

# Study on Links between Cerebral Infarction and Climate Change Based on Hidden Markov Models

Hiroshi Morimoto<sup>1</sup>

<sup>1</sup>Graduate School of Environmental Study, Nagoya University, Nagoya, Japan

Correspondence: Hiroshi Morimoto, Graduate School of Environmental Study, Nagoya University, Nagoya 464, Japan

Received: July 29, 2015

Accepted: August 14, 2015

Available online: August 19, 2015

doi:10.11114/ijsss.v3i5.1045

URL: <http://dx.doi.org/10.11114/ijsss.v3i5.1045>

## Abstract

Several studies have shown that variability in weather has been linked to stroke occurrence. However, the association has not been clear. The purpose of this study was to determine the effect of the change of weather pattern on stroke onset using Hidden Markov Models (HMMs). The daily data of emergency transport from 2002 to 2005 in Nagoya City in Japan were used. As "weather patterns", we used 11-classes classified by Japan Weather Association (JWA), including "East high pressure and West Low pressure", "Two centers of Low pressure" and so on. We proposed a new type of application of HMMs. Normally, HMMs were used to predict a hidden state from observed data. In this paper, HMMs were applied to show the existence of influence of weather variability on stroke occurrence. Our HMM was compared with the stochastic process which ignores the influence of weather states. A statistical test led to conclude that weather influenced the occurrence of stroke. These findings suggest that further research is possible to give more detailed study of the relation between weather states and stroke occurrence, if we use more specified data of weather elements such as daily temperature, sea level pressure and so on.

**Keywords:** cerebral infarction; stroke; biometeorology; hidden Markov process; weather; weather pattern; stochastic process; Markov process

## 1. Introduction

Many observational studies were devoted to assess the links between changes in weather and disease occurrences. Among diseases, stroke (or cerebral infarction) was one of the commonly known targets to be discussed. However, the links stayed unclear (Field et al. 2002, McDonald et al. 2012, Biller et al. 1988, Jimenez-Conde et al. 2008). Many studies used regression model to identify quantitatively a threshold of weather elements, such as temperature, humidity and so on, against onset of cerebral infarction (Chen, 1995, Ebi 2004). Regression models were quantity models and were too critical to determine whether diseases were influenced by weather or not.

The purpose of this study was to assess the effect of the change of weather pattern on the onset of cerebral infarction, providing a new approach based on Hidden Markov Models (HMMs). Normally HMMs were used for forecast, but here in this paper, we used them to assess the existence of the influence of weather states on occurrence of cerebral infarction.

We defined two types of stochastic models: one is HMM with consideration on weather change, the other is a simple MM (Markov Model) with no regards to weather change. These two methods were compared.

Our HMMs were described in Section 2 (Method section). Section 3 (results section) evaluated the simulation of HMM and MM. In Section 4 (discussion section) we discussed the differences of HMM and MM.

## 2. Data and Method

The daily data of emergency transport from 2002 January up to 2005 December in Nagoya City in Japan were used. The data contained the number of patients of many diseases other than stroke who were transported by ambulance to some hospitals in Nagoya City for each day.

The data of "weather pattern" were extracted from Japan Weather Association (JWA), which were called "11-types of weather patterns" by JWA. Typical patterns of these 11-types were the pattern of "high pressure", "a cold front" and so on. All the patterns were listed in Table 1.

Table 1. 11 classes of weather patterns

Weather state	
1	North high pressure and South low pressure
2	East high pressure and West low pressure
3	South high pressure and North low pressure
4	Strong West high pressure and East low pressure
5	West high pressure and East low pressure
6	High atmospheric pressure
7	Weather front
8	Typhoon
9	Strong low pressure
10	Weak low pressure
11	Two centers of low pressure

We constructed our database collecting the above data together. Among them the data from 2002 January up to 2004 December were used for model development (training data), and those of 2005 were used for model evaluation (test data).

In this paper, we used Hidden Markov Models (HMMs) for modelling the occurrence of cerebral infarction associated with the changes of weather patterns. HMMs were a well-known tool for models which incorporated randomness, and were applied in numerous fields. For example, statistic models for precipitation were used to relate atmospheric circulation to local rainfalls (Hughs and Gutorp 1994). They proposed a HMM over "weather states" which brought out rainfall occurrences as observed data.

