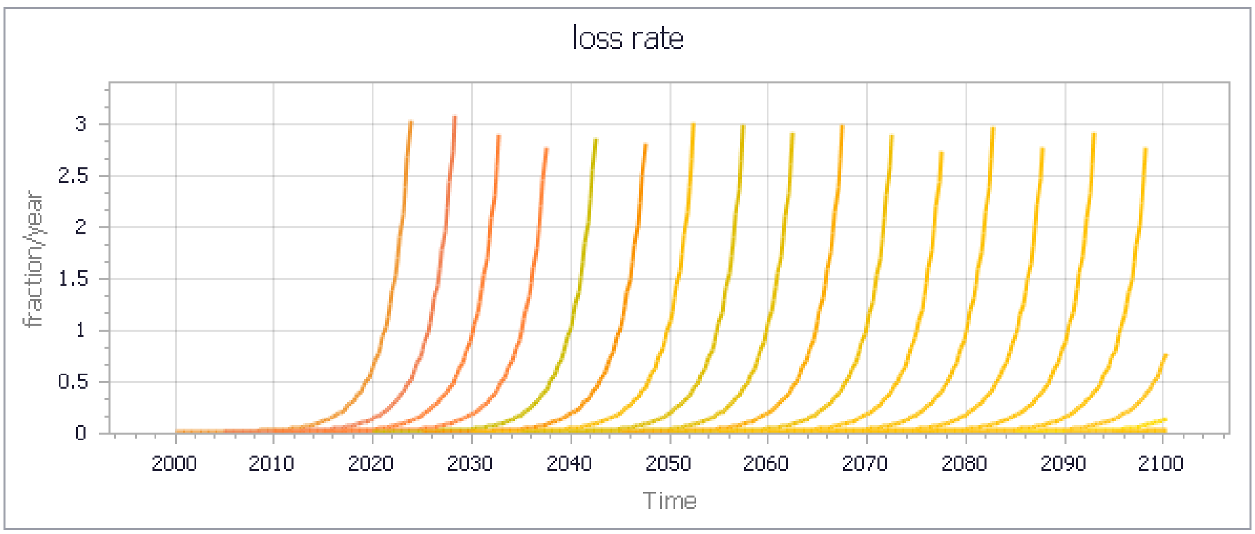


**Figure 13 Cohort loss rate.**

Each FleetCohort has an attribute that identifies the Fleet entity to which it belongs. The collection of FleetCohorts by Fleet can then be used to calculate the aggregate number of vehicles, scrapping and other features of the cohorts comprising a given Fleet.

The Fleet itself has two important features (Figure 14). First, it guides the flow of new vehicle Selling to the Vehicles stock of the FleetCohorts. In order to do this, it must have a way to identify the current "head cohort," i.e. the youngest cohort into which sales should be directed. This information is provided by an attribute that populates a reference to the needed cohort. Second, it creates new cohorts as needed, according to a timer, as in the population example (Figure 8).

