Environmental Risk Mapping
Mixture Models for Conditional Gaussian
Abstract: This paper proposes the use of Gaussian mixture models to estimate conditional probability density functions in environmental time series. The Gaussian mixture model is used to identify the different modes of the data and to estimate the probability density function of each mode. The results show that the Gaussian mixture model performs well in estimating the probability density function of the data, and it can be used to identify the different modes of the data.

IEEE International Workshop on Neural Networks for Signal Processing (NNSP), 2002.

Nicholas Chihali, Mark Mandel, and Sunny Denghi
\[
\frac{(x)^d}{(x^d)} = (x^d)
\]

2 Gaussian Mixture Models

2.1 Algorithms Description

Conditional GMM for local PDF estimation:

In this section, we present the algorithms for local PDF estimation. We introduce the GMMs and the GMMs with mixtures of Gaussian distributions. We also present the algorithms for local PDF estimation.

To evaluate the performance of this method, we compare it to the well-known Gaussian mixture model (GMM).

PDFs are used to model the probability distribution of a random variable. In this section, we focus on methods that can be used to estimate the PDF of a random variable.

In Gaussian mixture models, the PDF is approximated by a weighted sum of Gaussian distributions. This is done by fitting a mixture of Gaussian distributions to the data.

In this section, we present methods for estimating the parameters of a Gaussian mixture model. These methods are based on maximum likelihood estimation.

The Gaussian mixture model is a powerful tool for modeling data. It is widely used in many fields, including machine learning, computer vision, and signal processing.

The Gaussian mixture model is a flexible model that can be used to model a wide range of data. It is also a powerful tool for clustering, classification, and regression.

In summary, the Gaussian mixture model is a powerful tool for modeling data. It is widely used in many fields and is a flexible model that can be used to model a wide range of data.

Environmental surveys

Environmental surveys are used to collect data on environmental variables. These surveys are important for understanding the environment and for making decisions about environmental management.

Environmental surveys are typically conducted using a variety of methods, including field surveys, remote sensing, and modeling.

Field surveys are conducted by collecting data directly from the environment. This can be done using a variety of methods, including direct measurement, sampling, and monitoring.

Remote sensing is used to collect data from the environment using satellites or other sensors. This can be used to collect data on a wide range of environmental variables, including temperature, humidity, and precipitation.

Modeling is used to predict environmental variables based on data collected from the environment. This can be done using a variety of methods, including regression analysis, machine learning, and artificial intelligence.

Environmental surveys are important for understanding the environment and for making decisions about environmental management. They are used to collect data on environmental variables, and this data is used to inform decisions about environmental policies and practices.
In a Gaussian mixture model, the covariance matrix of each Gaussian is diagonal. Hence, for each Gaussian:

\[ \mathbf{\Sigma} = \text{diag}(\mathbf{\Sigma}_i) \]

Based on the Gaussian mixture model, the conditional and posterior sampling distribution are defined as:

\[ p(x|\theta) = \sum_i p(x|\theta_i) \pi_i \]

Each mixture component is then estimated using the expectation-maximization algorithm (EM). However, for the purpose of this discussion, we assume that the full posterior distribution is known.

The posterior Gaussian model is then used to compute the weights of a linear regression method called... (continued)

\[ p(\theta|x) \propto p(x|\theta)p(\theta) \]

where \( p(\theta|x) \) is the posterior distribution of \( \theta \) given the data. The posterior distribution of \( \theta \) is then used to compute the weights of a linear regression method called... (continued)

\[ p_\theta(x) = \frac{1}{(2\pi)^{d/2} |\mathbf{\Sigma}|^{1/2}} \exp\left\{ -\frac{1}{2} (x-\mathbf{\mu})^T \mathbf{\Sigma}^{-1} (x-\mathbf{\mu}) \right\} \]

where \( \mathbf{\mu} \) is the mean of the vector \( \mathbf{\mu} \) and \( \mathbf{\Sigma} \) is the covariance of the data.

Sequential Gaussian Simulations

\[ p(\theta|x) = \frac{1}{(2\pi)^{d/2} |\mathbf{\Sigma}|^{1/2}} \exp\left\{ -\frac{1}{2} (x-\mathbf{\mu})^T \mathbf{\Sigma}^{-1} (x-\mathbf{\mu}) \right\} \]

where:

- \( x \): data
- \( \mathbf{\mu} \): mean of the vector
- \( \mathbf{\Sigma} \): covariance matrix

The method developed for these experiments was found to be similar to the discrete Forward Algorithm... (continued)


3.2 GCS Experiments Protocol

3.3 GCM Experiments Protocol
Table 1: Number of data points inside the various subsets of SWIND and GRISON.

<table>
<thead>
<tr>
<th>Subset</th>
<th>Total</th>
<th>Reference</th>
<th>Test</th>
<th>Train</th>
<th>SWIND</th>
<th>GRISON</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100</td>
<td>9625</td>
<td>1000</td>
<td>900</td>
<td>900</td>
<td>900</td>
</tr>
</tbody>
</table>

In bold cases, results were generated using the same seed and a reference set. Table I gives the number of points inside each set.

