Comput Softw Big Sci (2024) 8:19
https://doi.org/10.1007/s41781-024-00121-4

Research

The CMS Statistical Analysis and Combination Tool: COMBINE

CMS Collaboration*
CERN, Geneva, Switzerland

Received: 10 April 2024 / Accepted: 27 June 2024

Abstract This paper describes the COMBINE software package used for statistical analyses by the CMS Collaboration. The package, originally designed to perform searches for a Higgs boson and the combined analysis of those searches, has evolved to become the statistical analysis tool presently used in the majority of measurements and searches performed by the CMS Collaboration. It is not specific to the CMS experiment, and this paper is intended to serve as a reference for users outside of the CMS Collaboration, providing an outline of the most salient features and capabilities. Readers are provided with the possibility to run COMBINE and reproduce examples provided in this paper using a publicly available container image. Since the package is constantly evolving to meet the demands of ever-increasing data sets and analysis sophistication, this paper cannot cover all details of COMBINE. However, the online documentation referenced within this paper provides an up-to-date and complete user guide.

Keywords CMS · Combine · Higgs · Higgs boson · Statistics

1 Introduction

The CMS statistical analysis software, COMBINE, is designed with two main features in mind. The first is to provide a command-line interface to several common workflows used in statistical analyses in high-energy physics, and the second is to encapsulate the statistical model using a human-readable configuration file—herein referred to as a "datacard". These features are intended to ensure consistency in statistical methodology and allow for efficient investigation of potential issues, without limiting the complexity of any single analysis. Perhaps the most important consequence is that the constructed likelihoods can be combined to produce a greater sensitivity in searches or measurements, provided that the data sets are statistically independent. The COMBINE analysis software is built around the ROOT [1], RooFit [2], and RooStats [2] packages.

The statistical methods in COMBINE, many of which were developed in the LHC Higgs Combination Group [3], were originally designed for searches for a Higgs boson in proton–proton collisions. The COMBINE tool was used in early searches [4] for, and the subsequent discovery of, the Higgs boson [5,6] by the CMS Collaboration. Historically, its main use case within the CMS Collaboration was in searches for the Higgs boson, hence many of the function names and variables used in COMBINE include "Higgs". However, the tool is not specific to these searches or measurements of Higgs boson properties but instead is a generic tool usable for various statistical analyses of LHC data. Since the Higgs boson discovery, many extensions have been included that have been used for the statistical analysis in numerous publications of the CMS Collaboration, including measurements of Higgs boson properties [7], searches for supersymmetry [8], and measurements of standard model parameters such as the top quark mass [9]. The COMBINE tool has also been used with data from the ATLAS and CMS experiments to produce combined measurements of the Higgs boson mass, production and decay rates, and coupling modifiers [10,11]. Furthermore, the COMBINE software includes several routines that provide diagnostic information regarding the statistical model and statistical analysis methods used in such publications.

This paper provides a summary of the statistical methods and capabilities of the COMBINE tool. For complete and up-to-date documentation, it is recommended that the reader consult the online documentation [12].

In this paper, command line instructions are indicated by the symbol `$` at the start of the line. Square brackets `[option]` indicate optional commands that alter the default behavior of COMBINE, while angular brackets `<option value>` indicate a value that must be specified by the user. The color scheme for the listings contained in this paper is as follows:

* e-mail: cms-publication-committee-chair@cern.ch (corresponding author)

Published online: 08 November 2024

```
Contents of a complete datacard.
```

```
$ Executable command line.
```

```
> Terminal output from the tool.
```

```
Snippet of a datacard or of Python
 ↪ code.
```

This paper is organized as follows: the "Installation" section details the dependencies of the software and instructions for its installation. The statistical model that is constructed by COMBINE is described in "The Statistical Model" section, followed by detailed explanations of the analysis types available in the tool and instructions on how they are implemented in the "Supported Analysis Types" section. The "Physics Models" section provides instructions on the use of physics models in COMBINE, with several examples given. This section also provides a concrete example of a full statistical model constructed in COMBINE. The "How to Run COMBINE" section provides instructions on how to run COMBINE. It demonstrates several common statistical procedures using the tool, for example, the calculation of maximum likelihood estimates and confidence or credible intervals, and performing goodness-of-fit tests. Finally, a summary is given in the "Summary" section.

