

SA 法による浜松市の企業仮想データの生成

○岩田恵人 李皓（静岡大学）

Generation of Virtual Company Data in Hamamatsu by Simulated Annealing Method

* K.Iwata and H. Lee (Shizuoka University)

概要一 日本において少子高齢化や空き家問題が進行している。これらの問題は浜松市においても例外ではなく、対策を行う必要がある。これら都市問題の把握や政策検討にマイクロシミュレーションの導入が進んでおり、政令指定都市の中で最も面積が広い故に、それぞれの地域で異なる都市問題を抱える浜松市への適用は大きな意味を持つ。マイクロシミュレーションの適用には、その第一段階として基盤となるデータを取得する必要がある。取得できない場合は何らかの手法で仮想データを生成する必要がある。本研究では浜松市の人口動態や不動産需給の変化をマイクロシミュレーションで分析するために必要となる企業仮想データを SA 法を用いて生成する。

キーワード: SA 法, 個票生成, 最適化, マイクロシミュレーション

1. はじめに

近年、日本では少子高齢化が進行し、今後の人口推計¹⁾では、Fig. 1 で示すように、総人口が2030年（平成42年）の1億1,662万人を経て、2048年（平成60年）には1億人を割って9,913万人となり、2060年（平成72年）には8,674万人になると見込まれている中、現時点では地方を中心に人口減少が始まっている。少子化は人口減少による生産性・労働力の低下、高齢化は社会保障費の負担の増加といった影響を与え、若年層の生活が圧迫されていく。若年層が年金制度や介護などの費用を大きく負担することになれば、今後もさらに国民1人当たりの経済的負担が増して、子どもを産み育てる環境がなくなっていくと考えられる。

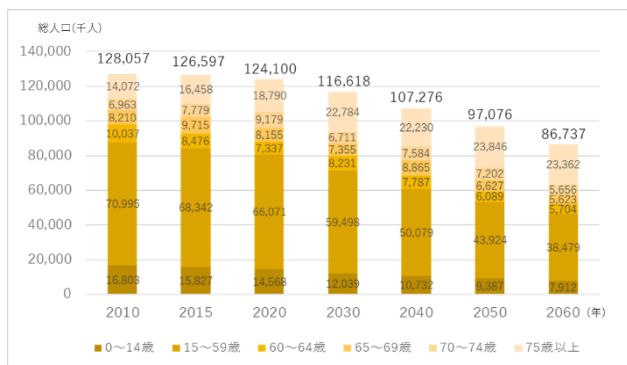


Fig. 1: 年齢区分別将来人口推計。

内閣府「将来推計人口でみる50年後の日本」¹⁾を基に作成

このように少子高齢化は経済面や社会面、生活水準などの様々な領域に影響を与える。そして、その中でも大きな問題の一つとなるのが空き家の増加である。Fig. 2で示すように、平成25年において、空き家数は820万戸となり、平成20年に比べて8.3%増加した。総住宅数に対する空き家率は平成10年に初めて1割を超えて11.5%となり、平成25年には13.5%と空き家数・空き家率共に過去最高となった。

空き家率が上昇する原因として、総住宅数が平成25年において6063万戸と5年前より5.3%増加し、少子高齢化が進展しつつある現状がありながらも、新築物件が供給され続けていること、住宅所有者の高齢化によって住人の死亡や入院で空き家となってしまうことなどが挙げられる。³⁾

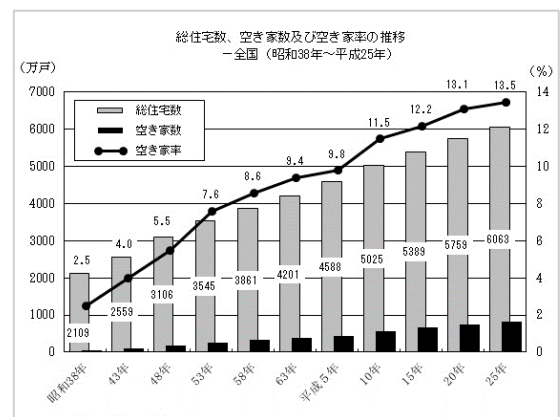


Fig. 2: 総住宅数、空き家数及び空き家率の推移.²⁾

増え続ける空き家によって起こる問題は、建物の老朽化や害虫の大量発生による景観トラブルや、ごみの不法投棄による衛生の悪化など様々であり、早急に対策を講じる必要がある。空き家発生の原因とそれにより生じる問題に関する自治体へのアンケート調査の結果をまとめた図をFig. 3, 4で示す。日本ではこのような問題に対して、平成26年11月に地方公共団体が空き家対策を行いやすくする「空家等対策特別措置法」の成立、国土交通省による空き家の解体やリフォームに対して援助が受けられる「空き家再生等推進事業」の実施といった空き家の増加を抑制しようとする動きが見られる。

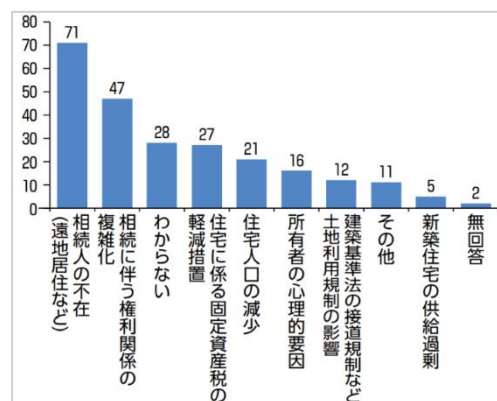


Fig. 3: 空き家発生の主な原因 (N = 116, 複数回答).³⁾

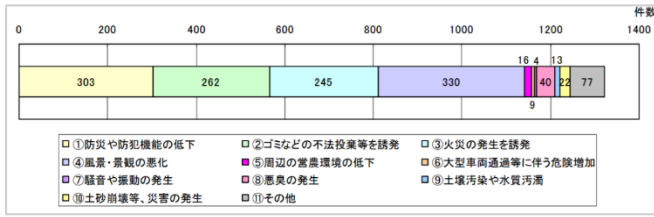


Fig. 4: 空き家発生により生じる影響 (N = 1,321, 複数回答).⁴⁾

前述した少子高齢化の影響は浜松市においても例外ではない。Fig. 5, 6で示すように、浜松市のほとんどの区において人口が減少すると共に65歳以上の世帯が増加していく傾向にあり、今後も少子高齢化が進展していくと推定されている。⁵⁾これを受けて浜松市では少子高齢化社会や人口減少、それに伴う空き家の急増などの社会情勢の変化に対応するための計画の策定・推進を行っている。⁶⁾

全国各地で空き家に対する取組みが行われる中、都心部、郊外住宅地、中山間地域それぞれで空き家問題を抱える浜松市は様々な対策を行う必要がある。市内では平成30年に静岡県内で初となる「空家等対策特別措置法」に基づく略式代執行で空き家一棟に対して一部強制撤去がなされた。また市内における空き家の流通を促すための「空き家バンク」が設置されるなど活発な取組みが見られる。

