EUROPEAN CONSORTIUM FOR MATHEMATICS IN INDUSTRY

ECMI Modelling Week 2018
University of Novi Sad
Department of Mathematics and Informatics, Faculty of Sciences

FORECASTING ALLERGEN'S CONCENTRATIONS

Instructor: Marko Nedeljkov (University of Novi Sad)

Nataša Tešić (University of Novi Sad)
Joanna Śmieja (Wrocław University of Science and Technology)
Samuel Medina Leva (Autonomous University of Barcelona)
Jenna Ritvanen (Lappeenranta University of Technology)
Florian Beiser (University of Kaiserslautern)

14 October 2018

Contents

1 Introduction

Common ragweed (Ambrosia artemisiifolia L.) is a highly allergenic plant, spread throughout Serbia (Šikoparija et al. 2009). As many people suffer from allergic reactions to its pollen and need to adjust their medication, there is an urgent need of predicting pollen concentration, both on a daily basis, and in the long-term run (Prank et al. 2013).

To make a realistic model, physiology and phenology of the plant has to be taken into account. It is an invasive species, that originates from North and Central America (CABI 2018). Its seeds can terminate in the ground for a few years, waiting with sprouting for a year with better weather conditions. Therefore a lower number of plants in the previous season does not have to result in a lower number of plants in the next season (Willemsen 1975). In different regions of the world different aspects of the weather conditions seem to be important when regarding the pollen levels (Franz et al. 2015, Deen, Hunt, and Swanton 1998,Prank et al. 2013, CABI 2018).

We will propose a model of Ambrosia pollen concentration prediction, concerning the area of Novi Sad, Serbia. The data analyzed comes from one measuring station in Novi Sad and covers years 2000-2017. It complies of concentration values from every two hours and daily weather conditions (temperature, wind, humidity, etc.).

The paper is divided as follows. First, the possessed data of pollen concentration and weather condition will be analyzed using statistical methods (section 2.1). Based on it, the long-term model (prediction of the whole season, its offset, peak, end and general curve) will be proposed (section 2.2). Next, a PDE model simulating the concentration of pollen over the area of Novi Sad in the next 24 hours will be described (section 2.3). Here are used the results from previous sections and the exact locations of the plant in the city. The work is then summarized in section 3.

2 Report content

2.1 Data analysis

First of all, it is important to know which are the data that we used in order to create our models. We have collected data of pollen concentration from 2002 to 2015 and we found the meteorological data corresponding to that period. With the described data, we tried to understand the relation between the pollen concentration and the most important meteorological parameters (as temperatures, rain or wind).

The pollen concentration of each year and the average one (in red) were represented and it is seen we have approximately a Gaussian, so that model is going to be used in the next section as the base curve of the distribution of the pollen concentration.

