1 **An Enhanced and Efficient Population Synthesis Approach to Support Advanced Travel**
2 **Demand Models**
3
4 **Ramachandran Balakrishna***
5 Senior Transportation Engineer
6 Caliper Corporation
7 Newton, MA, USA
8 Email: rama@caliper.com
9
10 **Srinivasan Sundaram**
11 Senior Transportation Engineer
12 Caliper Corporation
13 Newton, MA, USA
14 Email: srini@caliper.com
15
16 **Jim Lam**
17 Director, TransCAD Applications and Support
18 Caliper Corporation
19 Newton, MA, USA
20 Email: jimlam@caliper.com
21
22
23 *Corresponding author
24
25
26 Word Count: 4,969 words + 6 figures and 1 table (250 words per figure/table) = 6,719 words
27
28
31
32
33
34
35 **To be reviewed by the Transportation Planning Applications Committee (ADB50)**
36

1 **ABSTRACT**
2 An enumerated population (comprised of all households and their residents) is a key input to advanced
3 travel demand models of the tour-based, activity-based and hybrid genres. A synthetic population allows
4 the model to be sensitive to person-level behavioral heterogeneity and facilitates the use of demographic
5 and other variables (such as gender) that are otherwise not intuitive in aggregate models. Synthesis starts
6 from a seed sample of households and the persons living therein, estimates the number of each household
7 type per geographic unit, and samples from the seed according to weights. Such a population must
8 however reflect the distribution of key variables in the study region, captured by marginals collected at
9 the household and person levels. Popular synthesis techniques have generally focused on matching
10 household marginals but do not explicitly control for person-level distributions. While attempts have been
11 made to extend these methods to simultaneously fit person marginals, the results appear to be
12 experimental, create more data-related problems, and take a long time to run on large-scale, real-world
13 data. In this paper, we review the state of the art and state of the practice of population synthesis methods,
14 identify the key limitations, and propose simple techniques to overcome the same. We also demonstrate
15 the novel use of third-party data sources to correct errors in the marginals. The enhanced approach is
16 applied on two large, real-world examples in the USA: Las Vegas, Nevada and the Central Coast,
17 California. Empirical evidence supports the significant improvement in the quality of the synthesized
18 population.
19
20 **Keywords:** Population synthesis, Activity-Based Models (ABM), Iterative Proportional Updating (IPU),
21 Person totals

1 **INTRODUCTION**
2   Travel demand models form the backbone of technical analyses that inform public policy and
3 infrastructure investment decisions. Underlying each model is a set of assumptions pertaining to the
4 population's desire to travel, the ensuing interactions with the transportation network, and the
5 consequently emergent patterns of congestion and environmental effects. Advanced travel demand
6 models such as Activity-Based Models (ABMs) and hybrid varieties that include disaggregate predictions
7 can allow for more refined sensitivities to the behavioral heterogeneity that inevitably exists even between
8 seemingly similar households. This increased modeling fidelity can be attributed to the disaggregate (or
9 person-level) treatment of travel demand, which permits the assessment of person-specific variables such
10 as gender, age and driver license ownership when simulating personal and household decision outcomes
11 for the population. Such variables are generally difficult to handle in more traditional aggregate
12 approaches such as trip- and tour-based travel demand models, which work at the level of zonal averages.
13   A crucial requirement for advanced travel demand models is a robust population synthesis
14 process. This step generates a complete (but synthetic) enumeration of the study region's population in
15 such a way that their characteristics sum up to measured aggregate totals at various levels of geography.
16 In addition, the characteristics can be measured either at the level of households, or individuals. Examples
17 of such characteristics include household size, vehicle ownership and income; and household residents'
18 age, gender and work industry category. The target totals, also called marginals, may be drawn from
19 various data sources such as the Census (e.g. American Community Survey, ACS).
20   An accurate synthetic population, one that captures the real-world distributions of the key
21 variables in each geographic element of the study region, can serve as a high-fidelity laboratory to
22 forecast every individual's daily activity participation choices, the travel decisions they make to support
23 these choices, and the network-level impacts of the same. Such a laboratory provides a powerful platform
24 to validate the forecasting models against the current reality as well as test the system's wide-ranging
25 sensitivity to changes in infrastructure, policy and consumer behavior. Critically, these tests may be
26 conducted before expensive investment decisions are finalized, and without negatively impacting the real
27 population through potentially sub-optimal or counter-productive policies.
28   Population synthesis methods in practice have remained relatively simple in their underpinning
29 mechanics despite being in use for several decades now. These methods are also associated with several
30 technical problems, some due to the mathematical assumptions embedded within the approach and others
31 related to the quality and quantity of the available input data. In the following sections, we review the
32 predominant synthesis methods documented in the literature and prevalent in practice, along with their
33 most important limitations. This review includes methods proposed to circumvent some of these
34 problems, though they often create new problems that must then be resolved during real-world
35 deployment. We then propose a heuristic technique that eliminates or minimizes these problems by
36 simplifying a prior approach. We make a key contribution by illustrating the use of third-party data
37 sources to correct for errors in the marginals/control totals. We then demonstrate, via two separate case
38 studies on real-world data, the significant improvements to both accuracy and running times when
39 compared to the state-of-the-art and state-of-the-practice. We conclude with some future directions to
40 improve the use of the tool in practice.
41
42 **LITERATURE REVIEW**
43   Most popular population synthesis methods adopt a framework such as that shown in **Figure 1**.
44 They start with a sample of household (HH) and person records, typically obtained from a survey of the
45 study region, and tagged to a high-level geography such as Public Use Microdata Areas (PUMAs). An
46 Iterative Proportional Fitting (IPF) step then estimates the number of each type of household (identified in
47 the survey) to be generated for each of the sub-zones (e.g. blocks) contained within a zone. This is
48 achieved by identifying the number of households of each type that will add up to various household
49 marginals at the sub-zone level. The term "Nested" indicates that the household marginals may
50 themselves be specified at different (but nested) geographies: for example, the vehicle ownership
51 marginals may be at the traffic analysis zone (TAZ) level while the income marginals may be at the level