Here in this paper, we proposed to develop a new type of application of HMMs, not only to make a model for random series of observed data (i.e., the number of patients of cerebral infarction) but also to prove the existence of links between risk of cerebral infarction and the change of weather patterns.

In general, HMM was a tool for representing random process over sequences of observations. The observation at time  $t$  was represented by the variable  $R_t$ . A hidden Markov model consisted of two random processes. First, it assumed that the observation  $R_t$  at time  $t$  was generated randomly by some process whose state  $S_t$  was hidden from the observer. Second, it assumed that the state  $S_t$  was determined randomly from the state  $S_{t-1}$ . Both random processes were assumed to be Markov process (See Figure1). For basic elements of HMM, see Ghahramani, Z (2001).

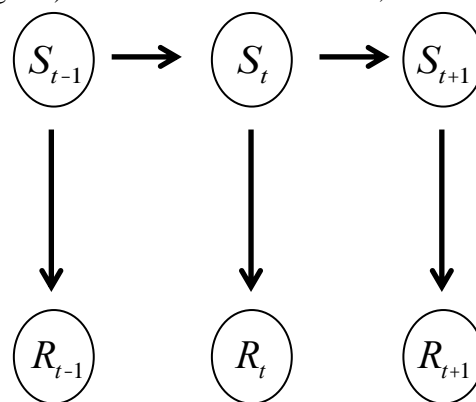


Figure1. Sequences of states and observations in a hidden Markov model

We considered the set of "weather states" as hidden states  $S_t$  and supposed each weather state resulted stroke occurrences  $R_t$  as observation data. The weather patterns used in this paper were those proposed by "Japan Weather Association"(JWA). They suggested that the weather patterns in Japan were classified as the following 11 classes in Table 1. These classes were abbreviated as  $S_1, S_2, \dots, S_{11}$ . Each weather pattern was called "weather state" in this paper.

We defined S to be the set of all these 11-classes

$$s_1, s_2, \dots, s_{11}$$

and called it the set of "states". They were numbered as 1,2,...,11 by Japan Weather Association(JWA), therefore the set S could be considered as the set of numbers from 1 to 11. A weather state in S was supposed to change to another weather state in S daily. Let  $S_t$  be the weather state at time  $t$ .

Let  $R_t$  be the number of persons (the number of risk) who were taken to hospitals by ambulance at time  $t$  in Nagoya City. Thus we have two kinds of random series  $\{S_t\}$  and  $\{R_t\}$ .

We then calculated the probability  $P_{ij}$  from one weather state  $s_j$  to another weather state  $s_i$ ,

$$P_{ij} = P(S_t = s_i | S_{t-1} = s_j)$$

extracting the daily data of 11-classes from 2002 January to 2004 December. This gave us 11x11 number of probabilities, thus forming 11 x 11-matrix

$$P = (P_{ij}), 1 \leq i \leq 11, 1 \leq j \leq 11.$$

This matrix was called "transition probabilities".

We defined the probability of the occurrence  $R_t$  of ci (cerebral infarction) for given weather state  $S_t$  as follows:

$$P(R_t | S_t)$$

Let  $m$  be the maximum number of  $R_t$  during the period considered. Then  $R_t$  takes the value from 0 to m, and  $S_t$  from 1 to 11. Let  $R$  be the set of numbers from 1 to m. The above probability forms together  $(m + 1) \times 11$ -matrix:

$$Q = (q_{ij}), 1 \leq i \leq m + 1, 1 \leq j \leq 11$$

$$q_{ij} = P(R_t = i - 1 | S_t = s_j)$$

Each column of the above matrix gives the distribution of the occurrence of ci for given weather state.

As a consequence, we have a set  $\{S, R, P, Q\}$

$$S = \{s_1, \dots, s_{11}\} = \{1, \dots, 11\}$$

$$R = \{r_1, \dots, r_m\} = \{1, \dots, m\}$$

$$11 \times 11\text{-matrix } P$$

$$(m + 1) \times 11\text{-matrix } Q$$

This set was called a Hidden Markov Model. We calculated these two matrices P and Q from the data from 2002

January to 2004 December in the next section.