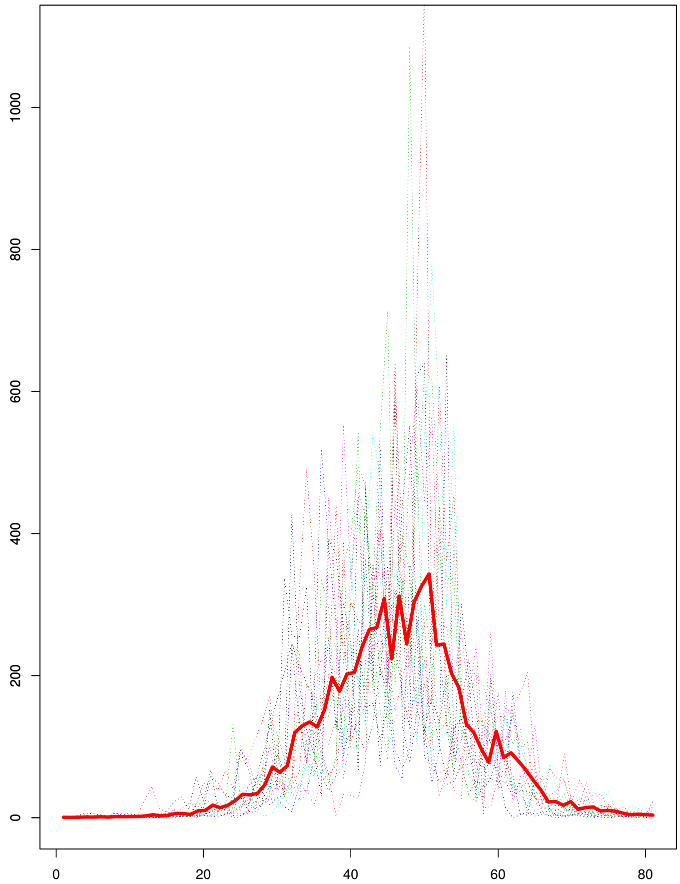


Figure 1: Pollen concentration each year and the average.

Then, it is interesting to see the representation of the rain in a year and understand if some relation exists between that and the pollen concentration. All the years in the period 2002-2015 were represented as we see in the results to 2003, which are shown in the figure 2. When the results were compared with pollen concentration we saw that, basically, the amount of rain the 4 days before the start of the season is the most important to describe the amount of pollen.

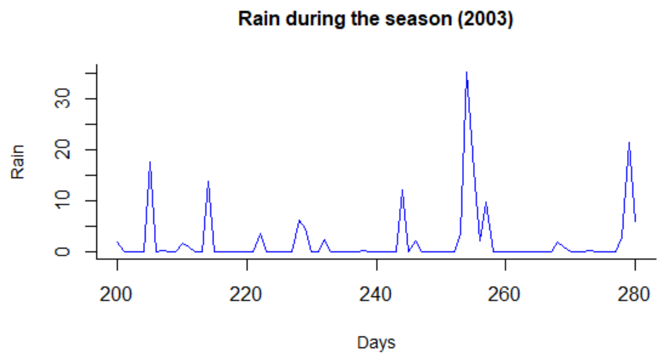


Figure 2: Rain in 2003.

The effect of other weather events is described in the next section when the long term model is explained and determined, but it is important to say that the way to analyze them was the same (graphic representation of each year and comparing with the pollen concentration).

Last, the night temperatures were considered, and it was seen (figure 3) that when there were lower night temperatures the average concentration of pollen was low too, so it is important to include that fact in our long term model.

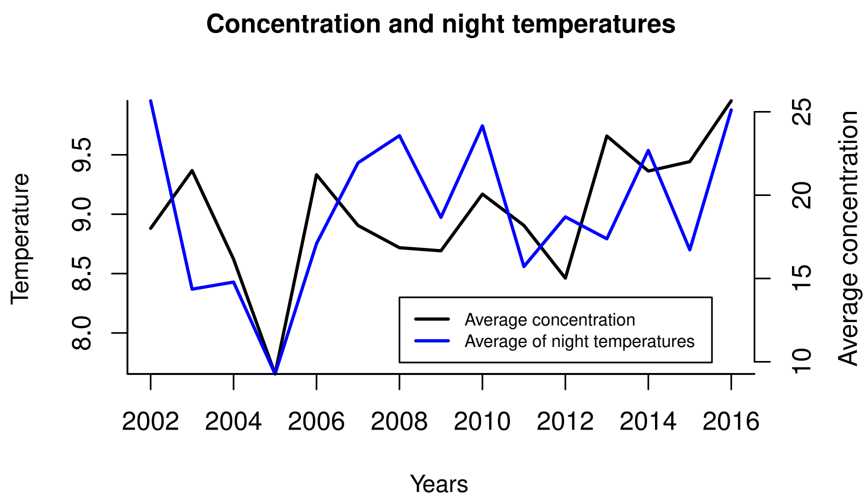


Figure 3: Night temperatures.

2.2 Long term model

The aim of the long term model is to predict the average daily pollen concentration during the main pollen season in the measurement point. Based on the data analysis described in the previous section, the basic form of the model was chosen to be multiplicative, consisting of a base curve and perturbations caused by weather events.

The base curve was first fitted as a Gaussian curve to the average daily concentrations. As concluded in the data analysis, the mean of the fitted curve depends on the total amount of rain during the four days before the start of the season. The fitted function for the mean is written as

$$\mu(r) = 45.8 - 0.2 \sum_{i=1}^{4} r_i \tag{1}$$

where r_i is the amount of rain on the ith day from the start of the season. The standard deviation of the Gaussian was approximated directly from the data, and the acquired value was $\sigma = 9.212$.