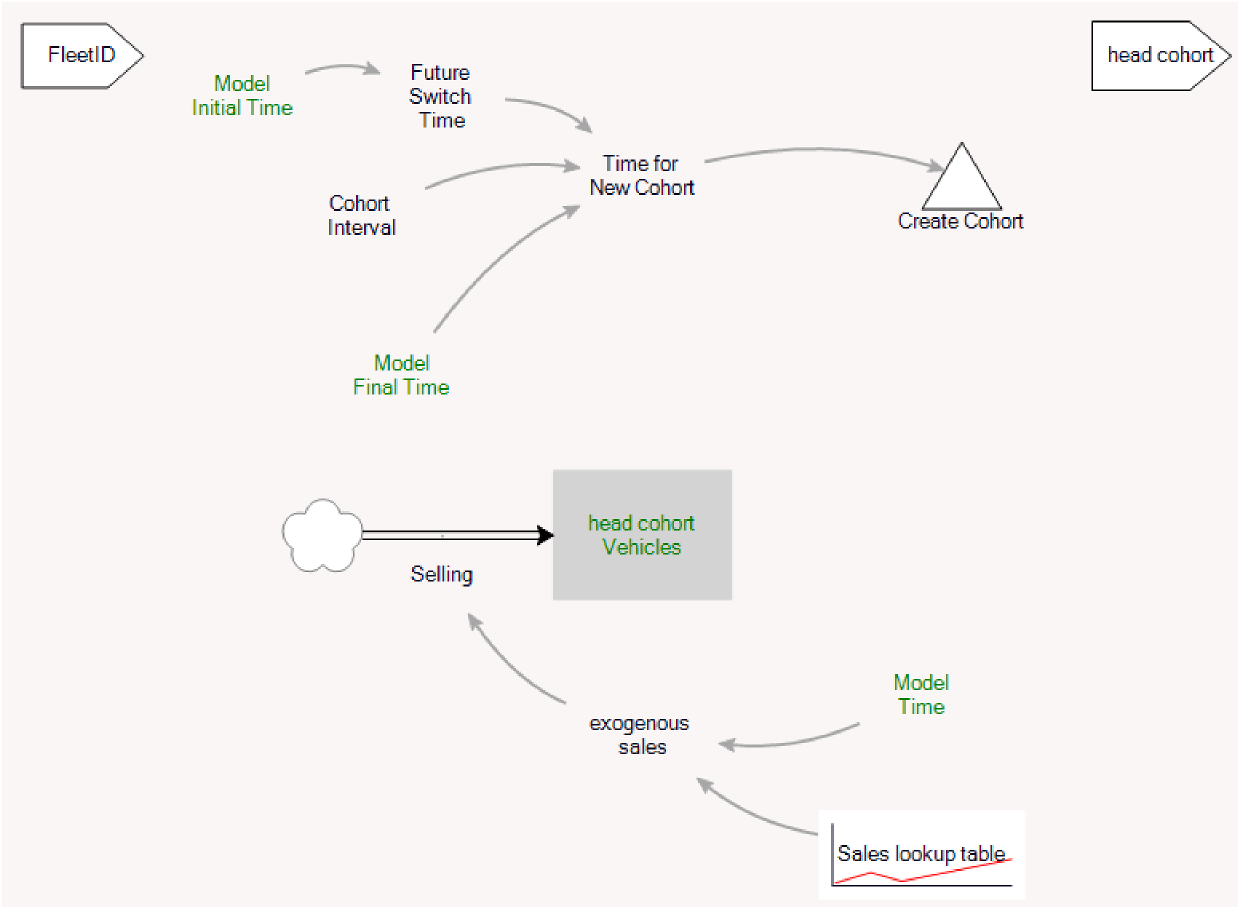


Figure 14 Fleet entity type.

The Fleet creates cohorts by invoking an action (Figure 15). The action has no dynamics, because it has no existence in continuous time. Instead, it initializes the states (stocks and attributes) of the new FleetCohort. In particular, it sets the FleetCohort's Fleet ID attribute to point to the Fleet that created it (given by the Invoker reference). It also updates the Fleet's head cohort attribute to point to the newly-created FleetCohort.

