

# 年齢-時代平面上における癌死亡リスクの視覚化

加茂 憲一<sup>1</sup>・富田 哲治<sup>2</sup>・佐藤 健一<sup>2</sup>

(受付 2010 年 12 月 28 日; 改訂 2011 年 7 月 19 日; 採択 7 月 21 日)

## 要 旨

癌の動向に対する適切な統計的手法を用いた評価は、癌対策の根幹をなす重要な情報となる。本論文では、癌死亡リスクを年齢と時代を座標とする空間内の曲面(リスク曲面)として視覚化することを目的とするが、実際には2次元の年齢-時代平面上において癌死亡リスクの高低を表現することになる。このように、ある平面上の関数曲面を推定する方法としては地理的加重回帰、あるいは離散変数にも適用可能な地理的加重一般化線形モデルが広く利用されており、本目的に対しては、年齢と時代の組み合わせを仮想的な位置情報とみなすことで癌死亡リスク曲面の推定が可能となる。このモデルはノンパラメトリックであり自由度が極めて高い。これに対して、別のパラメトリックモデルも考察する。本論文では、年齢と時代に関する交互作用を用いたパラメトリックモデルによる癌死亡リスク曲面の推定も試みる。更に、癌の時系列解析や予測において良く用いられる年齢・時代・コホートモデルに基づく解析結果を用いて癌死亡リスク曲面を推定することも可能であり、その結果も併せて紹介し、各モデルより得られた結果を比較検討する。解析には統計フリーソフトウェア R を用い、解析スクリプトおよび Web より入手可能なデンマーク肺癌死亡データを用いた解析例も併せて紹介する。本研究の目的である「視覚化」について達成されたかどうかに対する絶対的な評価は、現時点では不可能である。しかし、特性が既知であるデータに対して本手法を適用し、実際にその特性が視認できるかという点での検証は可能である。そこで、その検証の一例として「昭和一桁生まれ世代の高リスク」という出生コホート効果の示唆される日本の肝臓癌死亡データに適用し、その効果が視認できるかに着目した結果を示す。最後に、これらのモデルにおける数理的な面に着目し、その特性や問題点をまとめ、今後の改良点や応用について議論する。

キーワード：癌，リスク曲面，視覚化，地理的加重一般化線形モデル，交互作用モデル，年齢・時代・コホートモデル，コホート効果。

## 1. はじめに

ある集団の癌の経年変動を解析し精確に評価することは、今後の癌対策の立案にも繋がる重要なテーマである。本論文では癌死亡や罹患に関するリスクを、年齢と時代を座標とする空間内の曲面(以下、このような曲面を「リスク曲面」と表す)として視覚化する数理モデルの構築およびその解釈を与えることを目的とする。しかしながら、この曲面は3次元空間内のものであるため表現が難しい。そこで、年齢-時代からなる2次元平面上に色の濃淡と等高線を用いて曲面を表現することによる視覚化を試みる。本研究においては、特に癌死亡に着目する。

<sup>1</sup> 札幌医科大学 医療人育成センター：〒060-8556 北海道札幌市中央区南1条西17丁目

<sup>2</sup> 広島大学 原爆放射線医科学研究所：〒734-8551 広島県広島市南区霞1-2-3

このような研究目的に対しては、地理的加重一般化線形モデル (Geographically weighted generalized linear model : 以下「GW モデル」と表す) の適用が期待される。GW モデルは地理情報を有するデータの解析で多用されるが、本研究テーマにおいては年齢と時代の組み合わせを仮想的な位置情報とみなす。そして、カーネル平滑化の要領で、固定した年齢と時代の組み合わせ周辺のデータに対する局所的な重み付きポアソン回帰を繰り返すことにより、癌死亡リスク曲面が推定される。

一方で、ノンパラメトリックな GW モデルに対して、年齢と時代の交互作用項を導入したパラメトリックなモデル (以下「交互作用モデル」と表す) にも着目する。パラメトリックなモデルは自由度が低いが、自由度の高い GW モデルによる結果をどの程度再現可能か、癌死亡リスクの視覚化という点に注目し比較する。

更には、経時的に観測された癌データの解析や予測において広く用いられている年齢・時代・コホートモデル (Age-period-cohort model : 以下「APC モデル」と表す) についても、同様の研究目的に対して得られた結果を紹介する。

本論文は以下のように構成されている。まず第2章においては、本論文で取り扱う3種類のモデル、GW モデル、交互作用モデルおよび APC モデルの理論的な背景を紹介する。また、読者による解析結果の再現を可能にするために、統計フリーソフトウェア R のパッケージ Epi に実装されているデンマークにおける男性の肺癌死亡データ lungDK を用いた解析結果をスク립トと共に付加する。次に、第3章においては、実際に日本における癌死亡データに適用した結果を示す。本論文では、「昭和一桁生まれ世代の高リスク」という出生コホート効果の存在が示唆されている肝臓癌に着目し、その効果が上記モデルにより視覚化できるかという点に注目する。最後に、第4章においては、第2, 3章での解析例の結果をもとに本論文で取り扱った3手法の比較検討を行い、今後の実用上の課題について議論する。

## 2. 年齢・時代平面上における癌死亡リスクの視覚化

本章では、癌死亡リスクを、年齢と時代を座標とする空間内における曲面として視覚化することを目的とした理論について説明する。癌死亡の発生は、全人口に対して稀な事象と考えられるので、人口(あるいは人年)をオフセットとするポアソン回帰モデルを適用する。具体的には、時代  $p$  において年齢  $a$  である人口(あるいは人年)を  $z_{ap}$  人とし、そこから  $y_{ap}$  人の癌死亡が発生したとすると、

$$(2.1) \quad y_{ap} \sim \text{Poisson}(z_{ap}\lambda_{ap})$$

となる。ここで  $\lambda_{ap}$  は時代  $p$  において年齢  $a$  である人の癌死亡リスクを表す未知パラメータである。従って、本研究においては、この未知パラメータ  $\lambda_{ap}$  の年齢・時代に関する特性を視覚化する。

### 2.1 地理的加重一般化線形モデル

通常的回帰モデルにおいては未知パラメータを定数とするが、GW モデルでは、これを位置情報によって変化する関数として扱うのが特徴である (Brunsdon et al., 1998; Fotheringham et al., 2002; 谷村, 2010 など)。地理的に近いパラメータ間には関連性がある一方で、解析対象の全地域でパラメータが一樣でないことも多いためである。

モデルに含まれる未知パラメータは、観測地点間の距離をもとに局所的な重み付けを行い推定される。この重みを定義する関数は距離低減関数と呼ばれ、推定値が周辺地域から受ける影

響を距離が遠くなるに従って減少させるように設定される。例えばガウス型の距離低減関数は、

$$(2.2) \quad W(d) = \sqrt{\exp\left(-\left(\frac{d}{h}\right)^2\right)}$$

を定義される (Brunsdon et al., 1998)。ここで  $d$  は地点間の距離を表し、 $h$  は関数形を決定するパラメータでありバンド幅と呼ばれる。バンド幅は、Stone (1974, 1977) による交差検証法を用いて最適化される。すなわち

$$\sum_{i=1}^n (y_i - \hat{y}_{i[-i]})^2$$

を目的関数として、これを最小にするように求められる。ここで  $y_i$  は地点  $i$  における観測値、 $\hat{y}_{i[-i]}$  は地点  $i$  を取り除いたときの  $y_i$  の推定量 (ジャックナイフ推定量) である。

GW モデルにおける位置情報として 2 次元平面上の座標  $(x, y)$  を考えると、未知パラメータ関数は 3 次元空間内の曲面となる。経年的な癌死亡データは、死亡年別・年齢階級別という形式で構成されることが一般的であるので、GW モデルにおける  $(x, y)$  座標に対応する仮想的な座標として (年齢, 時代) を考察することにより、癌死亡リスク曲面の推定が可能となる。いま、(2.1) における未知パラメータを、年齢  $a$  と時代  $p$  の関数として

$$(2.3) \quad \log \lambda_{ap} = \beta_0(a, p)$$

と仮定する。この  $\beta_0(a, p)$  はノンパラメトリックに推定される。具体的にはカーネル平滑化の要領で、固定した年齢と時代の組み合わせ近傍データに対する局所的な重み付きポアソン回帰を繰り返し適用する。その際には、年齢・時代に関する近さを (2.2) における  $d$  とし、重み関数を (2.2) で定義する。

以上は、説明変数無しのモデルであるが、GW モデルにおいては説明変数をモデルに組み込むことも可能である。その一例として、一般的な癌死亡データが性別となっていることに着目し、性別を変数として組み込むことを考えよう。 $m$  を性別を表すダミー変数 (男性なら 1, 女性なら 0) として

$$(2.4) \quad \log \lambda_{ap} = \beta_0(a, p) + \beta_1(a, p)m$$

とモデルを構築する。ここで  $\beta_1(a, p)$  は性差の効果 (ダミー変数は男性を 1 としているので、女性をコントロールとした際の男性の相対分) を表す未知パラメータ関数である。

## 2.2 年齢と時代に関する交互作用モデル

GW モデルにおいて、未知パラメータは (2.3) の通り陽に与えられないため、ノンパラメトリックに推定することとなる。ノンパラメトリックモデルは関数形に制約が少ないため適用範囲が広い一方で、パラメトリックモデルは適用範囲が狭い分様々な出力結果が得られる。そこで、前節の GW モデルによる結果をパラメトリックモデルでどの程度再現可能であるかに着目する。ここでは、佐藤 他 (2009) や Satoh and Yanagihara (2008) による経時データに対する変化係数の推測法および、富田 他 (2010) による空間データに対する変化係数曲面の推測法を参考に、年齢と時代を座標とする空間内での癌死亡リスク曲面の推測法を考察する。GW モデルにおける未知パラメータ (2.3) に対して、年齢と時代に関する  $q$  個の基底関数を表す説明変数  $\mathbf{x}(a, p) = (x_1(a, p), \dots, x_q(a, p))'$  を考え、次の線形構造を仮定したモデルを考える：

$$\beta_0(a, p) = \boldsymbol{\theta}' \mathbf{x}(a, p).$$

ただし  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)'$  は  $q$  次元未知パラメータベクトルである。基底関数  $\mathbf{x}(a, p)$  として 3 次元データの曲面補間で用いられる多項式曲面 (Ripley, 1981) を利用する場合、1 次式なら

ば  $\mathbf{x}(a,p) = (1, a, p, ap)'$  と、2次式ならば  $(1+a+a^2)(1+p+p^2)$  の各項を基底とするので  $\mathbf{x}(a,p) = (1, a, p, a^2, p^2, ap, a^2p, ap^2, a^2p^2)'$  と設定する. 1次式の基底における  $ap$  あるいは2次式の基底における  $ap, a^2p, ap^2, a^2p^2$  の4項は、年齢と時代の交互作用を表す部分となっている.