1 of Census blocks or block groups. Nested synthesis will drill all the marginals down to a common
2 geography (usually the smallest/finest among all input geographies). In **Figure 2**, for instance, the
3 background fill colors demarcate Public Use Microdata Areas (PUMAs) that are the typical geographic
4 resolution for the Public Use Microdata Sample (PUMS) survey that seeds the population synthesis
5 process. Marginals are generally available at finer resolutions such as block groups, traffic analysis zones
6 (TAZs) or blocks. The synthesis procedure must identify the number of each type of PUMS household
7 (and its constituent residents) to pick for each small area (say a block) so as to match various
8 characteristic marginals available at that (block) level.

9
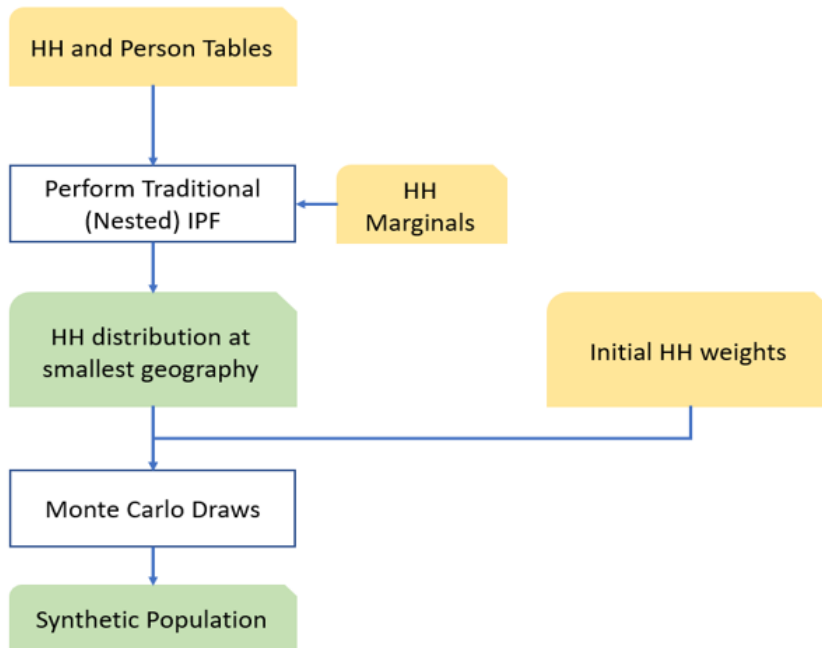
10 **Figure 1 General population synthesis process flow**
11
12   Finally, the predicted number of households in each sub-zone are sampled from the surveyed
13 households using externally supplied (initial) weights. All persons living within a chosen household are
14 automatically copied into an output file, which eventually becomes the synthetic population. This general
15 approach has also been called *synthetic reconstruction* (*1*).
16
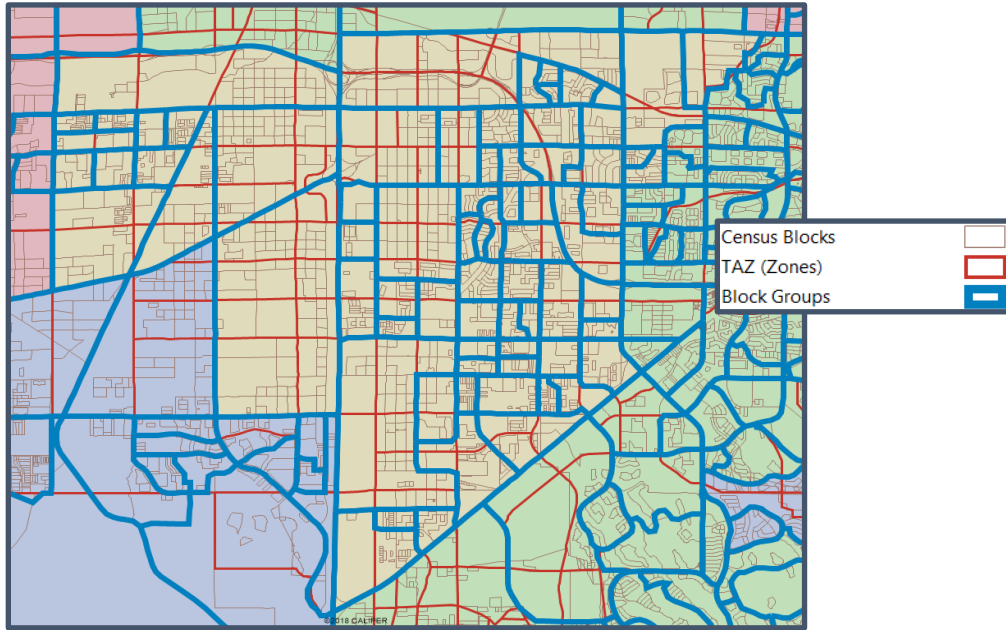
Figure 2 Nested geographies: PUMAs, block groups, TAZs and blocks