2 Installation

Aside from ROOT and ROOFIT, and their dependencies, several additional libraries are used for optimized algebraic calculations, such as vectorization libraries VDT [13], the GNU scientific library GSL [14], and EIGEN [15]. Additional common libraries are used, which include BOOST [16], and GZIP [17]. The COMBINE package may be compiled either within a CMS software (CMSSW) environment that provides a versioned set of all dependencies, or as a standalone package. Details of different installation instructions are regularly updated in the online documentation. A precompiled version of COMBINE is available as a DOCKER [18] container image:

```
$ docker run [--platform linux/amd64]
 ↪ --name combine -it gitlab-registry.
 ↪ cern.ch/cms-cloud/combine-
 ↪ standalone:v9.2.0
```

At the time of writing this paper, the latest version of COMBINE is v9.2.0, and all of the example datacards and the inputs necessary to run COMBINE can be found in the `data/tutorials/CAT23001` directory, which is available in the container image. For statistical calculations that make use of random sampling, the results obtained by a reader are expected to be consistent with, but not identical to, those provided in this paper.

3 The Statistical Model

The primary task of COMBINE is to produce a statistical model, $p(\text{data}; \vec{\Phi})$, which encodes the probability density for the observed data parameterized by the model parameters $\vec{\Phi}$, and is subsequently used for the statistical analysis.

For numerical efficiency, it is useful to factorize the statistical model $p(\text{data}; \vec{\Phi})$ as much as possible with respect to both observables and parameters. The parameter space is partitioned into parameters of interest $\vec{\mu}$ and nuisance parameters $\vec{\nu}$. Nuisance parameters are used to model various uncertainties of theoretical and experimental origin, such as those involved in the prediction of process cross sections or associated with luminosity calibration. Furthermore, the observable space is partitioned into primary observables $\vec{x}$, defined as those that appear in components of the model containing the parameters of interest, and auxiliary observables $\vec{y}$ that appear only in components of the model containing nuisance parameters. Each nuisance parameter ν_k (each element of the vector $\vec{\nu}$) is paired with a corresponding auxiliary observable y_k (element of $\vec{y}$) in a component of the statistical model that provides information about how well the nuisance parameter is known. Therefore, the statistical model constructed using COMBINE is factorized into the primary and auxiliary components of the probability as:

$$p(\vec{x}, \vec{y}; \vec{\Phi}) = p(\vec{x}; \vec{\mu}, \vec{\nu}) \prod_k p_k(y_k; \nu_k), \tag{1}$$

where $p(\vec{x}; \vec{\mu}, \vec{\nu})$ is the probability distribution of the observables for the primary analysis, and $p_k(y_k; \nu_k)$ are the probability distributions of the auxiliary observables. A likelihood function is constructed for a particular data set of independent identically distributed observables $\{\vec{x}_d\}$ as:

$$\mathcal{L}(\vec{\Phi}) = \prod_d p(\vec{x}_d; \vec{\mu}, \vec{\nu}) \prod_k p_k(y_k; \nu_k), \tag{2}$$

where the subscript d runs over all entries in the data set and the likelihood function is used in both Bayesian and frequentist calculations. Early searches for the Higgs boson by the CMS collaboration reported upper limits on the Higgs boson cross section using both Bayesian and frequentist methods [4] with COMBINE.

In COMBINE, the probability density terms associated with the auxiliary observables, $p_k(y_k; \nu_k)$, can also be reinterpreted as posterior distributions for the nuisance parameters, $p_k(\nu_k|y_k)$, resulting from the outcome of measurements of, or otherwise justified constraints on, the auxiliary observables

y_k, through the relationship

$$p_k(\nu_k|y_k) \propto p_k(y_k;\nu_k)\pi_k(\nu_k), \tag{3}$$

where $\pi_k(\nu_k)$ are the nuisance parameter priors. This procedure provides probability distributions from which nuisance parameter values can be sampled when generating pseudo-data sets that COMBINE uses in certain statistical calculations, as described in the "Pseudo-Data Generation" section. For all types of nuisance parameters in COMBINE, the priors are always assumed to be uniform [3], so all of the $\pi_k(\nu_k)$ are constants.