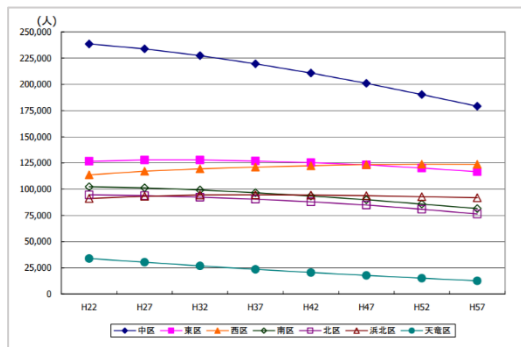


Fig. 5: 浜松市の行政区別人口の推移.⁵⁾

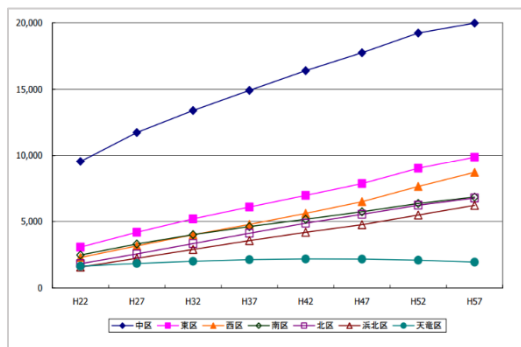


Fig. 6: 浜松市の行政区別 65 歳以上単独世帯の推移.⁵⁾

このような世帯構成の変化や建物の流通といった都市全体の動きを把握し、政策立案および検討を行うにあたってマイクロシミュレーションの導入が進んでいる。例としては、アメリカ合衆国の各州、フランスなどで実際に政策検討に用いられたUrbanSim⁷⁾が挙げられる。マイクロシミュレーションは都市構造が複雑化していく現在において、都市政策の効果や影響を詳細に

分析するために有効なアプローチとなる。都市政策は実施にあたって、現在発生している問題を解決していくことは当然のこと、事後的ではなく事前の予測によって問題の発生を抑止できるような対策を実施していかななくてはならない。

浜松市は産業や経済の中心となる都心部、農地が広がる郊外部、空き家の増加が顕著に現れる中山間部など特徴ある様々な地域を持ち、少子高齢化問題、空き家問題を抱える。浜松市の人口動態や建物の流通等を分析するマイクロシミュレーションを導入することは問題を解決していくにあたって大きな意味を持つ。

マイクロシミュレーションやエージェントベースシミュレーションを行うために、本研究では、その第一段階としてモデル化の対象にあたる市民や事業所などといった活動主体を生成する。政府が国勢調査を通して収集する市民や事業所に関するデータを個票といった形で用いることができれば、現実に近いシミュレーション結果を得ることができると考えられるが、プライバシーの観点から公開されていないのが現状である。そのため、政府や行政が一般に公開している統計データから個票データを復元することはシミュレーションモデルの有効な構築手法となる。

本研究において、統計データに基づいた市民や事業所の個票データの復元を行うために用いる手法が池田ら⁸⁾やMurata, Haradaら⁹⁾¹⁰⁾が用いるSA法 (Simulated Annealing) を用いたデータ復元手法である。個票データの復元を実現するための手法は数多く提案されており、例えば、個票データのサンプルをもとにIPF (Iterative Proportional Fitting) を行うことにより統計データに近いデータを生成するSR法 (Synthetic Reconstruction) がある。本研究におけるSA法を用いたデータ復元手法は、SR法の一つであり、サンプルを用いない手法となる。具体的には一般に公開されている統計データと初期生成した仮想の個票データとの誤差を計算する目的関数を設計し、SA法によって誤差の値を最小化していく手法である。

本研究の最適化の対象は従業者数とし、SA法を用いて、浜松市が公開している事業所統計¹²⁾に適合した仮想データを生成する。

2. SA法を用いた仮想データ生成

本研究では梶井らが提案した手法¹¹⁾を基に、浜松市の企業に関する仮想データを生成する。梶井らは、池田ら⁸⁾の手法を軸に、市民の年齢や世帯数に関するより多くの統計データに適合した復元手法を提案している。梶井らは、池田らの手法の目的関数に関する改良を提案している。式(1)が梶井らが提案する目的関数である。

$$f_s(A) = \sum_{j=1}^{G_s} |c_{sj}(A) - \text{Round}(r_{sj} \cdot m_{sj}(A))| \quad \dots (1)$$

A: 復元データ

s: 統計データの種類

G_s : 統計データの項目数

c_{sj} : 統計データ s の条件 X_{sj} と Y_{sj} を満たす、復元データの人・組の数

r_{sj} : 統計データ s の項目 j の正しい割合の値

m_{sj} : 統計データ s の条件 X_{sj} を満たす、復元データの人・組の数

式(1)の目的関数では $r_{sj} \cdot m_{sj}$ を四捨五入で整数値に丸めることで、小数点以下の誤差を考慮する必要がない。また、重み付けを行わないため、新たに統計データを追加する際に重み付けの値を再検証する必要がないといった特徴がある。式(1)で求められる目的関数値は統計データとの誤差を表すため、この値が0に近いほど復元データが統計データの割合に一致していることになり、復元データの市民の年齢を変更し、目的関数値を変化させ、統計データの割合に近い復元データを生成することが目標となる。本研究では、式(1)の目的関数を基に統計データとの誤差の最適化を行う。

2.1. 企業仮想データ概要

企業仮想データは産業分類、町丁・大字、従業員数を示す属性で構成されている。テーブルはcompany, industry_type, town, district, cityがある。companyテーブルは、企業を一意的な値で示すcompany_idと町丁・大字単位の立地を示しシミュレーション上の建物に割り振るためのtown_code、産業分類を示すindustry_type_id、従業員数を示すemployeeで構成される。

企業仮想データの構成を表すE-R図をFig. 3に示す。companyテーブルは初期生成により、town_code, industry_type_idが割り振られる。employeeは従業員割り振りで付与され、SA法による最適化の対象となる。townは外部キーとしてdistrict_codeを持ち、districtは外部キーとしてcity_codeを持つことで、townとdistrict, cityを関連付けている。

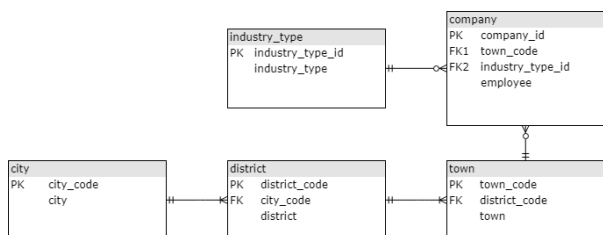


Fig. 7: 企業仮想データの E-R 図.

各テーブルをTable 1～Table 5で示す。各テーブルの項目数はcompany:37528, industry_type:17, town:547, district:51, city:7となっている。

Table 1: company テーブル.

company_id	town_code	industry_type_id	employee
0	22131001000	2	10
1	22131001000	2	5
2	22131001000	2	2
3	22131001000	2	1
4	22131001000	3	12
⋮	⋮	⋮	⋮
37523	22137077000	14	1
37524	22137077000	14	0
37525	22137077000	14	305
37526	22137077000	16	3
37527	22137077000	16	2

Table 2: industry_type テーブル.

industry_type_id	industry_type
0	農林漁業
1	鉱業、採石業、砂利採取業
2	建設業
3	製造業
4	電気・ガス・熱供給・水道業
5	情報通信業
6	運輸業、郵便業
7	卸売業、小売業
8	金融業、保険業
9	不動産業、物品賃貸業
10	学術研究、専門・技術サービス業
11	宿泊業、飲食サービス業
12	生活関連サービス業、娯楽業
13	教育、学習支援業
14	医療、福祉
15	複合サービス事業
16	サービス業

Table 3: town テーブル.

town_code	district_code	town
22131001000	12	相生町
22131002001	10	葵西一丁目
22131002002	10	葵西二丁目
22131002003	10	葵西三丁目
22131002004	10	葵西四丁目
⋮	⋮	⋮
22137073000	47	横川
22137074000	47	横山町
22137075000	47	両島
22137076000	47	緑恵台
22137077000	47	渡ヶ島