A detailed description and example of IPF mechanics are presented in (*1*). Briefly, iterations are performed to scale first the rows and then the columns of a seed (or reference) matrix to match row and column totals (the marginals) respectively. Every correction to the rows (columns) needs a re-evaluation of the fit to columns (rows), and the process continues until a satisfactory fit to both marginals is obtained. The seed matrix is typically constructed by counting the records of a household survey that fall into each combination of demographic variables. IPF provides a very simple mathematical formulation to the above problem and can be easily implemented in a wide array of software tools and scripts. However, the above framework has serious limitations. The most important drawback is the lack of control on person marginals. Since the algorithm only controls for household variables, the obtained distribution of person attributes in any given sub-zone is largely a lottery. There is no guarantee that the chosen households will yield, say, the correct distribution of age, gender, etc. in the synthetic population. Conversely, if IPF is performed on the person marginals, the fit at the household level is uncontrolled. From a demand modeling perspective, IPF is unable to simultaneously fit the distributions of both the household and person variables, which can have significant impacts on the results of the model.

An additional critique of IPF is the common set of sampling weights used for all zones (and sub-zones) in the region. Since all sub-zones must pick from the same pool of candidate survey households falling within their parent zone, the Monte Carlo step assumes identical distributions of those household types in all sub-zones. This is likely to be highly erroneous in practice, and a scheme that allows the weights to vary across sub-zones must be preferred over the default approach relying on common weights alone.

Other practical issues have also been encountered in deployed instances of population synthesis. The reliance on potentially ad hoc starting weights is a critical drawback, since they are traditionally calculated using simple heuristics. Realistic survey sample sizes also tend to be on the smaller side, which can introduce zeroes in the seed matrix when no households of certain types were sampled. Such structural zeroes persist through the IPF's scaling process and remain zeroes in the final solution. The zero-cell problem has received substantial research focus, as seen in (*2*).

Since advanced demand models are based on numerous disaggregate choice models that make heavy use of demographic variables, errors in person attributes can have a significant impact on the accuracy of the model's predictions of activity patterns and travel levels by mode and purpose. For instance, incorrect forecasts of the number of households with working-age adults or school-age children will result in biased estimates of mandatory travel related to work, school and child-care. It is therefore desirable to include person marginals in the population synthesis process.