Each element of $\vec{x}$ is referred to as a "channel" and is statistically independent from all other elements of $\vec{x}$. For example, each element of $\vec{x}$ could be the event counts in different reconstructed final states of some data set, or a continuous observable such as the invariant mass of a pair of final state particles. The term $p(\vec{x};\vec{\mu},\vec{\nu})$ in Eq. (1) becomes a product over the channels,

$$p(\vec{x};\vec{\mu},\vec{\nu}) = \prod_i p_i(x_i;\vec{\mu},\vec{\nu}), \tag{4}$$

where i runs over the channels that comprise the primary analysis and p_i is the probability density function (pdf) for the observable x_i.

The likelihood function constructed by COMBINE assumes that all y_k are statistically independent from each other and from the primary observables. The user must specify the observables and their pdfs using the datacard as described below.

4 Supported Analysis Types

A configuration file in plain text format is required for COMBINE to define the observables $\vec{x}$ and $\vec{y}$, and their pdfs $p(\vec{x};\vec{\mu},\vec{\nu})$ and $p_k(y_k;\nu_k)$. This file is the datacard and is the primary input to COMBINE. The file is also used to specify the observed data needed to define the likelihood function, whether the analysis is a simple counting experiment or a more complex analysis using binned or unbinned distributions of the data with histograms or parametric functions to describe the pdfs.

The package includes a script `text2workspace.py` that COMBINE uses to convert the user defined inputs (in the form of the datacard) into a binary representation of the statistical model. The script can be run before running COMBINE itself to produce a binary ROOT file containing the statistical model in the form of a ROOFIT `RooWorkspace` object. The script is automatically run if the datacard is provided as the input to COMBINE. The script is run with the following command:

```
$ text2workspace.py [-m <mass>] [-o <
  ↪ datacard>.root] <datacard>.txt
```

The output ROOT file is given the same name as the input datacard, with the extension modified to `.root` unless the option `-o` is specified. The value of `mass`, a parameter widely used in searches for new particles, is interpreted by the `text2workspace.py` script to specify the datacard keyword `$MASS`, as described in the "Template-Based Shape Analyses" section.

It is possible to combine several datacards into a single datacard using the `combineCards.py` script:

```
$ combineCards.py Name1=card1.txt Name2
  ↪ =card2.txt .... > <combined card>.
  ↪ txt
```

This allows for building complex statistical models, while retaining the readability of individual components (datacards) of the model. Multiple instances of any nuisance parameter, sharing the same name, are treated as a single parameter of the statistical model with a single corresponding auxiliary observable y, provided that the pdf specified for y is the same in each instance. The rest of this section describes the preparation of datacards and associated inputs for use with COMBINE.

The first line of the datacard is a declaration of the number of channels, `imax`, that are present in the statistical model:

```
imax <number of channels>
```

For a single-channel analysis the datacard entry would be `imax 1`. If the value of `imax` is specified as "`*`", COMBINE automatically determines the number of channels.

The next lines in the datacard declare the number of processes to be considered, `jmax`+1, and the number of nuisance parameters, `kmax`:

```
jmax <number of processes minus one>
kmax <number of nuisance parameters or
  ↪ sources of systematic uncertainties
  ↪ >
```

For datacards with a single signal process, `jmax` is the number of background processes. Datacard lines starting with "#" are ignored by COMBINE and any amount of whitespace is allowed to separate columns and lines in the datacard. These features can be used to include descriptive comments in the datacard. The next sections of the datacard have a different syntax depending on whether the datacard represents a counting or shape analysis.

4.1 Counting Analyses

A counting analysis is one for which the statistical model can be cast in the form of Eq. (1) with only one primary observable, namely the total event count in a single channel that

includes multiple sources of signal and background. In the following, the primary observable is labeled n. The probability to observe n events is described by a Poisson distribution:

$$\mathcal{P}(n;\lambda) = \lambda^n \frac{e^{-\lambda}}{n!}, \tag{5}$$

for which the expected value, λ, can be a function of one or more parameters, and represents the total number of expected signal and background events.

