Simon Holmgaard Yu Song Yu

## Gaussian Markov Random Fields

## Applied to Image Segmentation



# Simon Holmgaard and Yu Song Yu* 

## GAUSSIAN MARKOV RANDOM FIELDS APPLIED TO IMAGE SEGMENTATION

Preprint 1982 No. 1

## Abstract.

Gaussian Markov Random Fields (GMRF) can be used to describe a family of textures (*). A solution to the segmentation problem is presented together with a brief summary on GMRF.

## Keywords

Image processing,<br>Markov Random Fields.<br>Image Segmentation<br>Cluster Analysis<br>Variable Selection Methods.

(*) The term texture is used in image processing as a name for whatever structure there is present in a part of a picture which are homogeneous in some sense.

## Introduction.

A series of images of "Malmbjerget" in the Mestersvig area of Greenland has for some time been used for research in the field of digital image processing, see [5].

The images were originally obtained for geological purposes by "Grønlands Geologiske Undersøgelser" in order to try to locate areas where minerals are present.

This project began as an attempt to use statistical models for image processing, models which also take into account some spatial dependence between image points (pixels).

The work was carried out while the second author was visiting the Institute of Datalogy, University of Copenhagen. All programmes used have been written by Yu Song Yu, while S. Holmgaard is responsible for the statistical part and the presentation of paper.

This report will describe an attempt to solve the segmentation problem for greylevel images where the textures can be described as realizations of isotropic one dimensional Gaussian Markov Random Fields.

## Description of the Segmentation Problem.

We first start out with some notation: Let A be the index set for the pixel values which form the image.

The image is then $\left\{\mathrm{X}_{\mathrm{t}}\right\}_{t \in A}$ where $\mathrm{A} \subseteq \mathrm{z}^{2}$, and the $\mathrm{X}_{\mathrm{t}}$ 's are the measured pixel values.

In order to make segmentation meaningful we assume that there exist $A_{1}, \ldots, A_{N}$ such that $A={ }_{n}^{N}{ }_{1} A_{n}$ and $A_{i} n_{A_{j}}=\varnothing$ if iキj where each $A_{n}$ only consists of one type of texture, i.e. each $A_{n}$ is homogeneous in some sense.

The segmentation problem ..is now simply the problem of finding $A_{1}, \ldots, A_{N}$.

In our case the textures are something like snow, ice, sediment, granite, etc.

The segmentation problem can also be described as a special kind of clustering problem where one also uses the "geographical" information when making clusters.

Gaussian Markov Random Field (GMRF).
Denote by $\mathbf{z}^{2}$ the lattice in the plane, so that each lattice point corresponds to a pixel and thereby describes a location where a one dimensional normally distributed random variable is measured.

Consider a Gaussian Random Field:
$x=\left(X_{t}\right)_{t \in Z^{2}}$ where each finite set of $X_{t}^{\prime}$ s has a normal distribution.

We call X Markov, or say X has the Markov property, if for every finite $\Lambda \subset \mathbf{Z}^{2}$

$$
\begin{aligned}
& \mathscr{L}\left(\left(x_{t}\right)_{t \in \Lambda} \mid\left(x_{t}\right)_{t \in \Lambda} c\right) \\
= & \mathscr{L}\left(\left(x_{t}\right)_{t \in \Lambda} \mid\left(x_{t}\right)_{t \in \partial \Lambda}\right)
\end{aligned}
$$

Here $\partial \Lambda$ is the edge of $\Lambda$ in the following sense

$$
\partial \Lambda=\left\{t \in \Lambda \Lambda^{c}|\exists u \in \Lambda:|t-u|=1\}\right.
$$

where $t=\left(t_{1}, t_{2}\right)$,

$$
|t|=\left|t_{1}\right|+\left|t_{2}\right| .
$$

If we further assume that the GMRF is isotropic (translation and rotation invariant), that is assume

$$
\begin{align*}
\mathrm{EX}_{t} & =\xi, \quad \forall t \in \mathrm{Z}^{2}  \tag{1}\\
\operatorname{Cov}\left(\mathrm{X}_{\mathrm{t}}, \mathrm{X}_{\mathrm{u}}\right) & =\mathrm{R}(|\mathrm{t}-\mathrm{u}|) \quad \forall t, u \in \mathrm{Z}^{2}
\end{align*}
$$

