<u>Meeting notes (Fri 28th June 2024) – Graham, Dave, Luke, Rob</u>

1. Discussion of PSOR paper submission

- **Ignoring randomness in the fluid flow model**: we do not necessarily need to ignore randomness – as long as the 'server-always-busy' assumption holds, we can still use the fluid flow model and include the variance of the queue length in the computation of the expected squared queue length.
- **Negative Q lengths**: along with the understanding that in reality the Q would not come close to zero, it would be OK to use a surrogate model which could technically have a negative Q length and optimise using this surrogate model. This is essentially what we have done as there is currently nothing to rule out a negative queue length in our fluid flow model.
- **Definition of functional zero homelessness**: It would be good to revisit this and consider the potential to use this in an optimisation formulation.
- **Service times – complications to acknowledge (we may nonetheless decide to not change how we model service times)**
    - in queueing settings, a service time is typically dependent on the customer characteristics. In our setting, there may also be external factors (e.g. policies to influence how long people may stay in housing) which may also play a role.
    - Given the non-zero administration time associated with housing someone, which we absorb into the service time distribution, an exponential service time distribution may not be suitable since this gives non-zero weight to a service time of zero
    - Given that the longest someone may stay in housing is the rest of their life, an exponential distribution may not be suitable given it gives non-zero weight to very long service times.

2. Future research

- Open question: are deterministic models sufficient for the system we are modelling? **If:**

a) the complexities we want to include (e.g. re-entries to the system) can be incorporated into the fluid flow model **and**

b) all types of housing are assumed to always be occupied **and**

c) we are only interested in mean queue lengths in an objective function,

then a stochastic simulation model may not be necessary and we may be able to continue with our fluid flow model.

- Of course, even with a deterministic model, there is high uncertainty in model inputs such as arrival rates and service rates. An appropriate way forward to deal with this uncertainty may be **robust optimisation** using uncertainty sets – i.e. if we can model our uncertainty in the model inputs, we may formulate a problem where we want to:
    - Minimise the average performance over the uncertainty set
    - Minimise the worst case performance across our uncertainty set

3. Immediate next steps:

- To address the two points above in 'future research' – what further system complexities do we want to model and might the deterministic fluid flow model be enough? what might a robust optimisation formulation look like and what are current methods for modeling uncertainty using uncertainty sets?

<u>Do we really need to use discrete-event simulation?</u>

a) Complexities we may want to model:
   - **Priority queueing:** customers arrive with a certain class. Q prioritisation rules based on class. Simulation then necessary if we include waiting times in objective function.
   - **Non-Markovian service time distributions.** Simulation then necessary if we include Q-length variance or distribution in objective function.
   - **Treating shelter as a server** with a non-zero service time distribution. Leads to tandem Q with blocking. Simulation then necessary if there is a meaningful chance of all customers in shelter still receiving service when a house becomes available (leading to an empty house for some time).
   - **Converting shelter to housing:** in practice this could involve waiting for a group of shelters to become unoccupied, not allowing new customers into these shelters, taking them offline for some period of time during conversion, then bringing them back online as housing units. Simulation necessary to model this.

b) When would shelter/housing units ever be unoccupied?

   - **Time for moving:** in reality, once someone has left housing, there will likely be time before the next person moves in from shelter, as this person remains in shelter, preparing to move.
   - **Conversion**: during conversion from shelter to housing (see above).
   - **Capacity exceeds demand**: if the number of shelter/housing units rises to the point where there is a meaningful chance of capacity exceeding demand for some time.

c) What may we be interested in beyond mean Q lengths in an objective function?

   - **Q-length**: variance and distribution (the former allowing us to compute expected squared Q-lengths, the latter enabling us to compute the probability of a Q length rising above some acceptable limit).
   - **Waiting times**: mean/variance/distribution (the latter enabling us to compute the probability that someone waits longer than some acceptable time – as discussed with the concept of "functional zero").

Meeting notes (Thursday 18th July 2024) – Graham, Dave, Luke

Discussion of 'do we need stochastic simulation or is a fluid flow model enough?'

- **Priority queues**: a compartmental fluid flow model (with different compartments corresponding to different priority groups) could approximate queue dynamics.
- **Time dependent service rate**: if a certain customer group had a different service rate, a surge of arrivals from that group would lead to a change in the overall mean service rate. Since these surges would occur randomly, stochastic simulation necessary to model.
- **Waiting time distribution**: a compartmental fluid flow model could approximate this, but stochastic simulation needed for more accuracy.
- **Conclusion**: No definitive answer: fluid flow model could be enough for certain approximations, but combined effect of lots of approximations justifies stochastic simulation.
- **Note**: we do not necessarily need to model all of the complexities we discuss, but we need to consider them to understand which type of model is most appropriate.