また、GWモデルと同様に交互作用モデルにおいても、年齢と時代以外の変数も組み込むことが可能である. モデル(2.4)と同様に、性別を変数として組み込む場合は、性差に関する曲面  $\beta_1(a,p)$  に関して線形構造

$$\beta_1(a,p) = \phi' \mathbf{x}(a,p)$$

を仮定し、

$$(2.5) \quad \log \lambda_{ap} = \beta_0(a,p) + \beta_1(a,p)m = \boldsymbol{\theta}' \mathbf{x}(a,p) + \phi' \mathbf{x}(a,p)m$$

というモデルを構築すれば良い.

交互作用モデルはパラメトリックモデルであり、様々な仮説検定が可能であるという利点がある. 例えば、性別を変数とするモデル(2.5)において実際に性差が存在するかどうかを調べるためには、帰無仮説「性差を表す曲面=0」に対する検定を行えば良い. つまり、帰無仮説  $H_0$  は

$$(2.6) \quad H_0 : \beta_1(a,p) = 0 \quad \text{for } \forall a, p$$

である. 富田 他(2010)における変化係数曲面の推測法を適用すると、仮説(2.6)に関する検定統計量は

$$(2.7) \quad T_1 = \hat{\phi}' \hat{\Omega}^{-1} \hat{\phi}, \quad \Omega = \text{cov}(\phi)$$

となり、 $T_1$  の漸近分布が自由度  $q$  のカイ2乗分布に従うことを用いる. ここで  $q$  は基底の数(例えば、2次式の基底では9つ)である.

あるいは、モデル(2.5)において性差の一様性に着目する場合は、帰無仮説「性差を表す曲面=定数」つまり

$$(2.8) \quad H_0 : \beta_1(a,p) = c \neq 0 \quad \text{for } \forall a, p$$

に対する検定を行えば良い. 仮説(2.8)に対する検定統計量は

$$(2.9) \quad T_2 = \hat{\phi}'_{[-1]} \hat{\Omega}_{[-1]}^{-1} \hat{\phi}_{[-1]}, \quad \Omega_{[-1]} = \text{cov}(\phi_{[-1]})$$

となり、 $T_2$  の漸近分布が自由度  $q-1$  のカイ2乗分布に従うことを用いる. ここで  $\hat{\phi}_{[-1]}$  は  $\hat{\phi}$  の第1成分を除いた  $q-1$  次元ベクトルである.

### 2.3 年齢・時代・コホートモデル

癌統計データ、特に経時測定データを解析するにあたって多用される手法に APC モデル(例えば Holfold, 1983)がある. これは、特に癌の動向の将来予測において多用される. この将来予測に関しては、短期予測(例えば、Tiwari et al., 2004 や Pickle et al., 2007)と長期予測(例えば、Kaneko et al., 2003 や大島 他, 2004)のどちらからも異なる観点からのニーズがあるが、APC モデルは主に長期予測において用いられる. その理由は、コホート効果を明確にすることにより、同一コホートの時系列内での動きを的確に表現できるからである. ここでは、この APC モデルを癌死亡リスクの視覚化に適用することについて考察する.

APC モデルとは、癌の死亡率を年齢、時代およびコホートの3成分で記述するものであり、(2.1)における未知パラメータ  $\lambda_{ap}$  の対数値が

$$\log \lambda_{ap} = \beta + \beta_a + \beta_p + \beta_c$$

で表現されているものである。ここで  $\beta$  は総平均効果,  $\beta_a$ ,  $\beta_p$  および  $\beta_c$  はそれぞれ年齢  $a$  の効果, 時代  $p$  の効果およびコホート  $c$  の効果を表す。3つの変数  $a, p$  および  $c$  には線形従属な関係  $a+c=p$  が存在するため, 前述の効果を一意に推定することができない。この問題は APC モデルにおける識別問題と呼ばれており, 様々なアプローチによる克服が試みられており (Holfold, 1983; 中村, 2000, 2005; 丹後, 2002 など), 上記3成分が分解可能, あるいは推定可能な部分のみを用いた予測などが行われている。ここでは, 3成分の推定結果を再結合して得られる  $\lambda_{ap}$  の推定値を用いて癌リスクの視覚化を試みる。

#### 2.4 デンマーク肺癌死亡データを用いた解析例

前節で考察対象とした GW モデル, 交互作用モデルおよび APC モデルは, 統計フリーソフトウェア R に解析のための関数が用意されている。そこで, 本章では読者の解析マニュアルとして役立つことを想定し, R のパッケージ Epi に実装されているデンマーク肺癌死亡データ解析結果を, スクリプトと併せて紹介する。このデータは lungDK というファイル名で実装されており 9 種類の項目から成るが, 今回の解析で用いるのはこの中の 4 項目 (年齢, 観測年, 死亡数, 人年) のみである。Web 上 <http://staff.pubhealth.ku.dk/bxc/APC/MPIDR-2009/data/> にこれら 4 項目のみのデータが公開されており, 今回はそれを用いた。男性は lung5-M.txt, 女性は lung5-F.txt が対応するファイルである。以下, スクリプトの説明の簡便化のため, 男性データを変数名 Male, 女性データを変数名 Female としておく。また Male と Female を合わせて, 性別を表すダミー変数 male (男性=1, 女性=0) を付加したデータセットを変数名 Whole として準備しておく。ここで, Male および Female における項目 A, P, D および Y はそれぞれ, 年齢, 観測年, 死亡数および人年を表す。図 1 は, 性別・年齢階級別の肺癌死亡数の経年変動を表すグラフであり, (a) が男性を, (b) が女性を表す。これらを年齢と時代を座標とする平面上で表現すると (c) および (d) となる。ここで (c) は男性, (d) は女性であり, リスクの高低を色の濃淡で表現している。色の濃い箇所が高リスクである。

##### 2.4.1 地理的加重一般化線形モデルの適用

まず最初に GW モデルによる解析結果を紹介する。GW モデルの解析において最初に行うのは, 最適なバンド幅の決定である。男性の肺癌死亡データ (Male) を用いる場合, 以下によりバンド幅が決定される。

```
library(spgwr); library(akima)
ap <- cbind(Male$A, Male$P)
bw <- ggwr.sel(D~1+offset(log(Y)), coords=ap, Male, family=poisson)
```

まず, GW モデル解析において用いるパッケージ spgwr と, 線形補完のためのパッケージ akima を用意する。更に, 位置情報(今回は年齢と時代)を変数名 ap として準備する。次に最適なバンド幅を関数 ggwr.sel により決定し, 変数名 bw として保存する。いま, 最適なバンド幅は, 男性 0.883, 女性 0.908 と推定された。ここで, 距離低減関数としてはガウス型 (2.2) が用いられている。このバンド幅の推定値 (bw) を用いて次のスクリプトを実行すると図 2 が描ける。ここで (a) は男性を, (b) は女性を表す。

```
gw <- ggwr(D~1+offset(log(Y)), coords=ap, Male, bandwidth=bw, family=poisson)
rr <- interp(Male$A, Male$P, exp(gw$SDF$X.Intercept.), linear=T)
image(rr)
contour(rr, add=T)
```

関数 ggwr によって未知パラメータ  $\lambda_{ap}$  および相対癌死亡リスク  $\exp(\lambda_{ap})$  を推定し, 変数名

