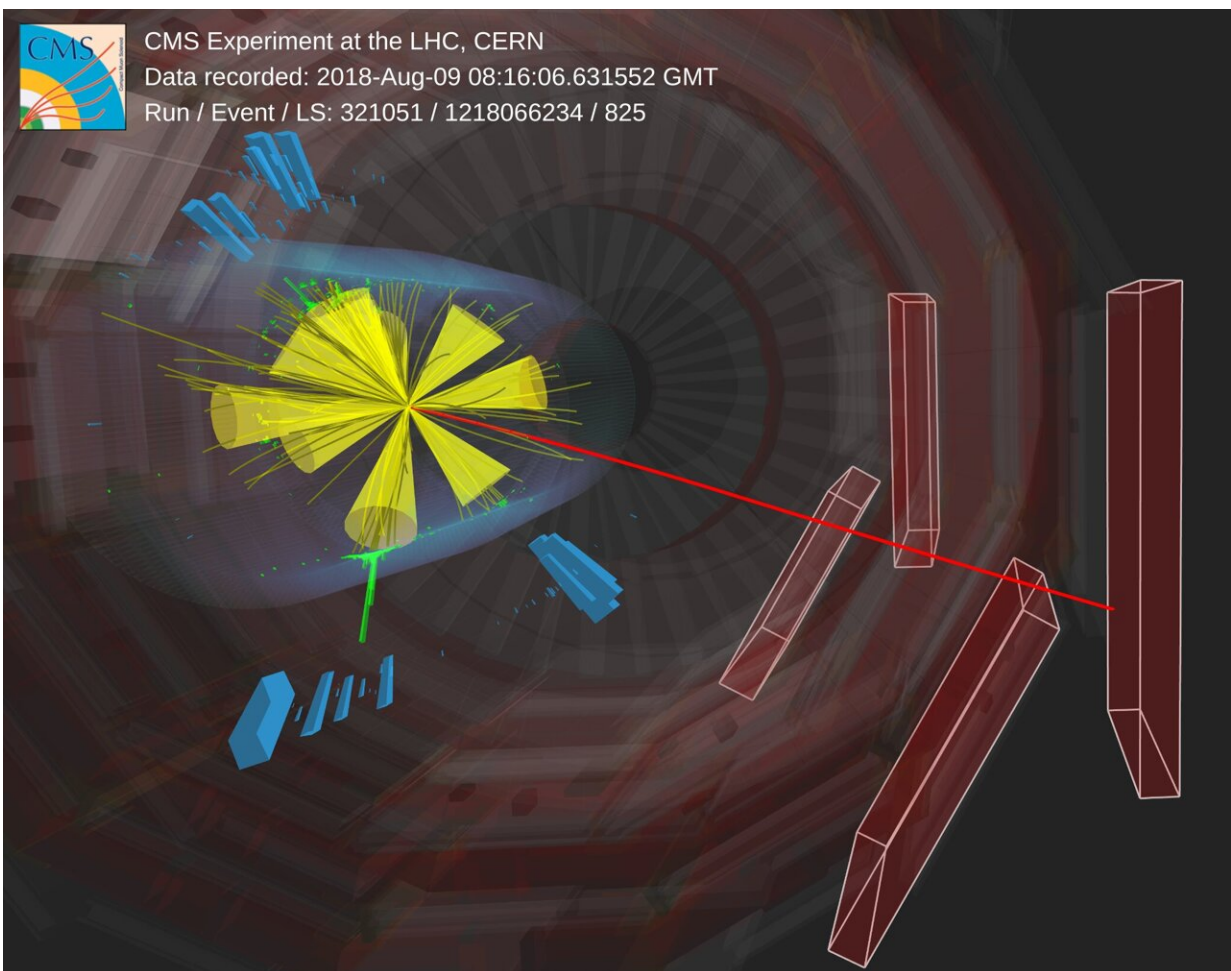# Re-analyzing LHC Run 2 data with cutting-edge analysis techniques allowed physicists to address old discrepancy

June 27 2024

A candidate collision event with two top quarks and multiple reconstructed jets (shown as yellow cones). Credit: CMS collaboration

Supersymmetry (SUSY) is an exciting and beautiful theory that answers some of the open questions in particle physics. It predicts that all known particles have a "superpartner" with somewhat different properties. For example, the heaviest quark of the Standard Model, the top quark, would have a superpartner called the top squark, or simply the "stop."

In 2021 the CMS collaboration analyzed the entire set of collision data collected from 2016 to 2018 and found features suggesting that it might contain stop particles. In that case, "might" meant that there was less than a 5% chance that data containing only known particles could look like what was observed.

Instead of waiting many years to collect more data with the hope of reproducing this behavior, the CMS collaboration decided to reanalyze the same data with upgraded analysis techniques.

The new analysis looks for the simultaneous production of pairs of stops. Each stop decays into a [top quark](#) accompanied by several lighter quarks or gluons, which then form bound states known as hadrons, ultimately creating clusters of particles reconstructed in the detector as "jets." The signal footprint is therefore two top quarks and multiple jets.

What makes the analysis challenging is that a very similar footprint is produced by one of the most common Standard Model processes in the LHC: the pair production of top quarks. Top quark production with many accompanying jets is a process that is difficult to accurately simulate, so to have a reliable determination of this background, it must be estimated from observed data.