Each process comes with a specified reference rate and one or more sources of uncertainty that are referred to as "systematic", even if they are of statistical origin. In COMBINE the definitions of each systematic uncertainty require the functional form for the probability distribution $p_k(y_k;\nu_k)$ to be specified. These typically reflect calibration measurements that often result in log-normal or gamma distributions, and are implemented in COMBINE through multiplicative factors that propagate the effect of systematic uncertainties to the statistical model.

Datacard 1 is an example with all of these elements, representing a counting experiment with one channel having a signal process ppX with reference rate of 1.47, two background processes WW and $t\bar{t}$, and three nuisance parameters that model systematic uncertainties in both the signal and background rates. In this example, the integrated luminosity uncertainty (`lumi`), assumed to be log-normal, results in an uncertainty in the expected signal rate, as well as in the expected rates of the WW and $t\bar{t}$ backgrounds. The reference rates of the signal and background processes are determined using simulation. A log-normal type uncertainty in the signal rate (`xs`) is included to account for the uncertainty in the predicted cross section of the signal process. A limited number of simulated events are available to determine the rate of the WW background. The statistical uncertainty due to this limited number of simulated events (`nWW`) propagates as a gamma distribution to the rate of WW.

The datacard lines immediately following the `imax`, `jmax`, and `kmax` lines describe the number of events observed in each channel. Line number 5, starting with `bin`, defines the label that should be used for each channel. In this example there is one channel, labeled `ch1`. Line number 6, starting with the word `observation`, indicates the number of observed events, which is 0 in Datacard 1. For analyses in which the data are binned in a histogram, a template-based datacard can be used instead of treating each bin of the histogram as a separate channel. There are typically several processes that contribute to the overall signal or background expected yields. Lines 8–11 in Datacard 1 describe the number of events expected for each channel and process, arranged in columns. The first column in each row identifies the information expected in the remaining columns. The number of columns beyond the first column must be equal to the total number of processes across all channels, i.e., to the product `imax`(`jmax` + 1). Line 8 starting with `bin` indicates that this row specifies the channel that each column refers to. In this case, since there is only a single channel, the number of columns in addition to the first one is equal to the number of signal and background processes in this channel. Lines 9 and 10 starting with `process` indicate that these rows refer to the labels and types of the various processes. Line 9 provides a label for each process and line 10 defines the type of the process which is either a positive number for a background process, or `0` or a negative number for a signal process. Line 11, starting with `rate`, indicates the expected event yield in the specified channel and process. This value should be considered as a reference rate for the process, assuming predetermined values for theoretical cross sections, detector acceptance and selection efficiencies, and integrated luminosity of the data set used in the analysis. The rates can be modified by the parameters of the statistical model $\vec{\Phi}$. In the simplest statistical model available in COMBINE, a single parameter of interest, the signal strength r, multiplies the rate of every signal process in the datacard as described in the "Physics Models" section.

The remaining lines 13–15 contain the description of systematic uncertainties that are to be included in the statistical model. Each of these systematic uncertainties is associated with a dedicated nuisance parameter ν. The systematic uncertainties section of the datacard is structured as follows:

- The first column indicates the name used in the binary representation of the statistical model for identifying the uncertainty. This is the name given to the ROOFIT `RooRealVar` object that encodes the corresponding nuisance parameter in the statistical model.
- The second column identifies the effect of the associated nuisance parameter and the form of $p(y;\nu)$ to be included in the statistical model. For gamma type nuisance parameters, this column has two entries, which is explained in the following.
- Finally, there are columns describing the effect of the systematic uncertainty on the rate of each process in each channel. The number of columns is the same as for the previous lines declaring channels, processes, and rates. If a process is unaffected by a nuisance parameter, the corresponding column entry for that nuisance parameter is "–".