**3. Results**

We defined, in the previous section, a Hidden Markov Model  $\{S, R, P, Q\}$ :

$$S = \{s_1, \dots, s_{11}\} = \{1, \dots, 11\}$$

$$R = \{r_1, \dots, r_m\} = \{1, \dots, m\}$$

11 × 11- matrix  $P$

$(m + 1) \times 11$ - matrix  $Q$ .

We calculated these two matrices P and Q from the data from 2002 January to 2004 December. Data were daily series of the occurrences of ci and daily series of weather states. From this data we counted the probability that weather state  $S_j$  changed to  $S_i$  as follows. First fix j, then we counted the numbers of the dates where the states changed from  $S_j$  to  $S_i, i=1, \dots, 11$ .

$$S_j \rightarrow S_1, \dots, S_j \rightarrow S_{11}$$

Dividing these numbers by the total number of occurrence of  $S_j$ , we obtained the distribution function from state  $S_j$  to another state. This gave the j - th column of the matrix P as in Table 2.

Table 2. Transition Matrix P of Weather States

Weather type (next day)	Probability (current day)										
	1	2	3	4	5	6	7	8	9	10	11
1	0.26	0	0.04	0	0.05	0.07	0.08	0.03	0	0.14	0.07
2	0.15	0.26	0.02	0	0.02	0.1	0.02	0.03	0	0	0
3	0	0.04	0.2	0	0.05	0.05	0.03	0	0	0.05	0
4	0	0.07	0.02	0.25	0.04	0.02	0.02	0.07	0.38	0.08	0
5	0.05	0.04	0.12	0.58	0.41	0.07	0.1	0.13	0.13	0.27	0.6
6	0.31	0.14	0.18	0.12	0.37	0.44	0.26	0.16	0.19	0.17	0.27
7	0.06	0.21	0.27	0.02	0.02	0.12	0.32	0.19	0.25	0.17	0
8	0.05	0.06	0.04	0	0	0.01	0.03	0.36	0	0.01	0
9	0.02	0.01	0	0	0.01	0.03	0.01	0	0	0	0
10	0.1	0.17	0.1	0.02	0.02	0.09	0.12	0.03	0.06	0.1	0
11	0.01	0	0	0.02	0.02	0.02	0.01	0	0	0.01	0.07

Recall that Q was defined by the  $(m + 1) \times 11$ -matrix:

$$Q = (q_{ij}), 1 \leq i \leq m + 1, 1 \leq j \leq 11$$

$$q_{ij} = P(R_t = i - 1 | S_t = s_j),$$

where each column  $j$  of the above matrix gives the distribution of the occurrence of  $c_i$  for given weather state  $s_j$ .

Fix  $j$  and suppose that the weather state is  $S_t = s_j$  at time  $t$ .

From the daily data from 2002 up to 2004, select only data, the weather state of which is  $s_j$ . Calculating the distribution of the occurrence of  $c_i$  of these data, we obtained the  $j$ -the column of the matrix  $Q$ . Changing  $j$ , we had the whole  $Q$  as in Table 3. This defined the Hidden Markov Model  $\{S, R, P, Q\}$ .

Table 3. Result of the matrix  $Q$  of 14 rows and 11 columns

	11 classes of weather state										
the number of patients	1	2	3	4	5	6	7	8	9	10	11
1	0	0	0	0	0	0.01	0.01	0.03	0.06	0.02	0
2	0.07	0.08	0.1	0.06	0.02	0.05	0.05	0.13	0	0.03	0
3	0.12	0.08	0.12	0.04	0.07	0.1	0.15	0.06	0.13	0.11	0.13
4	0.15	0.19	0.12	0.06	0.12	0.16	0.16	0.29	0.13	0.18	0.13
5	0.15	0.26	0.1	0.17	0.21	0.22	0.19	0.13	0.25	0.17	0.13
6	0.12	0.11	0.24	0.25	0.21	0.15	0.18	0.13	0.13	0.16	0.06
7	0.11	0.14	0.06	0.08	0.15	0.14	0.09	0.06	0.19	0.1	0.31
8	0.11	0.1	0.1	0.15	0.08	0.07	0.04	0.06	0.06	0.11	0.06
9	0.11	0	0.1	0.08	0.05	0.04	0.07	0.06	0	0.03	0.06
10	0.04	0.03	0.04	0.06	0.03	0.03	0.01	0.03	0	0.03	0.06
11	0.01	0	0	0	0.02	0.01	0.04	0	0	0.01	0
12	0.01	0	0	0.06	0.01	0.01	0.01	0	0	0.01	0
13	0	0	0	0	0	0	0.01	0	0	0	0.06
14	0	0	0	0	0	0	0	0	0.06	0.01	0