1   The need for person marginals in the synthesis has been acknowledged in the literature, and several
2   ideas have been presented to fill this gap. Proposed solutions typically fall within two broad categories:
3   those based on mathematical optimization frameworks, and those extending the IPF philosophy to include
4   person variables in addition to household variables. While detailed reviews of population synthesis methods
5   may be found elsewhere (*1,3*), we present brief ideas of some of the relevant papers below.
6   An entropy-maximizing objective function was proposed by (*4*), in which the household weights
7   are adjusted in a "list-balancing" framework. First, the household weights are adjusted to improve the fit to
8   both household and person marginals while remaining as close as possible to the initial weights.
9   Subsequently, utilities are executed to systematically round any fractional weights. A slightly modified
10   version of the above algorithm is deployed by (*5*). Both methods argue that the least amount of new
11   information must be added to the initial (seed) solution, which however limits the ability to correct large
12   errors in the seed. A combinatorial optimization methodology is adopted in (*6*), where an initial list of
13   households (or a prior population synthesis) is processed by selecting households for either replication,
14   deletion or swapping with another household/unit. If a starter list is unavailable (as is likely to be the case),
15   one is randomly assembled assuming that all households are equally likely. Such assumptions, as well as
16   the nonlinear and non-convex nature of the resulting formulation, can cause long run times, local optima,
17   and sensitivity to the scale of the objective function. Indeed, the authors report a run time of more than 11
18   hours for a problem with 5101 zones, and further indicate that the solution approach does not lend itself to
19   parallel computing.
20   Finally, the above methods tend to localize the estimated household weights around the initial input
21   values, which can be sub-optimal given that the initial weights are often of unknown vintage. While such
22   formally devised optimization methods do possess desirable qualities such as the ability to move away from
23   starting zeroes, they may not be generally applicable in a manner suitable for most practitioners. Extensions
24   to the approached outlined above are also in evidence (*7-8*). Another approach is the assembling of the
25   population one household at a time, using a fitness function and a "greedy" heuristic algorithm to determine
26   which household to add at each iteration (*9-10*). Neither study quantifies the running time, though.
27
28   **Iterative Proportional Updating (IPU)**
29   Iterative Proportional Updating (IPU) has been proposed as a way of handling person marginals
30   while essentially retaining the simplicity of the IPF framework (*11*). Here, each household's initial weight
31   is adjusted while explicitly considering its contribution to both household and person marginals through an
32   incidence table. For example, households with 4 residents and including children below age 5 years will
33   have their weights adjusted to better match the total number of such children predicted to live in a zone, in
34   addition to matching the number of 4-person households in that zone. Other households (which do not
35   feature such children) will not be impacted during this adjustment. Since there can be many marginals to
36   be matched in this process, the weights are adjusted sequentially (i.e. one marginal at a time), working with
37   only the relevant subset of households each time. After all marginals have been treated in this manner, the
38   procedure returns to the first marginal and continues to iterate until convergence (as defined by the modeler)
39   is reached. Note that the same household(s) may be weighted differently for different zones, which adds
40   precision compared to the use of *a priori* weights supplied externally. Further, this household re-weighting
41   may be applied at a different (usually more aggregate) geography than that used for simulating the
42   population.
43   The published version of the IPU procedure attempts to match the joint distribution of the marginals
44   of relevance, by exhaustively enumerating all possible combinations of their different levels. It starts with
45   a table that captures the incidence of each household record against every possible **combination** of variable
46   levels across both households and persons. If household auto ownership is divided into four levels (0, 1, 2,
47   3+), household income is split into three levels (low, medium, high), person age is divided into five
48   categories (0-5, 5-18, 18-25, 25-60, 60+), and person gender is divided into two levels (male, female), this
49   gives rise to $4*3*5*2 = 120$ columns. The inclusion of more variables and/or more levels within the
50   variables, a likely situation in practice, only increases this number rapidly and renders the incidence table
51   even sparser. A typical survey would thus have its records spread extremely thin across these numerous
52   columns, potentially over-extending (in a statistical sense) the information contained in the data. While this
53   joint distribution provides the theoretical possibility of a more accurate and finely profiled synthetic