The scaling of the base curve was determined to depend on the number of night frosts during the growing season. The function describing the dependency was found to be

$$s(N) = 8500 + 200N \tag{2}$$

where N is the number of night frosts during the growing season. Usually it is assumed that the plant is sensitive to frosts only in May; however, since most of the night frosts during the growing season occur in May, the total amount of night frosts was considered.

Therefore the base curve can be written as

$$BC(i) = s(N)\,\mathcal{N}(\mu(r), \sigma) \tag{3}$$

where i is the number of the day from the beginning of the season.

After fitting the base curve the effect of the weather events to the average pollen concentration was modeled. The negative effect of rain on the pollen concentration, as a proportional decrease, in was fitted qualitatively to be

$$W_r(r_i) = \begin{cases} -\min\{0.9, 0.2 + 0.05r_i\} & r_i > 0 \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

The positive, delayed effect of temperature drops as a proportional increase was described as

$$W_T(\Delta T_i) = \begin{cases} \min\{-100\Delta T_{i-1}, M(i)\} & (\Delta T_{i-1} \leq -2) \quad \text{and} \quad BC(i) > 30 \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where $M(i)$ is the curve fitted to the daily maximum values of pollen concentration, given by

$$M(i) = s(N)\,\mathcal{N}(\mu(r), 1.2\sigma) \tag{6}$$

Similarly, the positive effect of wind on the pollen concentration was described as

$$W_w(w_i) = \begin{cases} 0.20 + 0.02w_i & w_i \geq 4 \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

Finally, the long term model can be written as

$$WE(i) = [W_r(r_i) + W_T(\Delta T_i) + W_w(w_i)]\,BC(i) \tag{8}$$

The model was implemented using the Matlab mathematical computing software (MATLAB 2017).

Figures 4 and 5 show the predictions of the long term model on two pollen seasons compared to the actual concentrations. The figures also show the significant weather events. The year 2006 was selected to be shown since it is the only year in the data range that shows the effect of wind; the year 2003 was selected as it has a higher amount

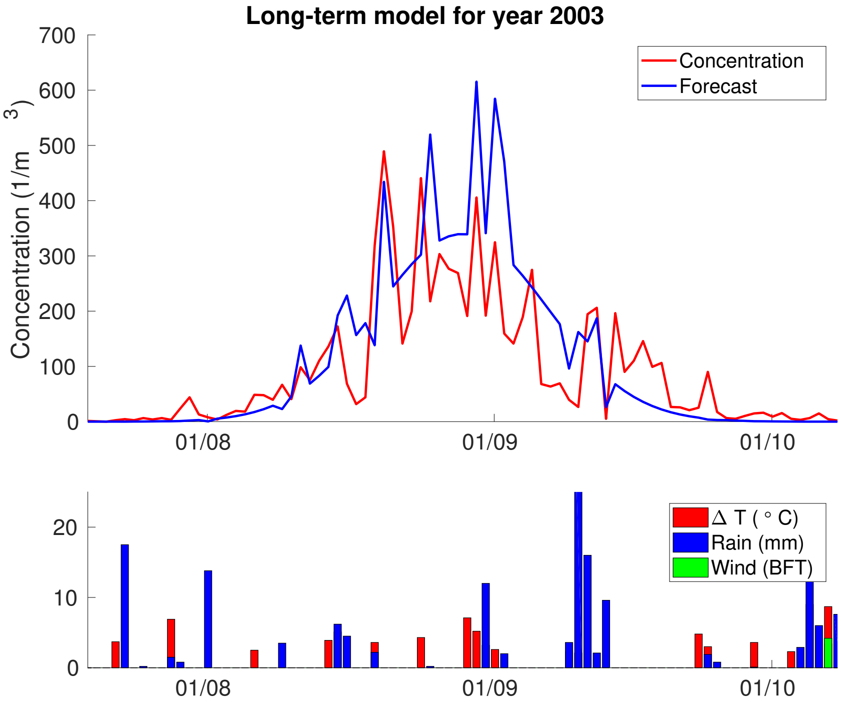


Figure 4: The long term model prediction for 2003.