A commonly used method of estimating backgrounds from data is called the "ABCD method." It requires two uncorrelated observables that can discriminate between signal and background. The data set can then be divided into four regions (A, B, C and D) depending on the value of each

observable being "signal-like" or "background-like."

The subdivision then provides a region dominated by the signal, a region dominated by backgrounds and two intermediate regions. The key feature of the ABCD method is that, following the mathematics of probabilities for independent events, one can estimate the background in the signal-dominated region using only the information from the other regions.

The problem with using this method for the stop search is that all simple variables are correlated in this search, making the method invalid. To overcome this issue, CMS physicists have implemented an innovative approach based on advanced machine-learning techniques to determine two variables with a minimal level of correlation. These two variables are then used to divide the data into the four aforementioned regions.

The figure below shows the correlation between the two variables for the signal and the background and demonstrates that the signal mostly lies in region "A."
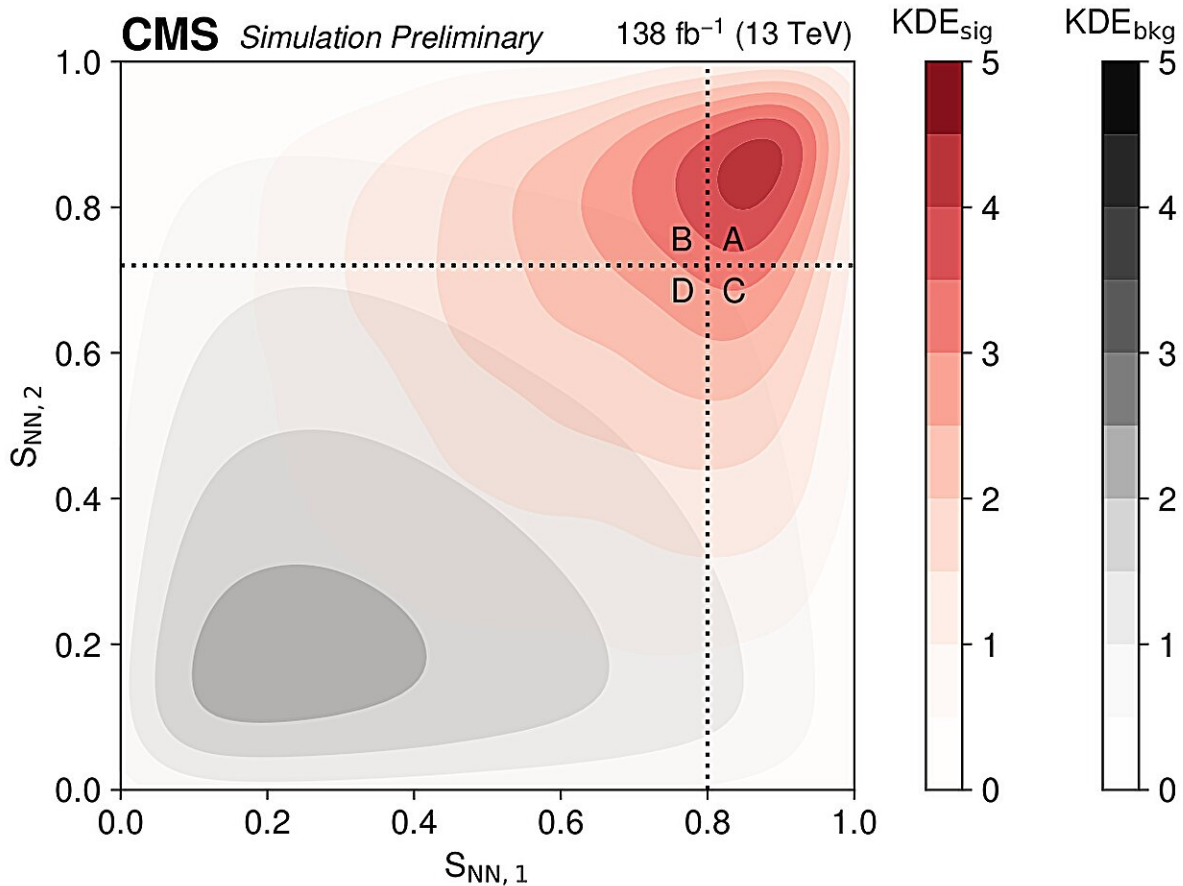
Distributions of signal (red) and background (gray) in the four (A, B, C and D) regions, defined based on two uncorrelated variables (SNN1 and SNN2) determined using machine learning. Credit: CMS collaboration

Using this novel method, the CMS collaboration was able to accurately predict the dominant background in this analysis from observed data, without relying on simulations with large uncertainties associated with the modeling of the jet multiplicity distribution.

This resulted in a large gain in analysis sensitivity. If the signal hinted at by the 2021 analysis was real, it would now have been observed without

any doubt. The fact that a signal was not seen in this analysis implies that, in specific SUSY scenarios, a stop decaying ultimately to top quarks and jets must have a mass greater than 700 GeV.

With a much more sensitive analysis method in place, the physicists are now eagerly looking forward to analyzing the data of the ongoing LHC Run 3 to go even further and to find where Nature hides its answers.

More information: The CMS Physics Analysis Summary: [cms-results.web.cern.ch/cms-re ... US-23-001/index.html](#)

Provided by CERN