1   population, it exhibits long run times and potentially exacerbates the zero-cell problem as acknowledged in
2   (*11*).
3   We propose a simplification of the procedure to reduce or even eliminate the zero-cell problem
4   when a reasonable sample size is available, while allowing for a highly efficient solution algorithm for
5   rapid application on large geographic extents. These are critical contributions toward the deployment of
6   general synthesis tools that can be widely used without case-specific customization of the solution
7   algorithm.
8
9   **METHODOLOGY**
10   The proposed technical approach is illustrated by the flowchart in **Figure 3**. The key difference
11   from **Figure 1** is the introduction of the IPU step which modifies the initial weights for each zone
12   independently and matches both household and person marginals.
13
14   **IPU enhancements**
15   The previously published work on IPU describes the the zero-cell problem, in which many columns
16   are largely filled with zeroes and hence causes the IPU to either struggle or fail. Attempts have been made
17   to overcome this difficulty by creatively re-using the available data, but this is statistically inefficient since
18   no new information is being introduced into the solution. Our adaptation largely alleviates the challenge by
19   only looking to match the marginals of individual variable levels and not their exhaustively enumerated
20   combinations. We consider each household and person variable and its levels, but not their combinations.
21   Under this setup, the example above would involve only 4+3+5+2 = 14 columns (instead of 120 columns
22   in the existing method), which gives the solution algorithm a significantly better chance at converging. An
23   added attraction is the greatly reduced running time, which is orders of magnitude lower than that expected
24   from the previously published IPU approach.
25
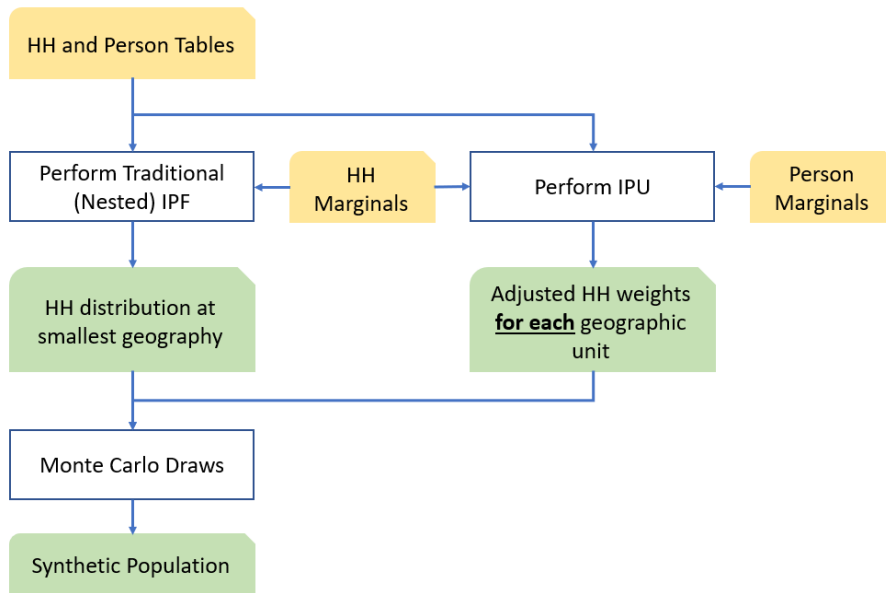
26   **Figure 3 Enhanced population synthesis with integrated IPU**
27

28   An example of households and their contributions to various household and person marginals is
29   shown in **Table 1**. The fit in each column is computed as a function of the column sum and the
30   corresponding target marginal. All column objective functions are then collected into a single overall
31   objective function that must be minimized to reach an optimum solution. The objective function
32   calculations are general and can take many forms. For example, the fit δ_j for column j may be computed
33   as:

$$\delta_j = \frac{\sum_i[w_i d_{ij}] - m_j}{m_j} \tag{1}$$

where w_i is the weight of household i; d_{ij} is the contribution of household i to column j; and m_j is the marginal for column j. The overall objective function δ could then be expressed as:

$$\delta = \sum_j (\delta_j)^2 \tag{2}$$

Other mathematical forms for equations (1) and (2) may be found, for example, in (*11*).

TABLE 1 Household incidence table for enhanced Iterative Proportional Updating

HH ID	Initial HH Weight	HH Veh 0	HH Veh 1+	Per Age 1	Per Age 2	Per Age 3
1	1	1	0	1	1	1
2	1	1	0	1	0	1
3	1	1	0	2	1	0
4	1	0	1	1	0	2
5	1	0	1	0	2	1
6	1	0	1	1	1	0
7	1	0	1	2	1	2
8	1	0	1	1	1	0
Column Sum		3	5	9	7	7
Marginals		35	65	91	65	104
Objective Function	4.563	0.914	0.923	0.901	0.892	0.933

It should be noted that the IPU step in **Figure 3** essentially re-weighs the relevant survey sample records for each zone. The traditional IPF step based on just household marginals is still used to determine how many of these households must be sampled into the synthetic population. This sampling will be achieved using the adjusted weights generated by IPU. The marginals for IPU and IPF can thus be specified at different geographic resolutions if necessary and may depend on the resolutions at which various marginals are available for the study region. For instance, if the person marginals are available only at the block group level, the IPU may be executed at this level. The IPF though, can still be executed at a finer block resolution using the household marginals.

The modified IPU approach was implemented in the TransCAD travel demand modeling platform to test its effectiveness from the perspectives of modeling accuracy and running times. The baseline for the tests was the standard IPF-based population synthesis already available in TransCAD (*12*). This baseline fits only the household marginals, while the person totals are essentially inherited through the residents of the selected survey households in each zone.

