

図 6-2 鉄道沿線別・社会増加率

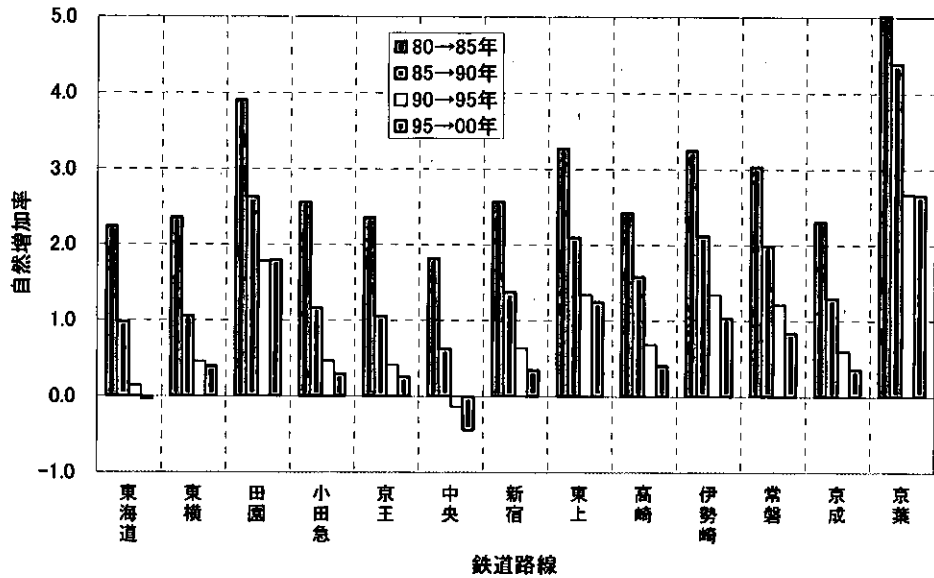


図 6-3 鉄道沿線別・自然増加率

V まとめ

本稿では東京都心から 30 km 圏を対象として、1980 年から 2000 年までの 20 年にわたる期間の人口動態について分析した。その主な結果は以下のようにまとめられる。

- ①都心回帰の主たる役割を担っているのは、20代後半から30代後半にかけてのコーホートである。非大都市圏あるいは郊外から進学のため都心に居住した人々の残留率が高まっていることに加え、非大都市圏から直接都心へ転入するパターンが増加していることが想定される。
- ②距離帯別に人口増加率を社会増加率と自然増加率に分解した結果、社会増加率は距離帯ごとに全く異なる動きを示したが、自然増加率は都心に近いほど低く、しかも期間を経るたびに一律に減少している。
- ③社会増加率と自然増加率を鉄道沿線別にみると、社会増加率については北東側を中心とする沿線と南西側を中心とする沿線の間で明瞭な差があり、北東側は南西側に約5年のタイムラグをもって追隨する動きがみられる。自然増加率については同一方向でも沿線別に差異が表れ、全体として古くから存在する鉄道の沿線ほど増加率が低い傾向がある。

今後の課題としては、20～30歳代を中心とする都心回帰によって都心部の自然増加率は今後どのような動きを示すのか、また地域人口推計には社会増減が決定的に影響するが(濱 1982)、社会増加率と自然増加率のそれぞれの動きから小地域の将来人口推計がどの程度まで可能であるか、などといった点を分析検討することが挙げられる。

現在、東京都区部に続いて大阪・名古屋・福岡・札幌などの各大都市において次々と都心回帰が確認されている。都心回帰の先行的な事例である東京都周辺の人口変動についてさらに分析を加えることによって、各都市の将来的な都心回帰の動向も占えるものと思われる。

本研究は平成14年度厚生科学研究費補助金(政策科学推進研究事業・課題番号 H12・政策・014)および東京大学空間情報科学研究センターの研究用空間データ利用を伴う共同研究(研究番号 26)による成果である。

注

1) 鉄道路線の名称は以下の通り。

東海道:JR 東海道本線, 東横:東急東横線, 田園:東急田園都市線, 小田急:小田急本線, 京王:京王本線, 中央:JR 中央本線, 新宿:西武新宿線, 東上:東武東上線, 高崎:JR 高崎線, 伊勢崎:東武伊勢崎線, 常磐:JR 常磐線, 京成:京成本線, 京葉:JR 京葉線。

参考文献

濱 英彦 1982.『日本人口構造の地域分析』千倉書房, 269p.

分担研究報告－2

東京大都市圏の鉄道沿線における距離帯別人口変化

－小田急線を例に－

江崎雄治

小池司朗

武者忠彦

小口 高

東京大都市圏の鉄道沿線における距離帯別人口変化

—小田急線を例に—

江崎雄治¹・小池司朗²・武者忠彦³・小口 高⁴

I はじめに

近年、大都市圏や地方中核都市において都心部での人口増加が指摘されるようになった。例えば東京都特別区部の1990年代の人口変化をみると、1990～95年においては2.4%の減少だったものが1995～2000年では2.1%の増加に転じた。特に都心3区においてはその変化が著しく、同じ2期間に千代田区は-11.9%から+3.6%に、中央区は-6.1%から+13.5%に、港区は-8.6%から+10.0%へと変化した。長らく転出超過基調にあった東京都心部に人口が戻り始めたことは社会的にも大きな関心を呼び、このような人口変化は「都心回帰」と称されるに至っている。

一方で、都心回帰ほどの注目度ではないものの、それと対をなすところの郊外の側での人口の伸び悩みに関する報告もみられるようになった¹⁾。たとえば木下ほか(2000)は京阪神大都市圏の50 km圏に位置する名張市において、1965年以降に開発された複数の住宅団地で人口が減少していることを報告している。また団地居住者に対するアンケート調査から、子供世代の住み替え希望が親世代よりも多く、通勤通学の不便さを理由に挙げるものが多いことが示されている。このように都心から遠方に位置する名張市では、利便性の欠如などにより今後の人口減少が予想される。また、ここから他の郊外都市において同様の状況が生じることも十分に考えられる。

以上から推測されるのは郊外住宅地の第2世代、つまり若年者を中心に郊外から都心に移り住む人が増加しつつある、という構図である。実際、矢部(2003)は港区における1996～2001年の年齢階級別人口変化から、2001年時点で25～39歳の年齢層が大きく増加していることを示している。また、アンケート調査の結果から神奈川、千葉、埼玉県からの転入者では高齢世帯はごくわずかであるとしている。一方で、都心の高層マンションの購入者の半数近くが50歳代以降で占められることなどから、「子育てを終えた郊外居住熟年層の都心回帰」を結論付けている報告もある(エコノミスト、2001年4月24日号)。したがって、都心回帰の担い手となっている年齢層については、他の地区に関する分析を含むさらに幅広い検証を行う必要があると考えられる。