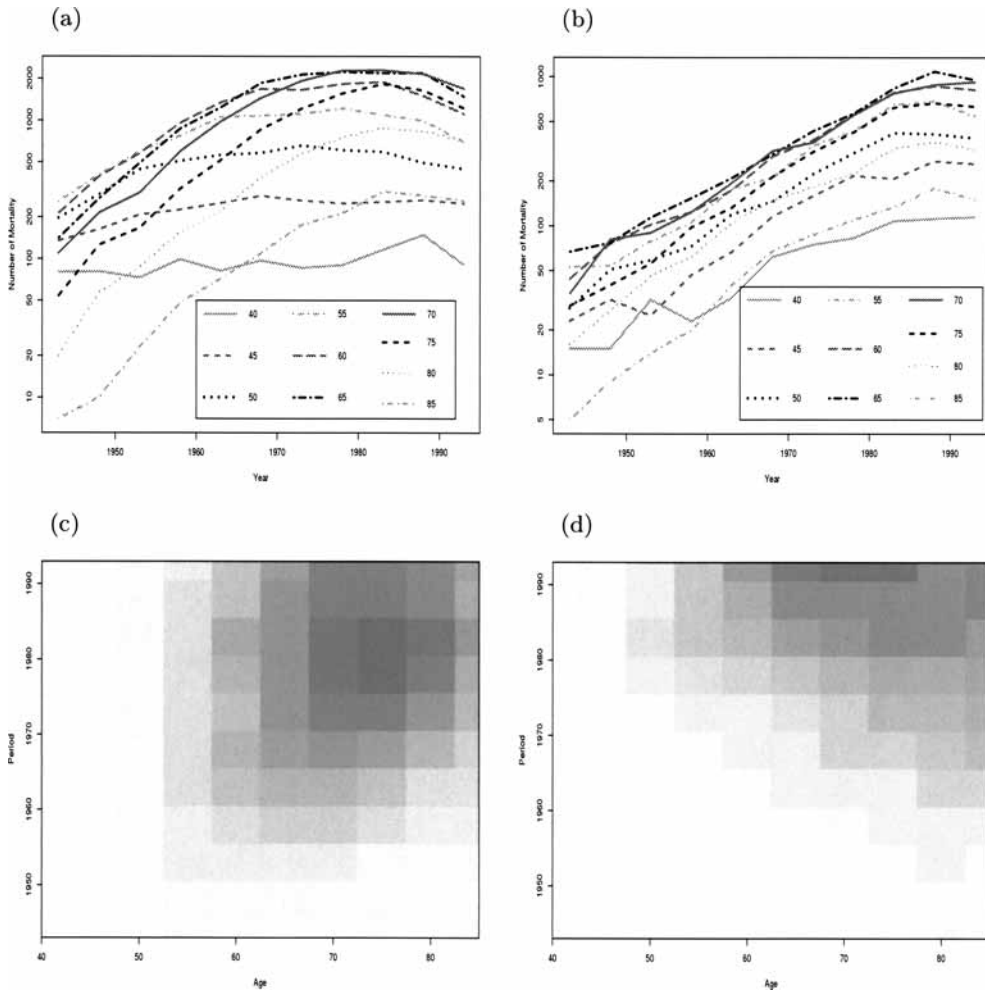


図1. デンマークにおける肺癌の死亡数. デンマークにおける肺癌の性別・年齢階級(5歳階級)別死亡数. (a)が男性, (b)が女性を表す. また(c), (d)はそれぞれ男性と女性について, 癌死亡率を色の濃淡を用いて年齢-時代平面上に表現したものである. 色の濃い箇所が高リスクである.

rrとして格納する. 次に, 関数 `image` を用いて年齢-時代平面上に癌死亡リスク曲面の高低を濃淡で表現した後に, 関数 `contour` により等高線を付加する. 図2によると, 男性に関しては(75歳, 1980年)辺りが最大癌死亡リスクとなっており, 等高線は楕円形の一部となった. 一方で女性に関しては, グラフ内にピークが見当たらないが, おおよそ(65歳, 1990年)に最大値が見受けられる. 等高線の形状は男性の楕円形よりはむしろ円形に近い.

以上は男女別に解析を行った結果であるが, 性別を変数とした包括的なモデル(2.4)を考察する場合, 解析は以下のスクリプトで行う.

```
library(spgwr); library(akima)
ap <- cbind(Whole$A, Whole$P)
```

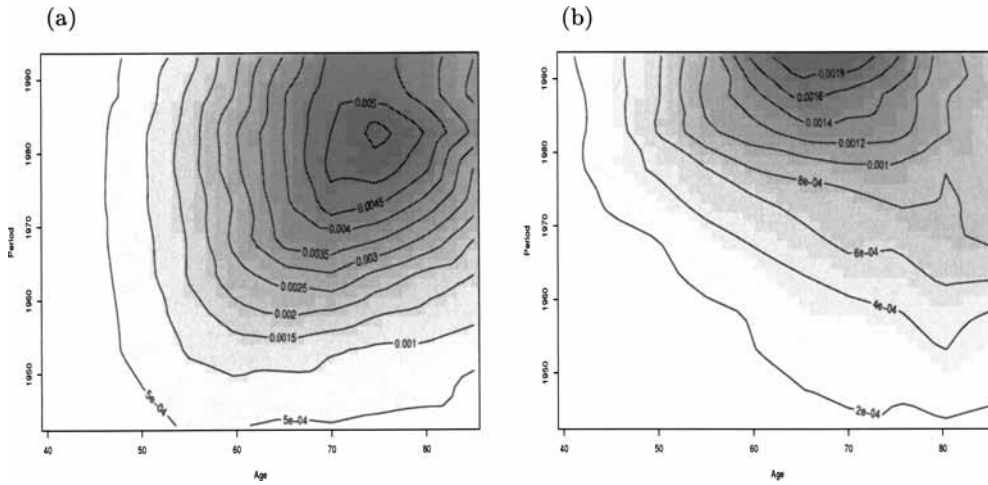


図 2. デンマーク肺癌の GW モデル解析結果(性別). デンマーク肺癌データに対する GW モデルによる推定結果. (a) と (b) はそれぞれ男性と女性の結果であり, 癌死亡リスク曲を年齢-時代平面上に表現したものである.

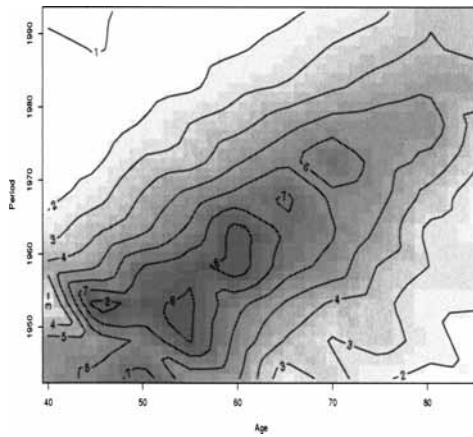


図 3. デンマーク肺癌の GW モデル解析結果(性差). デンマーク肺癌データに対して, 性別を変数として組み込んだ GW モデルによる癌死亡リスクの性差の推定結果. 女性の癌死亡リスクをコントロールとし, 男性の相対分を年齢-時代平面上に表現したものである.

```
bw <- ggwr.sel(D~1+offset(log(Y))+male,coords=ap,Whole,family=poisson)
gw <- ggwr(D~1+offset(log(Y))+male,coords=ap,Whole,bandwidth=bw,family=poisson)
rr.d <- interp(Whole$A,Whole$P,exp(gw$SDF$male),duplicate="mean")
image(rr.d); contour(rr.d,add=T)
```

この場合, バンド幅の推定値は0.914であった. また  $\exp(\beta_1(a,p))$  は性差の相対癌死亡リスクを表す部分であり, それを年齢-時代平面上に表したものが図3である. ただし, ダミー変数は男性を1としているので, 図3は女性をコントロールとした際の男性の相対分を表す. この

等高線についても、楕円形の一部に近い形であった。

#### 2.4.2 年齢と時代に関する交互作用モデルの適用

次に、交互作用モデルを考える。基底関数として3次式をフルモデルとし、Akaike (1973)による赤池情報量規準(AIC)に基づくステップワイズ変数減少法により、最適な変数の組み合わせを探索した。そのためには、以下のスクリプトを実行すれば良い。

```
library(MASS)
res <- glm(D~offset(log(Y))+(A+I(A^2)+I(A^3))*(P+I(P^2)+I(P^3)),Male,family
=poisson)
stepAIC(res)
```

AICに基づくステップワイズ関数 stepAIC はライブラリ MASS に用意されている。男性については  $a^3p$  の項を除いたモデル、女性についてはフルモデルが最適であった。この結果に基づき、癌死亡リスク曲面(女性)を描くには以下のスクリプトを実行すれば良い。

```
library(akima)
res <- glm(D~offset(log(Y))+(A+I(A^2)+I(A^3))*(P+I(P^2)+I(P^3)),Female,family
=poisson)
basis <- function(a,p)
{c(1,a,a^2,a^3,p,p^2,p^3,a*p,a*p^2,a*p^3,a^2*p,a^2*p^2,a^2*p^3,a^3*p,
a^3*p^2,a^3*p^3)}
theta <- res$coef
rr.int <- function(theta,data){
aa <- seq(min(data$A),max(data$A),by=1)
```

表1. デンマーク肺癌の交互作用モデルにおける未知パラメータ推定結果. デンマーク肺癌データに対して、年齢と時代に関する交互作用モデルを適用したときの、未知パラメータの推定結果。表中の1は定数項を、 $a$ は年齢の項を、 $p$ は時代の項を意味し、これらのクロス表で変数を表す。例えば、男性における変数  $ap^2$  に対する未知パラメータの推定値は、2行3列の  $-5.107 \times 10^{-2}$  である。また、基底のフルモデルを3次関数に設定したモデル選択の結果、男性における  $a^3p$  の項は不要であったので、記号“-”で表している。また、性差とは包括的なモデル(2.5)における男性の相対分に対する未知パラメータの推定値である。