it may be shown that $R(t)$, the covariance function, has the following form: (see [1])

$$
\begin{equation*}
R(t)=\sigma^{2} \sum_{n=0}^{\infty} \alpha n^{n}(0, t) \tag{2}
\end{equation*}
$$

where $\mathrm{P}^{\mathrm{n}}(0, t)$ is the $\mathrm{n}^{\prime}$ th iterate of the transition function for the symmetric random walk on $Z^{2}, \sigma^{2}>0$ and $|\alpha|<1$. From this one finds that for any $t \in Z^{2}$

$$
\begin{equation*}
\mathscr{L}\left(X_{t} \mid\left(X_{u}\right)_{u \in \partial t}\right)=N\left(\xi(1-\alpha)+\alpha Z_{t}, \sigma^{2}\right) \tag{3}
\end{equation*}
$$

where $Z_{t}=\frac{1}{4} u \in \mathcal{F}_{\mathrm{F}} \mathrm{X}_{\mathrm{u}}$.
Note that the condition $|\alpha|<1$ is necessary to ensure the existence of the GMRF over the infinite lattice, but $|\alpha| \geq 1$ is possible on finite subsets.

Estimation in one dimensional GMRF
Suppose that the GMRF is observed on a finite subset of $z^{2}$.

The aim is to estimate the parameters $\alpha, \xi$ and $\sigma^{2}$ In order to do this look at $\left\{\mathrm{X}_{\mathrm{t}}\right\}_{t \in \Lambda}$ for given $\left\{\mathrm{X}_{\mathrm{t}}\right\}_{t \in \partial \Lambda}$ and choose $\Lambda$ as the pixels denoted by "o" in the figure:

It is then seen that, conditional on $\left\{X_{t}\right\}_{t \in \partial \Lambda}, X_{t}$ and $X_{u}$ are independent for $t, u \in \Lambda$, $t \neq u$.
The likelihood function then becomes
(4) $L_{\xi, \alpha}, \sigma^{2}=\prod_{t \in \Lambda} \frac{1}{\sqrt{2 \pi}} \frac{1}{\sigma} e^{-\frac{4}{2}\left(\frac{X_{t}-\xi(1-\alpha)-\alpha Z}{} t^{\sigma}\right)}{ }^{2}$
writing $\xi(1-\alpha)+\alpha Z_{t}=\gamma+\alpha\left(Z_{t}-\bar{Z}\right)$
one gets the following estimates with direct reference to standara regression analysis. See [3] for example.
(5)

$$
\begin{aligned}
& \hat{\gamma}=\bar{x} \\
& \hat{\alpha}=\frac{t \sum_{\Lambda} x_{t}\left(Z_{t}-\bar{z}\right)}{t_{\Lambda} \sum_{\Lambda}\left(Z_{t}-\bar{z}\right)^{2}} \\
& \hat{\sigma}^{2}=\frac{1}{\# \Lambda-2} t \sum_{\Lambda}\left(x_{t}-\hat{\gamma}-\hat{\alpha}\left(z_{t}-\bar{z}\right)\right)^{2} \\
& \hat{\xi}=(\hat{\gamma}-\hat{\alpha} \bar{z}) /(1-\hat{\alpha})
\end{aligned}
$$

## Comparing GMRF's

Suppose that we have observed $n$ different GMRF's and we now want to see if they are equal in terms of distribution i.e. if they all can be described by the same $\xi, \alpha$ and $\sigma^{2}$.

It is then easy to see that if we condition on half the observations, e.g. the "x"'s we can use Bartlett'stest to test the hypothesis that all $\sigma^{2}$ are equal to a common $\sigma^{2}$.

If all $\sigma^{2}$ are equal, the task of comparing the $\alpha$ 's and the $\xi^{\prime} s$ is the same as comparing lines in linear regression analysis, which can be done using an $F$-test, see [3].

## Displaying Observations From a Lattice.

In order to get a visual impression of the images, plots are made in the following way:

Each pixel value is transformed into one of 16 different grey levels, where white $\sim l a r g e ~ v a l u e s ~ a n d ~ b l a c k ~ ~ s m a l l ~ v a l u e s . ~$

In each picture the grey level scale is linear and with maximum contrast, more precisely: the interval between the max and min pixel value is divided into 16 equally long subintervals, where each interval corresponds to a specific grey level.