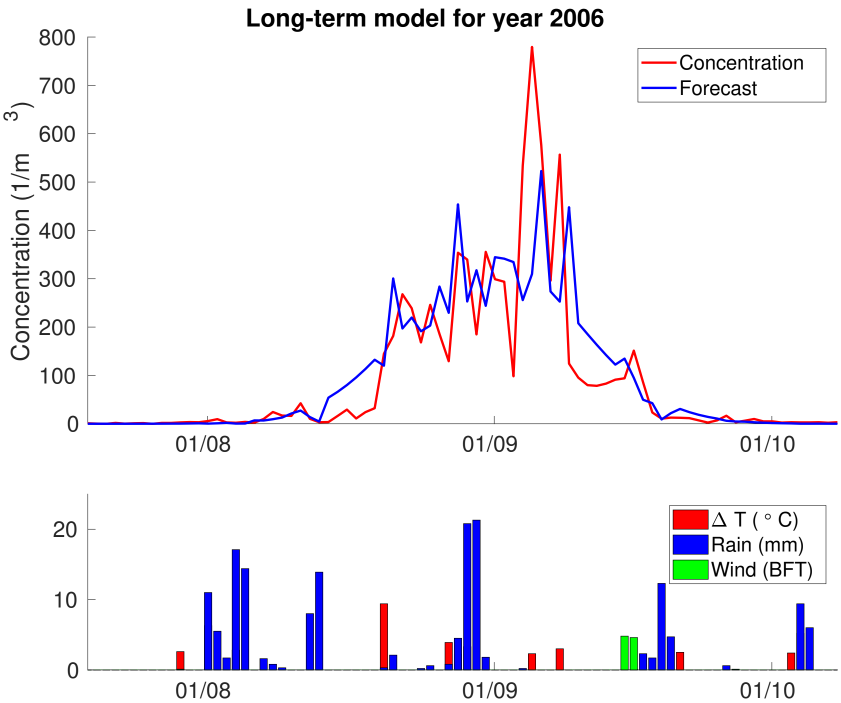


Figure 5: The long term model prediction for 2006.

of the other two weather events together as the other years, so it demonstrates the model well.

The figures clearly show, that the model manages to describe the main shape of the season. Most peaks in the pollen concentration are located, even if the predicted amount is smaller or bigger than the actual amount. However, the figures show that there are still some phenomena especially in the early and late season that the model cannot predict.

2.3 Short term model

In contrast to the long term model, the aim of the short term model is to predict the pollen concentration of the next day during the main season in the area of Novi Sad.

First, this new aim brings a significantly finer time scale than in the long term model but we will still use the general results from the previous subsections. So far, we just wanted to predict the concentration at a single point but now we get a new spacial dimension. We want to predict the concentration in the whole area of Novi Sad. Hence, we have to involve information about the positions of the Ambrosia plants in Novi Sad and the spacial spreading of the pollen starting from the plant.

The basic idea is to model the concentration with a advection-diffusion-model. On a microscopic scale we think about the behaviour of the single pollens and in our macroscopic model we consider the pollen concentration. The pollen are very light such that they are spread by any motion in the air in a small neighborhood around the plant. In our model this is represented by the diffusion part. Further, the spreading of the pollen is driven by wind, that is represented by the advection part in our model. For the concentration $c = c(t, x)$ our model equation for the time t from 0 to $24h$ reads

$$c_t - D\Delta c + w \cdot \nabla c = f$$

where D is a diffusion coefficient that we have to determinate, $w = w(t, x)$ the wind vector that we can get from the weather forecast, and f a source or sink term that we consider later. In our case the diffusion coefficient is often called 'turbulence factor' and depends on the humidity and slow winds.

For the initial condition, we assume that we start in the night at a time were almost all pollen are grounded, in particular $c(0) = 0$. If one has better information about the distribution at $t = 0$ one can set $c(0) = c_0$.

