# The forecast of bio-feedstock availability for material transition: modeling with time-series and machine learning

**Researchers**
Daoud Urdu, Xuezhen Guo, Marcel Kornelis

**Focus**
Combining different statistical methodologies, such as time-series analysis and machine learning provides a robust modeling tool to predict bio-feedstock availability while considering forecast accuracy and efficiency.

**Innovative idea and objective**
In order to produce materials from biomass sources to facilitate the bioeconomy transition, it is important to understand the future development of biomass availability. The success of model predictions for biomass availability in practice may depend on what is called *forecast efficiency*. This implies that the model can quickly adopt changes in the markets when they occur, and take them adequately into account when a new prediction is made. In other words, each time when new data becomes available a Decision Support System (DSS), including a forecast efficient model, is able to adopt the relevant information from this new data.
Recently, machine learning techniques have been developed to check for forecast efficiency of predictive models. By applying machine-learning techniques to the building of time-series models, we may develop predictive models with two strong features. Firstly, it may help to capture information in the available data, which was not found if only time-series techniques were used. Secondly, the model can still capture expert opinion in its functional form, as it remains a time-series model in its basis.
Up to the best of our knowledge, we are the first to consider the use of machine-learning techniques to test for the forecast efficiency of time-series models, within the context of biomass production.

**Relevance to the materials transition in textiles and/or building materials**
To facilitate the materials transition (i.e., replacing the fossil-based materials by the biobased materials), forecasting the availability of the biomasses especially from the major sources (crops) are essential. We believe that *forecast efficiency* is a relevant topic within this context, because the production patterns of these markets have to change in order to facilitate the bioeconomy transition. Consequently, useful predictive models should not only give accurate forecasts of the biomass commodities of interest, but also fast adopt information about potential changing patterns, so that predictions can quickly be revised.

**What we did**
We downloaded the FAOSTAT data on production and land use for cotton, flax and wood pulps for the five continents (Africa, America, Asia, Europe, Oceania), as well as for the world. We developed AutoRegressive (AR) models to predict each biomass variable of interest (a total of 36 models) . We developed Vector AutoRegressive (VAR) models to predict each biomass variable of interest (a total of 6 models) . We organized a workshop with WUR researchers where we discussed our approach. We compared the results of the single equation AR models with those of the VAR model with respect to model accuracy and efficiency. Developed the prototype of the dashboard to visualize the results.

**Key deliverables**
- Overview of data sources dataset
- A predictive model, in EViews and R, which applies the AR, VAR, and random forest techniques to be potentially developed into a DSS tool
- A draft manuscript to be submitted to a scientific journal
- Wireframe dashboard as a prototype to visualise results

## Main result, achievement and highlight

- The forecast accuracy differs substantially across biomass commodities and geographical regions. It is of interest to investigate what the underlying factors of these differences are. Various outcomes may be of interest. For example, it may be the case that region- and commodity specific drivers play an important role in the over-time development of the biomass commodities. It may also be the case, that actors in different regions respond differently to the same global drivers. In both cases, the use of forecast accuracy as a model-building criterion leads to more insights in specific underlying short- and long-run dynamics of biomass commodities.

- The forecast accuracy differs substantially across biomass commodities and geographical regions. It is of interest to investigate what the underlying factors of these differences are. Various outcomes may be of interest. For example, it may be the case that region- and commodity specific drivers play an important role in the over-time development of the biomass commodities. It may also be the case, that actors in different regions respond differently to the same global drivers. In both cases, the use of forecast accuracy as a model-building criterion leads to more insights in specific underlying short- and long-run dynamics of biomass commodities.

**Visual Abstract:**

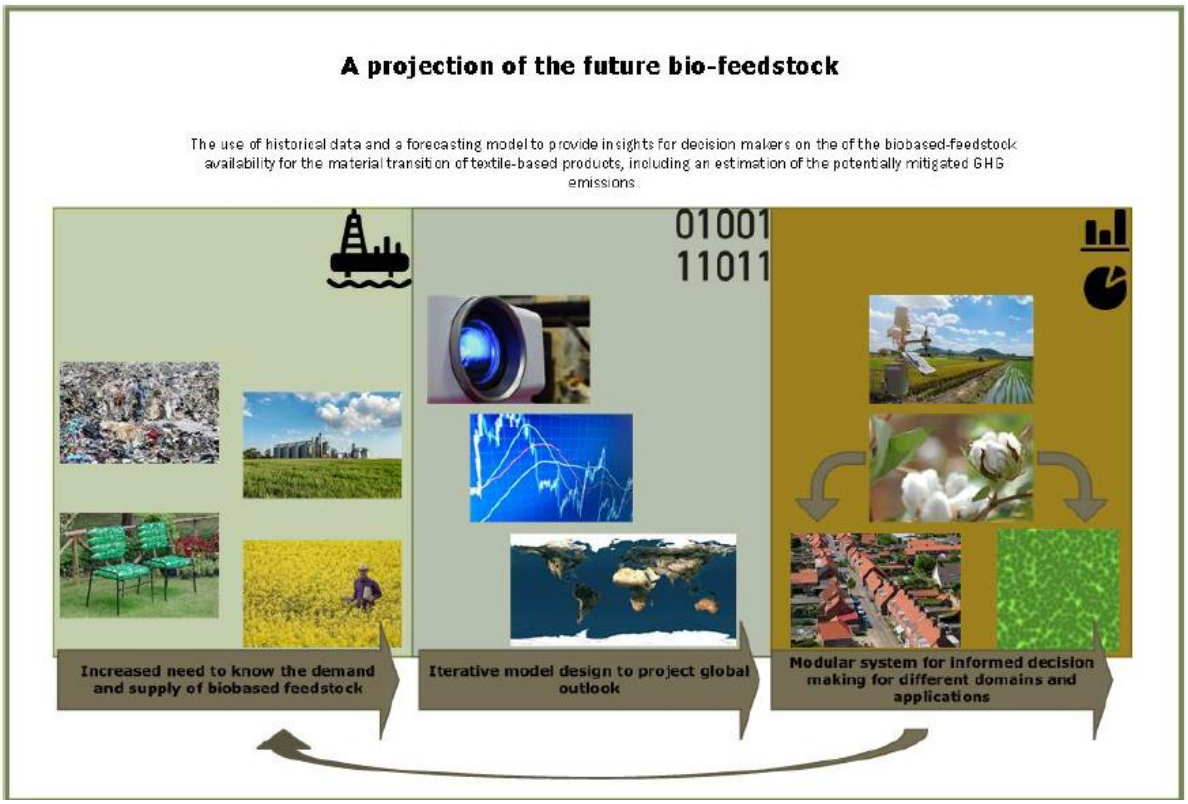


*Figure 1: Visual abstract of a projection of future bio-feedstock*