Generating Pseudo realizations of a GMRF.
Using the symmetry between the "o" and "x" one may generate pseudo GMRF realizations as follows:
(1) Take pseudo random numbers with mean $\xi$ and variance $\sigma^{2}$ for all "o 's i.e. conditioning on all $X_{t}=\xi$.
(2) The "o"'s are now known, so calculate the mean for all "x"'s as given by (3) and draw pseudo random numbers with the ppropriate mean and variance $\sigma^{2}$ for all "x"'s.
(3) Repeat (2) interchanging the roles of
(4) Repeat steps (2) and (3) a few times.

In order to do this in a proper way one must at each step drop all "x" or "o" on the edge because it is not possible to calculate their mean. Hence, if one wants a 40 by 40 pixel square interacting step (2) and (3) 3 times, one has to start out with a 52 by 52 square.

Figure 1 shows 15 generated images with different values for $\xi$ and $\alpha$ but all with $\sigma=12$. They were all generated with step (2) and (3) repeated 3 times.

It has not been proved that the above method converges in any sense, but when estimating from the generated pictures the estimates are quite reasonable.


Pseudo realizations of a GMRF with $\sigma=12$.

Image B with 5 subimages.


## The Images.

Figure 2 shows image $B$ from [5] together with 5 subimages which were pointed out by a geologist as being homogeneous in a geological sense.

One remark about subpicture $B 2$, which is a little suspicious because of the white edge in the bottom right half of the picture. This part of the picture has been analysed by Johansen and Yu in [5] using other methods and it was found that it consists of two different hornogeneous areas.

The Model.
We specify our model for the image in such a way that a segmentation is a hypothesis in this model.

Let $B_{1}, \ldots, B_{K} B_{i} \cap B_{j}=\varnothing$, i\#j be an initial segmentation i.e. each $\mathrm{B}_{\mathrm{k}}$ contains only one texture.

The segmentation now consists of finding $B_{k}$ 's with the same texture i.e.

$$
A_{n}={ }_{i=\frac{U^{U}}{n}}^{1} B_{m(n, i)} \quad n=1,-, N
$$

so that $A_{1}, \ldots, A_{N N}$ is a segmentation.
Expressed in terms of GMRF we assume that each $B_{k}$ follows the GMRF model with parameters $\xi_{k}, \alpha_{k}$ and $\sigma$.

If no natural initial segmentation exist it is reasonable to choose the $B_{k}$ 's as small as possible because this gives a detailed segmentation, the problem with small $B_{k}$ 's is : that estimates of the parameters are poor i.e. they have large variance.

Preliminary Image Analysis (Model check).
Before going deeper into the segmentation, some preliminary analyses were performed:
a) A goodness of fit test, for normality.
b) Checking the model, at local level, against a second order Markov model.
c) Checking that a fixed $\sigma^{2}$ and varying $\xi$ and $\alpha$ can describe image $B$.

## A Goodness of Fit Test for a GMRF.

In order to check the model, the usual goodness of fit test can be used to check the normality:

Choose a division of the real line

$$
-\infty=u_{0}<u_{1}<\ldots<u_{k}=\infty
$$

and denote by

$$
o_{i}=\neq\left\{x_{t} \in\left(u_{i-1}, u_{i}\right) \mid t \in \Lambda\right\} \quad i=1,-, k
$$

and

$$
E_{i}={ }_{t} \sum_{\Lambda} \int_{u_{i-1}}^{u_{i}} \frac{1}{\sqrt{2 \pi}} \frac{1}{\hat{\sigma}} e^{-\frac{1}{2}\left(\frac{y-\hat{\xi}(1-\hat{\alpha})-\hat{\alpha} Z}{t}\right)^{2}} \hat{\hat{\alpha}} d y
$$

and use the test statistic
(6) $T=\sum_{i=1}^{k} \frac{\left(O_{i}-E_{i}\right)^{2}}{E_{i}}$.

The asymptotic distribution of $T$ will lie between $\chi^{2}(k-1)$ and $\chi^{2}(k-4)$. See [7].