Each plant releases a specific number of pollens during the day that is going to be distributed in diffusive manner. Therefore we will add a source term to our model. To handle the spatial influence of the position of the ragweed plants, we construct a auxiliary function in space

$$a(x) = \begin{cases} C & x \text{ close to an Ambrosia plant} \\ 0 & \text{else} \end{cases}.$$

In time it gets more complicated. From the Data Analysis we get that the pollen release during a normal day within the main pollen season ('normal day' means in this case without any special incidents like rain or strong wind) shows qualitatively a similar behaviour like the probability density function of the log-normal distribution $f_P(t) = \frac{1}{\sqrt{2\pi}(0.2t)} \exp -\frac{1}{2}\left(\log(0.2t - 0.5)\right)$ for t from 0 to $24h$. So our source term so far looks $f_{pollution}(t,x) = f_P(t)c_0(x)$, but we have to include a scaling that takes the weather information into account: $f_{pollution} = P(t, \text{weather forecast}) f_P(t)a(x)$. Based on the weather forecast, we multiply the scaling function that respects forecasted weather events events like high temperature drops and strong wind positively, and rain negatively.

Conversely to the new pollution of pollen, the pollen in the air fall down to the air and the concentration in the air is reduced. We model this with $f_{falldown}$ as a ratio of the concentration as sink term, that is constant in space. Here the ratio depends on the forecasted humidity. We chose $f_{falldown}(t) = \phi_{forecasted}(t,x)0.07c$.

For the boundary condition, supplement homogeneous Dirichlet conditions $c = 0$ on $\partial\Omega$. This choice is justified if we choose a sufficiently big area around Novi Sad such that the pollen won't reach the boundary region and if the time horizon of consideration and the wind speed are small enough.

All together, we built the following model for a 24h prognosis within the area of Novi Sad:

$$
\begin{aligned}
c_t - D\Delta c + w \cdot \nabla c &= f_{\text{pollution}} - f_{\text{fall down}} \\
c(0) &= c_0 \\
c &= 0 \text{ on } \partial\Omega
\end{aligned}
$$