性別	変数	1	$p$	$p^2$	$p^3$
男	1	$1.878 \times 10^6$	$-2.872 \times 10^3$	1.465	$-2.489 \times 10^{-4}$
	$a$	$-6.499 \times 10^4$	$9.980 \times 10$	$-5.107 \times 10^{-2}$	$8.708 \times 10^{-6}$
	$a^2$	$5.345 \times 10^2$	$-8.312 \times 10^{-1}$	$4.305 \times 10^{-4}$	$-7.425 \times 10^{-8}$
	$a^3$	$6.874 \times 10^{-2}$	-	$-5.210 \times 10^{-8}$	$1.746 \times 10^{-11}$
女	1	$-8.002 \times 10^6$	$1.219 \times 10^4$	-6.191	$1.048 \times 10^{-3}$
	$a$	$4.175 \times 10^5$	$-6.364 \times 10^2$	$3.233 \times 10^{-1}$	$-5.476 \times 10^{-5}$
	$a^2$	$-6.686 \times 10^3$	$1.020 \times 10$	$-5.183 \times 10^{-3}$	$8.783 \times 10^{-7}$
	$a^3$	$3.380 \times 10$	$-5.157 \times 10^{-2}$	$2.623 \times 10^{-5}$	$-4.447 \times 10^{-9}$
性差	1	$1.030 \times 10^7$	$-1.571 \times 10^4$	7.981	$-1.352 \times 10^{-3}$
	$a$	$-5.037 \times 10^5$	$7.685 \times 10^2$	$-3.908 \times 10^{-1}$	$6.624 \times 10^{-5}$
	$a^2$	$7.568 \times 10^3$	$-1.156 \times 10$	$5.883 \times 10^{-3}$	$-9.980 \times 10^{-7}$
	$a^3$	$-3.559 \times 10$	$5.441 \times 10^{-2}$	$-2.772 \times 10^{-5}$	$4.708 \times 10^{-9}$



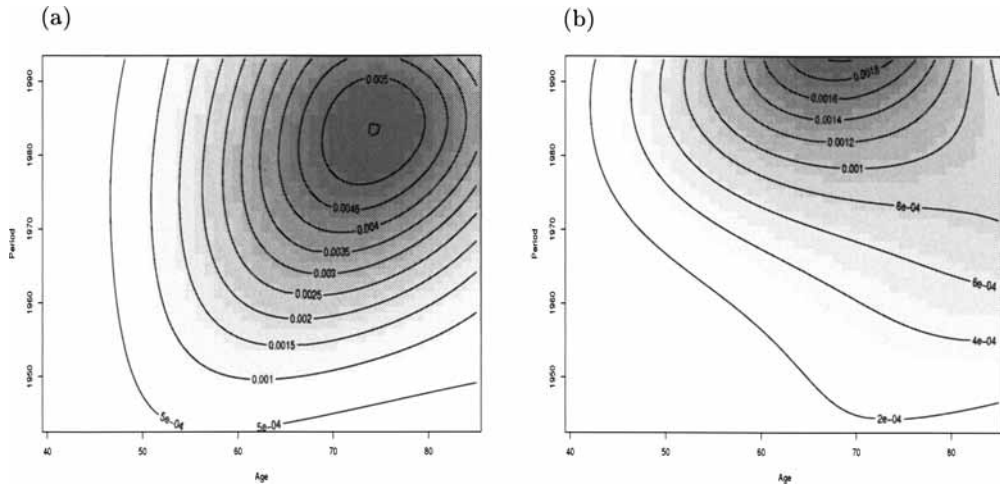


図 4. デンマーク肺癌の交互作用モデル解析結果(性別). デンマーク肺癌データに対して, 年齢と時代に関する交互作用モデルによる推定結果. ただし基底は, 男性は3次関数から  $a^3p$  の項を除いたもの, 女性は3次関数である. (a)と(b)はそれぞれ男性と女性の結果であり, 癌死亡リスク曲面を年齢-時代平面上に表現したものである.

```
pp <- seq(min(data$P),max(data$P),by=1)
rr <- matrix(NA,length(aa),length(pp))
for(i in 1:length(aa))for(j in 1:length(pp))
rr[i,j] <- exp(theta %*% basis(aa[i],pp[j]))
image(aa,pp,rr)
contour(aa,pp,rr,add=T) }
rr.int(theta,data=Female)
```

変数名 rr.int は, データと未知パラメータの推定値を引数とする自作関数であり, 癌死亡リスク曲面を表現するものである. 未知パラメータの推定結果および図 2 に倣った年齢-時代平面上における癌死亡リスク曲面の表現をそれぞれ表 1 と図 4 ((a)は男性, (b)は女性)に示す. このモデルはパラメトリックなモデルであるので自由度は低い, 自由度の極めて高い GW モデルによる結果である図 2 の特性を端的に捉えている.

次に, 性差を変数とした包括的なモデル (2.5) における  $\beta_1(a,p)$  を年齢-時代平面上に視覚化することを考える. 関数  $\beta_1(a,p)$  の推定結果を図 5 に示すが, これは以下のスクリプトを実行することにより描くことができる.

```
library(akima)
res <- glm(D~offset(log(Y))+(A+I(A^2)+I(A^3))*(P+I(P^2)+I(P^3))*male,Whole,
family=poisson)
basis <- function(a,p)
{c(1,a,a^2,a^3,p,p^2,p^3,a*p,a*p^2,a*p^3,a^2*p,a^2*p^2,a^2*p^3,a^3*p,a^3*p^2,
a^3*p^3)}
theta.d <- res$coef[grep("male",names(res$coef))]
rr.int(theta.d,data=Whole)
```

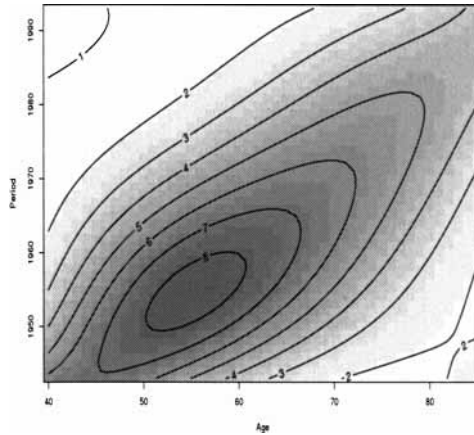


図5. デンマーク肺癌の交互作用モデル解析結果(性差). デンマーク肺癌データに対して, 性別を変数として組み込んだ交互作用モデルによる癌死亡リスクの性差に関する推定結果. 女性の癌死亡リスクをコントロールとし, 男性の相対分を年齢-時代平面上に表現したものである. ただし, 基底は3次関数である.

次に, 2.3節で触れた2種類の仮説検定(性差の存在性と一様性)を考える. これらの検定を行うには, 以下のスクリプトを実行すれば良い.

```
A0 <- Whole$A-mean(Whole$A)
P0 <- Whole$P-mean(Whole$P)
res <- glm(D~offset(log(Y))+(A0+I(A0^2)+I(A0^3))*(P0+I(P0^2)+I(P0^3))*male,
           Whole,family=poisson)
theta.d <- res$coef[ grep("male",names(res$coef))]
omega.d <- vcov(res)[grep("male",names(res$coef)),grep("male",names(res$coef))]
t1 <- theta.d %*% solve(omega.d) %*% theta.d
p1 <- pchisq(t1,df=length(theta.d),lower.tail=F)
t2 <- theta.d[-1] %*% solve(omega.d[-1,-1]) %*% theta.d[-1]
p2 <- pchisq(t2,df=length(theta.d)-1,lower.tail=F)
```

スクリプト中の変数名  $t_1$  と  $t_2$  はそれぞれ帰無仮説(2.6)と(2.8)に対する検定統計量(2.7)と(2.9)であり, 変数名  $p_1$  と  $p_2$  がそれぞれに対する  $p$  値である. また年齢と時代変数について, 年齢については40~85, 時代については1943~1993と大きな数値であるため, 平均値を減じる基準化を行った. このことにより, 分散共分散行列  $\Omega$  および  $\Omega_{[-1]}$  が特異になることを防いでいる.

性差の存在性に関する検定について, 検定統計量(2.7)は31383( $p < 0.001$ )であったので, 性差が存在することが分かる. また, 性差の一様性に関する検定について, 検定統計量は5591( $p < 0.001$ )であったので, 性差の影響は年齢や時代に対して一様でないことも分かる.

### 2.4.3 年齢・時代・コホートモデルの適用

最後にAPCモデルを考える. RにはAPCモデルに基づく推定関数 `apc.fit` がパッケージ `Epi` において既に用意されており (Carstensen, 2007), これを用いて, 例えば次のスクリプトにより解析が可能である.

```
library(Epi)
apc <- apc.fit(A=A,P=P,D=D,Y=Y,Male,model="factor")
apc.plot(apc)
```

パッケージ Epi の使用を宣言し、パッケージに含まれる関数 `apc.fit` による結果を変数名 `apc` として格納する。そして関数 `apc.plot` により図 6 の (a) と (b) が描かれる。ここで (a) は男性を、(b) は女性を表す。これらは 3 本の折れ線により構成されるが、左より年齢効果、コホート効果および時代効果を表すものである。これら 3 成分の推定結果を再結合することにより癌死亡リスク曲面が構築される。それが図 6 の (c) と (d) であり、これらは以下のスクリプトにより作製される。ここで (c) は男性を、(d) は女性を表す。

```
library(akima)
rr.apc <- function(apc,data){
  a <- apc$Age; p <- apc$Per; c <- apc$Coh
  aa <- unique(data$A); pp <- unique(data$P)
  rr <- matrix(NA,length(aa),length(pp))
  for(i in 1:length(aa))for(j in 1:length(pp))
  rr[i,j] <- a[a[,1]==aa[i],2]*p[p[,1]==pp[j],2]*c[c[,1]==pp[j]-aa[i],2]
  image(aa,pp,rr)
  contour(aa,pp,rr,add=T) }
rr.apc(apc,data=Male)
```