¹ 専修大学文学部

² 国立社会保障・人口問題研究所

³ 東京大学大学院総合文化研究科博士課程

⁴ 東京大学空間情報科学研究センター

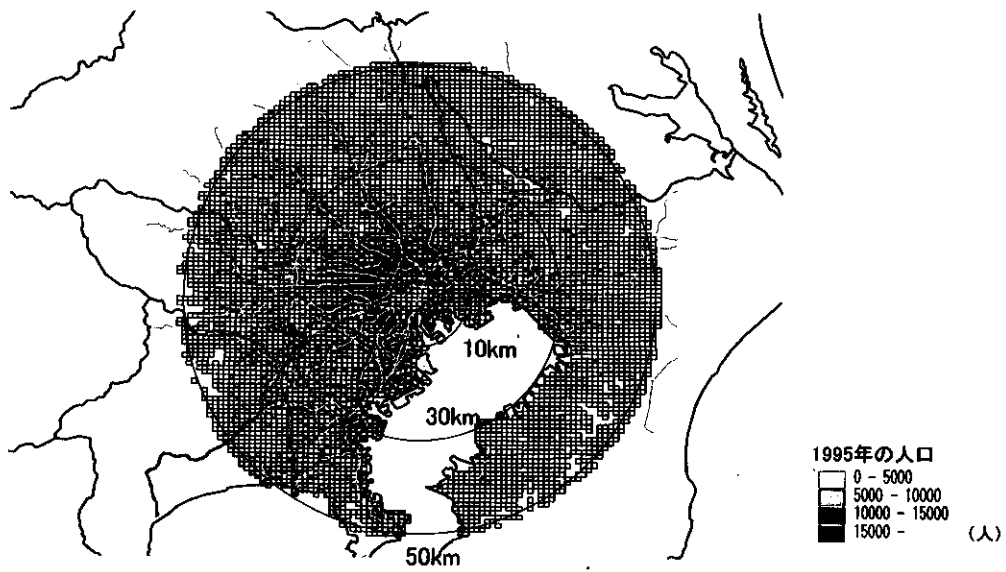
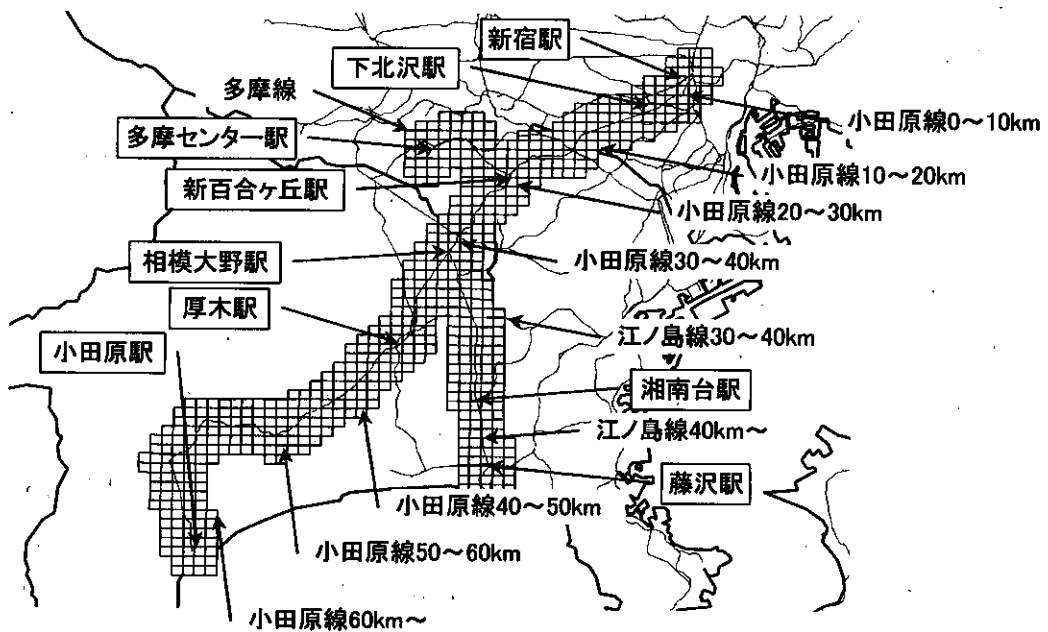


図1 東京大都市圏の鉄道網と小田急線の位置



※旧東京都庁からの距離に応じて区分

図2 小田急沿線地域の定義と距離帯区分

本章ではこれらの点をふまえ、小田急電鉄の沿線地域を事例に距離帯ごとの年齢別人口変化を観察することにより、都心および郊外の人口変化の内実について分析を行う。

II 対象地域と分析方法

本研究では、東京大都市圏のうち小田急電鉄の沿線地域を対象とした分析を行う。図 1 に示すように、小田急電鉄は本線である小田原線が都心からほぼ一直線に南西方面に伸びており、沿線地域の都心からの距離は多様である。

人口データについては、国勢調査の地域メッシュ統計を使用する。使用する年次については 1970～2000 年までの 5 年ごと 7 年次分とし、それぞれの 3 次(標準)地域メッシュを用いる。そして小田急電鉄の小田原、江ノ島、多摩の各線から 2 km 以内に含まれるメッシュを沿線地域と定義し、図 2 のように線区、距離帯ごとに区分した。なお、旧東京都庁からの距離に応じて距離帯を 10 km ごとに区分した。

III 小田急沿線の人口変化

1. 総人口の変化(1970～2000 年)

図 3 に 1970～2000 年における 5 年ごとの総人口の変化を示す。まず都心についてみると、1970～95 年まで人口が一貫して減少基調であり、とりわけバブル期以降の 1985～95 年では大幅な減少を記録していたが、1995～2000 年では一転して一部のメッシュを除いて大幅に増加したことがわかる。バブル経済終焉後の 1995 年が、いかに大きな転換点であったかがこの図からも読み取れる。

全体を見渡すと、沿線全域にわたって人口増加メッシュは減少しつつあるが、小田原線の都心からおおむね 40 km 以内では、近年においても人口増加メッシュが多いのに対して、40 km 以遠では人口減少に転じるメッシュが多くなっている。とりわけ直近の 1995～2000 年では 250 人以上という比較的大きな減少となっているメッシュもあり、人口減少が目立ってきている。すなわち 1995～2000 年においては、小田急沿線地域でも都心回帰現象およびそれと対照的な郊外の人口減少が、同時に発生していたのである。

2. 年齢階級別人口の変化(1995～2000 年)

次に年齢階級別人口をもとに、1995～2000 年におけるコーホート変化を観察した。対象としたのは「20～24 歳→25～29 歳」から「55～59 歳→60～64 歳」までの 8 階級である。未青年期の人口変化に