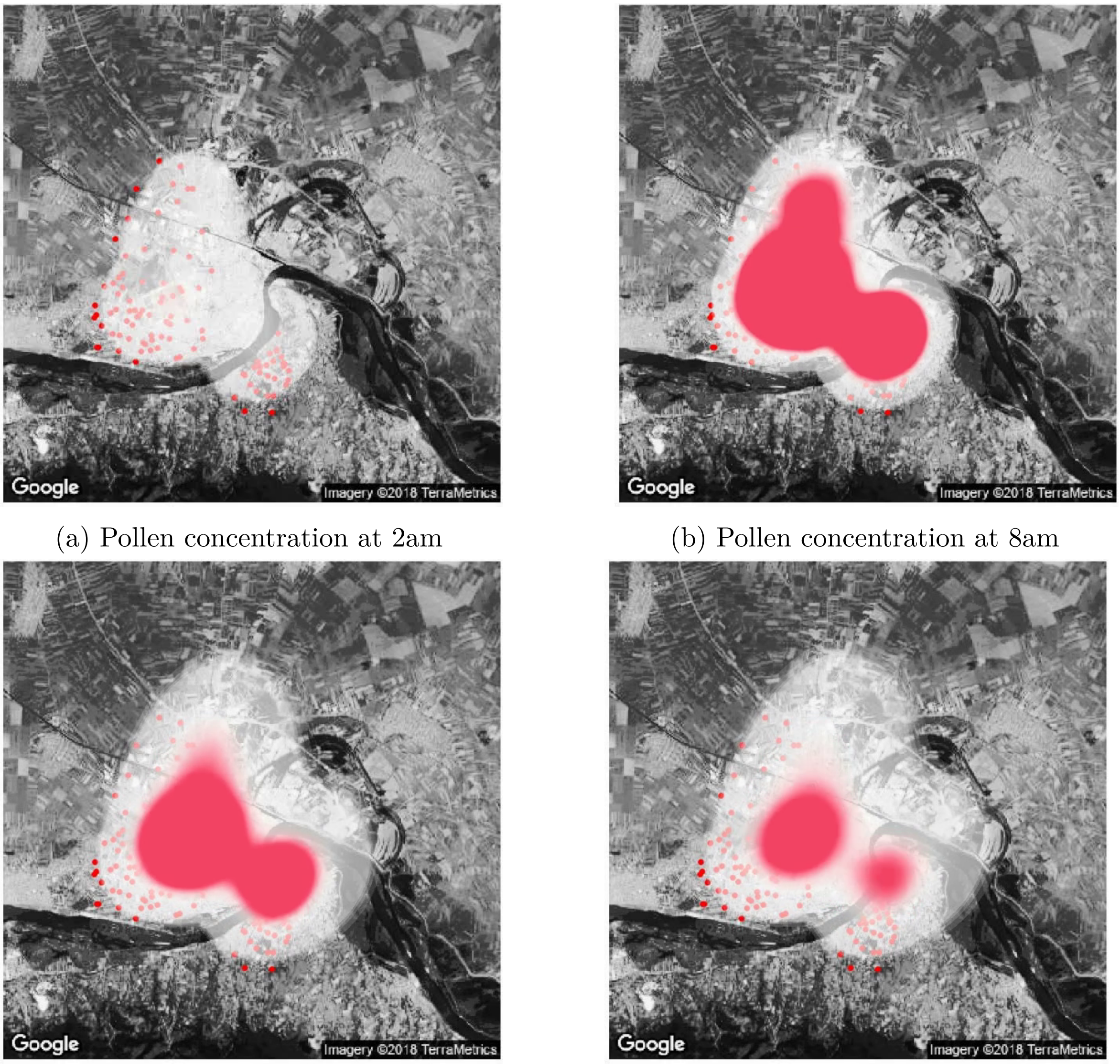
with

- diffusion coefficient $D(\text{humidity, slow wind})$
- wind vector w
- pollution of new pollen
  $f_{\text{pollution}}$(day, **time**, weather forecast: change of avg temp, rain, wind speed)
- falling down of pollen $f_{\text{fall down}}$(humidity).

We want to illustrate our numerical results for our model. For the computation, we have used the software FEniCS[1] in Python - we apply Newton's rule for the time stepping and use the FEniCS finite element solver for the elliptic PDE. In figure 6 the background shows the region around Novi Sad and the red dots are streets with an accumulation of ragweed plants. (The information about the positions of the plants we copied from pictures from the Biology Department.) The white color represents a low concentration and the red color a serious and a for health reasons critical concentration

[1] `https://fenicsproject.org/`

of ragweed pollen in the air. For our computations we chose slight wind from south west, no rain and a day during the main season with a high release of new pollen.



(a) Pollen concentration at 2am

(b) Pollen concentration at 8am

(c) Pollen concentration at 2pm

(d) Pollen concentration at 8pm

Figure 6: 24h prognosis: Modelled pollen concentration in the area of Novi Sad.

3 Discussion and conclusions

During this project, we studied the pollen concentration of the common ragweed in the area of Novi Sad, Serbia. We conducted data analysis on the pollen concentration and meteorological conditions. The results were applied in two models: a long term model for predicting the shape of the pollen season, and a short term model for predicting the spread of the pollen during one day.

For the long term model, we built a deterministic, data-driven model that depends on the weather forecast and observations. The model captures well the main features of the data for the years that were studied. However, further validation against independent data, such as the pollen concentration observations for August 2018, should be carried out. Due to the nature of the model, it cannot forecast random changes in concentration, so it provides more a qualitative forecast: is the concentration high or low during the following days.

For the short term prediction we built a qualitative model depending on the weather forecast. Due to the lack of detailed measurements, we could not verify the quality of our model. Anyway a stronger correlation of the long term and the short term model should be part of future research.

4 Group work dynamics

Starting the first day we analyzed the data we had at that time (the pollen concentration and temperature from 2000 to 2017) and read up on basic biological facts on pollen, Ambrosia, spreading and maturation of pollen. Further we considered climatic conditions in Novi Sad and tried to extract factors that could cause the higher concentration of pollen at some days in our data. To organize our group work we split up into two teams - one team, Samuel and Nataša, focused on the data analysis using R Studio and the other group members, Jenna, Joanna and Florian, focused on research on biological facts. To find a constructive model, the individual results were presented regularly to the rest of the group. At the end of the day, all members looked at the day-to-day work, made conclusions and planned in what direction the work will continue.