The $u_{i}$ 's must be chosen such that $E_{i}>5 i=1,-k$ in order to assure that the $\chi^{2}$ approximation is valid.

This test was used on images B1-B5, and all except B2 were accepted at significance level 5\%.

## Test of First-order Markov Property

 Against Second-order Markov Property.A given lattice point "o" can be defined to have an edge $\partial t$ with either 4 or 8 pixels, called the neighbour pixels.

| $\otimes$ | $X$ | $\otimes$ |
| :--- | :--- | :--- |
| $X$ | $O$ | $X$ |
| $\otimes$ | $X$ | $\otimes$ |

The figure shows the usual 4 neighbours denoted by "x" with the additional 4 neighbours denoted by " ". $^{\text {. }}$

The second-order or 8-neighbour model can now be described by its conditional distribution:

$$
\begin{equation*}
\mathscr{L}\left(X_{t} \mid\left(X_{u}\right)_{u \in \partial t}\right)=N\left(\alpha+\gamma Z_{t}+\delta Y_{t}, \sigma^{2}\right) \tag{7}
\end{equation*}
$$

where $Z_{t}$ are the average pixel values over the neighbour "x" and $Y_{t}$ the averages over the neighbour " $\otimes$ ".

The model with \# $\# t=4$ is called first-order Markov and the model with \#at=8 is called second-order Markov. For reference, see [2].

Estimation in this model can then be performed by conditioning on both "x" and "凶", that is $3 / 4$ of all pixels. The test of the first-order model against the second-order model is now the same as testing the hypothesis " $\delta=0$ " and this can be done using an F-test, see [2]. It is important that when estimating under " $\delta=0$ " one only uses the "O"'s as observations.

This test was again used on images B1-B5 and only B2 was rejected at 5\% level.

## Test For Uniform Variance Over Image B.

In order to check that a large image has the same $\sigma^{2}$ in every pant, image $B$ was divided into $1528 \times 8$ squares and $\xi$, $\alpha$ and $\sigma^{2}$ were estimated separately for each square. Then Bartlett's test is used to test whether all $\sigma^{2}$ are equal to a common value. The test on image $B$ was accepted at the $5 \%$ level.

One could also note that $s^{2}$, the estimated $\sigma^{2}$, has a $\sigma^{2} \chi^{2} / \mathrm{f}$ distribution for all small squares, so one can compare the empirical distribution function of the $s^{2}$ 's with the $\sigma^{2} \chi^{2} / \mathrm{f}$ distribution with $\sigma^{2}$ equal to the common estimated value.

The Suggested Solution to the Segmentation Problem.
The preceding analysis shows that the GMRF model fits quite nicely on selected areas, and assuming that it fits on small areas then the variance is the same anywhere on an image.

This then leads to the following algorithm: Start out by dividing the image into small areas and then merge areas which are most alike, etc.
More precisely:
(1) Compare all pairs of areas provided they are physical neighbours, (they have one cutting edge in common).
(2) If no pair of neighbours is sufficiently alike, then stop.

Otherwise merge the pair which is most alike, and let their union be a new area.
(3) Repeat from step (1).

In all cases, we start by conditioning on half the picture, the "x"'s, and cutting as follows:


Leaving four "o" in each small square.

This has the advantage that no cutting line goes through an "o", so when merging two neighbouring areas into one, no extra "o"'s are added, This would be the case if the cuts were horisontal or vertical.

The reason for choosing the squares with four "o" in each square is the wish to make the segmentation as detailed as possible.

The disadvantage of choosing the squares as small as post sible is of course that the parameters are badly determined. It is obvious that one should try with larger initial cuttings in order to get better initial estimates, but this has not been done.

Verbal Description of the Algorithm.
In view of the preceding description we only have to devise a reasonable distance measure and a way of choosing the stop criterion.

The suggestion is to use the $F$-statistic for two neighbouring areas having the same $\xi$ and $\alpha$ as the distance measure, and stop if all $F$ values are larger than the appropriate 1- $\delta$ fractile, e.g. $\delta=5 \%$.

There are no problems when merging the two first areas but, when merging for the second time one must decide whether to pool the variances or not.

The problem "pool or not" does not affect which pair to merge but only when to stop, because the ranking of the $F_{i j}$ values is the same.