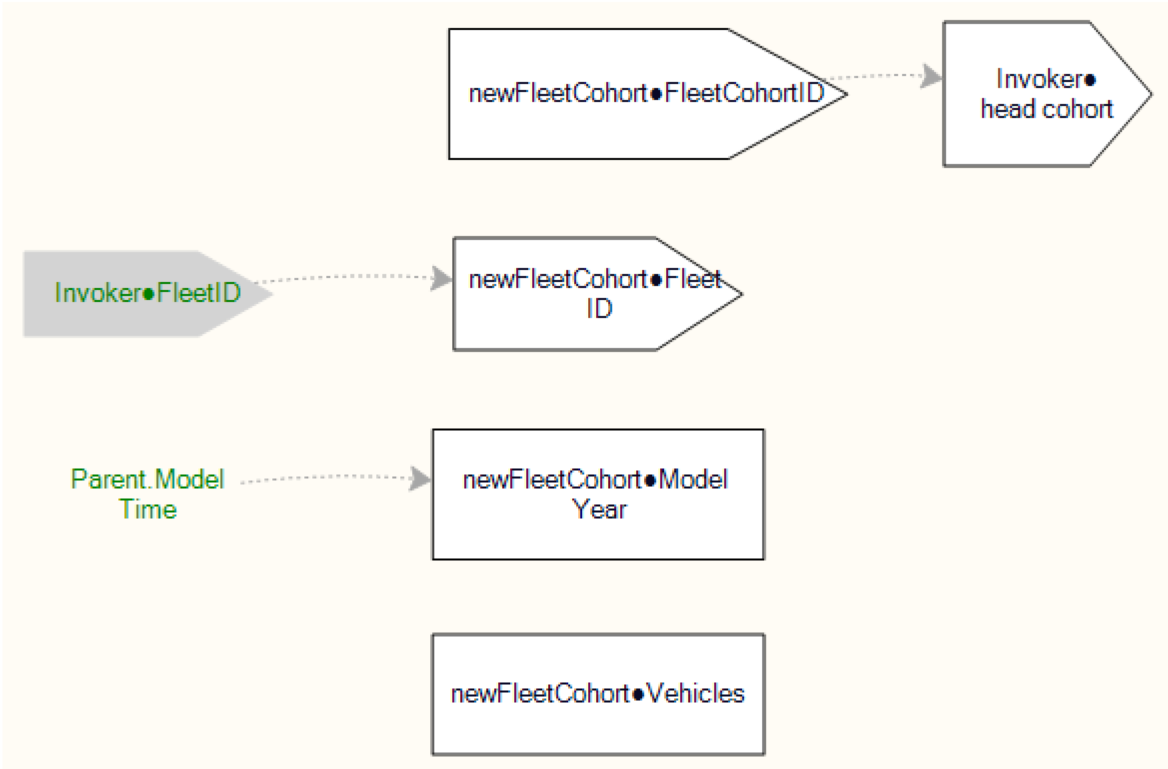


Figure 15 Action to create a FleetCohort.

Once the structure is in place, it can be populated with data and run. Because the model is generic, we do not need to supply data for historic cohorts, unlike the population example (Table 1). Instead, we supply only some scenario parameters that are global to the Model (not shown) and the Fleets we wish to simulate (Figure 16). Here, there are two fleets, with different Cohort Intervals of 1 and 5 years. Note that the 1-year cohort interval is left blank; it need not be specified because it is the default value in that constant's equation.

| Fleet | | | | | | FleetCohort | |
|---|---|---|---|---|---|---|---|
| | ☑ | Time | CalendarTime | FleetID | head cohort | Future Switch Time | Cohort Interval |
| 1 | ☑ | | | 1 year | | | |
| 2 | ☑ | | | 5 year | | | 5 |

Figure 16 Fleet initialization data.

By setting that single parameter value, we can now simultaneously experiment with two variants of the model, having different aggregation of the aging structure. Figure 17 compares the Vehicles stock for the two levels of detail. Note that the 5-year cohorts are roughly 5 times as large due to their greater model year range, but correspondingly 1/5 as numerous.