Table 4: district テーブル.

district_code	city_code	district
1	22131	中央
2	22131	西
3	22131	城北
4	22131	北
5	22131	東
⋮	⋮	⋮
47	22137	天竜
48	22137	春野
49	22137	佐久間
50	22137	水窪
51	22137	龍山

Table 5: city テーブル.

city_code	city
22131	中区
22132	東区
22133	西区
22134	南区
22135	北区
22136	浜北区
22137	天竜区

2.2. 使用する統計データ

本研究における企業仮想データの生成には、浜松市のホームページで公開されている平成26年浜松市事業所統計¹²⁾を用いる。本研究では立地する地域と産業分類、企業規模を対象とし、適合させる統計データを決定した。以下が本研究で使用する統計データである。

- (A) 産業分類, 町丁・大字別事業所数
- (B) 産業分類, 町丁・大字別従業者数
- (C) 行政区, 産業分類, 企業規模別事業所数
- (D) 行政区, 産業分類, 企業規模別従業者数
- (E) 産業分類, 企業規模(100人以上)別事業所数
- (F) 産業分類, 企業規模(100人以上)別従業者数

各統計データにおいて, 産業分類は統計法に基づいて定められた日本標準産業分類における産業大分類を用いている. 統計データ(A), (B)の町丁・大字は浜松市を547の区画に区分けしている. 統計データ(C), (D), (E), (F)の企業規模はそれぞれ1~4人, 5~9人, 10~19人... 300人以上といった区切りで示されている. 各統計データをTable 6~Table 11に示す.

Table 6: 産業分類, 町丁・大字別事業所数(A).

index	town_code	industry_type_id	number_of_companies
0	22131001000	0	0
1	22131001000	1	0
2	22131001000	2	4
3	22131001000	3	4
4	22131001000	4	0
⋮	⋮	⋮	⋮
9294	22137077000	12	1
9295	22137077000	13	1
9296	22137077000	14	17
9297	22137077000	15	0
9298	22137077000	16	2

Table 7: 産業分類, 町丁・大字別従業者数(B).

index	town_code	industry_type_id	number_of_employees
0	22131001000	0	0
1	22131001000	1	0
2	22131001000	2	18
3	22131001000	3	40
4	22131001000	4	0
⋮	⋮	⋮	⋮
9294	22137077000	12	104
9295	22137077000	13	61
9296	22137077000	14	544
9297	22137077000	15	0
9298	22137077000	16	5

Table 8: 行政区, 産業分類, 企業規模別事業所数(C).

index	city_code	industry_type_id	company_size	number_of_companies
0	22131	0	1~4人	4
1	22131	0	5~9人	2
2	22131	0	10~19人	2
3	22131	0	20~29人	1
4	22131	0	30~49人	0
⋮	⋮	⋮	⋮	⋮
947	22137	16	20~29人	1
948	22137	16	30~49人	0
949	22137	16	50~99人	2
950	22137	16	100人以上	0
951	22137	16	出向・派遣のみ	0

Table 9: 行政区, 産業分類, 企業規模別従業者数(D).

index	city_code	industry_type_id	company_size	number_of_employees
0	22131	0	1~4人	7
1	22131	0	5~9人	16
2	22131	0	10~19人	21
3	22131	0	20~29人	23
4	22131	0	30~49人	0
⋮	⋮	⋮	⋮	⋮
947	22137	16	20~29人	26
948	22137	16	30~49人	0
949	22137	16	50~99人	145
950	22137	16	100人以上	0
951	22137	16	出向・派遣のみ	0

Table 10: 産業分類, 企業規模(100人以上)別事業所数(E).

index	industry_type_id	company_size	number_of_companies
0	0	99人以下	177
1	0	100~199人	0
2	0	200~299人	0
3	0	300人以上	0
4	1	99人以下	11
⋮	⋮	⋮	⋮
63	15	300人以上	2
64	16	99人以下	2096
65	16	100~199人	33
66	16	200~299人	8
67	16	300人以上	12

Table 11: 産業分類, 企業規模(100人以上)別従業者数(F).

index	industry_type_id	company_size	number_of_employees
0	0	99人以下	2130
1	0	100~199人	0
2	0	200~299人	0
3	0	300人以上	0
4	1	99人以下	52
⋮	⋮	⋮	⋮
63	15	300人以上	1093
64	16	99人以下	16452
65	16	100~199人	4574
66	16	200~299人	2034
67	16	300人以上	5818

Table 6, 7では, 産業分類, 町丁・大字別事業所・従業者数を示している. Table 6を用いることで, 初期生成において, 産業分類, 町丁・大字の属性を持った企業を統計データ(A)との誤差0で生成することができる. それぞれの企業に属する従業者の割り振りはTable 6, 7を用いて行う.

Table 8, 9では, 行政区, 産業分類, 企業規模別事業所・従業者数を示している. 統計データ(C)から算出したTable 12に示す行政区, 産業分類, 企業規模別事業所割合を用いることで, 初期生成で割り振られた従業者数を最適化する. Table 8, 9, 12の「出向・派遣のみ」という項目は, 別経営となる他の事業所から出向・派遣された従業者のみで経済活動している事業所を示しており, これらの事業所は従業者数0人として計算される.

Table 12: 行政区, 産業分類, 企業規模別事業所割合.

行政区	産業分類	企業規模							出向・派遣のみ
		1~4人	5~9人	10~19人	20~29人	30~49人	50~99人	100人以上	
中区	農林漁業	0.400	0.200	0.200	0.100	-	-	-	0.100
中区	鉱業、採石業、砂利採取業	-	-	-	-	-	-	-	-
中区	建設業	0.582	0.241	0.110	0.030	0.017	0.014	0.007	-
中区	製造業	0.526	0.212	0.134	0.047	0.042	0.022	0.017	0.001
中区	電気・ガス・熱供給・水道業	0.333	0.500	-	-	-	-	0.167	-
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
天竜区	生活関連サービス業、娯楽業	0.862	0.083	0.041	0.007	-	-	0.007	-
天竜区	教育、学習支援業	0.854	0.073	0.049	-	0.024	-	-	-
天竜区	医療、福祉	0.333	0.317	0.158	0.058	0.058	0.055	0.025	-
天竜区	複合サービス事業	0.516	0.323	0.065	0.032	-	0.032	0.032	-
天竜区	その他サービス業	0.778	0.130	0.065	0.009	-	0.019	-	-

Table 10, 11では, 産業分類, 企業規模(100人以上)別事業所・従業者数を示している. 統計データ(E)から算出したTable 13に示す産業分類, 企業規模(100人以上)別事業所割合を用いることで, 従業者数が100人以上となる企業の従業者数を調整する.

Table 13: 産業分類, 企業規模(100人以上)別事業所割合.