For the evaluation of our HMM, we applied it to the daily data of 2005. The original data gave us real data of occurrences of  $c_i$ ,

$$c_1, c_2, \dots, c_T,$$

and we compared these real data series with data series produced by our HMM,  $\{S, R, P, Q\}$ .

We selected arbitrary month in 2005 and selected one day (for example, 1-st day or 15-th day and so on) as the initial value of HMM. We used also the weather state of the initial day as the initial value of HMM. The Figure 2 showed a typical sample of simulation of HMM during one month (November 2005).

30 days from 2005/11/1

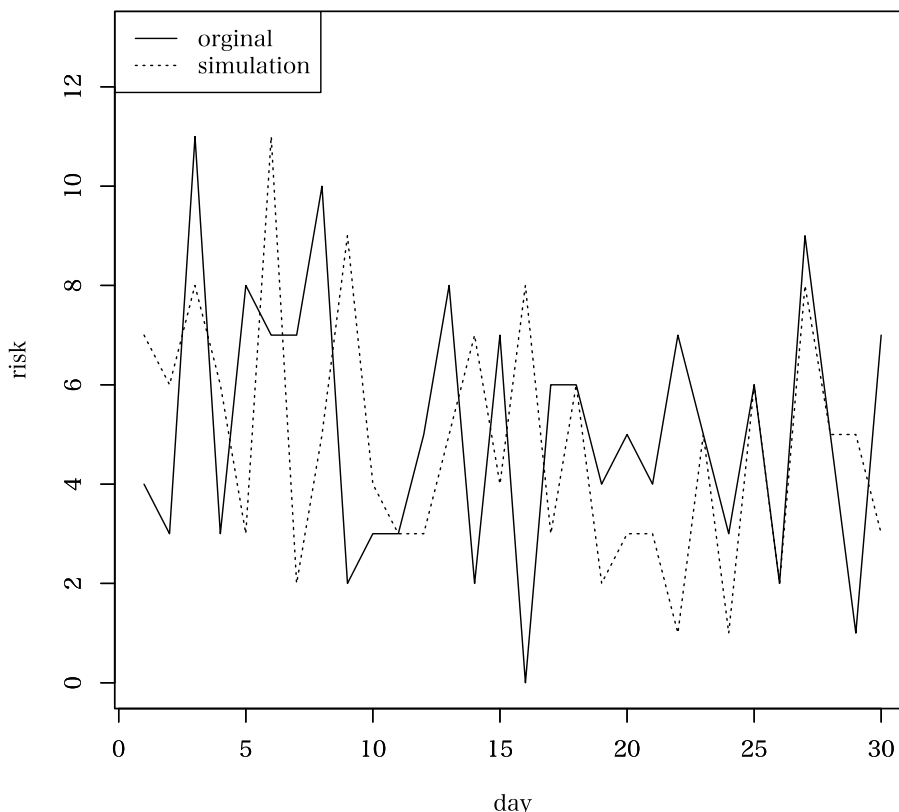


Figure 2. An example of simulation result of HMM and original data for November 2005.

We proceed our HMM for these initial values for the period of  $T = 3$  days. Then we had a time series of expected ci (expected ci-occurrence)

$$r_1, r_2, \dots, r_T$$

which could be compared with the original data.

$$c_1, c_2, \dots, c_T$$

For this comparison, we used the method of Root Mean Squared Error (RMSE).

In order to argue the existence of influences of weather states to the occurrences of ci, we constructed a simple stochastic model (MM: Markov Model) which did not use the information of weather states. This MM depended only on the distribution of the occurrences of ci.