The different types of systematic uncertainties that can be included in the datacard for counting experiments are shown in Table 1. Each of these types results in an associated probability term $p(y;\nu)$ which is either a normal distribution:

$$\mathcal{N}(y;\nu,\sigma_\nu) = \frac{1}{\sigma_\nu\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\nu-y}{\sigma_\nu}\right)^2}, \tag{6}$$

Datacard 1 Counting experiment datacard - `datacard-1-counting-experiment.txt`

```
imax 1
jmax 2
kmax 3
# A single channel - ch1 - in which 0 events are observed in data
bin             ch1
observation     0
# ----------
bin             ch1    ch1    ch1
process         ppX    WW     tt
process         0      1      2
rate            1.47   0.64   0.22
# ----------
lumi     lnN    1.11   1.11    1.11
xs       lnN    1.20     -      -
nWW      gmN 4    -    0.16     -
```

Poisson distribution,

$$\mathcal{P}(y;\nu) = \nu^{y}\frac{\mathrm{e}^{-\nu}}{y!}, \tag{7}$$

or uniform distribution,

$$\mathcal{U}(y;a,b) = \begin{cases} \frac{1}{b-a} & \text{if } y \in [a,b], \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

Equation (6), aside from being used directly for normally distributed observables, is also the building block for log-normally distributed observables, as discussed below. The Poisson distribution of y in Eq. (7) becomes a gamma distribution $\Gamma(\nu; y)$ when the observed value of y is substituted and the expression is interpreted as a likelihood function:

$$\Gamma(\nu; y) = \nu^{y}\frac{\mathrm{e}^{-\nu}}{y!}. \tag{9}$$

In Datacard 1 there are three sources of systematic uncertainty that affect one or more processes. Each of these lines results in a single nuisance parameter ν, auxiliary observable y, and associated probability density $p(y;\nu)$ being included in the statistical model. The nuisance parameters are $\nu_{\texttt{lumi}}$, $\nu_{\texttt{xs}}$, and $\nu_{\texttt{nWW}}$ with corresponding auxiliary observables $y_{\texttt{lumi}}$, $y_{\texttt{xs}}$, and $y_{\texttt{nWW}}$.

The first two uncertainties are log-normal types [19], detailed below using the integrated luminosity L uncertainty as an example. The rates of signal and background are typically proportional to L. The rates defined in the datacard are normalized to a reference value L_0, which represents the nominal value of the integrated luminosity. Deviations from this reference value are expressed through the dimensionless quantity $f = L/L_0$. An estimate $\hat{f}$ is available as a random sample from a pdf $p(\hat{f}; f)$. The probability distribution $p(\hat{f}; f)$ is log-normal, so that the sampling distribution of $\ln\hat{f}$ is normal, with mean equal to $\ln f$. The standard deviation of $\ln\hat{f}$ can be judiciously written as $\ln\kappa$, where κ is a positive constant that is specified in the datacard. The nuisance parameter ν that controls the systematic uncertainty in the integrated luminosity is then not considered to be L itself, but rather is defined as $\nu = \ln f/\ln\kappa$ from which it follows that the multiplicative factor in the rate corresponding to luminosity is $f(\nu) = \kappa^{\nu}$. The observable y is then a sample from $p(y;\nu)$, which is normal with mean ν as in Eq. (6), and unity standard deviation. The log-normal type is typically used when f is positive by definition, as is the case with the integrated luminosity, and $p(\hat{f}; f)$ continuously approaches zero as $\hat{f} \to 0$.

The first uncertainty in Datacard 1, `lumi` in line 13, represents the uncertainty in the measured integrated luminosity of the data set. The effect of this uncertainty is 11% on the rate of the `ppX`, `WW`, and `tt` processes. This means that the rates of all three processes are multiplied by a factor of 1.11 when the nuisance parameter $\nu_{\texttt{lumi}}$ is set to $+1$, and by a factor of $1/1.11 = 0.90$ when it is set to -1. The log-normal type is also useful when a systematic uncertainty is taken to be a factor of κ, also often expressed as percentage uncertainty of $\kappa - 1$. This means that high tails of the distribution of $\hat{f}$ above κf and low tails below f/κ each contain equal probability of roughly 16%. The second uncertainty in Datacard 1, `xs` in line 14, represents an uncertainty in the calculation of the theoretical cross section of the signal process. This uncertainty has an effect of 20% on the signal rate while leaving the background processes unaffected.