産業分類	企業規模			
	99人以下	100~199人	200~299人	300人以上
農林漁業	1.000	0.000	0.000	0.000
鉱業、採石業、砂利採取業	1.000	0.000	0.000	0.000
建設業	0.998	0.002	0.001	0.000
製造業	0.972	0.017	0.005	0.006
電気、ガス、熱供給・水道業	0.926	0.037	0.037	0.000
情報通信業	0.990	0.003	0.008	0.000
運輸業、郵便業	0.942	0.043	0.007	0.007
卸売業、小売業	0.995	0.004	0.001	0.000
金融業、保険業	0.986	0.013	0.002	0.000
不動産業、物品賃貸業	1.000	0.000	0.000	0.000
学術研究、専門・技術サービス業	0.996	0.004	0.000	0.001
宿泊業、飲食サービス業	0.996	0.002	0.001	0.001
生活関連サービス業、娯楽業	0.995	0.004	0.001	0.000
教育、学習支援業	0.993	0.005	0.000	0.002
医療、福祉	0.976	0.016	0.003	0.005
複合サービス事業	0.964	0.014	0.014	0.009
その他サービス業	0.975	0.016	0.004	0.006

2.3. 目的関数

本研究での目的関数は, Murata, Harada⁹⁾¹⁰⁾ がSA法を行う際に用いているものである. 式(2)が本研究における目的関数である.

$$f_s(A) = \sum_{j=1}^{G_s} |c_{sj}(A) - R_{sj}| \quad \dots (2)$$

R_{sj} : 統計データ s の条件 X_{sj} と Y_{sj} を満たす, 統計データの人・組の数

本研究で行う最適化は事業所の従業員数を対象としたものとなっているため, 従業員数を変更するとその都度, 目的関数値が変化する. 仮想データ全体の評価値は, 各項目の目的関数値を合計したものとなる. 式(2)の目的関数は仮想データ全体評価, 最適化の際に用い, 得られた目的関数値を SA 法により最適化する.

2.4. 仮想データ生成手法

企業仮想データの生成は, 初期生成, 仮想データ全体評価, 最適化の手順で行う. 全体のフローチャートをFig. 4に示す.

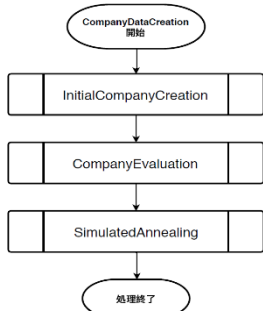


Fig. 8: 企業仮想データ生成フローチャート.

2.4.1. 初期生成

初期生成(InitialCompanyCreation)は, 企業生成, 従業員割り振りの順で実行され, 統計データ(A), (B)(Table 6, 7)を用いて行う. 企業生成では, Table 6に従い, 浜松市に立地する企業の数(37528)だけ産業分類, 町丁・大字の属性を持つ企業を生成するため, 統計データ(A)(Table 6)に関して誤差のない仮想データにすることができる. 従業員割り振りではTable 6, 7を用いる. 具体的な方法としては, 同じ町丁・大字の同じ産業分類間の従業員数を事業所数で割って, 商と余りを求め, その商をそれぞれの企業に割り振っていき, 最後尾の企業に余りを加えるようにし

て行う. これにより統計データ(B)(Table 7)と総従業員数において誤差なく割り振ることができる. 初期生成の概要を表すフローチャートをFig. 9に示す.

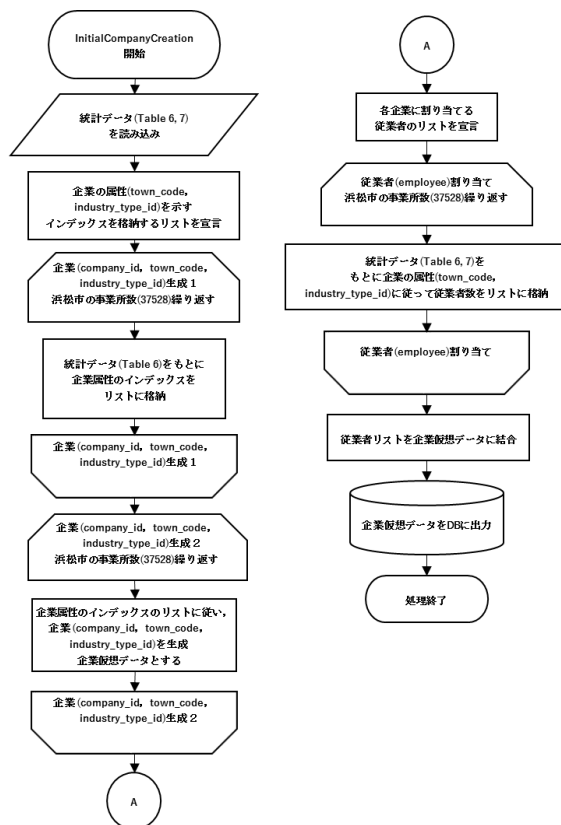


Fig. 9: 初期生成フローチャート.

2.4.2. 仮想データ全体評価

仮想データ全体評価(CompanyEvaluation)は, Table 7, 8, 9, 10, 11の事業所・従業員数の列を目標値として扱う. 仮想データのそれぞれの項目における事業所数, 従業員数を算出し, 評価値リストを作成, 目標値となる統計データの列を抽出しリストの形で切り取った目標値リストと比較することで全体の評価値を算出する. 仮想データ全体評価の概要を示すフローチャートをFig. 10, 11に示す.

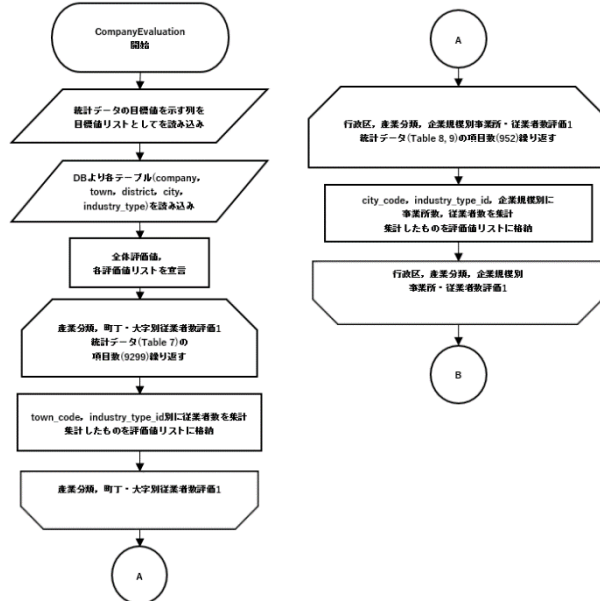
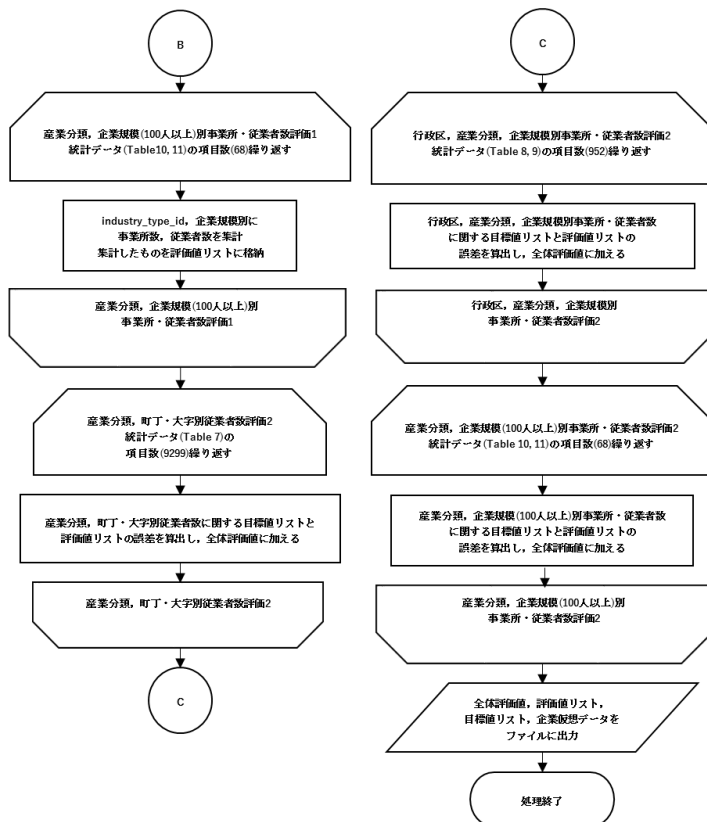


Fig. 10: 仮想データ全体評価フローチャート 1.