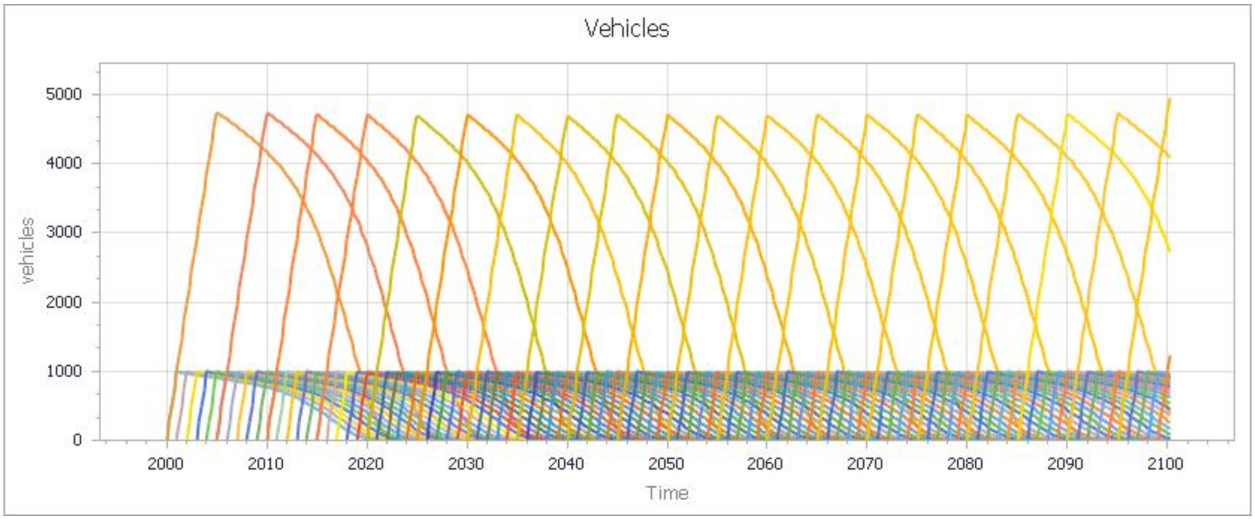


**Figure 17 Vehicle stocks with 1 and 5 year cohorts.**

The individual cohort dynamics are of course important to verify, but in most cases it is the fidelity of the aggregate outcome that is of interest. Turning to the collection of FleetCohorts belonging to the Fleet, we can see that the size of the fleet is nearly identical, irrespective of the cohort detail (Figure 18). However, total Scrapping shows a small oscillation when 5-year cohorts are used (Figure 19). This is not a dynamic oscillation arising from a delayed negative feedback loop. Instead, it is the superposition of many peaks in individual cohort lifecycles (Figure 20). Each individual cohort's scrap lifecycle peaks because its loss rate initially rises as the vehicles age and increase in mileage, but then eventually the vehicle fleet is depleted, so scrap must fall when few vehicles remain to be scrapped.

The oscillation might or might not be an acceptable condition for a given application, but there are several ways to deal with it. The obvious solution is to use 1-year cohorts, which remain stable, or to experiment with some intermediate granularity, like 2-year cohorts. Another possible approach is to recognize that the problem is largely due to the rapidly escalating scrap rates (Figure 13). For parameters that make the loss rate less steep or less nonlinear, the problem does not occur at all. Finally, in some cases one could sample the aging effect on condition over a range, rather than at the midpoint age of the cohort, recognizing that the vehicles in a cohort are in fact somewhat diverse.

