The objectives of the European Union with respect to renewable sources (RED II) are ambitious both in terms of energy uses and for use in transport. In particular for the transport sector, by 2030 it will have to be produced by renewable sources 14% of the share of fuels used. As part of renewable energy sources, residual biomass plays a significant role, because of its heterogeneity and the possibility of obtaining advanced biofuels and biochemicals through several thermochemical and biochemical processes, such as pyrolysis, gasification and aerobic digestion. The main challenge of using residual biomass is its properties heterogeneity (i.e., C, H, N, O content, ashes content, extractives, cellulose, hemicellulose, lignin and so on) according to type, geographic localization and seasonality. For this reason, properties and yields of the products are highly variable. Furthermore, the complexity of the processes involved in biomass valorization often makes the classical chemical-physical modelling approach very challenging.

In this framework, the research community as well as the industrial sector are focusing their attention on the adoption of new tools based on advanced statistical methods such as principal component analysis (PCA), machine learning techniques such as Artificial Neural Network (ANN) and operative research techniques for decision-making support. In particular, this work represents a first attempt to this “new modelling concept”, focusing its attention on the thermochemical valorization of residual biomass, specifically the fast pyrolysis process. A broad database with 450 observations is developed collecting data from scientific literature, which includes both qualitative and quantitative information regarding main biomass characteristics (e.g., type, proximate and ultimate analysis), pre-treatments (e.g., drying), operative parameters (e.g., temperature, pressure, type of reactor) pyrolysis products and byproducts (e.g., yields, quality). PCA is applied to a subset of the original data considering only fast pyrolysis process, non-catalytic, in a fluidized-bed reactor with a temperature range of 450-550°C. The aim of PCA is to obtain new inputs as a linear combination of the original ones keeping the information content of the original dataset almost unchanged and reducing computational time, therefore costs, related to the use of forecasting tools such as multivariate regression and machine learning techniques.