2.4.3. 最適化

最適化(SimulatedAnnealing)は、目的関数により算出された評価値をSA法で最小化するようにして行う。本研究で用いるSA法アルゴリズムの概要を以下に示す。

- (1) 2.4.1 で述べた手法での仮想データの初期生成
- (2) 規定試行回数を満たしていれば探索を終了
- (3) 仮想データ内から一つの企業をランダム選択
- (4) 企業から移動もしくは受け入れる従業者数を決定
- (5) 受け入れもしくは移動する二つ目の企業を対象企業群からランダム選択
- (6) 従業者移動判定
- (7) (2)に戻る

(4) では、Table 12, 13を用い、従業者数を決定する。(3) で決定された企業の産業分類、行政区を特定し、それぞれの組合せにおける企業規模の割合を抽出し、その割合に従い、遷移する企業規模を選択する。選択された企業規模となる従業者数になるように新たな従業者数を決定することで、移動・受け入れる従業者数を算出する。

(5) では、産業分類、町丁・大字が一つの企業と同一であり、移動に十分な従業者数を持つ企業群を抽出し、その中から一つの企業をランダム選択することで決定する。産業分類、町丁・大字が同じ企業間で従業者の移動を行うことで誤差0で初期生成した仮想データのTable 7との誤差を0で保ったまま他の統計データとの誤差を最適化することができる。

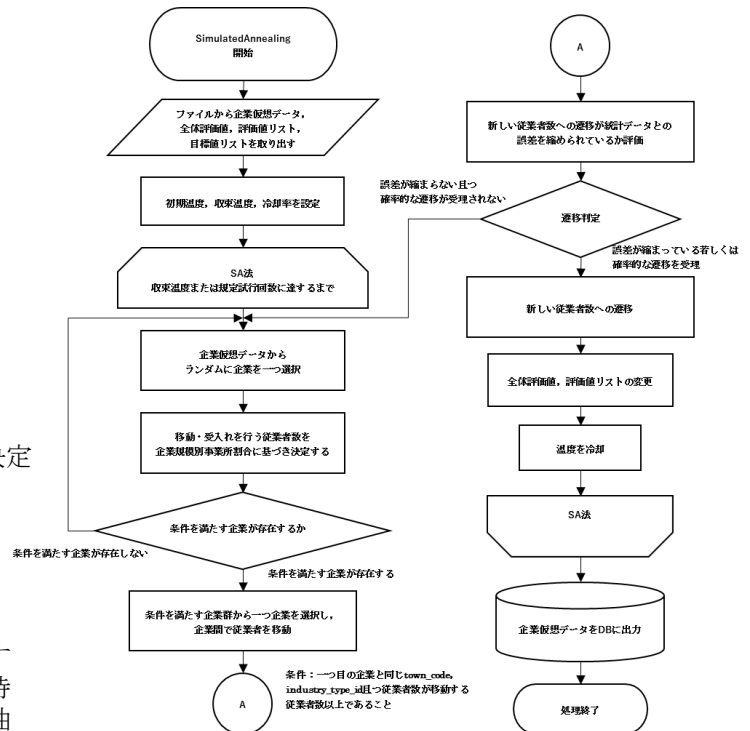
(6) では、移動の対象となった企業だけを評価し、現在の全体の評価値との差分として評価値を更新する。これにより、試行ごとに仮想データを全体評価することで生じる無駄が削減されるため、一回の解探

索にかかる時間を少なくすることができ、大きな探索回数を設定することが可能となる。判定は式(3)で示すメトロポリス基準に従い、改良する解はすべて受理、改悪する解は確率的に受理する。改悪解を受理する確率は温度が下がるにつれ低くなり、また現在の評価値との差が大きいほど低くなる。

$$Prob(accept) \begin{cases} 1 & \text{if } \Delta E \leq 0 \\ \exp\left(-\frac{\Delta E}{T}\right) & \text{otherwise} \end{cases} \dots (3)$$

$$\Delta E = (E' - E)$$

式(3)では、 $\text{Prob}(\text{accept})$ は、従業者数の遷移を受理する確率、 E は遷移前の誤差、 E' は遷移後の誤差、 T は温度パラメータを示している。 $\Delta E < 0$ の場合、新たな従業者数への遷移をすべて受理する。その他の場合でも、誤差の変化量と温度パラメータに応じて、改悪解を確率的に受理する。温度パラメータは初期温度として高い温度を設定し、試行ごとに一定の比率で温度を減少させる。最適化の概要を表すフローチャートをFig. 12に示す。



2.5. パラメータチューニング

本研究において調整を行うパラメータは、T(初期温度)とcool(冷却率)である。冷却スケジュールは指数冷却を用いる。本研究で行うSA法の停止条件は、規定回数に達した場合か、受理がほとんど起こらなくなった場合を採用するため、t(収束温度)は考慮しないものとする。パラメータチューニングを行った結果をTable 14, 15に示す。それぞれのパラメータにおける実行回数は10とする。停止条件は試行回数が5億回に達した場合か、受理が100万回連続で起こらなくなった場合とする。冷却率0.999999999以上は誤差が収束するまでの試行回数が5億回を超えるため、今回のパラメータ調整では対象外とする。

Table 14: 各パラメータにおける最終誤差.

		冷却率						
		0.9	0.99	0.999	0.9999	0.99999	0.999999	0.9999999
初期温度	10 ⁻¹	10020.2	9421.0	9463.0	9801.0	9427.4	8974.6	8371.8
	10 ⁻²	9822.2	9485.0	9047.8	9422.6	8911.4	8600.2	8338.6
	10 ⁻³	9637.4	9669.8	9397.0	9254.2	8760.6	8467.4	8369.4
	10 ⁻⁴	9429.4	9457.0	9200.2	9489.0	9036.6	8521.8	8202.6
	10 ⁻⁵	9244.6	8802.2	9151.8	8791.8	8450.6	8494.6	8199.4

Table 15: 各パラメータにおける収束までの時間[秒].