自作関数 `rr.apc` によって、APC モデルによる 3 成分の推定結果から各 (年齢, 時代) の癌死亡リスクを算出し、癌死亡リスク曲面を年齢-時代平面上に表現する。

#### 2.4.4 解析結果のまとめ

最後に、これらの解析結果に対する考察を行う。性別に描いた図については、癌死亡リスクが最大となる位置が存在し (図中に必ず存在するとは限らないが) そこより連続的に減少してゆく傾向にあった。これを等高線で表現すると、楕円形ないし円形 (の一部) に近い形状であった。もし、年齢効果と時代効果共に連続で単峰の凸性が存在するならば、等高線は楕円形 (あるいは円形) となる。その長軸は、年齢効果と時代効果における凸性の強弱に依存して、年齢軸か時代軸のどちらかに並行となる。しかし男性 (図 2, 4 および 6 の (a)) および性差 (図 3 および 5) に関しては、長軸はおおむね傾き 1 の方向であった。この方向は、同じ出生年に属するコホートに対する軸を表すものであるため、今後「同一コホート方向」と呼ぶこととする。もし癌死亡リスクに影響を与えるのが出生コホート効果のみであるならば、等高線は同一コホート方向に表れることになる。図 2, 4 および 6 の男性および図 3, 5 における等高線は楕円形 (の一部) に近い形状、かつその長軸が同一コホート方向であるため、同じ出生年のコホートは共通の特徴を持つという出生コホート効果の存在性が示唆される結果であった。

### 3. 日本における肝臓癌データを用いた解析

本章では、上記の手法を実際に日本のデータに適用した結果を紹介する。癌死亡および人口のデータは、人口動態統計 (厚生労働省大臣官房統計情報部編) より提供されており、国立がん研究センターのホームページ <http://www.ncc.go.jp/jp/> よりダウンロードできるが、これは観測年毎の性別・年齢階級別死亡数という形である。一般的に癌死亡や罹患 (Matsuda et al., 2011) に関するデータはこの形式であることが多い。

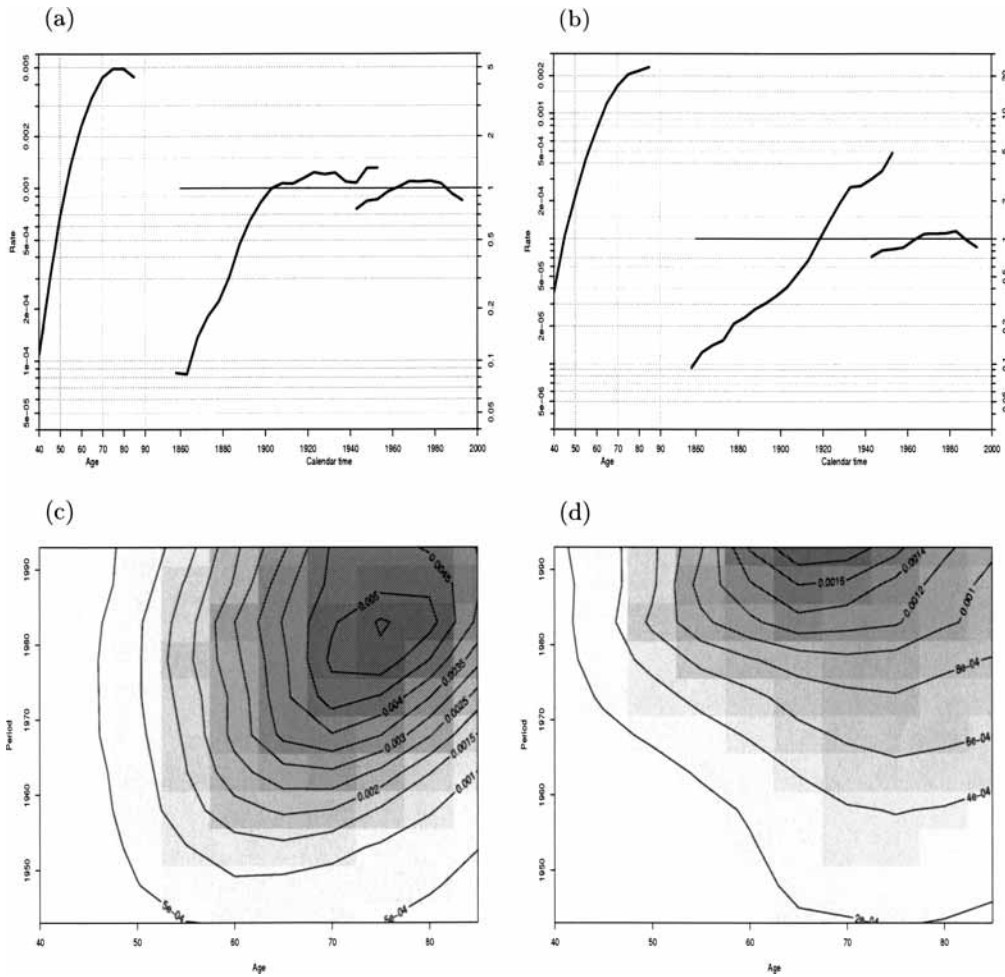


図 6. デンマーク肺癌の APC モデル解析結果. デンマークにおける肺癌データに対する APC モデルによる推定結果. (a) と (b) はソフトウェア R におけるパッケージ Epi の関数 `apc.plot` による推定結果で, (a) は男性, (b) は女性に関する結果である. 3 本の折れ線で構成されているが, 左より, 年齢効果, 生まれ年(コホート)効果, 時代効果を表す. また, (c) と (d) はそれぞれ, (a) と (b) の結果を用いて, 癌死亡リスク曲面を年齢-時代平面上に表現したものである. (c) は男性, (d) は女性の結果である.

本研究の目的である癌死亡リスクの視覚化について, 現時点では目視による確認を行うこととなる. そこで本章では, 年齢と時代に関連する特徴を持つことが先験的に知られている臓器に着目し, その特徴が視認できるかどうかをチェックすることとする. そのための一例として, 肝臓癌に着目しよう. 肝臓癌は 2009 年における臓器別死亡数 32725 人であり, 肺 (67583 人), 胃 (50017 人) に次いで第 3 位である主要な癌の 1 つである (ただし, 大腸 (42800 人) = 結腸 (28692 人) + 直腸 (14108 人) は, この順位から除いている). まず, 肝臓癌の性別・年齢階級別 (40~84 歳) の死亡率(粗率)の経年変動を図 7 に示す. ただし, (a) は男性を, (b) は女性を表す. これらを年齢と時代を座標とする平面上で表現すると (c) および (d) となる. ここで (c)

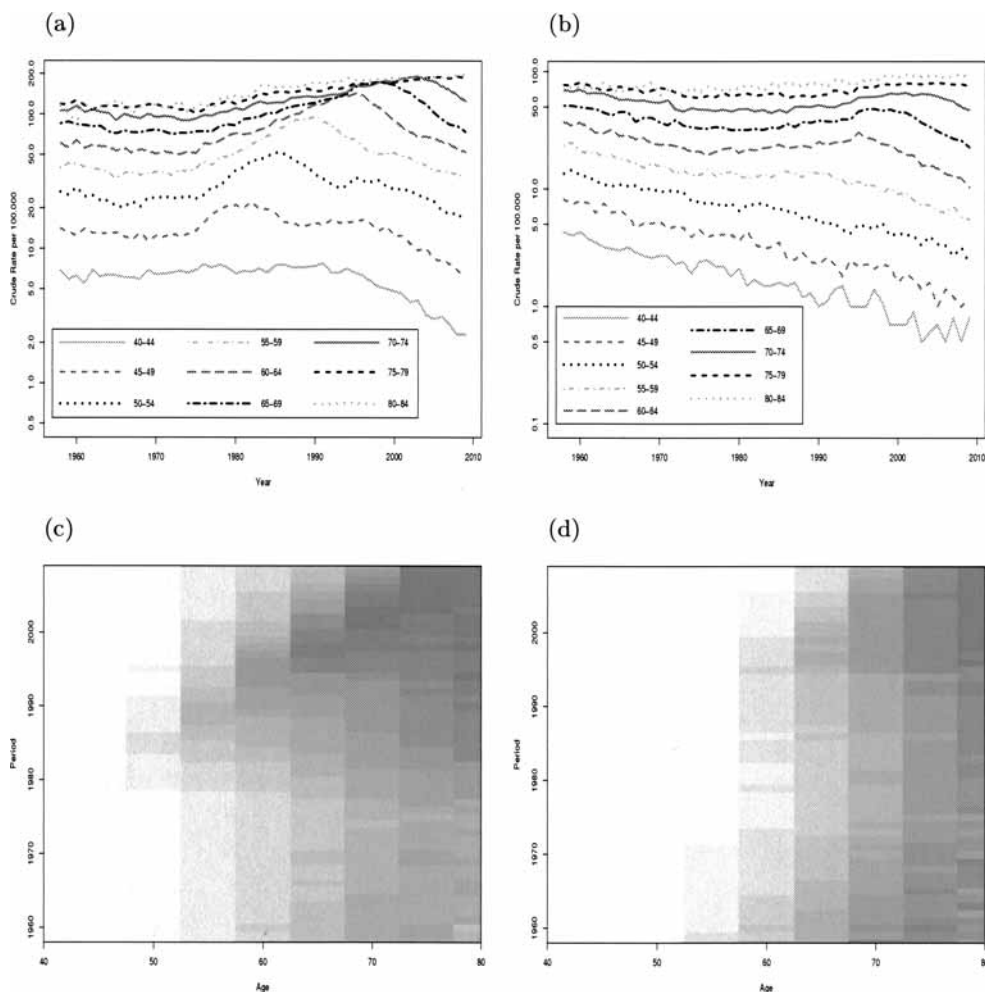


図7. 日本における肝臓癌の死亡率. 日本における肝臓癌の性別・年齢階級(5歳階級)別死亡率(粗率). (a)が男性, (b)が女性を表す. また(c), (d)はそれぞれ男性と女性について, 癌死亡率を色の濃淡を用いて年齢-時代平面上に表現したものである. 色の濃い箇所が高リスクである.