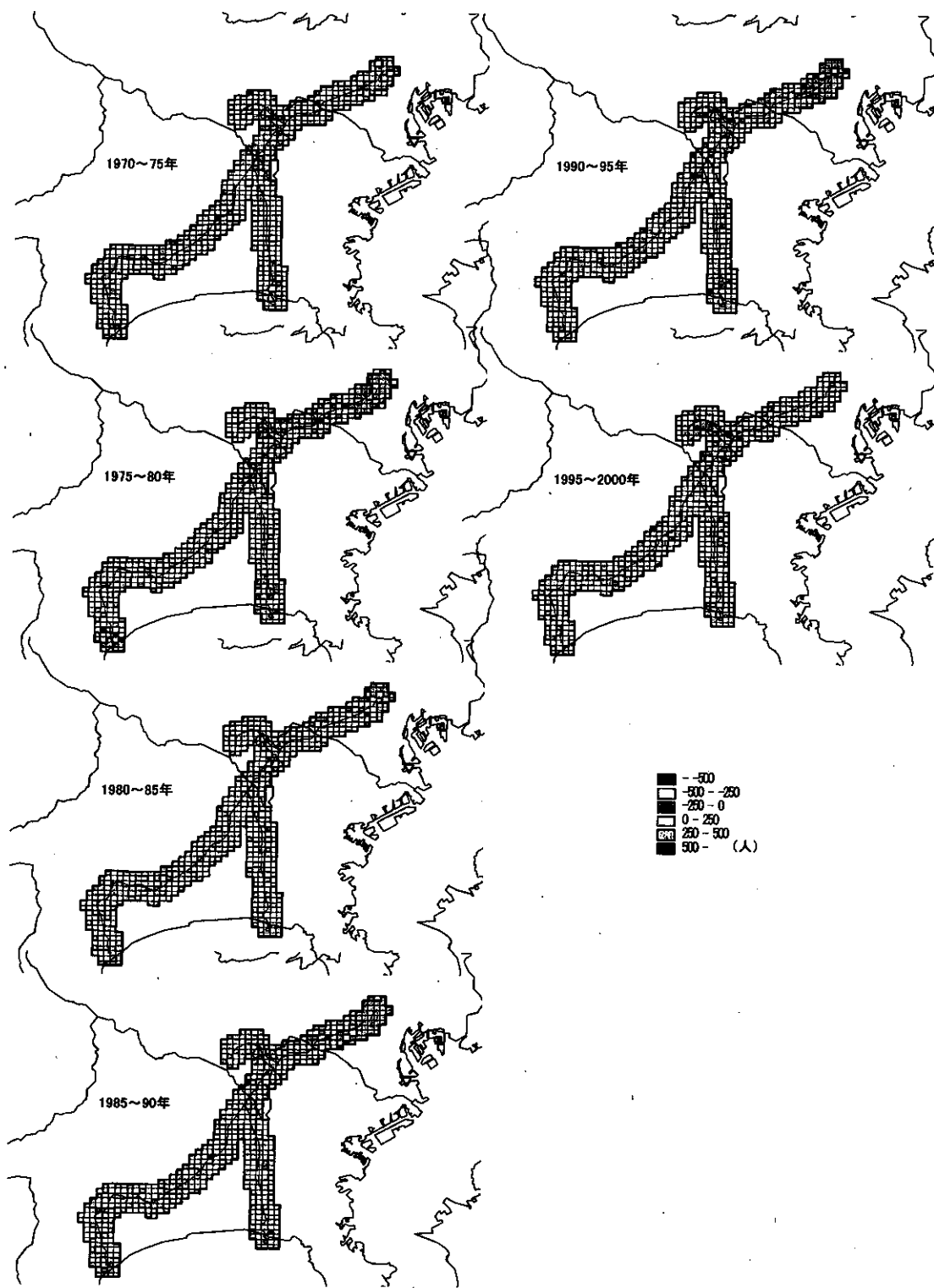


図3 小田急沿線の人口変化(1970~2000年)

