**Final Report EAACI Research Fellowship 2019 Yi-Ting Tseng**

**Predicting the severity of the next pollen season by using a "Hidden Markov Model"**

Research Fellow: Yi-Ting Tseng
Type of Fellowship: Short term Research Fellowship
Host supervisor: Prof. Jeroen Buters
Location: Zentrum Allergie und Umwelt (ZAUM),
Technische Universität München (TUM) und
Helmholtz Zentrum München
Duration: 3 months (mid-January 2019 – mid-April 2019)

Prof. Buters and I in front of the pollen indicator.


**Abstract of the research fellowship**

This research fellowship aimed to use a state-space model, Hidden Markov Model (HMM), to predict next year´s birch pollen season intensity. The model worked well in countries where birch trees showing significant masting behavior, but did not work well in Munich, Germany. We assume that this is due to a different innate flowering rhythm of birch trees in more southern locations like Munich. However, a SVM model (Support Vector Machine) with less parameters (only meteorological factors in previous summer) better explained the annual change in SPI. These results are now being prepared for publication.

Besides the research activity on pollen in Munich, I also visited two other labs in Europe known for their expertise in pollen and aerosols, Dr. Athanasios Damialis from UNIKA-T (TUM) and Dr. David Topping from Manchester University. I also got a rare chance to join the installation of an improved pollen sampler in the research station on the highest mountain of Germany (2650m a.s.l.), plus getting ample exposure to a different culture by lab activities and other extracurricular activities.

The stay proved to be very fruitful scientifically as we were able to develop a model for Munich, Germany to predict next year birch pollen load from this year climatic parameters. It also broadened my horizon to other cultures.

## 1. Introduction

### 1.1 Background

Allergenic pollen has a remarkable impact on human health, provoking pollinosis in sensitized patients with a high prevalence of 10–30% in the global population [1]. In addition, pollinosis is thought to be on the rise because of two main reasons: i) the interrelation and synergistic effect with other aerosols (biotic and abiotic ones) ii) changes in the seasonal timing of pollen season and increased allergen exposure resulted from climate change. The Seasonal Pollen Index (SPI), integral of pollen concentration over pollen season [2], is useful for determining the overall dispersal outcome in each year. It provides an important reference for clinical strategies and anti-allergy medicine production given that the severity of symptoms is closely related to allergenic pollen concentration [3-5].

### 1.2 Research problem

SPI is known to change greatly from year to year. Hence, accurate SPI prediction for next year would thus be of great value to allergy stakeholders. Previous studies made effort on predicting SPI by determining statistical relationships between observed pollen data and observed meteorological factors since the quantity of male flowers is closely related to meteorological variables from the previous summer [6, 7]. Multiple linear regression is the most common approach [6, 8-10], which is followed by time series analyses [10] , and other computational intelligence approaches [11]. In recent years, more studies also took physiological aspects into account for attaining better forecast for anemophilous trees considering the phenomena "masting behavior" (referring to both highly variable interannual flower and fruit production and intra-annual production synchrony within a population). However, considering the uncertainty in spatial representativeness of SPI and intrinsically stochastic natural in SPI time series, it is difficult to reproduce the intermittent peaks of SPI only with few environmental factors. Therefore, we proposed an unprecedented approach, a state space model, that allows us to account for uncertainties in the input data (SPI change and meteorological data) and recurrence of mast year to predict the future state via previous state and observation.

### 1.3 Aim of the project

This study proposed an unprecedented approach to predict SPI using stochastic network Hidden Markov Model (HMM). It considers the uncertainties in input data and it is able to capture the sequential transition characteristics of masting behavior and the relationship between the amount of male flower and the meteorological conditions in the previous summer through stochastic aspects.

In this project, we aimed to (1) train a HMM model with parameters describing the effects of masting behavior and the interrelationship between the meteorological conditions in the previous summer and the SPI and (2) construct a more generalized model to predict the SPI.

## 2. Work program and results


Fig. 1 Analysis process

The detailed information of each step of work process (Fig. 1) is as below:

**i) Data input**

The 24 years' input dataset from Munich, Germany contained the objective variable (Y) and explanatory variables (X).

Y: SPI data of birch pollen (1995-2018)

X: Meteorological data including minimum air temperature, mean air temperature, maximum air temperature, precipitation, relative humidity, sunshine duration, concentration of $CO_2$ in the previous summer (1994-2017)

The whole dataset was split into training a dataset (1995-2015) and a test dataset (2016-2018).

**ii) Feature selection**

Instead of using simple monthly meteorological variables, we used the average of each variable during optimized agglomeration periods preceding the flowering year. The optimal agglomeration periods were specified as the periods of occurrence of the maximum positive and minimum negative correlation coefficients of the meteorological variable and SPI, by moving the start date with changing window (the number of days). The start date moved from the 1st to the 30th of May, June, and July, and the moving window expanded from 14 to 30 days.

**iii) Dimension reduction**

In order to decompose the features into lower dimension without significant loss in variance, Principal Component Analysis was implemented. 12 features from feature selection step were decomposed into five dimensions with ability to explain ~90% variance in original data.

**iv) Prediction model**

The projected features in training data were then fit to the HMM to predict the season intensity. In addition, sensitivity analysis was carried out to test how the performance of the model changes when the criteria of each pollen level changes. Another approach, the Support Vector Machine (SVM) was implemented for performance comparison.