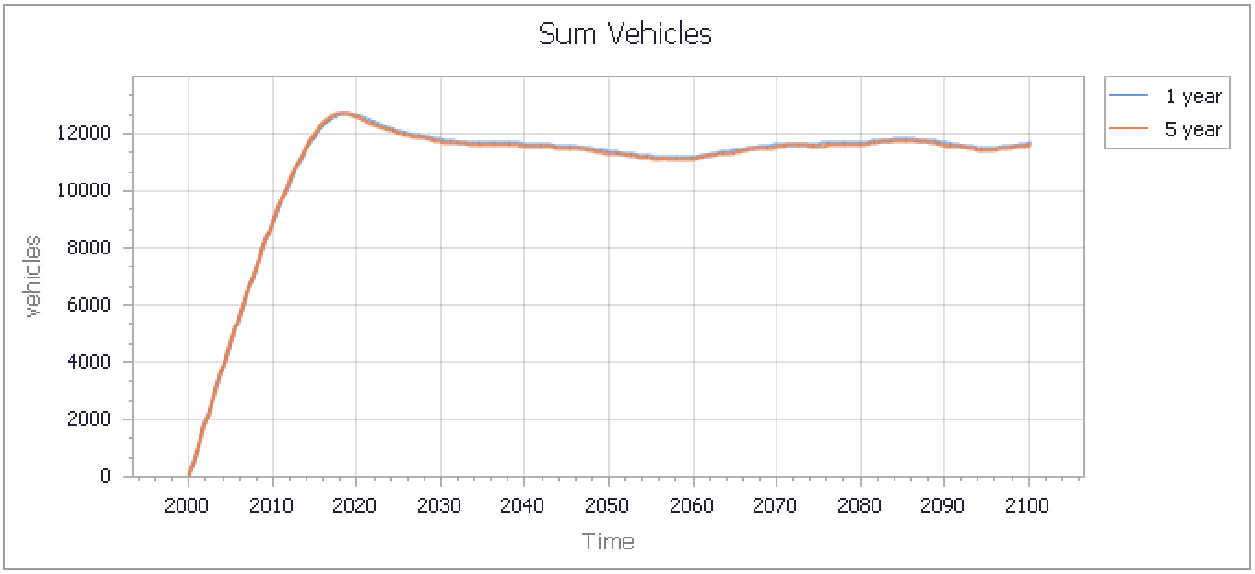


**Figure 18 Aggregate vehicle stocks for 1 and 5 year cohorts.**

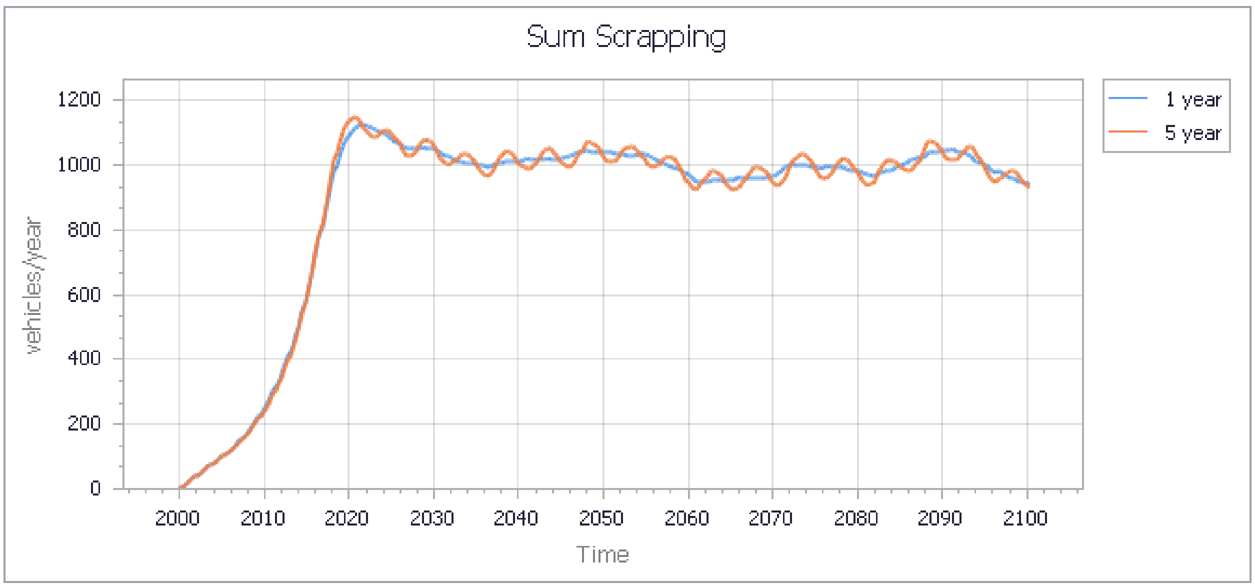


**Figure 19 Aggregate scrap rate for 1 and 5 year cohorts.**

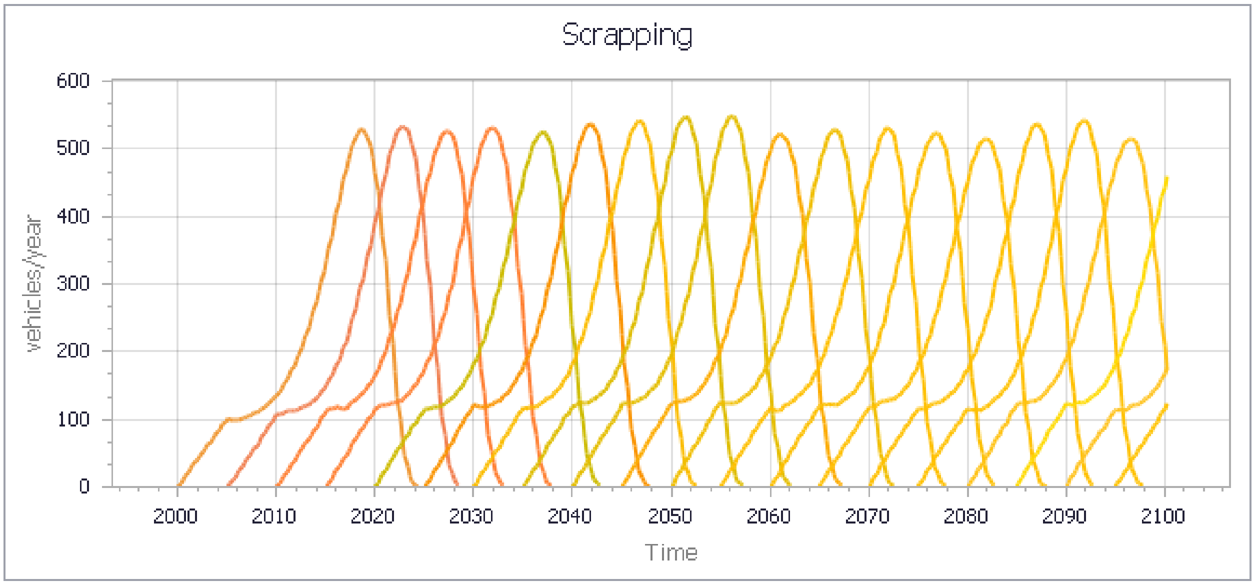


**Figure 20 Individual cohort scrap rates responsible for oscillatory aggregate scrapping.**

Once this cohort structure is in place, and the level of detail has been chosen by parameter experimentation, it is straightforward to extend the model. In this example, one could add:

- Transshipment of vehicles among interacting regions.
- Explicit dynamics of the erosion of fuel efficiency or other performance attributes.
- A more explicit maintenance process and state of repair of the vehicle.

These could also be implemented in an aging chain, but the dynamic cohort approach makes it easier, because there is no need for complex logic to flow conserved quantities into older stocks. A cohort entity represents a single, generic lifecycle of a group of similar vehicles, which is easy to understand and debug. Once perfected, it is easy to replicate the cohort as needed by creating new instances from data or via actions.

## Conclusions

As we have shown, dynamic cohorts provide a solution to the blending problem that does not entail a heavy computational burden. It permits the cohort detail to be chosen according to the requirements for representing nonlinearities and the heterogeneity of cohort members, rather than having it dictated by the time step required to represent any fast dynamics in the model.

For demography, the cohort structure naturally supports Age, Period and Cohort effects. The same generic framework applies equally to vehicles, perishable goods, and many other phenomena. Critically, the entity approach also makes it easy to connect a collection of cohorts to other more aggregate structures. For example, a health care system might be represented with a collection of patient cohorts, but a single (or a few) aggregate delivery centers with associated capacities (e.g., hospitals). This makes it possible to disaggregate where necessary, but to avoid providing detail for detail's sake.