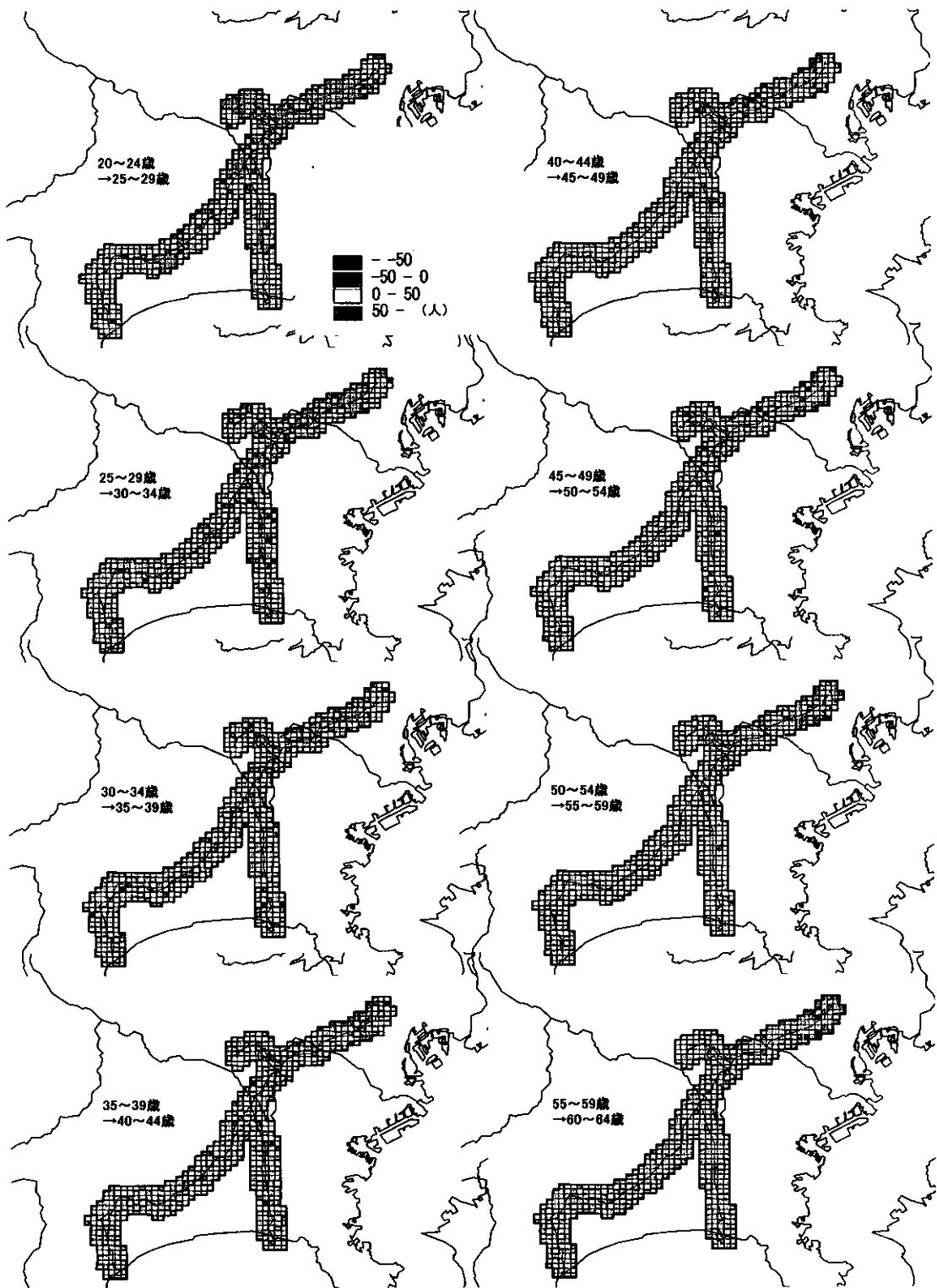


図4 1995~2000年のコーホート変化・男

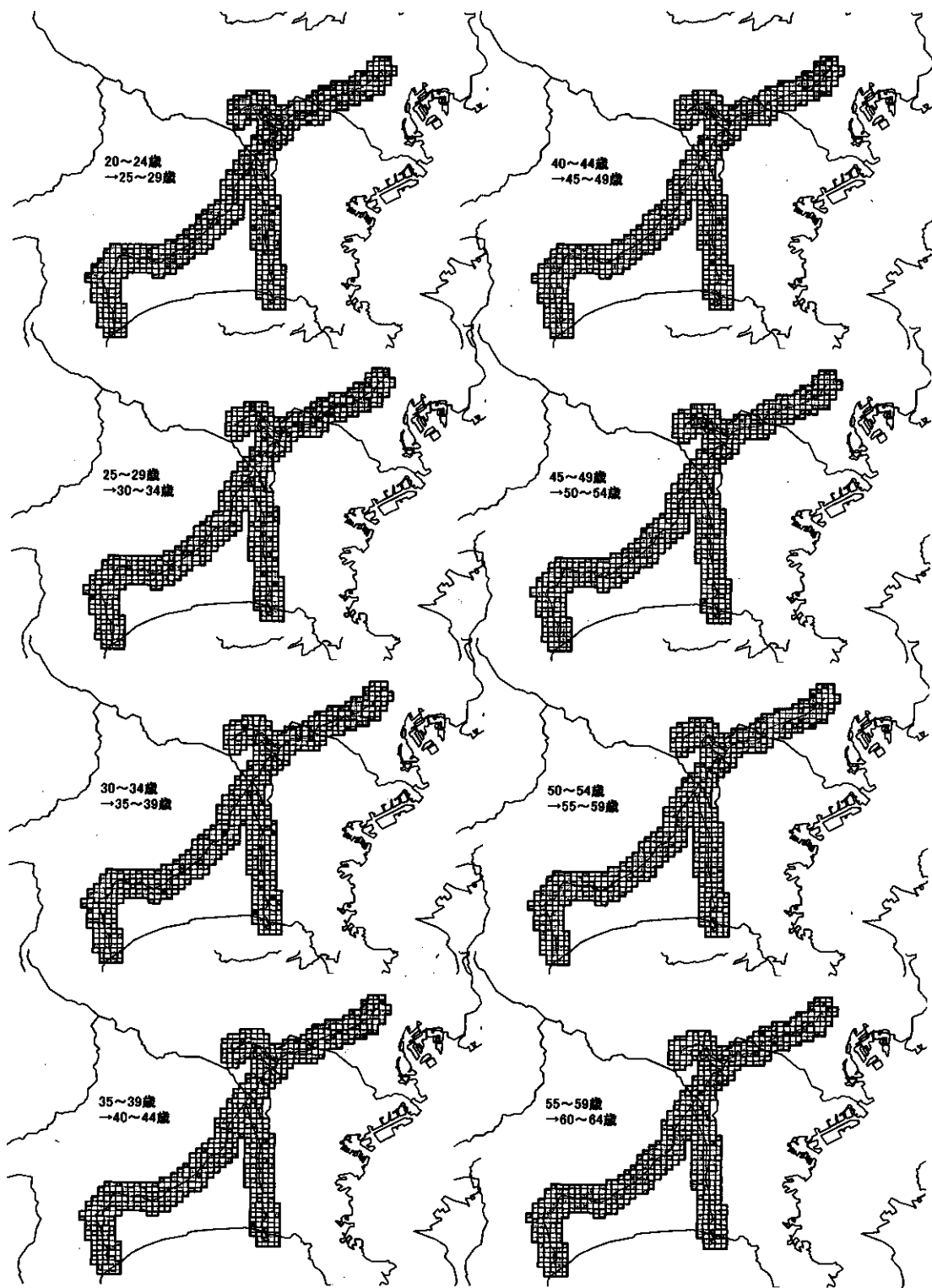


図5 1995~2000年のコーホート変化・女

については、随伴移動によって生じる場合が多く、親世代の変化をみることで代替しうるため、ここでは対象としなかった。また、ここでは移動によって生じる人口変化を把握することが主眼であるため、死亡によるコーホート規模縮小がコーホート変化全体に大きく寄与する高齢者層については分析を見送った。図4、5にそれぞれ男性と女性のコーホート変化を示す。両図より、男女間で異なる状況も一部で観察されるが、大勢としては共通の変化が生じていると判断できるため、ここでは図5の女性のコーホート変化について述べることにする。

まず「20～24歳→25～29歳」についてみると、全体として減少となったメッシュが多いが、これは大都市圏一般の状況として大学周辺の地域などに学卒時の転出者が多く存在するためである。なお全般に増加と減少のめりはりが大きい印象を受けるのは、第2次ベビーブーム世代を含むコーホートであることも影響していよう。

以降、年齢が上昇するにつれて全体としてコーホート変化は落ち着きを見せるようになっていく。ただ、総人口の観察で人口減少メッシュの多かった小田原線の40 km以遠を見てみると、「25～29歳→30～34歳」では減少メッシュが大勢となっており、50人以上の大幅な減少があったメッシュも散見される。「30～34歳→35～39歳」では「25～29歳→30～34歳」に比べると増加メッシュが多いものの、50人以上減少したメッシュも同程度存在するなど、若年層の人口流出を経験した地区が少なくないことがわかる。以降の年齢層ではおおむね増加と減少が均衡している状況となっており、50人以上減少したメッシュはみられない。以上のように、20歳代後半から30歳代前半の若年者が比較的多数流出する一方で、親世代を含めた彼女達よりも上の世代の流出はあまりみられないことから、40 km以遠の一部地域では郊外第2世代の離家によって1995～2000年の期間に人口減少がもたらされたことが推測される。