Technical Description of the Computational Procedure.
We must first introduce some notation.
Let $A \subseteq Z^{2}$ such that all "o" in the image are indexed by a $t \in A$, and $\operatorname{set} M=\# A$.

We look at the situation at some step in the algorithm, and assume that

$$
A={ }_{i}^{N} \underline{=}_{1}^{N} A_{i} \quad \text { and } A_{i}=n_{j}^{U_{i}^{i}} A_{i j}
$$

where $A_{i j}$ are the original small squares with $\# A_{i j}=4$ and $A_{i}$ are the areas which have been formed by the successive merging operations.

Let $\mathrm{X}=\left(\mathrm{X}_{\mathrm{t}}\right)_{t \in \mathrm{~A}}$ be the observations and $\dot{P}_{\mathrm{B}_{1}, \ldots, B_{\mathrm{m}}}(\mathrm{X})$ be the projection on the meanvalue subspace of $R^{N}$ where $\alpha$ and $\gamma$ take $m$ different values each belonging to a $B_{i}$, i.e. for $X_{t} \in B_{i}: E X_{t}=\alpha_{i}+\gamma_{i} Z_{t}$. Finally ${ }_{i} \underline{U}_{1}^{\mathrm{U}} B_{i}=A$ and $B_{i} \cap B_{j}=\varnothing$ if i $\ddagger j$.

Now define the following SSD's

$$
\begin{align*}
& \operatorname{SSD}_{i j}^{N}=t \sum_{A}^{N}\left(P_{A_{1}-r^{-}, A_{N}}(X)_{t}-P_{A_{1},-,} A_{i} U A_{j},-, A_{N}(X)_{t}\right)^{2}  \tag{8}\\
& =t \sum_{A_{i}} \cup A_{j}\left(P_{A_{1},-, A_{N}}(X) t^{\left.-P_{A_{1}},-, A_{i} \cup A_{j},-, A_{N}(X)_{t}\right)^{2}(i \neq j)}\right. \\
& \operatorname{SSD} \quad={ }_{t} \sum_{A}\left(X_{t}-P_{A_{11}},-, A_{1 n^{\pi}} A_{21},-, A_{N 1^{1}}-, A_{N n_{N}}(X)\right)^{2} \\
& \operatorname{SSD}_{S}^{N}={ }_{t} \sum_{A}\left(P_{A_{1}},-, A_{N}(X)_{t}-P_{A_{11},-, A_{N n_{N}}}(X)_{t}\right)^{2},
\end{align*}
$$

and note that in step N -1 we have
(9) $\quad \operatorname{SSD}_{S}^{N-1}=\operatorname{SSD}_{S}^{N}+\operatorname{SSD}_{i j}^{N}$
after merging $A_{i}$ and $A_{j}$.
We are now ready to write up the $F$ values:
i) pooled case:
(10) $F_{i j}=\frac{\frac{1}{2} S S D_{i j}}{\frac{1}{M-2 N}\left(S S D+S S D_{S}\right)} \sim F(2, M-2 N)$
ii) non-pooled case:

$$
(11) \mathrm{F}_{i j}=\frac{\frac{1}{\frac{1}{2} \mathrm{M}-2(\mathrm{~N}-1)}\left(\mathrm{SSD}_{i j}+\mathrm{SSD}_{\mathrm{S}}\right)}{\frac{1}{\mathrm{M}-\frac{1}{2} \mathrm{M}} \mathrm{SSD}} \sim \mathrm{~F}\left(\frac{1}{2} \mathrm{M}-2(\mathrm{~N}-1), \frac{1}{2} \mathrm{M}\right)
$$

The nice thing about this algorithm is that it is only necessary to compute some $S S D_{i j}$ in each step in order to choose which pair to merge, i.e. the $S S D_{i j}$ belonging to areas which are neigbours to the area formed by the last merge operation.

It is immediately seen that the ranking ot the $\mathrm{F}_{\mathrm{ij}}{ }^{\prime} \mathrm{s}$ is the same as that of the $S S D_{i j}$ 's both in case i) ans ii) so the difference between the two methods is only in the stop criterion.

In order to describe the different stop criteria, the following may be said:

In the pooled case, we stop if the next step is unreasonable provided that all previous steps are correct decisions.