For a priori or synthetic data models, the creation of cohorts by actions frees the modeler from rewriting array dimensions and re-aggregating data. Different levels of detail can be tested by variation of a single parameter, the interval at which cohorts are created. This turns what is often guesswork in the choice of detail in an aging chain into a simple matter of experimentation.

The entity approach used in Ventity also makes it straightforward to compare the cohort approach to a still more granular structure, with individual people or vehicles in an agent based model, using similar logic. The two formulations can even be run simultaneously within the same model for easy comparison and synthetic data experiments (using a detailed model to generate data to be estimated in an aggregate framework, for example).

We fully expect that further creative uses will arise to exploit the new capabilities of entities and dynamic cohorts as we accumulate experience with the new approach.

## Supplementary Material

- Original Eberlein & Thompson model of the population of Japan, with extensions, in Vensim. Requires the Vensim Model Reader, Vensim Pro, or Vensim DSS. http://vensim.com/vensim-model-reader/
- Dynamic cohort version of the Japan model, in Ventity. http://ventity.biz/
- Fleet model, in Ventity.
- Equation listings of the above.

## References

Robert L. Eberlein and James P. Thompson, 2013. Precise modeling of aging populations. *Syst. Dyn. Rev.* 29, 87–101.

Jay W. Forrester, 1961. *Industrial Dynamics*. Cambridge: MIT Press. Chapter 11.

Meadows et al. 1972. *The Limits to Growth*. New York: Universe Books.

John Hobcraft, Jane Menken and Samuel Preston, 1982. Age, Period, and Cohort Effects in Demography: A Review. *Population Index* Vol. 48, No. 1 (Spring, 1982), pp. 4-43 DOI: 10.2307/2736356

Ventana Systems, Inc. 2017. Ventity simulation software. http://ventity.biz.

Yeager, Larry with Thomas Fiddaman and David Peterson, 2014. A New Entity-Based System Dynamics Tool. *Proceedings of the International System Dynamics Conference*, Delft, 2014.