は男性, (d)は女性であり, リスクの高低を色の濃淡で表現している. 色の濃い箇所が高リスクである.

図7より, 特に男性で顕著に1935年前後の出生コホート(昭和1桁生まれ世代)における癌死亡リスクが他に比べて高いことが伺える. つまり, 同一の出生コホートが共通の特徴を持つ, あるいは特定の出生コホートが他と異なる特徴を持つという「出生コホート効果」の存在が示唆される癌であることが分かる. この点については大島(2003)など多くの論文で触れられているように, 日本における肝臓癌の多くは肝炎ウイルスの持続感染が原因であり, その抗体陽性率が特にこの出生コホートにおいて高いことがその理由として考えられる. このような肝臓癌における出生コホート効果が, GWモデル, 交互作用モデルおよびAPCモデルにより視覚化可能かどうかは非常に興味深いテーマであろう. なお, Marugame and Sobue (2004)や Imamura

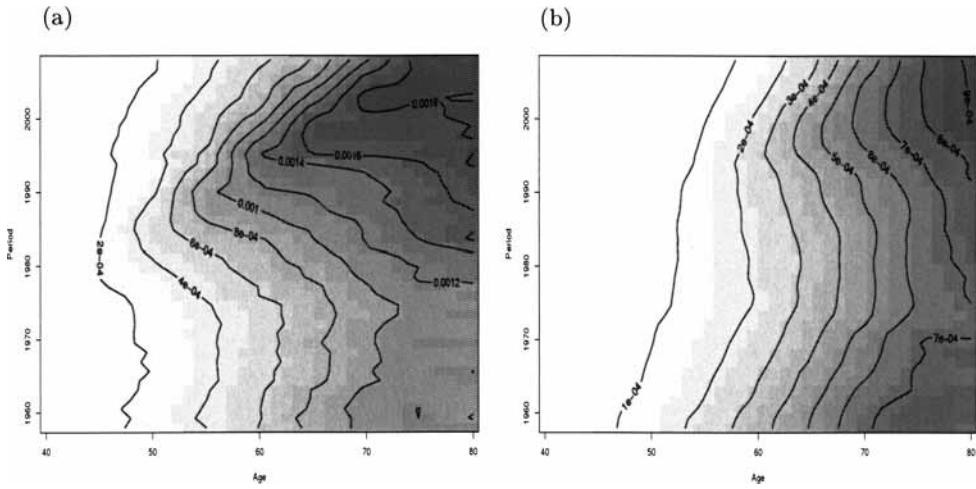


図 8. 日本肝臓癌の GW モデル解析結果. 日本における肝臓癌データに対する GW モデルによる推定結果. (a) と (b) はそれぞれ男性と女性の結果であり, 癌死亡リスク曲面を年齢-時代平面上に表現したものである. 全ての図は, デンマーク肺癌における図 2 に対応する.

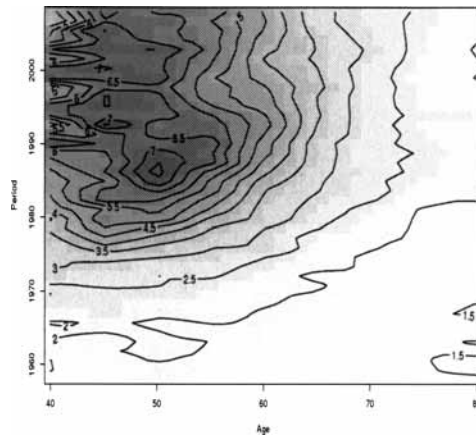


図 9. 日本肝臓癌の GW モデル解析結果 (性差). 日本における肝臓癌データに対して, 性別を変数として組み込んだ GW モデルによる癌死亡リスクの性差の推定結果. 女性の癌死亡リスクをコントロールとし, 男性の相対分を年齢-時代平面上に表現したものである. デンマーク肺癌における図 3 に対応する.

and Sobue (2004) で指摘されているように, 肝臓癌に限らず癌には出生コホート効果による差の存在が示唆されるものは少なくない.

解析に使用したデータは, 前述の通り人口動態統計 (厚生労働省大臣官房統計情報部編) により提供されている性別, 年齢階級 (5 歳階級) 別, 部位別の癌死亡および, オフセットとして人口について 1958 ~ 2008 年のものである. これらは, 国立がん研究センターホームページ <http://www.ncc.go.jp/jp/> よりダウンロードできる. ただし, 本解析においては結果の安定性

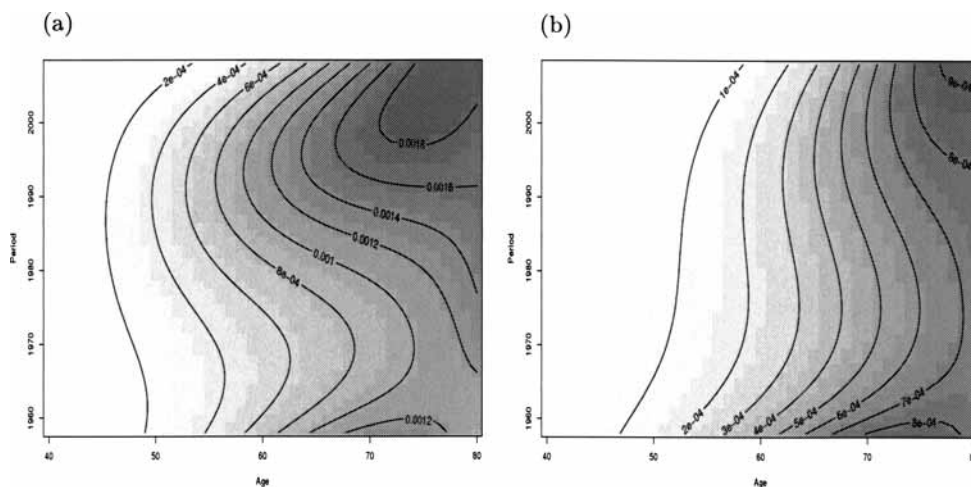


図 10. 日本肝臓癌の交互作用モデル解析結果(性別). 日本における肝臓癌データに対して、年齢と時代に関する交互作用モデルによる推定結果. ただし基底は、男性は3次関数から  $a^3p$  の項を除いたもの、女性は3次関数である. (a)と(b)はそれぞれ男性と女性の結果であり、癌死亡リスク曲面を年齢-時代平面上に表現したものである. 全ての図は、デンマーク肺癌における図4に対応する.

を図るため、年齢については40歳~84歳に限定することにより、若年および老年の影響を排除した.

GWモデル(2.3)による性別の推定結果を図8(ただし(a)は男性、(b)は女性)に示す. これは、前出の図2に対応する. まず、バンド幅の推定値は、男性で0.263、女性で1.116であった. 一方、性別を変数とした包括的なモデル(2.4)においてはバンド幅は0.548と推定された. そして、図3に倣って性差(女性をコントロールとした際の男性の相対分)を年齢-時代平面上に表現したものが図9である.

次に、交互作用モデルを考察する. 本データの解析においても、3次式の基底をフルモデルとした. つまり推定する未知パラメータは、フルモデルにおいては  $(1+a+a^2+a^3)(1+p+p^2+p^3)$  の各項における係数であり、全部で16個である. AICに基づく変数選択の結果、男性については  $a^3p$  の項を除いたモデル、女性についてはフルモデルが最適であった. 未知パラメータの推定結果を表2に示し、前出の図4に倣って癌死亡リスク曲面を表現したものが図10である. また、性差について図5に倣った結果が図11である. 次に、性差が存在するかどうかを調べるための、帰無仮説(2.6)に対する検定を行うと、検定統計量(2.7)は226050( $p < 0.001$ )であったので、性差は存在することが分かる. 次に性差の一様性については、帰無仮説(2.8)に対する検定統計量(2.9)が26527( $p < 0.001$ )であったので、性差の影響は、年齢や時代に対して一様でないことも分かる. しかし、図5と図11を比較すると、性差の特徴はデンマークの肺癌と日本の肝臓癌で大きく異なった. デンマークの肺癌については、1900年前後の出生コホートで極めて高い性差(男性の高リスク)が見られたような出生コホート効果の存在が特徴的であったが、日本の肝臓癌については、このような性差に関する出生コホート効果は見られなかった.

最後にAPCモデルを適用した結果を図12に示す. これは、前出の図6に対応するものである. 前の2つのモデルと同様な特性が表現できたが、年齢階級の影響でメッシュが粗くなっている.