We had the time series of real ci-occurrences from the original data from 2002 to 2004. Gathering these 3-year data gave us the distribution of ci-occurrences. That is the function of the integers,  $0, 1, \dots, m$  (maximum of ci-occurrences during three years) taking values of the probability. This simple model was described by a distribution function

$$f : [0, m] \rightarrow R$$

For given month, a simulation could produce a time series obeying to this  $f$ ,

$$s_1, s_2, \dots, s_T$$

starting from the same day like  $r_1, r_2, \dots, r_T$  and  $c_1, c_2, \dots, c_T$ .

Then we calculated RMSE for this series  $s_1, s_2, \dots, s_T$  compared with the original  $c_1, c_2, \dots, c_T$ . We totally obtained

RMSEs for  $r_1, r_2, \dots, r_T$  and  $s_1, s_2, \dots, s_T$  for every month of 2005 as in Table 4.

Whether the difference of these two RMSEs was statistically significant or not had been investigated by t-test. The p-values of t-test were described in the Table 4. That showed that samples generated by HMM and MM were statistically different for most of months. This implied that the influence of variability of weather states existed with respect to the occurrence of cerebral infarction.

**4. Discussion**

To develop a HMM, we classified weather states into 11-states, and calculated a transition matrix, which was composed of probabilities of change of a weather state to another state. We also calculated the distribution functions of ci-occurrence, given arbitrary weather state.

For the evaluation of our HMM, we applied it to the daily data of 2005. The original data gave us real data of occurrences of ci, and we compared these real data series with data series produced by our HMM, {S, R, P, Q}. We selected arbitrary month in 2005 and selected one day as the initial value of HMM (such as 1-st day or 15-th day).

For given month, a simulation of our HMM could produce a time series  $r_1, r_2, \dots, r_T$ . We also constructed a simple Markov Model (MM), which excludes the influence of weather change. Starting from the same day, suppose that this

MM produced  $s_1, s_2, \dots, s_T$ . Let the original ci-occurrence to be  $c_1, c_2, \dots, c_T$ . Then we calculated RMSE for series  $c_1, c_2, \dots, c_T$  and  $r_1, r_2, \dots, r_T$  compared with the original  $c_1, c_2, \dots, c_T$ . The results were given by the Table 4.

The statistical significance was investigated by applying t-test, and results were in the same table.

Table 4. RMSE (Root Mean Squared Error) of MM and HMM and p-value of t-test applied to the data of 2005

month	1	2	3	4	5	6	7	8	9	10	11	12
MM	2.64	3.20	2.77	2.62	3.30	2.31	2.83	3.24	2.51	3.70	2.33	2.39
HMM	2.56	3.13	2.89	2.59	3.24	2.24	2.86	3.17	2.48	3.72	2.21	2.48
p.value	0.09	0.11	0.00	0.45	0.22	0.08	0.53	0.14	0.41	0.71	0.00	0.05

From this t-test, we could conclude that the model HMM (with weather states) was better than the model MM (without weather states), by comparing RMSE. This fact was proven to be stochastically significant by t-test for RMSE. The general conclusion is that cerebral infarction is influenced by weather

June is a rainy season in Japan. In November, the weather is relatively good, but there are some days when the sudden cold wind rages over Nagoya City. These months could be considered as typical months when the change of weather state induces the risk of ci.

The result of March did not support our theory. This might be due to the fact that March is a typical month when the season changes from winter to spring. The weather state changes too quickly that our theory could not follow the variability of ci well.

December seemed critical in that more experiments by simulations showed sometimes that December fitted our theory.

Since we focused our efforts on the effect of variability of weather patterns, the following question may arise. Which pattern of weather change is threatening our health? In order to assess this question, we calculated the means of the number of patients of ci (transported by ambulance during 2002 up to 2004) with respect to all types of change of weather pattern. The results were listed in Table 5, where only the cases with means higher or equal to 7 were listed, associated with change of weather patterns. Most of cases happened when a weather pattern changed to low pressure type. Other cases happened when a pattern of pressure changed significantly.