CASE STUDIES

We tested our methodology on two real-world datasets corresponding to locations that are actively exploring advanced travel demand models. The first is the Regional Transportation Commission for Southern Nevada (RTCSNV) in charge of transportation modeling for the Las Vegas, Nevada region. The other is the Central Coast region of California, comprised of the Metropolitan Planning Organizations (MPOs) for Monterey Bay, Santa Barbara and San Luis Obispo. While the former is investigating a hybrid travel demand model that will selectively and judiciously introduce disaggregate destination choice models into their existing model, the latter is building a data-driven activity-based model (ABM) from state-wide survey data. Both the hybrid and the ABM paths must start from a realistic synthetic population, which motivated the current research. Preliminary results from these tests have been presented at recent conferences (*13, 14*).

1 **The Las Vegas application**
2
3   The PUMS data for the Las Vegas case study covered 13 PUMAs. IPU was performed at the block
4 group level while the IPF component ran on the blocks. The region has 1,294 block groups and 24,521
5 blocks. The household and person marginals and their levels are:
6
7   •   Household size (1, 2, 3, 4, 5+)
8   •   Household vehicles (0, 1, 2, 3+)
9   •   Household income (USD in thousands: 0-25, 25-50, 50-75, 75-100, 100-150, 150-200, 200+)
10   •   Person gender (Female, Male)
11   •   Person age (in years: 0-4, 5-14, 15-19, 20-24, 25-44, 45-64, 65-100)
12
13   The numerical results confirm that the no-IPU baseline can indeed match household marginals with
14 a high degree of accuracy. This is to be expected, as the baseline synthesis process explicitly controls for
15 these marginals. **Figure 4**, for example, shows the fit to household marginals for the number of households
16 with 5+ residents, with the horizontal and vertical axes representing the target and synthesized totals at the
17 block group level. The introduction of the modified IPU step does tighten the fit especially toward the
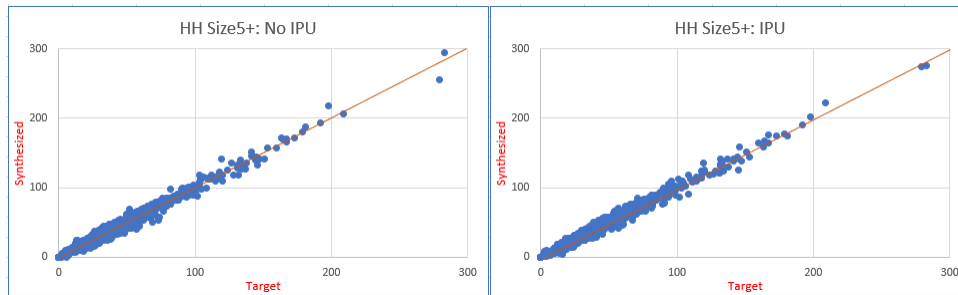18 higher end of the chart.

19 **Figure 4 Household size 5+ marginals (a) Without IPU and (b) With IPU**
20
21   Similar near-perfect fit is obtained both with and without IPU for all other levels of household size,
22 as well as each level of household auto ownership and income. However, the benefits of the IPU corrections
23 are realized when comparing person marginals. **Figure 5**, for instance, compares the fit to the number of
24 male residents across block groups, indicating a significant tightening of the solution when the proposed
25 IPU is employed. An even more distinct comparison is obtained when comparing the age marginals. The
26 power of IPU is made obvious by focusing on the number of people aged 65-100 years, where a systematic
27 bias is also corrected while improving the overall fit:
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42

1 **Figure 5 Gender (male) and Age (65+ years) (a) Without IPU and (b) With IPU**
2
3 **The Central Coast application**
4
5   The PUMS data for the California case study covered 10 PUMAs. IPU was again performed at the
6 block group level while the IPF component ran on the blocks. The region has 941 block groups and 39,660
7 blocks. The household and person marginals and their levels are:
8
9   • Household size (1, 2, 3, 4, 5+)
10   • Household vehicles (0, 1, 2, 3+)
11   • Household income (USD in thousands: 0-25, 25-50, 50-75, 75-100, 100-150, 150-200, 200+)
12   • Person gender (Female, Male)
13   • Person age (in years: 0-4, 5-14, 15-19, 20-24, 25-44, 45-64, 65-100)
14   • Person work industry (Agriculture, Manufacturing, Utilities/Transportation/Construction,
15   Wholesale, Retail, Finance/Insurance/Real Estate, Education, Healthcare, Service, Public
16   Administration, Non-Worker)
17
18   The list of marginals above is the same as that for Las Vegas, with the inclusion of worker status
19 by industry. Accurately predicting worker status by industry is highly beneficial for travel demand
20 modeling, since it can help narrow a person's potential work locations (during destination choice
21 calculations) to only those that have employment in her work industry.
22
23   The numerical evidence for the Central Coast dataset reinforces the findings from the Las Vegas
24 example. **Figure 6** summarizes the fit to person age marginals for the 20-24 and 65+ categories, indicating
25 that the enhanced IPU is again highly successful in generating more representative populations than the
26 initial weights alone.