The gamma type uncertainty is used to model uncertainties in the rate of a particular process due to the limited sample size used to predict the rate. The third uncertainty in Datacard 1, `nWW` in line 15, is a gamma type uncertainty. This line in the datacard specifies that the reference rate of the `WW` background process is determined from a sample of 4 simulated events, each with an event weight of 0.16.

Table 1 Available uncertainty types for counting experiments

Uncertainty type	Directive	Inputs	Multiplicative factor, $f(\nu)$	$p(y;\nu)$	Default values
Log-normal	`lnN`	`kappa`	κ^{ν}	$\mathcal{N}(y;\nu,1)$	$\nu = y = 0$
Asymmetric log-normal	`lnN`	`kappaDown`, `kappaUp`	$\left(\kappa^{\text{Down}}\right)^{-\nu}$ if $\nu < -0.5$, $\left(\kappa^{\text{Up}}\right)^{\nu}$ if $\nu > 0.5$, $e^{\nu K\left(\kappa^{\text{Down}},\kappa^{\text{Up}},\nu\right)}$ otherwise*	$\mathcal{N}(y;\nu,1)$	$\nu = y = 0$
Log-uniform	`lnU`	`kappa`	κ^{ν}	$\mathcal{U}(y, 1/\kappa, \kappa)$	$\nu = y = \frac{1}{2}(\kappa + 1/\kappa)$
Gamma	`gmN`	`N, alpha`†	ν/N	$\mathcal{P}(y;\nu)$	$\nu = N+1$, $y = N$‡

The second and third columns indicate the entries for the datacard required to specify the type, and the relative effect on the yield of each process in each channel. The fourth and fifth columns indicate the resulting multiplicative factor by which COMBINE scales the normalization of the relevant process in the specified channel, and the term $p(y;\nu)$ that is included in Eq. (1). Finally, the last column indicates the default values of ν and y. Where relevant, the value of $\kappa - 1$ can be interpreted as the relative uncertainty in the process normalization in a given channel

* $K\left(\kappa^{\text{Down}}, \kappa^{\text{Up}}, \nu\right) = \frac{1}{8}\left[4\ln\left(\kappa^{\text{Up}}/\kappa^{\text{Down}}\right) + \ln\left(\kappa^{\text{Up}}\kappa^{\text{Down}}\right)\left(48\nu^5 - 40\nu^3 + 15\nu\right)\right]$ ensures that the multiplicative factor and its first and second derivatives are continuous for all values of ν, and reduces to a log-normal for $\kappa^{\text{Down}} = 1/\kappa^{\text{Up}}$

† The rate value for the affected process must be equal to $N\alpha$

‡ The default value for the nuisance parameter is set to the mean of a gamma distribution with parameters $\kappa = N + 1$, $\lambda = 1$, as defined in Ref. [20]

The full statistical model that COMBINE produces by default using Datacard 1 is given in the "Physics Models" section.

4.2 Shape Analyses

A shape analysis is defined as one that incorporates one or more primary observables, beyond a single number of events, in the statistical model of Eq. (1). The datacard has to be supplemented with two extensions: a new block of lines defining the pdfs for the observables related to each process in each channel, and a block of lines defining systematic uncertainties that affect those pdfs.

The pdf can be parametric or template-based, depending on the inputs provided by the user. In the former case, the parametric pdf for each process has to be provided as a ROOFIT object that is derived from the `RooAbsPdf` class. These objects must be contained in a `RooWorkspace` object that is identified as an input workspace in the datacard. In the latter case, for each channel, histograms must be provided to represent the pdf for each process binned in the observable for that channel. These must be either ROOT `TH1` or `RooDataHist` histogram objects, for analyses in which the data are *binned*, or `RooDataSet` objects when the data are *unbinned*.

As with the counting experiment, the total reference rate of a given process must be identified in the `rate` line of the datacard. However, there are special options for shape-based analyses:

- A value of `-1` in the rate line indicates that COMBINE should calculate the rate from the input `TH1` object using the `TH1::Integral` method, or the `RooDataSet` or `RooDataHist` using the `RooAbsData::sumEntries` method.
- For parametric shapes defined as `RooAbsPdf` objects, if a parameter is found in the input workspace with the name `pdfname_norm`, the rate is multiplied by the value of that parameter.