At the second day, we mainly continued the tasks from the previous day and we included in the analysis of our problem factors that we thought would be relevant and examined the statistical significance of them. During the afternoon, we developed a problem-solving plan based on past results and determining tasks for the third day.

Copying rainfall data from 2002-2017 to an Excel file was a new task at the third day which was accomplished by Joanna and Nataša. Samuel conducted an analysis of these new data in R Studio. Parallel Jenna, Joanna and Florian started to build a short term and long term model based on the data analysis and biological facts.

The fourth day consisted mostly of continuation of the work like at the third day in two teams. On the one hand work on the long term and short term model was proceeded, on the other hand the data analysis in R Studio according to the needs of both models was finished.

In the morning of the last day, the end of the works on the models was reached and the presentation was prepared. The rest of the day, all members of the group worked on making a presentation to present our problem and model on Saturday.

For the report, we established an ‘overleaf’-project where we could collaborate on the

LaTeXdocument. Every group member and our instructor wrote a section (or subsection) and the others reviewed.

As mentioned above, most of the time we build small and agile teams within our group to organize and distribute the work load. It was of special importance to communicate clearly the team results with the entire group and to discuss the next step with all group members. Everyone was welcome to give new ideas. Once or twice a day we coordinated our work plan with our instructor (e.g. he could help us with more meteorological data).

5 Instructor's assessment

I am glad to say that the group succeeded to analyze the given problem much better than I expected. The problem is important but still unsolved and their contribution is respectable even in a scientific contest.

They have always had a very serious approach to their work, showing a great interest and efficiency in it. All members of the group have shown a high level of enthusiasm. They have look for more general model than expected in a direct approach to the problem under consideration. Their background knowledge of mathematics was sufficient. During the whole ECMI Modeling week event, they were ambitious and reliable, willing to go beyond what was required by the problem.

They communicated clearly and always demonstrated creativity. Moreover, the group finished their work on schedule as well as provided an interesting approach to the short term prognosis problem. As I already mentioned above, the problem itself is not yet completely solved. I have asked some experts in that field to look at their solution. They confirmed it was quite good could be used as a starting point for an efficient solution to the short term prognosis problem.

In conclusion on the aspects that I enumerated and explained above, I am highly satisfied with the group's work and the results they obtained through their work effort.

References

CABI, 2018 (2018). *Ambrosia artemisiifolia (common ragweed).In: Invasive Species Compendium. Wallingford, UK: CAB International.* URL: http://www.cabi.org/isc.

Deen, William, L. Anthony Hunt, and Clarence J. Swanton (1998). "Photothermal Time Describes Common Ragweed (Ambrosia artemisiifolia L.) Phenological Development and Growth". In: *Weed Science* 46.5, pp. 561–568. ISSN: 00431745, 15502759. URL: http://www.jstor.org/stable/4045962.

Franz, Essl et al. (2015). "Biological Flora of the British Isles: Ambrosia artemisiifolia". In: *Journal of Ecology* 103.4, pp. 1069–1098. DOI: 10.1111/1365-2745.12424. eprint: https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/1365-2745.12424. URL: https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/1365-2745.12424.

MATLAB (2017). *version 9.2.0 (R2017a)*. Natick, Massachusetts: The MathWorks Inc.

Prank, Marje et al. (2013). "An operational model for forecasting ragweed pollen release and dispersion in Europe". In: *Agricultural and Forest Meteorology* 182-183, pp. 43 –53. ISSN: 0168-1923. DOI: https://doi.org/10.1016/j.agrformet.2013.08.003. URL: http://www.sciencedirect.com/science/article/pii/S0168192313002104.

Willemsen, Roger W. (1975). "Dormancy and Germination of Common Ragweed Seeds in the Field". In: *American Journal of Botany* 62.6, pp. 639–643. ISSN: 00029122, 15372197. URL: http://www.jstor.org/stable/2441944.

Šikoparija, B. et al. (2009). "The Pannonian plain as a source of Ambrosia pollen in the Balkans". In: *International Journal of Biometeorology* 53.3, pp. 263–272. ISSN: 1432-1254. DOI: 10.1007/s00484-009-0212-9. URL: https://doi.org/10.1007/s00484-009-0212-9.