Clara Stegehuis
Faculteit EWI
Universiteit Twente
c.stegehuis@utwente.nl

Francesca Arici
Mathematisch Instituut
Universiteit Leiden
f.arici@math.leidenuniv.nl

Proof by example Portraits of women in Dutch mathematics

Annika Betken

In 'Proof by example', Clara Stegehuis and Francesca Arici portray women in Dutch mathematics. This edition portrays Annika Betken, assistant professor at the University of Twente, who was recently awarded a Veni grant for her research on the statistics of time series and change points. In this interview she tells about her research and her motivation to pursue a career in mathematics.

When did you first become interested in mathematics?
"For me, it was not always obvious that I wanted to be a mathematician. I liked the formal language in mathematics in elementary school or in high school, so mathematics was something that I always enjoyed. Still, I also liked other subjects, and in fact I was planning to study literature or philosophy. But when I visited the job center in Germany, they told me that these did not have great job prospects. They asked me if I had any other interests as well, and then it became clear that mathematics was a better choice. So then I ended up studying mathematics, but I still did a minor in Philosophy as well."

And then you moved into statistics?
"That was also not really a conscious choice. My statistics lecturer gave very interesting lectures, so I started following more of them. He already mentioned that he had some PhD positions when I was still in my bachelor program. By then, I did not know yet if I wanted to pursue a PhD, but in the end, I obtained one of these positions. Beside the interesting lectures I followed, I also like the link to applications that statistics has. The mathematics that you work with on a daily basis is of course still quite abstract, but for me it is motivating that I can always keep its applications in mind."

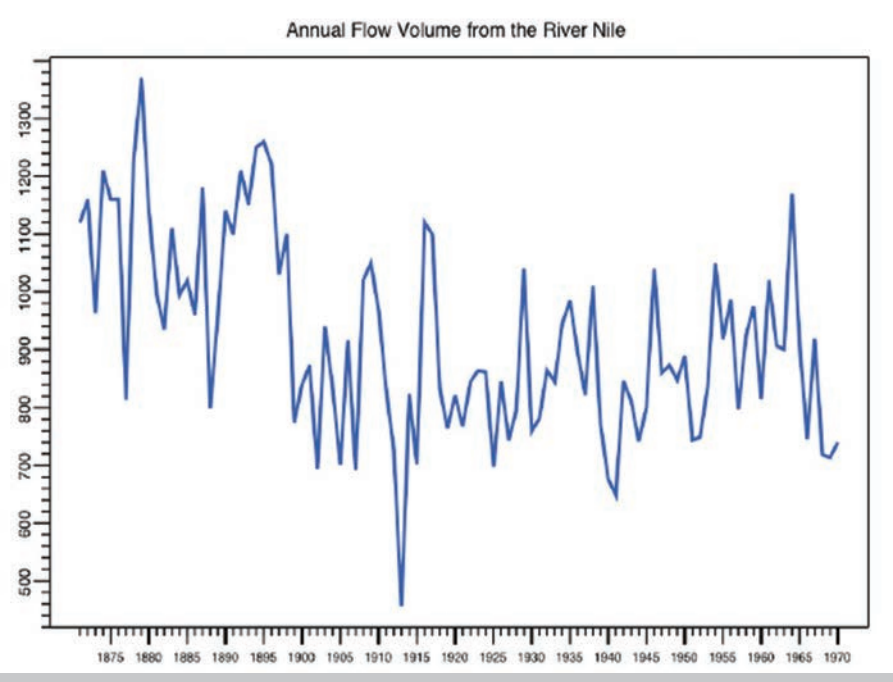

Annika Betken

What is your main topic of research?
"My research is about time series analysis, and in particular about change point detection. In change point detection, the main question is: does your data undergo a change at some time point, or are changes in the data over time just due to random fluctuations? In my research I investigate how we can answer this question when long-range correlations are present."

What are these long-range correlations?
"In time series analysis, you have measurements at different time steps. These time series are usually correlated: if your measurement was high in the previous time step, it is often likely to be high in the current time step as well. Still, when the time gap between two measurements gets larger, you expect to see less correlation between the measurements. How-

ever, in many applications, there is so-called long-range correlation. This means that the correlation between two measurements decays very slowly with the time gap. These long-range correlations make it impossible to use standard statistical methods such as central limit theorems. This makes it also more difficult to see whether there is a change point in your data, or whether this is caused by the long-range correlations. In my research, I try to come up with new strategies to still be able to detect these change points."

Can you give an example of data with such long range correlations?
"One famous application of long-range correlations in time series analysis is in the discharges of the river Nile. A researcher there was interested in the amount of water flowing though the Nile close to Cairo, and measured this flow on a regular basis. However, these measurements did not fit with standard theoretical predictions at all. Only when they took long-range correlations into account, they were able to match the theory with the actual measurements. This shows that when you want to investigate a property of such data, it is important to take long-range correlations into account."



The discharges of the river Nile have long-range correlations

What is a topic that you have recently been interested in?
"Together with Hongwei Wen, a PhD student I am working with, and Hanyuan Hang, a fellow assistant professor at the University of Twente, we are currently doing research more on the topic of machine learning. Suppose that you have one data set with labels, and another data set without labels, which is a 'shifted' version of your first data set. In practice, you can think of the first data set to be disease symptoms of persons when they get tested for example for COVID-19, before its major outbreak. The labels are classifications of whether the person with these symptoms is infected or not. The second data set is taken during the pandemic, when the distribution of infected persons differs from the first scenario. For example, the age distribution of these persons during the disease may be different than in the first data set. You can train a classification algorithm on the first labeled data set to find a classifier for the disease. But how well does this algorithm perform on the second data set? And how much information gets lost when the data set is shifted? This is important to know, because often the data on which you train a specific algorithm is slightly different than the data you will use the algorithm on, like in the example of COVID-19 infections at different time points. For very simple classification problems Hongwei already has some results, but we hope to extend these to more scenarios soon."

What is the result that you are most proud of?
"This is probably the first paper I wrote, based on the research I did during my master thesis. My supervisor asked me to think of a better statistic for something. So I gave it some thought, and then I came up with an idea. My supervisor said that it was actually quite a good idea, and that I should write this up as a paper. I had never written a paper before, so I had to learn this just by doing it. I am mainly proud of this paper because the main idea in this paper was probably more original than many of the ideas that I have now. As a master student, I think that your ideas can sometimes be more fresh than when you have been into a topic for a long time, as you then usually already have a particular research direction in mind."

What do you like best about doing research?
"In the research itself, I mostly like puzzling. You are stuck on a problem for a while, and have to come up with new solutions every time to solve a new problem. But other than that, I also really enjoy traveling for conferences and I especially like meeting new people. The people you meet at universities are usually very open minded, which I appreciate a lot. They are interested in new mathematical problems, and you can easily connect with them."

And what do you like less?
"Recently I got more overwhelmed with many small tasks. I really do not like this, because when you have so many different things to do, you cannot do all of them well anymore. In fact, you really have to choose what to focus on and what not to focus on. This is something that I have not mastered yet myself."