また0～5 km帯における都心回帰についてみてみると、まず年齢層については若年層で大幅な増加がみられる一方で、中高年ではむしろ減少メッシュのほうが目立っている。これは最も高年齢の「55～59歳→60～64歳」でも同様であり、「定年を機に都心居住を開始」といった転居パターンは、少なくともこれまで大勢であった郊外への転出を上回るほどは存在しないといえる。つまり都心の人口増加については、20、30歳代の転入超過の寄与が大きく、都心回帰現象の中心となっているのは若年者であると結論付けることができる。なお、0～5 kmは若年者を中心に増加基調となっているが、隣接する5～10 km付近では同じ年齢層で大幅な転出超過となっていることに注目する必要がある。これは都心回帰現象がターミナル駅を中心とした限られたエリアで発生していることを示しており、また5～10 km付近が40 km以遠と同様に若年人口の都心への供給地帯となっている可能性を示すものといえる。

IV 小田急沿線の将来人口推計

1. 小田急沿線の人口変化

前章における年齢別人口変化の分析をふまえて、ここではコーホート変化率法による小田急沿線の将来人口推計を行う。まず図2に示した沿線別、距離帯別地域区分に基づき国勢調査メッシュ人口を

合算し、各地域ごとに1995年および2000年の男女、年齢5歳階級別人口を求め、1995～2000年のコーホート変化率を求めた。本推計ではこの1995～2000年のコーホート変化率が2000～2005年、2005～2010年、…、2025～2030年にわたって一定であると仮定し、2005年から2030年まで5年ごとに推計人口を算出した。また出生力、出生性比の仮定については、各地域の2000年における子ども婦人比、および出生性比がそれぞれ2030年まで一定であるとした。

図6に推計結果を2000年までの実績値とともに示した。まず都心回帰傾向が注目される「小田原線0～10 km」では、2005年までは微増傾向となるものの、その後は減少局面となり、しかもその減少のペースが比較的大きいことがわかる。つまり、小田急沿線においても1995～2000年においては確かに都心回帰傾向が観察されるものの、今後転入超過がこの期間に観察された勢いを上回らなければ、都心回帰は早々に失速することを意味している。この理由の一つは、前章で観察したように1995～2000年の都心回帰を担っていたのは20、30歳代の若年者であり、第2次ベビーブーム世代がこの年齢層を通過した後は必然的に都心へ流入する若年層が減少するためである。また都心における出生力の低さも一因であろう²⁾。さらに都心においてはもともと老年人口の層が厚いために³⁾毎年の死亡数が大きいことも無視できない要因である。これらの要因によって、若年者などの流入傾向に今後さらなる加速がみられない限りは、人口の都心回帰は一過性のものに終わる可能性が高いと考えられる。

対照的に「小田原線10～20 km」「小田原線20～30 km」「小田原線30～40 km」では比較的順調な人口の伸びが予測されている。特に新百合ヶ丘を中心とする「小田原線20～30 km」、町田、相模大野を中心とする「小田原線30～40 km」で今後30年間の増加が大きい。これはいうまでもなく、これらの街とその周辺の住宅地が今後しばらくは順調に発展することを示すものである。しかしながらこれ以遠の小田原線沿線ではこのような人口増加は期待できない見通しとなっている。本厚木を中心とした「小田原線40～50 km」はゆるやかな増加基調を維持するものの、さきの分析で若年層の転出超過が目立った「小田原線50～60 km」では人口増加は期待できず、2030年にかけて横ばいからやがては減少へ向かうものと予測されている。

次に支線についてみると、大和を中心とする「江ノ島線30～40 km」においては2030年にかけて横ばいから減少へ向かう予測となっている。この地域では1970年から2000年にかけて人口を倍増させるなど大幅な人口増加を記録しただけに注目に値する。一方で同じ沿線でより遠方に位置する「江ノ島線40 km～」については、当面はゆるやかな増加基調となるものと予測されている。この地域は、中心となっている藤沢から東海道線によっても都心へのアクセスが可能であるため「江ノ島線30～40 km」と比べて利便性が必ずしも低くないこと、また湘南地域の中心部にあたるために住宅地としての人気が高いこと等が理由として考えられる。一方で「多摩線」については計算上は2030年まで一貫した人口増加となっているが、図3～5からもわかるように1995～2000年では大幅な増加と減少を記録したメッシュが混在しており、今後のこれらの勢力関係如何ではこの予測曲線から乖離した動きとなることも十分に予測されるため、同地域の今後については断定的な指摘を行うことは困難といえる。