4.1 Data Description

**Experiments**

The statistic is the mean of the D-statistics over the whole testing set.

- Compute the mean of the D-statistics over the whole testing set.
- Function are different.
- For each n, a distribution is fitted to verify whether two distribution $G(n)$ and $G(\text{test})$ are different.
- Compute the D-statistic, which is the greatest distance between $G(n)$ and $G(\text{test})$ at every location $n$.
- Estimate the cumulative PDF estimator.
- Consistently the conditional PDF estimator, we proceed as follows:

Other definitions and approaches are discussed in detail in the literature. Of course, this cumulative distribution is only an approximation of the true CDF. However, it is the best approximation one can expect without any a priori knowledge of the location. Therefore, when the cumulative distribution is not clear, the cumulative distribution is considered a cumulative distribution function at every location of the data.
The performance of the complete data set for SWAND (Ref) and GISONS (Ref) are shown in Figure 2. The baseline curves for the performance set, the dashed curve is the baseline set, and the dot curve the baseline set.

The null distribution of the complete data set is shown in Figure 3, the distribution is estimated directly from the local estimation. On the left, the baseline distribution is estimated from the local estimation. On the right, both methods are used to estimate the baseline distribution. On the right, both methods are used to estimate the baseline distribution. On the right, both methods are used to estimate the baseline distribution.
The smoothing tendency of GMM and SGS data set becomes obvious when
location, the cumulative from GMM and the one from SGS are very close to each other.

The similarity between GMM and SGS performance is also visible on Figure 2. For both sample
sets, the performance of GMM is better. On the other hand, GMM has a loe of differences to
improve the performance and SGS performance is slightly better than GMM. It is now clear
that SGS is not a better fit on this data set.

Table 3 shows that conditional GMM and SGS perform in a very similar way on the GRISON data
set.

Table 3: Comparison of cumulative distribution of 99 Gaussians from SWIND and SGS.

<table>
<thead>
<tr>
<th>SWIND</th>
<th>SGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>0.95</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Figure 3: Comparison of the cumulative distribution of all these estimations from conditional GMM.

There are the estimations from SGS.

These are the estimations from SGS.

Figure 4: Results from distribution calculation on SWIND and SGS data set between 20 locations.

Table 4: Results from distribution calculation on SWIND and SGS data set between 20 locations.

<table>
<thead>
<tr>
<th>SWIND</th>
<th>SGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>0.95</td>
<td>0.95</td>
</tr>
</tbody>
</table>
of multiple output variables, of a multivariate at any location of the studied area, and can easily handle joint distributions. A strong advantage of conditional GMM is that it needs less computationally intensive data to produce results, at least as efficient in terms of D-statistics, and even better when the distribution of variables is unknown. Conditional Gaussian mixture models proved to be efficient in estimating local probability density.

### Conclusion

Conditional GMM and SGS are performing similarly in terms of D-statistics, but in a completely different way. GMMs are performing better in terms of D-statistics as long as the density function is known, while SGS perform well in terms of D-statistics when the density function is not known. The conditional GMMs do not need information about the other variables to reproduce the conditional density.

**Figure 6**. Risk maps of the probability density under 23 elevations. The darker, the less probable. Conditional GMM risk map is on the left, GMM risk map is in the middle, and SGS risk map is on the right.

<table>
<thead>
<tr>
<th>GMM</th>
<th>SGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.08</td>
</tr>
<tr>
<td>0.96</td>
<td>0.08</td>
</tr>
<tr>
<td>0.08</td>
<td>0.96</td>
</tr>
</tbody>
</table>

**Table 3**: Results from D-statistics calculation on GMMs between the model conditional Gaussian mixture (GMM) and conditional Gaussian mixture (GS). The smaller the values, the better the model.
References


Acknowledgments

This work was supported by Swiss National Science Foundation (CARTHAN project: FN 30-692300) and the Swiss National Office for the Environment (ENV 30-692300). The GMD experiments were conducted using the Earth Observation Laboratory, the SGS, and the Geostatistical Laboratory.