1
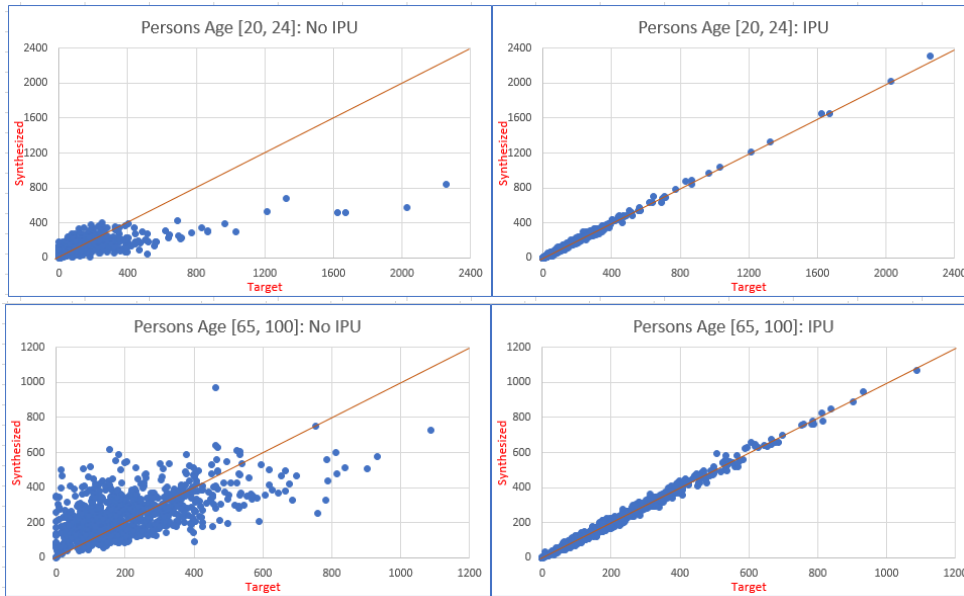
Figure 6 Age 20-24 and 65+ years (a) Without IPU and (b) With IPU

Computational efficiency

While the goodness of fit provided by the enhanced and simplified population synthesis approach has been demonstrated above, a critical additional contribution of this research is the drastic reduction in running times. Each of the two case studies ran in less than 10 minutes on standard desktop hardware generally available to planning agencies around the US. This is an encouraging benefit compared to the state of the practice, which either reports run times of 11-16 hours (*6, 10*), or requires expensive hardware and large amounts of RAM to achieve more practical times (*5*). An advantage of small run times is the flexibility to test the synthetic population's sensitivity to perturbations in the input marginals, or when several scenarios need to be run on different sets of population assumptions.

A note on the consistency of marginals

The success of the IPU procedure (whether the previous published version or our proposed enhanced version) relies greatly on the consistency of the various marginals involved. It could also be argued that such consistency is required for any synthesis approach, since the existence of a feasible solution(s) depends on it. In practice, however, it is highly likely that the household and person marginals do not always add up. This situation was noted in both test cases described above. When such inconsistencies exist in the data, it is not surprising that IPU can match either the household marginals or the person marginals, but not both.

The primary cause for such inconsistencies was identified as the highly skewed average household size estimates for the last household size category (in our case, these were the households with seven or more residents). The Census estimates for the average household size for 7+ households were consistently in the range of 17 to 42, which is extremely high for reasonable households. Our hypothesis is that multi-family and senior living situations were combined into single households, which tends to skew the average occupancy upwards. Group quarters were not a part of these data. For the purpose of travel demand modeling though, these units should be treated separately, since their daily activities, constraints and behaviors are expected to be mutually exclusive. Since these data were inherently inconsistent, external information was necessary in order to resolve the errors in a realistic manner.