		冷却率						
		0.9	0.99	0.999	0.9999	0.99999	0.999999	0.9999999
初期温度	10 ⁻¹	2550.7	3681.0	2335.8	3066.2	4065.0	4097.3	10511.8
	10 ⁻²	2562.8	2639.1	3943.0	3122.3	3491.3	4278.8	13333.5
	10 ⁻³	2423.9	2565.3	3566.8	2868.0	3765.7	5332.4	21263.2
	10 ⁻⁴	3152.7	3078.7	2360.6	2543.8	4190.8	5639.7	25927.4
	10 ⁻⁵	2333.4	3265.7	4734.5	3670.1	4258.0	7472.5	31618.0

以上の結果から、初期温度と冷却率を高く設定するほど、誤差を縮小できることが分かる。最良結果は初期温度、冷却率ともに最大の設定によるものとなっており、誤差が最大となった初期温度10⁻¹、冷却率0.9の設定に比べ、約2000の差がついた。また、最良結果のパラメータでの実行時間が一番時間がかかる結果となり、2番目に良い結果となる初期温度10⁻⁴、冷却率0.9999999に比べ、約6000秒の差がついた。二つのパラメータの間には、収束時間に大きな差があるのに対して、最終誤差における差はほとんどなかった。仮想データ生成の効率を考えると、できるだけ短い時間で収束するパラメータを設定する必要がある。以上を踏まえ、本研究におけるパラメータ設定は初期温度10⁻⁴、冷却率0.9999999とし、仮想データの最適化を行う。また、本研究の実行環境は以下の通りである。

- ・プロセッサ：3.5GHz Intel Core i5
- ・メモリ：8GB
- ・OS：macOS 10.14
- ・開発環境：PyCharm Version 2018.2.4

3. 結果と考察

3.1 仮想データの生成結果

2.4 で示した仮想データ生成手法を用いて、浜松市の企業仮想データの生成を行った。初期温度10⁻⁴、冷却率0.9999999の設定で、3億回の試行を行ったところ初期誤差473543から最終誤差8183と約98.3%の誤差の縮小に成功した。SA法の実行結果をFig. 13、各種統計データとの初期誤差及び最終誤差をTable 16に示す。

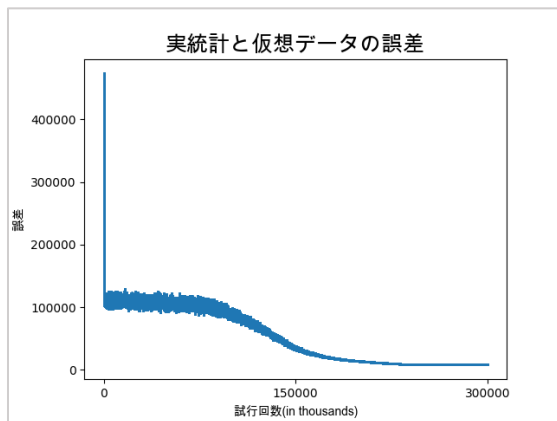


Fig. 13: SA 法実行結果(3 億回).

Table 16: 各統計データとの初期誤差・最終誤差.

統計データ	初期誤差	最終誤差
統計データ(1)	0	0
統計データ(2)	0	0
統計データ(3)	25177	519
統計データ(4)	297466	7162
統計データ(5)	570	42
統計データ(6)	150330	460
全体	473543	8183

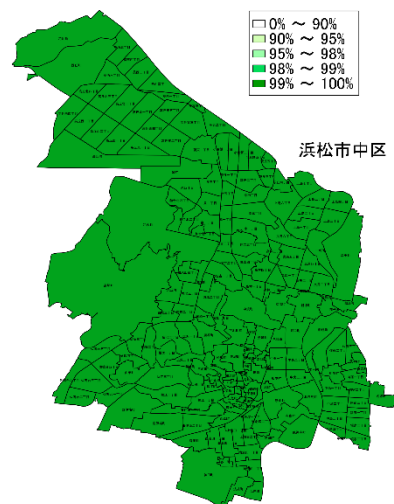
Fig. 13 からはSA法実行直後、急激に誤差が縮小し、その後、誤差の拡大・縮小を繰り返しながら、誤差を収束させていることが分かる。Table 16 からは同じ産業分類、町丁・大字間での従業員の移動により統計データ(B)との誤差が0のまま他の統計データとの誤差を縮小できていることが分かる。

各統計データと仮想データの最適化前後における適合率をTable 17に示す。適合率は仮想データにおいて、真に正しかったものの割合を指し、 $[(\text{統計値} - \text{統計データとの誤差}) / \text{統計値}]$ で求めるものとする。本研究で得られた仮想データは統計データに対して全体で99%以上適合しており、精度の高い仮想データが生成できているといえる。各統計データにおいても高い水準で統計データを復元できている。

Table 17: 各統計データと仮想データの適合率.

統計データ	適合率	
	最適化前	最適化後
統計データ(1)	100.00%	100.00%
統計データ(2)	100.00%	100.00%
統計データ(3)	32.87%	98.62%
統計データ(4)	23.01%	98.15%
統計データ(5)	98.48%	99.89%
統計データ(6)	61.09%	99.88%
全体	62.76%	99.36%

各町丁・大字、行政区における適合率をFig. 14, 15, 16に示す。統計データ(B)に対して全ての町丁・大字で100%適合した仮想データが生成できており、統計データ(C), (D)に関しては、行政区単位でみると人口密度の最も低い天竜区において、比較的適合率が低い結果となった。浜松市全域においては精度の高い仮想データが生成できていることが分かる。

Fig. 14: 統計データ(B)適合率(浜松市中区).¹³⁾

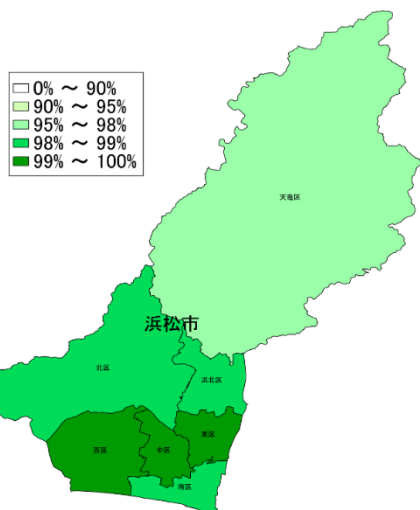


Fig. 15: 統計データ(C)適合率.¹⁴⁾

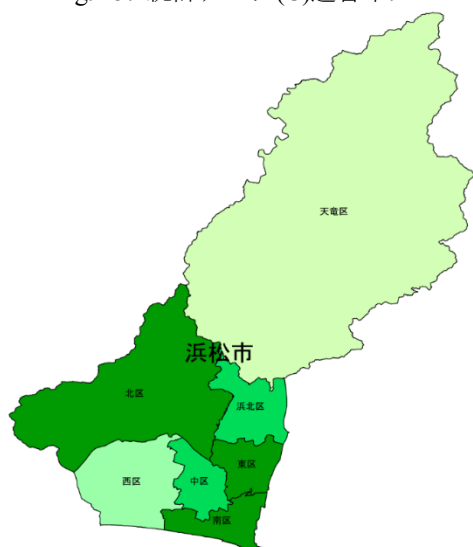


Fig. 16: 統計データ(D)適合率.¹⁴⁾

「国勢調査結果」(総務省統計局)。
「数値地図」(国土地理院)¹³⁾¹⁴⁾を基に作成

3.2 本研究で用いた手法の有効性検証

本研究で用いた仮想データ生成手法における、従業者数の遷移方法と最適化手法の有効性を検証するために、従業者数を変更する手法と従業者を企業間で移動させる手法、局所探索法を用いた最適化手法と SA 法を用いた最適化手法を比較する。局所探索法とは、SA 法と異なり、改悪する解への確率的な遷移を行わず、改善される解のみの遷移を受理し最適解を探索する手法である。それぞれ 1 億回を 10 試行行い、その平均値を示している。パラメータ設定は、初期温度 10^3 、冷却率 0.9999999 とする。

従業者数の遷移方法に、一つの企業の従業者数を単に変更する手法と従業者を企業間で移動させる手法をそれぞれ適用する。最適化手法はどちらにおいても SA 法を用いている。また、従業者数の変更では、統計データ (B) の最適化も行われるため、本研究で行っている統計データ (B) に従い誤差 0 で従業者を割り振る手法 (以下、統計 (B) 割当) に加え、統計 (C)、(E) の割合に従った割当手法 (以下、統計 (C) (E) 割当) も行うこととした。

統計 (C) (E) 割当とは、統計データ (C) (E) の従業者割合 (Table 12, 13) に従い、企業規模を選択し、選択された企業規模になるように従業者数を決定する割当手法である。統計 (C) (E) 割当は初期誤差が平均 442652.2 で従業者を割り当てることができる。統計 (B) 割当は統計データ (B) を初期誤差 0 で適合させることができるものの初期誤差 473543 と前者の割当手法に比べて大きな初期誤差がでる手法となる。各割当手法により生じる初期誤差を Table 18 に示す。

Table 18: 各割当手法における初期誤差.