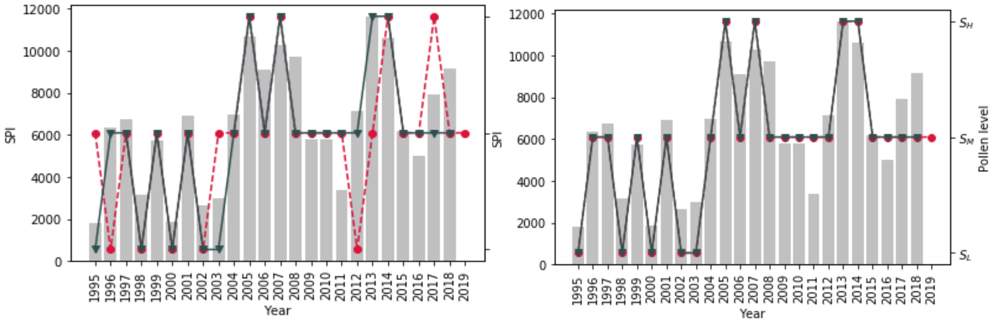


Fig. 2 Prediction results of annual change in SPI in Munich (left: HMM, right: SVM).

We assumed the lower performance of HMM was because of no obvious masting behavior found in Munich. Hence, the other model SVM with smaller size of parameters gave a better explanation for annual change in SPI. These results are planned to be published in the near future with notation of EAACI's research fellowship project.

## 3. Acknowledgement

First, I would like to thank to Dr. Jeroen Buters for providing the valuable collaborative study chance in Munich. I realized further the importance of my study by being in an environment where people are more (or also) aware of the pollinosis issue than in Asia. Every time passing by the pollen indicator, I deeply get motivated to probe into the question "How to make a better prediction". Also, special thanks to Dr. Jose Oteros for giving me considerable useful technical suggestions for improving the prediction models. During this stay, I really got considerable kind support from them and the team members in study, living, and experiencing different cultures.

Settled the improved Hirst-type sampler in Zugspitze.

Experienced "schlittenfahren" for the first time in my life with members in Dr. Buters' lab.

Second, I would like to thank to my professor Dr. Shigeto Kawashima for encouraging me to achieve international short-term research fellowship experience during my doctoral course. His kind support also strengthens my passion in solving the allergy-related problems.

Third, I would like to thank to Dr. Athanasios Damialis from UNIKA-T (TUM) for giving me substantial introduction of the studies and Dr. David Topping from Manchester University for giving me advice for algorithm. Based on visits in host institutes and the two, I realize the common trends in aerobiology: (i) interrelation and synergistic effect of airborne pollen, spore, and air pollutants (ii) requirement for better aerosol samplers and data quality (iii) climate change impacts (iv) "smartilize" the process from pollen data acquirement to pollen prediction or even allergic reaction prediction by combining personal data.

Finally, I would like to thank to EAACI Headquarters for giving me the opportunity to implement the collaborative study in international research environment. This stay fulfilled my doctoral course and encouraged me continuing the study for better quality of life of human.

## 5. References

1. World Allergy, O., *The World Allergy Organization white book on allergy: 2013 update*. 2017.
2. Galan, C., et al., *Recommended terminology for aerobiological studies.* Aerobiologia, 2017. **33**(3): p. 293-295.
3. Bastl, K., et al., *First comparison of symptom data with allergen content (Bet v 1 and Phl p 5 measurements) and pollen data from four European regions during 2009–2011.* Science of The Total Environment, 2016. **548-549**: p. 229-235.
4. Brito, F.F., et al., *Olea europaea pollen counts and aeroallergen levels predict clinical symptoms in patients allergic to olive pollen.* Annals of Allergy Asthma & Immunology, 2011. **106**(2): p. 146-152.
5. D'Amato, G., et al., *Allergenic pollen and pollen allergy in Europe.* Allergy, 2007. **62**(9): p. 976-990.
6. Ranta, H., et al., *Male flowering of birch: Spatial synchronization, year-to-year variation and relation of catkin numbers and airborne pollen counts.* Forest Ecology and Management, 2008. **255**(3): p. 643–650.
7. Yasaka, M., et al., *Prediction of birch airborne pollen counts by examining male catkin numbers in Hokkaido, northern Japan.* Aerobiologia, 2009. **25**(2): p. 111-117.
8. Latałowa, M., M. Miętus, and A. Uruska, *Seasonal variations in the atmospheric Betula pollen count in Gdańsk (southern Baltic coast) in relation to meteorological parameters.* Aerobiologia, 2002. **18**(1): p. 33-43.

9. Subiza, J., et al., *Prediction of annual variations in atmospheric concentrations of grass-pollen - a method based on meteorological factors and grain crop estimates.* Clinical and Experimental Allergy, 1992. **22**(5): p. 540-546.
10. Garcia-Mozo, H., et al., *Statistical approach to the analysis of olive long-term pollen season trends in southern Spain.* Science of the Total Environment, 2014. **473**: p. 103-109.
11. Zhang, Y., et al., *Bayesian analysis of climate change effects on observed and projected airborne levels of birch pollen.* Atmospheric Environment, 2013. **68**(Supplement C): p. 64-73.