In the non-pooled case, we stop if the next step gives a total segmentation which is unreasonable compared with the initial segmentation.

Note that $M$ very well can be very large e.g. for images P1-P15:800 and for image $B: 5120$. This means that $F_{i j} \approx F(2, \infty)$ $=X^{2}(2) / 2$ in the pooled case, and $F_{i j} \approx F(\infty, \infty)=\varepsilon(1)$ in the non-pooled case where $\varepsilon(1)$ is the measure with mass 1 in 1 .

This means that it is only possible to direct the stop criterion using the significance level in the pooled case, because the non-pooled case ends with stop when all $F_{i j}>1$.

## Experiments.

In the following pages results from experiments are shown. In figure 3 the first picture shows the result of running the algorithm on $P 1$ from figure 1. The remaining 3 examples in figure 3 are made form joining the middel part of 2 pictures from figure 1 and then running the algorithm.

One immediately notices that stopping at the 5\% level in the pooled case seems quite unreasonable, but good enough in the non-pooled case.

Pooled


P1-P15


Before the first merge:<br>Image D1 from [5].



Plot of estimated $\gamma$ values.

Plot of estimated $\alpha$ values.


```
Plot of F Fij values.
```



Top Part of Image B.



Result of Segmentation Non-pooled
case and stop at the 5\% level.


$$
\text { Plot of initial } F_{i j} \text { values. }
$$

Results of Segmentation Pooled case, with stop at:
$F(2, \infty)=4.61(1 \%)$
$F(2, \infty)=100$
$F(2, \infty)=1000$



In figure 3 note also that when trying to separate images P1 and P13, which have the same $\xi$ but different $\alpha$ 's, the algorithm fails.

This can be explained by the fact that $\alpha$ is not determined very exactly in the early stages of the algorithm.

$$
\begin{align*}
& V(\hat{Q})=E\left(V\left(\hat{\alpha} \mid z_{t} ; t \in \Lambda\right)\right)  \tag{12}\\
+ & V\left(E\left(\hat{\alpha} \mid z_{t} ; t \in \Lambda\right)\right) \\
= & E \frac{\sigma^{2}}{\sum_{\Lambda \Lambda \Lambda}\left(z_{t}-\bar{z}_{t}\right)^{2}} \approx \frac{4}{\# \Lambda-3}
\end{align*}
$$

where " $\approx$ " by assuming that

$$
\sum_{t \in \Lambda}\left(Z_{t}-\bar{z}\right)^{2} \approx \frac{\sigma^{2}}{4} x^{2}(\# \Lambda-1),
$$

and because \# $\Lambda=4$ in the initial stages of the dlgorithm it will be quite impossible to find which observations belong where.

It seems that this does not matter much because most of the information necessary for segmentation is present in $\xi$, see Figure 4.

All the same, $\alpha$ is not the same for the whole picture, a test for this hypothesis on image D1 rejects it.

The Plots of the Initial $\mathrm{F}_{\mathrm{ij}}$ values.
Plots of the initial $F_{i j}$ values were made in order to find out how good they are at indicaring edges.

The results from Fig. 4 and 5 suggest, that one could perhaps start by combining all areas with $F_{i j}$ below a certain level. This, however, would give the following uniqueness problem: What should be done if we have the following situation


## Conclusion and Final Comments.

I think that the presented method works quite nicely especially when using the non-pooled stop criterion. The problem of being unable to segmentate areas with the same $\xi$ but different $\alpha$ does not seem to be a serious drawback, and could be solved by enlarging the initial square size.

The fact that there is actually no information in the $\alpha ' s$ also suggests that one estimates a single $\alpha$ for the whole picture, and then starts segmentation on the basis of $X_{t}-\hat{\alpha} Z_{t}$, which can be considered as the observations compensated for spatial effects.

This paper mainly serves as an Ilustration of the method, the practical importance lies in solving the multivariate case.