統計データ	統計(B)割当	統計(C)(E)割当
統計データ(A)	0	0
統計データ(B)	0	293345
統計データ(C)	25177	2426.4
統計データ(D)	297466	97117.6
統計データ(E)	570	126.4
統計データ(F)	150330	49636.8
合計	473543	442652.2

それぞれの手法を適用した結果を Table 19, Fig. 17, 18, 19 に示す。適用結果から従業者を企業間で移動させる手法が最も誤差を縮小できていることが分かる。また、それぞれの手法の適用結果に対して両側の分散が異なることを仮定した t 検定を行ったところ、有意水準 1% で有意な差がみられた。このような結果になった要因としては、統計データ (C)、(D)、(E)、(F) のみの最適化に焦点を当てることができたこと、項目数が多く全体で大きな誤差が発生しやすい統計データ (B) の誤差を 0 に固定できたことが挙げられる。

Table 19: 従業者数遷移手法の比較.

従業者数遷移手法	従業者数の変更				従業員の移動	
	統計(C)(E)割当		統計(B)割当		統計(B)割当	
初期従業者割当	平均値	標準偏差	平均値	標準偏差	平均値	標準偏差
最終誤差						
統計データ(A)	0.0	0.0	0.0	0.0	0.0	0.0
統計データ(B)	13736.9	951.4	13749.1	286.8	0.0	0.0
統計データ(C)	948.8	25.5	911.0	17.3	539.5	27.1
統計データ(D)	3281.7	159.2	2718.8	77.7	7201.6	83.7
統計データ(E)	64.6	6.2	69.4	10.8	43.2	3.0
統計データ(F)	669.7	87.9	666.5	52.5	490.9	35.5
全体	18463.3	1100.7	17935.0	333.2	8275.2	94.7

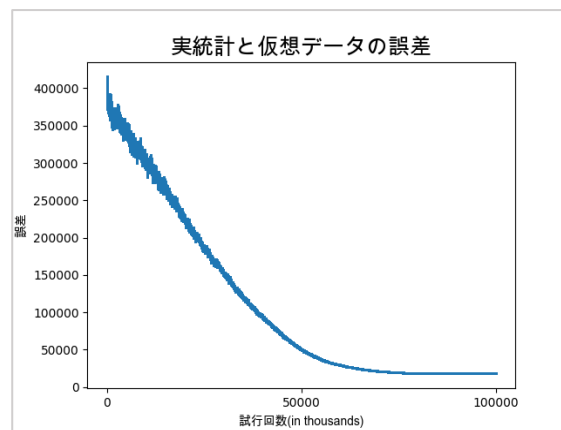


Fig. 17: 従業者数変更・統計(C)(E)割当(1 億回).

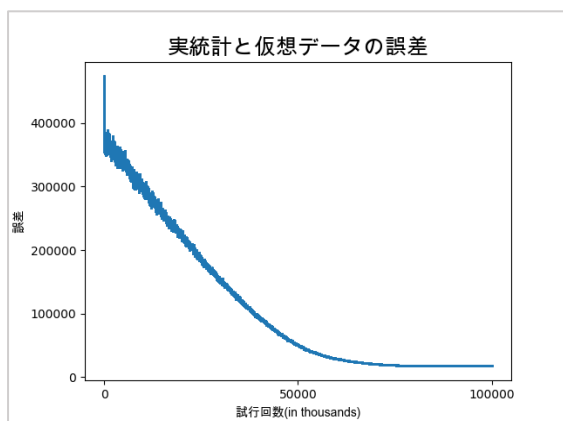


Fig. 18: 従業者数変更・統計(B)割当(1 億回).

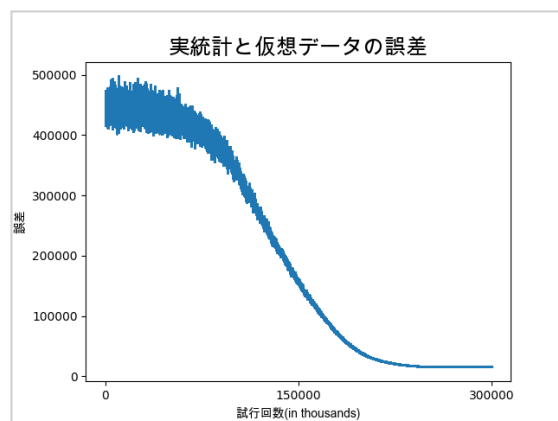


Fig. 20: 局所探索法による最適化(1 億回).

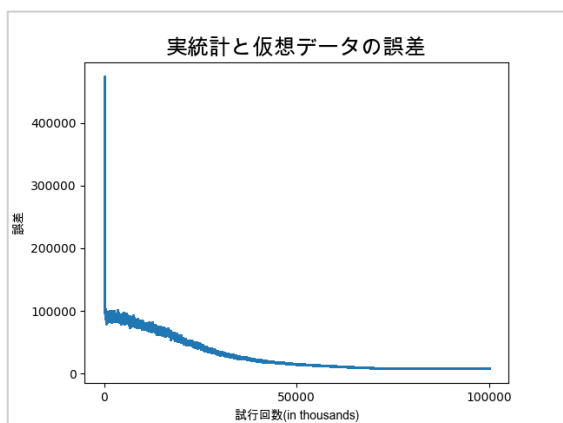


Fig. 19: 従業者移動・統計(B)割当 (1 億回).

それぞれの手法を 3 億回試行した結果を Table 21, Fig.21, 22, 23 に示す. SA 法における各パラメータは初期温度 10^4 , 冷却率 0.999999 とする.

Table 21: 各手法の最終誤差(3 億回).

	従業者数変更		局所探索法	SA法
	統計(3)(5)割当	統計(2)割当		
最終誤差	17149	15570	16633	8183

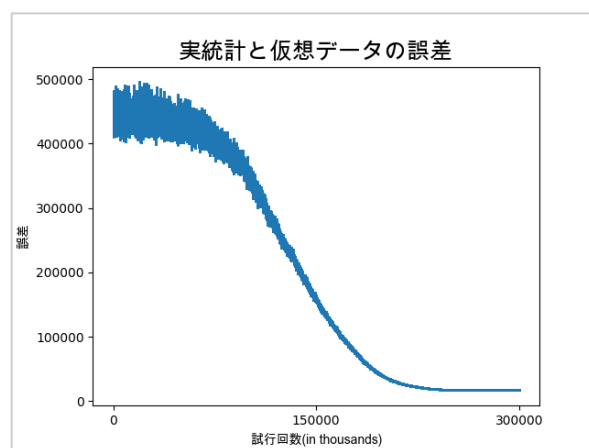


Fig. 21: 従業者数変更・統計(C)(E)割当 (3 億回).