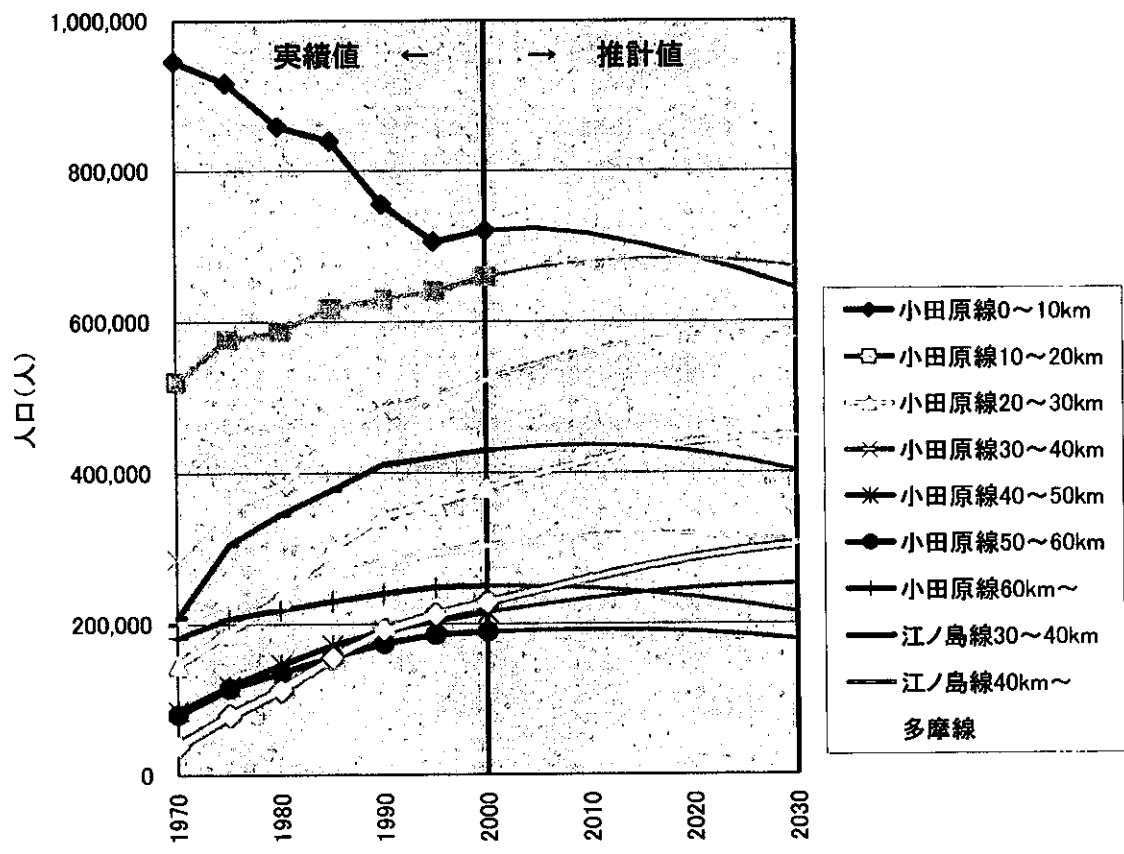


図6 小田急沿線の人口変化(～2000年)と
 コーホート変化率法に基づく推計人口(2000～2030年)

表1 推計人口に基づく老年人口割合

	2000年	2010年	2020年	2030年
小田原線0～10km	16.8	20.1	22.2	23.5
小田原線10～20km	14.4	19.1	21.9	23.4
小田原線20～30km	13.1	19.9	24.2	25.8
小田原線30～40km	12.5	19.9	24.3	25.4
小田原線40～50km	10.5	16.5	21.2	22.1
小田原線50～60km	11.9	19.5	27.6	29.5
小田原線60km～	15.7	22.2	27.8	29.4
江ノ島線30～40km	12.4	20.6	25.6	26.5
江ノ島線40km～	14.9	20.1	23.6	25.3
多摩線	10.4	18.6	26.4	29.5

(%)

※2000年は実績値

2. 小田急沿線の高齢化

前節で述べた人口推計の結果を用いると、年齢別人口の変化についても知ることができる。ここでは沿線の高齢化の進行状況を見通すべく、2030年までの老年人口割合を算出した(表1)。とりわけ注目されるのは、2000年時点では比較的老年人口割合の小さい地域で今後急速な高齢化が見込まれる点である。たとえば「小田原線 50～60 km」においては、2000年時点では11.9%と比較的小さい部類に属していた老年人口割合が、2030年には最も高水準の29.5%にまで急上昇する。既に述べたように「小田原線 50～60 km」では若年層について転出超過となっているメッシュが多い一方、中高年層については目立った変化が見られない。このような年齢別人口変化パターンが続くことを前提とした推計では、必然的に多地域よりも急ピッチで老年人口割合が上昇することとなる。

一方でこれまで相対的に高齢化が進行していた「小田原線 0～10 km」では、1995～2000年にみられたような若年人口の転入超過基調が続いた場合、高齢化の進行は比較的緩やかなものとなる。表1からわかるように、2030年の老年人口割合23.5%は沿線で最も低水準といつてよく、「小田原線 50～60 km」において予測される29.5%とは6%もの差が生じることとなる。

V おわりに

人口の都心回帰現象については、これまでマスコミ等において「郊外で子育てを終えた熟年層が都心の高層マンション等に転居している」といった報告がみられたことから、中高年が都心回帰の主たる担い手であるかのような理解がなされることがあったが、本研究における分析の結果、都心の人口増加を支えているのは20、30歳代の転入超過であることが示され、一方で中高年層の寄与はほとんどみられないことが確かめられた。

また、都心における人口増加が注目される一方で比較的取り上げられることの少なかった郊外の人口変化についても分析したところ、大都市圏周縁部においては都心回帰を担っているのと同じ20、30歳代の減少が目立ち始めており、ここから郊外第2世代の一部が都心へと移動していることが推測される。一方で親世代の年齢層には大きな人口変化がないことが確かめられた。このような「若年層の流出、親世代の居住継続」といった図式が続けば、大都市圏周縁部においては今後急速かつ大幅な人口高齢化の到来が不可避となると考えられる。

国立社会保障・人口問題研究所の推計では、日本の人口は2006年にピークを迎え、以後は長期の人口減少局面に入るとされる。また都道府県ごとの格差が今後拡大するとみられており、地方圏では今後30年で大幅な人口減少となる県もある一方で、東京都や神奈川県については人口維持が予測されている。しかしながらそれぞれの県内部での地域差が存在することはいうまでもなく、本稿で示したように東京都と神奈川県にまたがって路線を有する小田急線の沿線に限ってみても、地域差は大きい。さらにいえば本章で定義した沿線別、距離帯別の各地域の内部においても、たとえば多摩線沿線地域のようにメッシュごとの違いが非常に大きいところもある。このような地域について将来の人口変化を見通

すためには、各メッシュについて年齢別人口変化を綿密に分析し、さらには現地調査等による各地区の転入、転出状況の観察等を通じて人口動態を詳細に把握することが必要となろう。

本研究は平成 14 年度厚生科学研究費補助金(政策科学推進研究事業・課題番号 H12-政策-014)および東京大学空間情報科学研究センターの研究用空間データ利用を伴う共同研究(研究番号 26)による成果である。

注

- 1)ただし郊外から都心へと人口が流入することのみが都心回帰の要因ではないことに留意する必要がある。たとえば江崎ほか(2002)は都道府県単位の分析から、東京都の社会増加については周辺 3 県(埼玉, 千葉, 神奈川県)に対する転入超過数よりも、その他の 43 道府県に対する転入超過数の寄与が大きいことを示している。さらに東京都と周辺 3 県(埼玉, 千葉, 神奈川県)との間の転出入の観察から、東京都から周辺 3 県への転出者数は 2000 年までの間に減少しつつあるが、周辺 3 県からの転入者数はほとんど変化がないことを示している。
- 2)47 都道府県中最低の出生率である東京都の中でも都心付近の出生率はとりわけ低い。1995 年の東京都の合計特殊出生率は 1.11 であるが、同年の新宿区、世田谷区の合計特殊出生率はそれぞれ、0.91, 0.86 であった。
- 3)小田急沿線地域に限ってみても、「小田原線 0~10 km」における 2000 年の老年人口割合は 16.8% と最も大きい。

参考文献

- 江崎雄治・西岡八郎・大場 保・小林信彦(2002):人口移動統計からみた近年の移動性向. 人口学研究, 31, 90-94.
- 木下貴弘・浦山益郎・小川宏樹・神吉順子(2000):大都市圏周縁都市における郊外団地の住み替え構造に関する研究. 都市計画, 49(4), 89-94.
- 矢部直人(2003):1990年代後半の東京都心における人口回帰現象—港区における住民アンケート調査の分析を中心にして—. 人文地理, 55, 277-292.