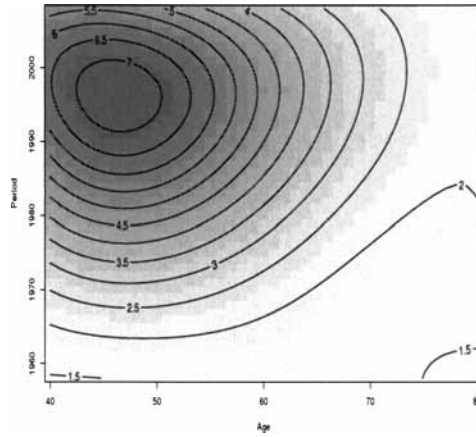


図 11. 日本肝臓癌の交互作用モデル解析結果(性差). 日本における肝臓癌データに対して, 性別を変数として組み込んだ年齢と時代に関する交互作用モデルによる癌死亡リスクの性差に関する推定結果. 女性の癌死亡リスクをコントロールとし, 男性の相対分を年齢-時代平面上に表現したものである. ただし, 基底は3次関数である. デンマーク肺癌における図5に対応する.

表 2. 日本肝臓癌の交互作用モデルにおける未知パラメータ推定結果. 日本における肝臓癌データに対して, 年齢と時代に関する交互作用モデルを適用したときの, 未知パラメータの推定結果. 表中の1は定数項を,  $a$ は年齢の項を,  $p$ は時代の項を意味し, これらのクロス表で変数を表す. 例えば, 男性における変数  $ap^2$  に対する未知パラメータの推定値は, 2行3列の  $1.458 \times 10^{-1}$  である. また, 基底のフルモデルを3次関数に設定したモデル選択の結果, 男性における  $a^3p$  の項は不要であったので, 記号“-”で表している. また, 性差とは包括的なモデル(2.5)における男性の相対分に対する未知パラメータの推定値である.

性別	変数	1	$p$	$p^2$	$p^3$
男	1	$-4.916 \times 10^6$	$7.521 \times 10^3$	-3.834	$6.516 \times 10^{-4}$
	$a$	$1.851 \times 10^4$	$-2.845 \times 10^2$	$1.458 \times 10^{-1}$	$-2.489 \times 10^{-5}$
	$a^2$	$-1.531 \times 10^3$	2.349	$-1.247 \times 10^{-3}$	$2.161 \times 10^{-7}$
	$a^3$	$-2.803 \times 10^{-1}$	-	$2.139 \times 10^{-7}$	$-7.192 \times 10^{-11}$
女	1	$-7.619 \times 10^6$	$1.158 \times 10^4$	-5.862	$9.895 \times 10^{-4}$
	$a$	$3.329 \times 10^5$	$-5.063 \times 10^2$	$2.566 \times 10^{-1}$	$-4.335 \times 10^{-5}$
	$a^2$	$-4.375 \times 10^3$	6.662	$-3.382 \times 10^{-3}$	$5.722 \times 10^{-7}$
	$a^3$	$1.756 \times 10$	$-2.681 \times 10^{-2}$	$1.364 \times 10^{-5}$	$-2.313 \times 10^{-9}$
性差	1	$1.829 \times 10^6$	$-2.734 \times 10^3$	1.362	$-2.261 \times 10^{-4}$
	$a$	$-1.033 \times 10^5$	$1.543 \times 10^2$	$-7.686 \times 10^{-2}$	$1.276 \times 10^{-5}$
	$a^2$	$2.101 \times 10^3$	-3.146	$1.570 \times 10^{-3}$	$-2.611 \times 10^{-7}$
	$a^3$	$-1.381 \times 10$	$2.071 \times 10^{-2}$	$-1.035 \times 10^{-5}$	$1.725 \times 10^{-9}$

#### 4. おわりに

本論文では, 癌死亡リスクを年齢と時代を座標とする空間内の曲面として視覚化するという目的に対する統計モデルおよび解析結果を紹介した. ここでは, GW モデル, 交互作用モデルおよび APC モデルの3つを用いて癌死亡リスク曲面を推定し, その結果を年齢-時代からなる



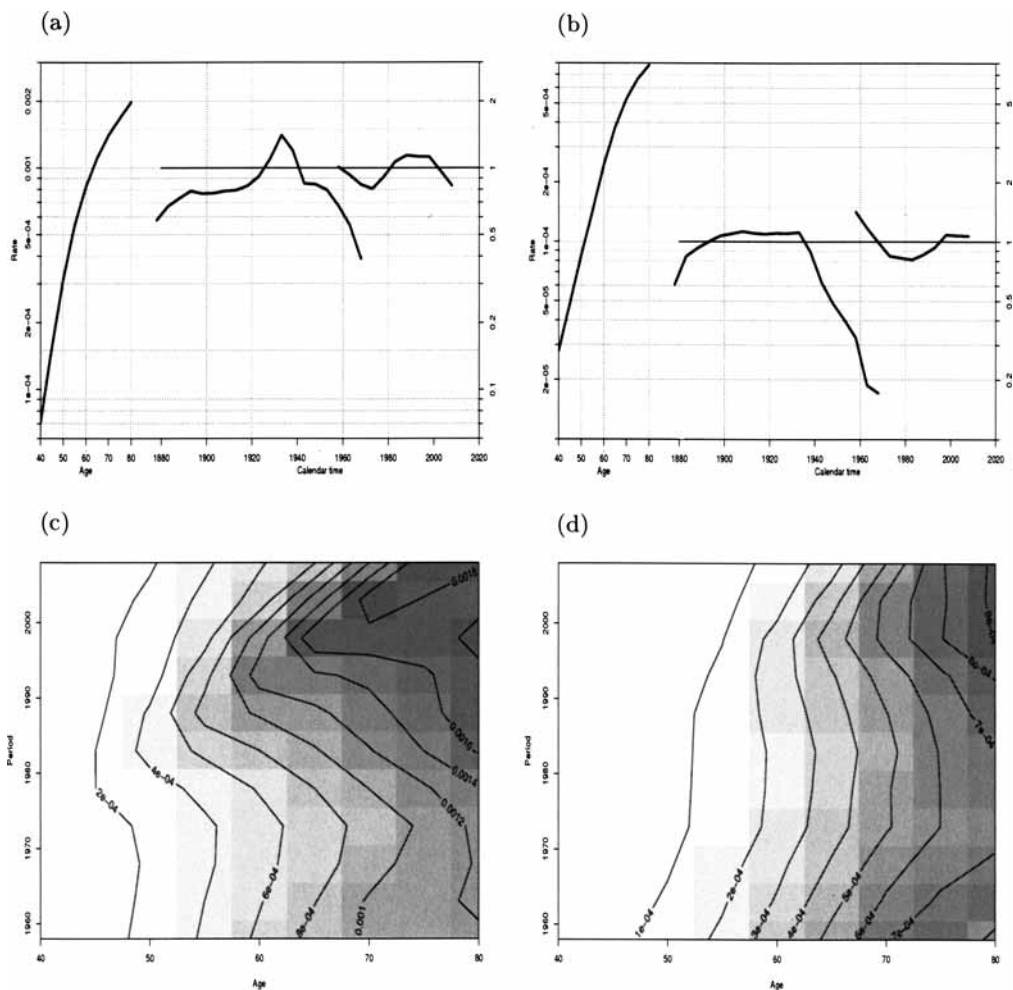


図 12. 日本肝臓癌の APC モデル解析結果. 日本における肝臓癌データに対する APC モデルによる推定結果. (a)と(b)はソフトウェア R におけるパッケージ Epi の関数 apc.plot による推定結果で, (a)は男性, (b)は女性に関する結果である. 3本の折れ線で構成されているが, 左より, 年齢効果, 生まれ年(コホート)効果, 時代効果を表す. また, (c)と(d)はそれぞれ, (a)と(b)の結果を用いて, 癌死亡リスク曲面を年齢-時代平面上に表現したものである. (c)は男性, (d)は女性の結果である. 全ての図は, デンマーク肺癌における図 6 に対応する.

2次元平面上に濃淡と等高線を用いて視覚化した.

まず始めに GW モデルについて考察する. このモデルの概念は, カーネル平滑化に基づいており, 距離低減関数を用いて近隣情報に重みをつけた推定を行うため, 自由度は非常に高い. 従って, リスクの集積性など, 局所的な特性を表現することが可能となる. 今回, 年齢-時代を仮想的な位置情報とみなすことにより癌死亡データへの適用が可能となった. また, 年齢と時代以外の変数を組み込むことも可能であり, 例えば性別を変数として組み込むと図 3 や 9 のように性別の効果(性差)についても, 年齢-時代平面上に視覚化が可能である.