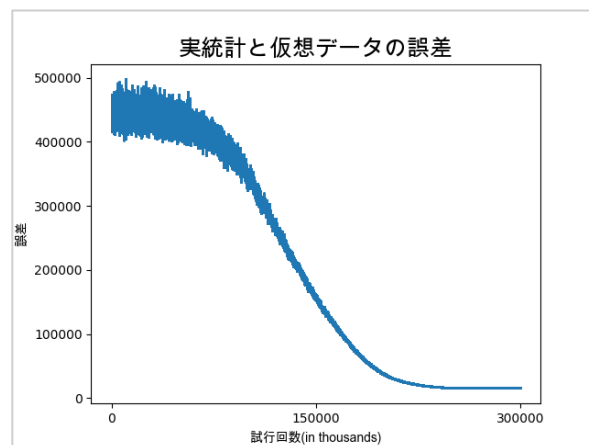


Fig. 22: 従業者数変更・統計(B)割当 (3 億回).

局所探索法と SA 法をそれぞれ適用した結果を Table 20, Fig.20 に示す. 従業者数の遷移方法はどちらの最適化手法においても, 従業者を企業間で移動させる手法を用いている. 適用結果から SA 法は局所探索法より誤差を縮小でき, 効率的な最適解の探索が行われていることが分かる. 各統計データにおいても SA 法の方がより小さな値まで誤差を縮めることができている, 改悪解への確率的な遷移が仮想データの最適化に効果的であるといえる. また, それぞれの手法の適用結果に対して両側の分散が異なることを仮定した t 検定を行ったところ, 有意水準 1% で有意な差がみられた.

Table 20: 最適化手法の比較.

最適化手法	局所探索法		SA法	
	平均値	標準偏差	平均値	標準偏差
統計データ(A)	0.0	0.0	0	0.0
統計データ(B)	0.0	0.0	0	0.0
統計データ(C)	1032.7	44.0	539.5	27.1
統計データ(D)	10830.7	513.0	7201.6	83.7
統計データ(E)	83.8	7.6	43.2	3.0
統計データ(F)	1720.8	277.3	490.9	35.5
全体	13668.0	725.2	8275.2	94.7

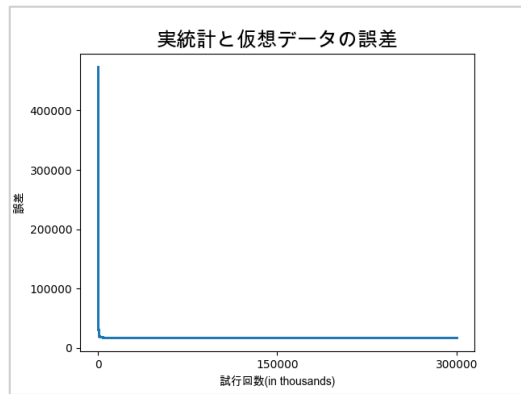


Fig. 23: 局所探索法による最適化 (3 億回).

以上の結果から 3 億回試行した結果においても、本研究で採用する従業者数の遷移方法に企業間の従業者の移動を用いる手法と SA 法による最適化手法を適用した最適化が最も誤差を縮小できているといえる。

4. おわりに

本研究では、今後浜松市へのマイクロシミュレーションを実行するために必要となる企業仮想データの生成を行った。マイクロシミュレーションから正確な結果を得るためには、その前処理として初期状態の個人や世帯、企業などの属性を出来る限り現実 に即した形で再現することが必要であり、本研究においては、統計データとの誤差が最小となる企業仮想データの生成に挑戦した。

企業仮想データ生成の結果は、初期生成された仮想データの誤差の約 98% を解消し、統計データとの適合率 99% 以上の仮想データが生成できた。具体的には、統計データ(A)(B)には 100%，統計データ(C)(D)には 98%，統計データ(E)(F)には 99% 以上適合した仮想データが生成され、極めて現実に近い仮想データが生成できたといえる。

また、本研究で行った生成手法を実行するにあたり、他の生成手法との比較を行い、その有効性を検証した。有効性を確認したのは、従業者遷移の方法と最適化手法選択の部分である。

従業者遷移は、本研究では企業間の従業者の移動によって行っている。これを企業の従業者数を単に変更する手法と比較した。比較の結果からは、従業者の移動の方がより誤差を縮小でき、総従業者数の誤差がない仮想データを生成できていることが示された。

最適化手法選択においては、その比較対象として、局所探索法を挙げた。検証の結果は、SA 法の方がより誤差が少ない解の探索ができており、SA 法の特徴である改悪解への確率的な遷移が、局所最適解からの脱出に作用し、より統計データに適合した仮想データの生成に貢献できていることが確認できた。これらの有効性の検証を通して、より誤差が最適化できる手法を選択できたのではないかと考える。

課題としては、全体的には統計データとの適合率が高い仮想データを生成できたが、まだ誤差が残っている部分があることが挙げられる。特に天竜区での適合率が低く、浜松市の行政区で最も空き家率の高い天竜区¹⁵⁾における誤差は、今後のシミュレーションに影響を与える可能性があるため、より誤差を解消できる手

法を考えていく必要がある。

今後は、企業仮想データの town_code を基に各建物に割り振り、より詳細な企業の立地を決定する。その上で、本研究の成果物である企業仮想データと Murata, Harada⁹⁾¹⁰⁾による世帯と個人、建物の仮想データを用い、浜松市の人口動態を推定するマイクロシミュレーションを行う。マイクロシミュレーションから得られた推定結果の妥当性を検証し、将来不動産需給推定シミュレーションを実施することを目標とする。

参考文献

- 1) https://www8.cao.go.jp/kourei/whitepaper/w-2013/zenbun/sl_1_1_02.html
- 2) <https://www.stat.go.jp/data/jyutaku/topics/topi861.html>
- 3) <https://www.tama-100.or.jp/cmsfiles/contents/0000000/431/12-13.pdf>
- 4) <http://www.ecool.jp/news/images/ecool-090423-13.pdf>
- 5) <https://www.city.hamamatsu.shizuoka.jp/kikaku/zinkousuikiei/documents/jinkousuikiei.pdf>
- 6) https://www.city.hamamatsu.shizuoka.jp/jutaku/home_tochi/home/juseikatsu/documents/juseikatsukihonkeikakuhonpen.pdf
- 7) P.Waddell : UrbanSim:Modeling Urban Development for Land Use, Transportation and Environmental Planning, journal of the American Planning Association, 68:3, 297/314 (2007)
- 8) 池田, 喜多, 薄田 : 「地域人口動態シミュレーションのためのエージェント推計手法」, 第 43 回システム工学部会研究会資料, 11/14 (2010)
- 9) T. Harada, T. Murata and D. Masui : Comparing Transition Procedures in Modified Simulated-Annealing-Based Synthetic Reconstruction Method without Sample, SICE Journal of Control, Measurement, and System Integration, Vol. 10, No.6, 513/519 (2017)
- 10) T. Murata and T. Harada : 「Projecting Households of Synthetic Population on Building Using Fundamental Geospatial Data」, SICE Journal of Control, Measurement, and System Integration, Vol. 10, No.6, 505/512 (2017)
- 11) 梶井, 村田 : 「統計データからの市民の属性復元のための進化計算と SA による 2 段階最適化」, システム制御情報学会論文誌, Vol. 30, No.6, pp. 216-227 (2017)
- 12) https://www.city.hamamatsu.shizuoka.jp/gyousei/library/cyousakekkasyo/2_jigyousyo-kigyoutoukei/cyousakekka/020_h26_keizaisensasukisochousa.html
- 13) <https://www.e-stat.go.jp>
- 14) <http://nfltp.mlit.go.jp/ksj/index.html>
- 15) https://sumaity.com/town/ranking/shizuoka/vacant_house/