分担研究報告－3

高速アクセス可能で GIS に応用可能な
AMeDAS 観測年報データベースの作成

大場 保
小池司朗
小口 高
江崎雄治

高速アクセス可能で GIS に応用可能な AMeDAS 観測年報データベースの作成

大場 保¹・小池司朗¹・小口 高²・江崎雄治³

I はじめに

死亡率の研究において、それらがその時々気象条件とどういう関係にあるのかは大いに興味のあるところである。例えば、大場(1994)は、死亡発生頻度時系列を MEM によってフーリエ変換し、そのパワースペクトルを調べた。これは 1 時間ごとの死亡発生頻度時系列のみを分析したものである。そこで得られたスペクトルがどうして得られるのか、その原因として第一に想定された物が気象条件である。個々の死亡とそのときの気象条件との分析が必要と考えられた。

気象条件の一方で、それがどの場所であるかも重要である。というのも、例えば同じ気温であっても北海道と沖縄とでは普段のベースとなる気温が異なっているため、ある特定の温度に対する死亡発生の応答がそもそも違ふと考えられるからである。

このような、それぞれの場所での時系列的な気象条件と例えば毎時の死亡データを対比させて分析するためには、それぞれの場所の座標とセットになった時系列的な気象データが必要であるが、このような GIS 的用途に供することが可能な気象データはこれまでになかった。

そこで、本研究では、GIS 研究に応用可能で、様々な座標点における時別値あるいは日別値の気象データのデータベース開発を行うこととした。また、この成果は、単に死亡発生頻度研究に役立つのみならず、多くの分野で利用可能な貴重な物になるはずである。

さて、GIS 研究に利用可能で時別値の得られる気象データの候補として、気象台および測候所からの物(気象庁年報)と「地域気象観測システム(Automated Meteorological Data Acquisition System; 以下 AMeDAS)」による物が挙げられる。どちらも全国的な観測網があり、気象庁から CD-ROM でのデータ提供もなされているが、それぞれ一長一短がある。

前者は多くの項目のデータが得られ、また、観測期間も 1961 年¹⁾から現在までと長い。観測点が全国で 150 ヶ所とやや少ない。一方、AMeDAS は 1976 年以降²⁾しか公表されているデータはなく、また、湿度や気圧という項目がないが、観測点が全国に約 1300 ヶ所(約 17 km 間隔)もあり、うち約 850 ヶ所(約 21 km 間隔)では、降雪量に加えて風向・風速、気温、日照時間を観測しており、さらに雪の多い地方の約 280 ヶ所では積雪の深さも観測されている。AMeDAS によるデータは、気象台と測候所のデータに比べて距離的な分解能が高いので GIS による分析にはより適していると言えよう。

¹ 国立社会保障・人口問題研究所 人口構造研究部

² 東京大学空間化学情報センター

³ 専修大学文学部

一方で、AMeDAS を始めとした時系列の気象データは、その量が膨大であり、その中から簡単に必要なデータを抽出して研究に用いるのは容易ではない。AMeDAS 観測年報 CD-ROM については、後述のように、2000 年以降ファイル形式が変更され、GIS 的な用途にはそのままでは扱いにくくなっている。

そこで本研究では、まず新形式の AMeDAS データを、多くのデータに対してより早くアクセス可能なデータ形態へと変換するとともに、データアクセスの方法についても、高速で利便性の高いインターフェースを実現することを目的として、独自形式の AMeDAS データベースの開発を開始した。

II 気象庁による AMeDAS 観測年報 CD-ROM について

AMeDAS の観測値は、気象庁の AMeDAS 観測年報の CD-ROM として入手可能である。この CD-ROM は、2003 年 3 月現在、1976 年から 2001 年までの観測データが(財)気象業務支援センターから販売されている。このうち、1976～2000 年まではバイナリ形式で 1 ヶ月ごとにファイルを区切って全観測点のデータが収納されていた。

これが 2000 年以降は形式が変更され、テキストデータの CSV 形式で記録したファイルとなった³⁾。新形式のデータファイルは、「観測点ごと、月ごとに、1 時間ごとの全データをまとめたもの(以後「特別値」ファイルと呼ぶ)」と「観測点ごと、月ごとに、1 日での最高値、最低値などの日ごとデータをまとめたもの(以後「特別値」ファイルと呼ぶ)」の 2 種類となった。これらが、年月フォルダ(1976～2001 年で $26 \times 12 = 312$ 個)とさらにその下のエリアフォルダ(北海道を支庁別になっており、年により 61～63 地域)に分かれて格納されている。

この変更に伴い、2001 年版の CD-ROM には、1976～1999 年のデータを旧バイナリ形式から新テキスト(CSV)形式へと変換するプログラムが添付された。

さらに新形式の 1976～2000 年のデータに対して、旧形式 CD-ROM 発売以降の修正値で更新するプログラムも添付された(これにより、修正されたデータを利用するためには、旧形式を新形式に変換しなくてはならなくなった)。

新形式となったことによって、観測所ごと、1 ヶ月ごとに特別値ファイルと日別値ファイルの 2 つのファイルとなったため、ある一時点における一つの観測点でのデータを検索してテキストエディタや表計算ソフトですぐ見られるという利点が生まれた。

その一方で、1976～2001 年合計で 770,335 個のファイルが 19,902 個のフォルダに分かれて保存されているという状態であるため、ハードディスクに占める容量は 10GB を超え、全部のファイルにアクセスして読むという基本的な操作に 8 時間以上(筆頭著者のパソコン環境の場合)かかってしまうという欠点も生じている。これは、ある時点における全観測点のデータを見たい、というような GIS 的な用途には厳しい状況である。

III 開発目標とするデータベースの仕様

例えば先述の死因発生頻度時系列で用いたデータとの対比という場合には、日本における 12 年間の全死亡の 1/2, 約 400 万件の死亡について、個々の死亡時刻と死亡地(県)における気温が何度であるのか、瞬時に検索可能であることが求められる。