## Suggestions for Further Work.

1) Try to make a supervised version of the algorithm i.e. start out with training areas and then try to classify the small squares. Of course the poorly determined $\alpha$ 's will also give problems.
2) Try to solve the multidimensional problem, using the presented algorithm but with multivariate response in each pixel. However, this is not just an ordinary multivariate analysis of variance: The estimation problem will be the following: Estimate $\underline{\xi}$, $\underline{=}$ and $\underline{\underline{\sum}}$ from

$$
\begin{aligned}
& E\left(\underline{X}_{t} \mid\left(\underline{X}_{u}\right)_{u \in \partial t}\right)=(\underline{\underline{I}}-\underline{\underline{A}}) \underline{\xi}+\underline{\underline{A}}_{\underline{Z}} \underline{t} \\
& V\left(\underline{X}_{t} \mid\left(\underline{X}_{u}\right)_{u \in \partial t}\right)=\underline{\underline{E}}
\end{aligned}
$$

 positive definite.

The last condition follows from the fact that

$$
\underline{\underline{R}}(t)=\sum_{n} \sum_{0}^{\infty} P^{n}(0, t) \quad \stackrel{A}{A}^{n} \underline{\underline{n}}
$$

must be symmetric and positive definite for $t=0$. This can be solved using an iterative procedure which shifts between estimating one of $\underset{\underline{A}}{\underline{A}}$ and $\underset{\underline{\underline{N}}}{ }$, keeping the other fixed. But then there is still the test problem.
4) Try to use better methods for estimation see e.g. [4]. for approximations to the likelihood function.
5) It would be wise to study some of the literature on variable selection methods and cluster analysis.

## Acknowledgement.

I would like to thank Martin Jacobsen, Peter Johansen and S申ren Johansen for their participation in this work.

Also I would like to thank Graham Leeson for helping me with the English language.

## References.

[1] Spitzer, F. (1973)
Introduction aux processeus de Markov a parametre class $Z_{\nu}$. Ecole d'Ete de probabilités de Saint-Flour III 1973. Lecture Notes in Mathematics 390, Springer. (See also [6]).
[2] Besag, J. (1974)
Spatial Interaction and The Statistical Analysis of Lattice Systems.
J.R. Statist. Soc. B, 36, 192-236.
[3] Scheffé, H. (1959)
The Analysis of Variance. Wiley, New York.
[4] Künsch, H. (1981) Maximum Likelihood with edge correction for Gaussian Markov Fields. (to appear).
[5] Johansen, P. and Yu, Y.S. (1981) Digital Processing of Multi Spectral Image. Institute of Datalogy, Univ. of Copenhagen.
[6] Rozanov, Yu.-A. (1967)
On Gaussian Fields with given conditional distribution. Theory of Prob. and its Appl. 12.
[7] Chernoff, H. and Lehmann, E.L. (1954)
The Use of the Maximum Likelihood Estimates in $\chi^{2}$ Test for Goodness of Fit. Ann. Math. Stat., 25, 579-586.

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN $\varnothing$, DENMARK。

No. 1 Johansen, Søren: Asymptotic Inference in Random Coefficient Regression Models.

No. 2 Asmussen, Sфren: On the Role of a Certain Eigenvalue in Estimating the Growth Rate of a Branching Process.

No. 3 Lauritzen, Steffen. Time Series Analysis in 1880. A Discussion of Contributions made by T.N. Thiele.

No. 4 Asmussen, Søren: Conditioned Limit Theorems Relating a Random Walk to its Associate, with Applications to Risk Reserve Processes and the GI/G/I Queue.

No. 5 Johansen, Søren: The Statistical Analysis of a Markov Branching Process.

No. 6 Asmussen, S申ren: Time-Dependent Approximations in some Queueing Systems with Imbedded Markov Chains Related to Random Walks.

No. 7 Skovgaard, Ib M.: A Second-Order Investigation of Asymptotic Ancillarity.

No. 8 Rootzén, Holger: The Rate of Extremes of Stationary Normal Sequences.

No. 9 Skovgaard, Ib M.: Large Deviation Approximations for Maximum Likelihood Estimators.

No. 10 Jensen, Ulla Funck: A Stochastic Projection Model with Implications for Multistate Demography and Manpower Analysis.

No. 11 Johansen, Søren: An Extension of Cox's Regression Model.
No. 12 Skovgaard, Ib M.: Edgeworth Erpansions in Statistics.

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN $\varnothing$, DENMARK.

No. 1 Holmgaard, Simon and Yu, Song Yu: Gaussian Markov Random Fields Applied to Image Segmentation.