Table 5. Pattern of weather change and means of patients of ci transported by ambulance (per day).

weather of previous day	weather of current day	mean of patients
North High, South Low	East High, West Low	7
East High, West Low	Strong West High, East Low	7
East High, West Low	Weak Low pressure	7
South High, North Low	Weak Low pressure	7
Strong West High, East Low	High pressure	8
West High, East Low	East High, West Low	9

West High, East Low	Strong Low pressure	9
West High, East Low	Two centers of Low pressure	9
High pressure	North High, South Low	11
High pressure	Strong Low pressure	7
High pressure	Weak Low pressure	7
Strong Low pressure	Strong West High, East Low	8
Weak Low pressure	Strong West High, East Low	7

Discription: “North High, South Low” means “North high pressure and South low pressure” and so on.

Our theory argued the dependency of the risk of ci on variability of weather patterns. But there is a risk of ci-occurrence, even when weather pattern does not change at all. When the weather is stable, then combining our theory with other method (such as regression model, decision tree or support vector machine and so on) will be useful to develop more detailed analysis in future.

We assumed that the state of this hidden process satisfied the Markov property: that is, given the value of  $S_{t-1}$ , the current state  $S_t$  is independent of all the states prior to t-1. This rule does not hold in general. For example, the weather state of two days before current day might affect the weather state of the current day. In such cases, we only have to change the daily data of states to a series of couples or triples of weather states. Then the assumption of Markov property would hold.

Our theory did not use weather elements, such as temperature and so on. In this meaning, our theory appears very rough, but this means that we could develop further detailed theory, taking weather elements in consideration. Given the weather pattern, the detailed analysis of weather elements may be possible in order to forecast the occurrence of cerebral infarction (cf., Kojima et al 2014).

## References

- Chen, Z. Y., Chang, S. F., & Su, C. L (1995). Weather and Stroke in a Subtropical Area: Ilan, *Taiwan Stroke*, 26,569-572. <http://dx.doi.org/10.3174/ajnr.A3059>
- Ebi, K. L., Exuzides, K. A., Lau, E., & Kelsh, M. (2004). ‘Weather changes associated with hospitalizations for cardiovascular diseases and stroke in California, 1983–1998’, *Int J Biometeorol*, 49, 48–58. <http://dx.doi.org/10.1007/s00484-004-0207-5>
- Field, T. S., & Hill, D. (2002). *Weather, Chinook, and Stroke Occurrence’ Stroke*, 33, 1751-1758. <http://dx.doi.org/10.1161/01.STR.0000020384.92499.59>
- Ghahramani, Z. (2001). ‘An Introduction to Hidden Markov Models and Bayesian Networks’, *Journal of Pattern Recognition and Artificial Intelligence*, 15 (1), 9 - 42. <http://dx.doi.org/10.1142/S0218001401000836>
- Hughes, J. P., & Guttorp, P (1994). ‘A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena’, *Wat. Resour. Res.*, 30, 1535-1546. <http://dx.doi.org/10.1029/93WR02983>
- Hughes, J. P., & Guttorp, P. (1999). ‘A non-homogeneous hidden Markov model for precipitation occurrence’, *Appl. Statist*, 48(1), 15-30. <http://dx.doi.org/10.1111/1467-9876.00136>
- Jimenez-Conde, J., Ois A., Gomis, M., Rodriguez-Campello, A, Cuadrado-Godia, E, Subirana, I., & Roquer, J. (2008). ‘Weather as a Trigger of Stroke’, *Cerebrovasc Dis.*,26, no.4, pp.348–354. <http://dx.doi.org/10.1159/000151637>
- Kojima, M., Ogasawara, N., & Morimoto, H. (2014). ‘Relationship between onset of cerebral in-farction and classification of meteorological elements’, *Open Journal of Applied Sciences*, 4, 242-248. <http://dx.doi.org/10.4236/ojapps.2014.45024>
- McDonalda, R. J., McDonaldb, J. S., Bidae, J. P., Kallmesb, D. F., & Cloft, H. J. (2012). ‘Subarachnoid Hemorrhage Incidence in the United States Does Not Vary with Season or Temperature’, *Am. J. Neuroradiol*, 33, 1663-1668. <http://dx.doi.org/10.3174/ajnr.A305>



This work is licensed under a [Creative Commons Attribution 3.0 License](https://creativecommons.org/licenses/by/3.0/).