このような条件を考慮して、最終目標とするデータベースの仕様は、つぎのようなものとした。

1. 最初に全ての AMeDAS データをメモリ上に読み込み、専用プログラムで必要なデータだけ高速に抽出できること(理由:全データをメモリ上に読み込むことは、レコードへの高速アクセスには欠かせない。例えば、死亡した県と年月日時が分かっている死亡個票データ 1000 万件について、個々の死亡時における気温を求めたい、などという処理要求に対して、短時間で応答できるようにするためには必須である)。
2. Windows パソコンにメモリ増設(2~4GB)するだけで利用可能となること(理由:現在開発しているデータベースが、今現在は一部のパソコンにメモリ増設して動き、開発が済んで世に広まるころ(3~5 年先)には、そのとき販売されている一般的なスペックにおいても実行可能となることを想定している)。
3. OS は、上記メモリ容量が利用可能な Windows 2000 Professional, Windows 2000 Server, Windows XP Professional, Windows Server 2003 を想定する。
4. 気温に関しては、国内のほとんどの地域における任意の座標における値を、周囲の AMeDAS 観測点からの内分点として重み付けして推定できること(理由:国内の任意の座標における気温が扱えたら便利である。例えば県別の何らかのデータとそのときの気温を比較したいというとき、県の代表値を県庁所在地の座標にするとか地理的重心の座標などが考えられるが、これらの座標における気温が必要となるからである。多くの場合、それらの県を代表する座標が AMeDAS 観測所と重なることはないため、この座標における気温を推定する必要がある。他の観測値についても単純な推定ができれば用いたい。降水量は高度が変わると雲の高さによって事情が変わるし、日照時間も同様である。風向風速も単純な計算ではかなり雑な推定しかできないので、今回はこれらについては推定しないことにする)。
5. 2002 年 4 月の改正測量法施行に準じるため、「日本測地系:ベッセル楕円体」→「世界測地系:GRS80 楕円体」への座標変換に対応すること。
6. DLL 形式などの専用プログラムを用意し、Microsoft Excel からこのプログラムを操作して必要なデータだけ高速に抽出できること。

上記の 1 を実現するためには、データレコードから冗長なビットを削減するための記録構造の大幅な変更が必要になる。というのも、上記 OS における最大ユーザーズメモリ空間は 2GB であるが、テキスト形式での AMeDAS データはその限界を超えている。そのため、テキスト形式のままでは、

アプリケーションプログラムがすべてのデータをメモリ上に読み込むことはできず、データレコードにアクセスするたびにディスクを読みに行かねばならないからである。

データレコードから冗長なビットを除くためには、それぞれの変数について値域を調べて、何ビットあれば表現できるのかを押さえておく必要がある。そのためにはすべてのファイルを読んで、その値域がどこからどこまでなのかを調べなくてはならない。そのため、77 万個に分割されている特別値と日別値の 2 種類のファイルを、テキスト形態のままカンマと改行コードを略して、それぞれ 1 本にまとめてしまうこととした。仮に容量が同じ 10GB であっても、単一のファイルであればハードディスクからハードディスクへコピーするのに数分しかかからないし、もちろん全レコードを読むために必要な時間も数分で済むはずである。

また、こうしてまとめられたファイル自体、元の物よりはるかにコンパクト化されており、MIFES 等の市販のテキストエディタですぐ見られるし、以下に示すフォーマットを参考にすれば、プログラムを作成することで、データベースとしてそのまま利用可能である。

そこで、第一段階として手を付けることとしたのは、分散しているデータファイルの保存状態の変更である。テキスト形式はそのままにし、77 万個のファイルに分散している状態を、特別値と日別値の 2 つのファイルにまとめる。こうして通常の読み書きに要する時間を大幅に減少させ、次の段階として、さらなる冗長ビットの縮減を行うこととした。

そのため、本年度(平成 14 年度)では、まず、すべてのデータファイルを読み込んで、レコードをすべてコンパクト化し、ひとつの大きなデータファイルとすることとした。

IV 新形式ファイルの集約

約 2 万のフォルダに分散した約 77 万のファイルを 2 本に集約するという作業は、手作業ではできない。当然、プログラムを作成することになる。この作業が本稿の結果を導くためになされた労力の半分を占めており、このプログラムについて簡単に記しておく。

このプログラムは、OS のディレクトリ検索ルーチンを呼び出して、一定の順序に従って対象となるすべてのディレクトリを順次検索して移動し、各ディレクトリに存在する AMeDAS の特別値あるいは日別値の CSV ファイルを、そのファイル名から認識し、それらを順次オープンする。

次いで、オープンしたファイルのヘッダー部を読み飛ばし、データレコードの文字列を 1 レコードごとにバッファに読み出す。バッファに読み込んだ文字列を、年月日、観測所番号、各気象データ等の各項目ごとに決まっているバイト数に従って切り出し、カンマを除いたそれぞれの文字列のうち、必要なものだけつなぎ合わせて新しいレコード文字列とし、この新レコード文字列を、特別値であれば特別値用のまとめファイル、日別値であれば日別値用のまとめファイルに書き出す。各ファイルの全レコードについて処理が終わったらそのファイルをクローズし、次の AMeDAS のファイルを検索する、という動作を最後のディレクトリの最後のファイルまで繰り返す物である。

V まとめられた特別値ファイルのフォーマット

特別値ファイルの一つのレコードは、観測点番号、年月日からなるヘッダー部と、午前1時から1時間ごとの降水量、風向風速、気温、日照時間、積雪量からなる1組の特別観測値が、24組連なったデータ部という構造とした。1976～2001年までのすべての特別値レコードを一つのテキストファイル(ファイル名:AMeDAS_Hourly.txt)にまとめた結果、447Bytesのレコードが11,725,243個集まって約5,118MBという巨大なファイルになった。

表1にレコード構造を示す。一つの特別値レコードの最初の13バイトは、レコードを識別するための観測所番号と年月日を記したヘッダー部である。以後、計18バイトの1時間観測値が24個並び、最後に2バイトの改行コードを付け⁴⁾、合わせて447バイトとした。

表1 特別値ファイルAMeDAS_Hourly.txtのレコードフォーマット

ヘッダー部(計13バイト)	
1 - 5	: 観測所番号
6 - 9	: 西暦
10 - 11	: 月
12 - 13	: 日
(14-31桁は、午前1時の観測値。以下23回繰り返し。18バイト×24=432バイト。)	
14 - 16	: 1時間降水量 (1mm. 計測は0.5mm単位で行われ、記録は1mm単位。)
17 - 18	: 観測時10分前から観測時までの風向の平均値 (0: 静穏(観測時前10分間の平均風速が1m/s未満のとき), 1-16: 北北東(=1)から時計回りに北(=16)まで。)
19 - 20	: 風速(1m/s)
21 - 25	: 気温(0.1℃)
26 - 28	: 日照時間(0.1時間)
29 - 31	: 積雪の深さ(1cm)
.	
.	
.	
446 - 447	: 改行コード(2バイト)

VI まとめられた日別値ファイルのフォーマット

日別値ファイルは、観測所番号、年月日を記したヘッダー部と、一日の総降水量、最大の時間あたり降水量などの、特別値から得られたデータのうち、1日についてのデータを集めて記したデータ部とした。1レコードのサイズは72バイトとし、カンマも改行コードもない。ファイルサイズは約850MBのテキストファイル(ファイル名:AMeDAS_Daily.txt)である。