Discussion of next steps of research

- **Decision making**: note that in practice, public sector organisations typically have a relatively small number of genuine options so simulating all options can be enough to aid decisions
- **Optimisation**: in public sector is usually multi criteria. Often performance vs. cost. We have two key performance indicators (long-term throughput with housing vs. short-term unsheltered queue) as well as cost.
- **Uncertainty in inputs**: Because of long-term modelling, we can't simply collect more data on arrivals/service-completions to improve input models. To account for this genuine uncertainty in simulation optimisation would need something different from current practice of input uncertainty where sampling distributions are available.
- **Similar applications** of long-term capacity planning: in justice (don't know future crime rates) and pandemic preparedness (don't know future transmission rates)
- **Fluid flow model vs. stochastic simulation:** fluid flow approach could lead to a more applied PhD project working with our particular system. Simulation approach could lead to a more general setting and potentially more work on methodology.

STORi Forum follow up (Friday 19th July 2024)

- Good suggestion to include a flow out of the system from shelter (i.e. abandoning queue)
- Suggestion to check how well our model performs against reality. I.e. if we had right data from Alameda County for last 10 years ago and we knew the decisions they made in that time, we could check our model agreed with what actually happened. While quality assurance is an important part of any modelling exercise, probably not an immediate priority given how difficult it may be to acquire the right data (if available). However, this may be a good way to quantify the performance of our "low-fidelity" fluid flow model in comparison to a high-fidelity stochastic simulation model.
- Some further thoughts/questions on points we are already considering (e.g. priority queueing, time-dependent weighting of the penalty on shelter in our objective function)

Next steps (Graham brainstorming w/c 22nd July)

Continuation of analysis with fluid flow model

Our recent work for the PSOR submission optimised a quadratic objective function subject to shape constraints on housing/shelter. This gave us an optimal solution (see Figure 1 below) which ramped up shelter initially (to bring the unsheltered queue down) and then decommissioned some shelter, making funds available for the continued increase in housing which the shape constraints required. The downside was the rise in the unsheltered queue while decommissioning shelter.

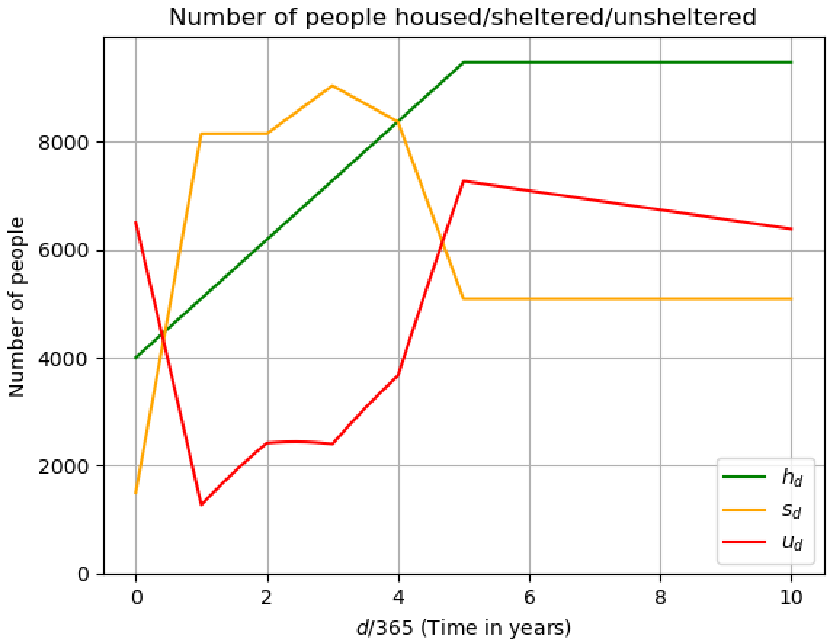

Figure 1: Optimal Solution from PSOR submission

We may be able to address this downside and find better solutions with a simpler approach. We first note that during this analysis it became clear that a "good" solution had two characteristics:

a) At the end of the building phase (5 years), to have enough housing that the total service rate exceeded the arrival rate (achieving stability from queueing perspective).
b) While the housing was being built, to minimise the unsheltered queue using shelter.

With this in mind, we note that previous work of Singham[1] took the approach of:

1) first estimating the long-term capacity requirement for housing for a steady state system (using estimates of arrival rate, mean service time and a required server utilisation of 0.95)
2) dealing with current backlog using double the housing capacity from step 1 for three years, and the capacity from step 1 thereafter. Shelter capacities differed in different scenarios.