We adopted a novel corrective process in which a third-party data source, one sourced from retail and other transactional records, was used to arrive at more reasonable and realistic estimates of average household size for the 7+ case. This was used as a control variable to re-distribute "excess" Census households into lower-size bins proportional to their incidence in the Census. Such corrections were found

1 to be critical in ensuring data consistency for IPU. In addition, realistic household sizes are also essential
2 in downstream components of ABMs, since they can result in more plausible interactions between the
3 members of smaller households. Models of joint activity participation among household members can also
4 become tractable and obviate the need for some of the simplifying assumptions seen in practice.
5
6 **CONCLUSIONS**
7 In this paper, we identify key limitations in the handling of person marginals in currently
8 documented population synthesis techniques. We propose enhancements that address these gaps, while
9 simplifying the problem and resolving critical issues (such as the zero-cell problem) raised in the existing
10 body of work. Our proposed methodology was implemented and tested on two large-scale and real-world
11 case studies, and the empirical analysis clearly shows the efficacy of the enhanced methodology. We also
12 provide a qualitative discussion on algorithmic efficiency and show that our simplifications lead to more
13 robust convergence behavior. The enhanced population synthesis is expected to form the backbone of
14 future hybrid and activity-based model development for the two regions. Additionally, the enhanced IPU
15 can also be used to (re-)weigh existing surveys for more traditional trip-based model calibration, thus
16 providing benefits across the spectrum of travel demand models. An interesting avenue for future work is
17 to compute the joint distributions of the marginals from the synthesized population and compare the same
18 against those found in the survey.
19

26 **AUTHOR CONTRIBUTIONS**
27 The authors confirm contribution to the paper as follows: study conception and design: R. Balakrishna, S.
28 Sundaram; data collection: S. Sundaram, J. Lam; analysis and interpretation of results: R. Balakrishna, S.
29 Sundaram; draft manuscript preparation: R. Balakrishna. All authors reviewed the results and approved
30 the final version of the manuscript.
31

32 **REFERENCES**
33 1. Muller, K. and K.W. Axhausen, K.W. Population synthesis for microsimulation: State of the art,
34 Presented at the TRB annual meeting, 2011.

35 2. Abdoul-Ahad, C. and A.R. Mamdoohi. Population synthesis using iterative proportional fitting (IPF): A
36 review and future research. Transportation Research Procedia, 2016. Vol. 17, pp. 223-233.

37 3. Barthelemy, J. and E. Cornelis. Synthetic Population: Review of the existing approaches, FUNDP –
38 University of Namur, Belgium, 2012. Working paper no. 2012-8.

39 4. Vovsha, P., J.E. Hicks, B.M. Paul, V. Livshits, P. Maneva, P. and K. Jeon. New features of population
40 synthesis. Presented at the TRB annual meeting, 2015.

41 5. Paul, B.M., J. Doyle, B. Stabler, J. Freedman and A. Bettinardi. Multi-level population synthesis using
42 entropy maximization based simultaneous list balancing. Presented at the TRB annual meeting, 2018.

43 6. Abraham, J.E., K.J. Stefan and J.D. Hunt. Population Synthesis Using Combinatorial Optimization at
44 Multiple Levels. Presented at the TRB annual meeting, 2012.

45 7. Pritchard, D.R. and E.J. Miller. Advances in population synthesis: fitting many attributes per agent and
46 fitting to household and person margins simultaneously. Transportation, 2012. Vol. 39, Issue 3, pp. 685-
47 704.

48 8. Muller, K. and K.W. Axhausen. Multi-level fitting algorithms for population synthesis. Working paper,
49 2012. Paper No. 821, Institute for Transport Planning and Systems (IVT), ETH Zurich, Zurich.

9. Hafezi, M.H. and M.A. Habib. Synthesizing Population for Microsimulation-Based Integrated Transport Models using Atlantic Canada Micro-Data. Procedia Computer Science, 204. Vol. 37, pp. 410-415.

10. Srinivasan, S., L. Ma and K. Yathindra. Procedure for Forecasting Household Characteristics for Input to Travel-demand Models. Report prepared for the Florida Department of Transportation, 2008. TRC-FDOT-64011-2008.

11. Ye., X., K. Konduri, R.M. Pendyala, B. Sana and P. Waddell. A methodology to match distributions of both household and person attributes in the generation of synthetic populations. Presented at the TRB annual meeting, 2009.

12. Caliper. Travel Demand Modeling with TransCAD. User manual, 2018.

13. Balakrishna, R. and S. Sundaram. Practical population synthesis enhancements to support advanced and activity-based travel demand models. Presented at the World Conference on Transport Research (WCTR 2019), 2019.

14. Balakrishna, R., S. Sundaramand J. Lam. A Deployed Population Synthesis Approach that Matches Person Control Totals and Avoids Known Pitfalls. Presented at the TRB National National Transportation Planning Applications Conference, 2019.