次に、交互作用モデルについて考察する。このモデルは、汎用的な一般化線形モデルがベースとなっているのが利点である。実際 R においては、GW モデルについて推定関数 `ggwr`、APC モデルについては推定関数 `apc.fit` が用意されているが、現時点では他の多くのソフトウェアには実装されていない。しかし、一般化線形モデルについては殆どの統計ソフトウェアに実装されているので、交互作用モデルについては、多くの読者は手持ちのソフトウェアで実行可能であろう。また、今回用いた多項式基底では自由度が極めて低いものの、信頼区間の構築や検定が実行可能であるのも利点である。実際の解析結果(デンマークの肺癌と、日本の肝臓癌)を見ても、3次式をフルモデルとする基底でほぼ特徴は再現できていたので、癌死亡リスクのおおまかな特徴を知るにはこのようなパラメトリックなモデルで十分であると考えられる。

最後に APC モデルについて考察する。APC モデルの最大の利点は、年齢・時代・コホート効果の定量的な評価が可能にある点にある。例えば、今回解析例として取り上げた日本の肝臓癌における昭和一桁生まれの高リスク出生コホート効果については、図 12 (a) のように明確なピークとして捉えられる。しかし、本論文の目的である、年齢-時代平面上における癌死亡リスクの視覚化という点に関しては、説明変数が飽和状態となるほどの柔軟なモデルであるが、結果に関しては GW モデルほどの柔軟さは見られない。一方で、モデルのシンプルさという点に関しては交互作用モデルほどではないという位置付けになる。

次に、これらの手法を出生コホート効果の存在が示唆されている日本の肝臓癌に適用した結果について考察を行う。大島(2003)など多くの先行研究により、1935年前後(昭和一桁世代)の出生コホートにおいて、他に比べて癌死亡リスクが高いことが指摘されており、この点を表現できるかが1つの鍵であった。GW モデルによる結果である図 8 においては、男性における1935年前後の出生コホートにおける特異的な高リスク効果を視認することができた。それは、この出生コホートが辿る同一コホート方向において、あたかも山の尾根のように高リスクなラインが走っているからである。性差について年齢-時代平面上に表現した図 9 は、45歳-1995年周りにピークを持ちそこから均等に低くなる傾向が見られた。また、交互作用モデルにおいても(3次関数をフルモデルとして)1935年出生コホート効果および、その他全体的な特性の再現が可能であった。ただし自由度が低いため、1935年出生コホート効果は若干丸められて表現されている。

最後に APC モデルに関しては、同様の特性が観測されたが、年齢階級設定の影響によりメッシュが粗くなる。

以上、GW モデル、交互作用モデルおよび APC モデルについて、R のスクリプトを交えながら説明し、日本の肝臓癌データに適用した結果についても議論を行った。本研究の目的は、統計モデルによる癌死亡リスクの視覚化であり、癌死亡リスク曲面を年齢-時代平面上に表現することを試みた。実解析として、日本の肝臓癌における「昭和一桁世代の高リスク」というコホート効果の存在も、具体的な年の確認までには至らなかったが視覚化した。これらの結果を踏まえて、3つのモデルを2つの視点(結果の精密さ、モデルの汎用性)から比較する。最も精密かつ柔軟な結果(リスク曲面の推定結果)が得られたのは GW モデルであった。一方で、最も柔軟性を欠く結果は交互作用モデルによるものであったが、ソフトウェアの利用や検定の可能性等を鑑みると、最も汎用性の高いモデルであるといえる。

これらの推定結果および解釈はあくまで目視に依存するものであり、解析者の主観や先入観がある程度混入することが考えられる。そこで、主観に頼らず自動的に癌死亡リスクの特性を検出する手法の開発が今後のテーマの1つとなる。

また、今後の研究テーマとして癌の時系列解析で多用される「予測」に着目する。この点に関して、GW モデルにおいては外挿の問題が発生する。本来 GW モデルはホットスポットと呼ばれる癌死亡リスクの集積する地点の検出など、考察対象地域内における特性を検出するこ

とに優れており、予測を対象地域外に適用(外挿)することは基本的に想定されていない。また交互作用モデルにおいても、自己回帰モデルのように同一コホート内の経時的な相関を記述するなどの工夫が必要である。従って、これら2つのモデルについては、現時点では予測に適したモデルとは言い難い。一方でAPCモデルについては、本来予測において多用される手法でもあり、予測方法はある程度確立されている。従ってGWモデルと交互作用モデルに関しては、予測の可能性についての議論およびモデルの改良が今後のテーマである。

本研究においては、癌死亡リスクの視覚化、特にそれを年齢と時代を座標とする空間内における曲面として表現する手法に関する考察を行った。地図を描く際に有効なGWモデル、汎用性の高い交互作用モデル、癌の時系列解析で頻出のAPCモデルを用いた解析を2種類(デンマークの肺癌と日本の肝臓癌)行った。本論文では、モデルの特性や結果の解釈および今後の発展の可能性等を総合的に判断して、これらの優劣を完全に決する段階には至らなかった。そこで、現時点では各モデル固有の長所や短所を勘案しつつ、実際の解析においては、手法を限定せず複数を併用して評価を行うのが有効であると考えられる。

## 謝 辞

査読者より大変有益なコメントを頂きましたこと、感謝申し上げます。

## 参 考 文 献

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, *2nd International Symposium on Information Theory* (eds. B. N. Petrov and F. Csáki), 267–281, Akadémiai Kiadó, Budapest.
- Brunsdon, C., Fotheringham, S. and Charlton M. (1998). Geographically weighted regression — Modelling spatial non-stationarity, *Journal of the Royal Statistical Society: Series D*, **47**(3), 431–443.
- Carstensen, B. (2007). Age-period-cohort models for the Lexis diagram, *Statistics in Medicine*, **26**, 3018–3045.
- Fotheringham, A. S., Brunsdon, C. and Charlton, M. E. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, Wiley, Chichester.
- Holford, T. R. (1983). The estimation of age, period and cohort effect for vital rates, *Biometrics*, **39**, 311–324.
- Imamura, Y. and Sobue, T. (2004). Mortality trend of colon, rectal, liver, “gallbladder and biliary tract” and pancreas cancer in Japan by birth cohort, *Japanese Journal of Clinical Oncology*, **34**(8), 491–493.
- Kaneko, S., Ishikawa, K. B., Yoshimi, I., Marugame, T., Hamashima, C., Kamo, K., Mizuno, S. and Sobue, T. (2003). Projection of lung cancer mortality in Japan, *Cancer Science*, **94**(10), 919–923.
- 国立がん研究センター ホームページ (2011). がん情報サービス, <http://www.ncc.go.jp/jp/>.
- Marugame, T. and Sobue, T. (2004). Mortality trend of mouth and pharynx, esophagus, stomach, larynx and lung cancer in Japan by cohort, *Japanese Journal of Clinical Oncology*, **34**(7), 432–438.
- Matsuda, T., Marugame, T., Kamo, K., Katanoda, K., Ajiki, W., Sobue, T. and Japan Cancer Surveillance Research Group (2011). Cancer incidence and incidence rates in Japan in 2005: Based on data from 12 population-based cancer registries in the Monitoring of Cancer Incidence in Japan (MCIJ) project, *Japanese Journal of Clinical Oncology*, **41**(1), 139–147.

- 中村 隆(2000). 質問項目のコウホート分析—多項ロジット・コウホートモデル—, 統計数理, **48**(1), 93–119.
- 中村 隆(2005). コウホート分析における交互作用効果モデル再考, 統計数理, **53**(1), 103–132.
- 大島 明(2003). 日本での慢性肝疾患, 特に肝癌の疫学的特徴, 第123回日本医学会シンポジウム記録集 ウイルス肝炎, 13–19.
- 大島 明, 黒石哲生, 田島和雄(2004). 『がん・統計白書』, 篠原出版社, 東京.
- Pickle, L. W., Hao, Y., Jemal, A., Zou, Z., Tiwari, R. C., Ward, E., Hachey, M., Howe, H. L. and Feuer, E. J. (2007). A new method of estimating United States and state-level cancer incidence counts for the current calendar year, *CA A Cancer Journal for Clinicians*, **57**(1), 30–42.
- Ripley, B. D. (1981). *Spatial Statistics*, Wiley, New York.
- 佐藤健一, 柳原宏和, 加茂憲一(2009). 離散分布の経時測定データにおける線形な変化係数の推測について, 応用統計学, **38**, 1–11.
- Satoh, K. and Yanagihara, H. (2008). Estimation of varying coefficients for a growth curve model, *American Journal of Mathematical and Management Sciences* (in press).
- Stone, M. (1974). Cross-validated choice and assessment of statistical predictions, *Journal of the Royal Statistical Society: Series B*, **36**, 111–147.
- Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion, *Journal of the Royal Statistical Society: Series B*, **39**, 44–47.
- 丹後俊郎(2002). 『医学データデザインから統計モデルまで』, 共立出版, 東京.
- 谷村 晋(2010). 『地理空間データ分析』, 共立出版, 東京.
- Tiwari, R. C., Ghosh, K., Jemal, A., Hachey, M., Ward, E., Thun, M. J. and Feuer, E. J. (2004). A new method of predicting US and state-level cancer mortality counts for the current calendar year, *CA A Cancer Journal for Clinicians*, **54**(1), 30–40.
- 富田哲治, 佐藤健一, 柳原宏和(2010). 空間データに対する交互作用モデルを用いた変化係数曲面の推測について, 応用統計学, **39**(2&3), 1–12.

## Cancer Mortality Risk Visualization on Age-period Plane

Ken-ichi Kamo<sup>1</sup>, Tetsuji Tonda<sup>2</sup> and Kenichi Satoh<sup>2</sup>

<sup>1</sup>Center for Medical Education, Sapporo Medical University

<sup>2</sup>Research Institute for Radiation Biology and Medicine, Hiroshima University

Cancer information obtained by a suitable method becomes a basis for planning an effective cancer control program. Our purpose in this paper is to visualize cancer mortality risk as the curved surface on the age-period plane. To achieve this objective, a geographically weighted generalized linear model (GW model) may be suitable by regarding the information for age and period as the geographical one on a two-dimensional plane. On the other hand, we consider a parametric interaction model by age and period as a rival of the non-parametric GW model. Moreover, we consider one more model, that is, an age-period-cohort model, which is widely used in the analysis and prediction of longitudinal cancer data. These models are described with the script which can be executed in R and results for lung cancer mortality in Denmark for reappearance by the reader.

To verify the achievement of our objective, that is, cancer mortality risk visualization, we also apply these methods to data of liver cancer mortality in Japan. Liver cancer is said to have a cohort effect such that the birth cohort around 1935 suffers a high risk. We check whether this effect can be visualized.

Finally, we discuss not only the statistical aspect, but also future problems for our research objective. Each model has its merit, so it may be effective to use plural methods jointly to estimate the properties for cancer mortality risk.