We could adopt step 1 above into our PSOR formulation with a similar calculation of long-term capacity requirement and we could incorporate the steady build up to this capacity in our shape constraints (meeting objective (a) above). We could then work with our shelter capacity in order to control the unsheltered queue while housing was ramping up. An illustrative example of such a solution is shown below (**Note**: no optimisation is performed here, solution not necessarily meeting our PSOR constraints, only 5-year model run displayed).

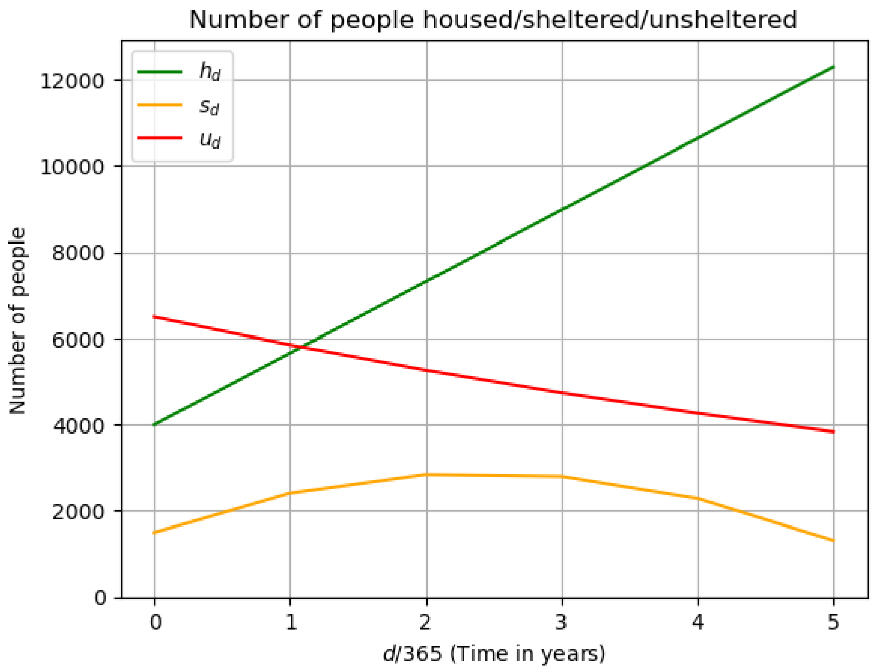

Figure 2 Illustrative solution with shape constraints on unsheltered queue

Here housing steadily increases to a sufficient level and a shallow unimodal shelter capacity function ensures the unsheltered queue drops by 10% each year. Finding a feasible solution like this could potentially be done by imposing shape constraints on the unsheltered queue.

This work could contribute to the literature on time-varying capacity planning for queueing systems and to the application of function estimation problems.

[1] Singham D, 2023, Estimating quantile fields for a simulated model of a homeless care system, WSC

Robust optimisation

Our work with the fluid model assumes knowledge of key inputs such as arrival and service rates. To incorporate uncertainty in these inputs (using e.g. robust optimisation) would improve the ability of the optimal solution to be successful in an uncertain future. I do not propose a formulation or research contribution here, given my limited experience in this area. However, in early September I will be in a better position to make suggestions following a 5-day course in 'Decision making with multiple criteria and uncertainty' which I am attending as part of the STORi training programme.

Stochastic constraints

The concept of functional zero homelessness (FZH):

Prob [wait for housing longer than some threshold] < alpha,

lends itself to a formulation using stochastic constraints where we could minimise the costs of housing/shelter capacity over time subject to achieving FZH by some realistic deadline. A further stochastic constraint on the wait in the unsheltered queue could balance the need for housing (to bring down overall waiting times) and shelter (to give some temporary relief to those waiting for housing). Here I do not suggest specific research contributions but do acknowledge that this is an active area of research in the simulation community, as illustrated in recent work by Eckman et al[2] who propose several metrics for evaluating feasibility in this setting.

[2] Eckman DJ, Henderson SG, Shashaani S, 2023, Stochastic Constraints: How Feasible Is Feasible? WSC

Multi fidelity simulation optimisation:

We have considered reasons why a stochastic simulation model may be justified and would give a higher fidelity view of the homelessness care system (compared to deterministic models). We have a deterministic optimisation problem which captures key elements of the time-varying capacity planning problem (note: there is of course scope to amend this formulation as discussed in PSOR submission). If we can only estimate our objective function with stochastic simulation, then this problem becomes a simulation optimisation (SO) problem. Several SO methods would then have relevance to our problem but there is scope for research contributions given our problem setting:

- For SO problems with integer-ordered decision variables, RSPLINE[3] and discrete stochastic approximation[4] move from a discrete space to a continuous space and move along gradients in the continuous space where those gradients are constructed based on multiple simulation replications at nearby integer points. A research contribution here could use multi-fidelity models to more efficiently estimate gradients.
- The Gaussian Markov random field approach[5] can be viewed as a Bayesian optimisation routine on a discrete space, where the objective function across the solution space is modelled as a Gaussian Random Field and the Markov property stipulates that the performance of each solution is independent of all solutions except those in a small neighbourhood. In this setup, no prior information is assumed about the structure of the problem. A research contribution could be to enforce structure (using prior distributions) given what we know about our problem using low-fidelity models. Furthermore, information from multi-fidelity models could be incorporated into the expected improvement criterion to make better informed decisions on where to simulate next.
- For SO problems with continuous-valued decision variables, multi-fidelity SO approaches typically model the objective function using a low-fidelity estimate plus corrective terms. The corrections may be modelled using polynomials[6] or Gaussian Processes (Multi-Fidelity Sequential Kriging Optimisation[7]). The expected information criterion in the latter considers both where to simulate next and with what degree of fidelity. I do not have a suggested research contribution here but it may be interesting use these methods on our problem.
- For problems with discrete decision variables, Ordinal Transformation followed by Optimal Sampling[8] offers an approach to make use of low-fidelity models (to order solutions in one-dimension) and high-fidelity models (to sample using stochastic simulation) to efficiently search for an optimal solution. In the optimal sampling step, a research contribution may be to introduce a multi-armed bandit as an alternative to OCBA to efficiently balance exploration and exploitation when deciding where to simulate next.

[3] Wang H, Pasupathy R, Schmeiser BW. 2013 Integer-ordered simulation optimization using R-SPLINE: Retrospective search with piecewise-linear interpolation and neighborhood enumeration
[4] Lim, E. 2012. Stochastic approximation over multidimensional discrete sets with applications to inventory systems and admission control of queueing networks
[5] Salemi, P., Nelson, B. L., and Staum, J. 2014. Discrete optimization via simulation using Gaussian Markov random fields
[6] Osorio, C. and Chong, L. 2015. A computationally efficient simulation-based optimization algorithm for large-scale urban transportation problems.
[7] Huang D, Allen TT, Notz WI, Miller RA. 2006 Sequential kriging optimization using multiple-fidelity evaluations
[8] Xu J, Zhang S, Huang E, Chen CH, Lee LH, Celik N. 2016 Mo2tos: Multi-fidelity optimization with ordinal transformation and optimal sampling

Input uncertainty analysis: Aforementioned SO approaches do not incorporate the uncertainty in model inputs. A research contribution would be to incorporate this uncertainty when looking for an optimal solution given that we cannot simply collect more data to improve our input models (as is typically done in the current input uncertainty literature), given that there is genuine high uncertainty in our system as we model far into the future. Different forecasting models make different assumptions and use different methods and we would like to capture this range when looking for an optimal solution. In thinking about how to tackle this difficult problem we may take inspiration from the sensitivity analysis discussed by Kleijnen[9].

[9] Kleijnen, J. P. C. 1994. Sensitivity analysis versus uncertainty analysis: When to use what?

Meeting notes (Tuesday 30th July 2024) – Graham, Dave, Luke, Rob, Dashi

- DO formulation: there may be better ways to capture trade-offs we're looking for, but not a high priority for now.
- Compartmental fluid model: given, say, two class of customers (using same resources but with different service times), a compartmental model could track the number of customers in each class over time and adjust the average service time accordingly during a model run. Graham to look at example from Dave to understand in more detail.
- Dashi positive feedback on ideas such as stochastic constraints (functional zero) and suggestion to consider a sample average approximation approach for SO.
- Simheuristics: a multi-fidelity approach where a heuristic method could identify good candidate solutions (using a low-fidelity model) and a high-fidelity simulation model could be used to evaluate candidate solutions to find an optimal solution. Rob to share review paper.
- Multi-fidelity SO (MFSO):
    - Careful thought needed in how to suitably apply above MFSO methods to our specific problem. A starting point could be to consider a more generic housing-waiting list problem and match it to one of the discussed MFSO methods, where a low-fidelity model was available and potentially helpful.
    - Above MFSO methods are all black-box methods (do not stipulate type of problem) but in using a low-fidelity queuing model we would utilise knowledge of the queueing problem. Challenge will be in how best to incorporate this knowledge.
    - SO with multi-fidelity modelling and SO with input uncertainty are both challenging areas – will be best to tackle them one at a time.
- Next steps: Graham to get stuck into the nuts and bolts of relevant SO methods.