For shape analyses, the statistical model constructed by COMBINE is factorized where possible into normalization and shape terms that provide significant gains in computation time. Examples of this factorization can be seen in Eqs. (11) and (18).

4.2.1 Template-Based Shape Analyses

The majority of statistical analyses performed by the CMS Collaboration are template-based analyses. This choice of analysis is particularly common when there is no physically motivated parametric function to describe the pdfs for the primary observables. Example analyses include the observations of Higgs boson decays to bottom quarks [21], which uses the output of a deep neural network, and the production of four top quarks in proton–proton collisions [22], which uses boosted decision trees to construct the observable.

A template-based shape analysis is one in which the observable in each channel is partitioned into N_B bins. The number of events n_b in the data that fall within each bin b (with b running from 1 to N_B) is considered as an independent Poisson process. The observable x in Eq. (4) in each channel is replaced by the set of observables n_b for the purpose of constructing the statistical model in COMBINE. For each channel, COMBINE constructs the term $p(x;\vec{\mu},\vec{\nu})$ as a

product of Poisson probabilities, yielding

$$p(x;\vec{\mu},\vec{\nu}) = \prod_{b=1}^{N_B} \mathcal{P}\left(n_b;\lambda_b(\vec{\mu},\vec{\nu})\right), \tag{10}$$

where the total expectation for a given bin is denoted as λ_b, and the observed event count in data in a given channel and bin is denoted by n_b. In the case of $N_B = 1$, Eq. (10) reduces to the pdf that COMBINE constructs for a single counting analysis, cf. Eq. (5). For each channel, histograms must be provided that specify, for each bin, the observed data and the expected yield for each process that contributes. These can be provided as either ROOT `TH1` objects or ROOFIT `RooDataHist` objects. Within each channel, all histograms must use the same partitioning (binning) of the observable. When using `RooDataHist` objects, the observable name must be the same for all processes within that channel. An explicit check is made to ensure the normalization of the data histogram corresponds to the number of observed events and that the normalization of the histograms matches the rates provided in the datacard in each channel.

Template-based analysis datacards contain one or more rows in the form:

```
shapes <process> <channel> <file> <histogram>
 ↪ [<
 ↪ histogram_systematic_uncertainty_variation
 ↪ >]
```

In this datacard line, `process` is any of the process names or "*" for all processes, or `data_obs` for the observed data, and `channel` is any of the channel names, or "*" for all channels. The value of `file` gives the name of the ROOT file. The labels histogram and `histogram_systematic_uncertainty_variation` identify either the names of the `TH1` objects or the `RooWorkspace` and `RooDataHist` objects within that file. Several keywords in the datacard line are reserved for automatic substitution when constructing the statistical model:

- `$PROCESS` is substituted with each process label or `data_obs` for the observed data.
- `$CHANNEL` is substituted with each channel label.
- `$SYSTEMATIC` is substituted with the name of each nuisance parameter (with an additional suffix `Up` or `Down`) that appears in the lines at the end of the datacard.
- `$MASS` is replaced with a mass value, which is passed as a command line option.

These lines are interpreted sequentially in the order they appear in the datacard, which allows for multiple instructions with wild card characters and keywords to define the pdfs for each process across the channels. The `$MASS` keyword was originally intended to allow for a single datacard to be used for different mass hypotheses of the Higgs boson m_H. In analyses such as the search for $H \to WW^* \to 2\ell 2\nu$ [23], a different set of histograms for the signal at specific values of m_H are chosen by specifying this keyword on the command line. In the $H \to ZZ^* \to 4\ell$ and $H \to \gamma\gamma$ analyses, this keyword makes it possible to measure m_H using the `MH` parameter, which can be varied continuously [10]. This keyword is useful in searches for new particles whose masses are not known. The COMBINE package supports arbitrary user-defined keywords in the datacard, the values of which can be set at runtime with the command line. This is particularly useful for analyses in which the probability distributions depend on more than one parameter.

For each channel, the expected value for the total rate in a given bin b is expressed as a sum over processes p:

$$\lambda_b(\vec{\mu},\vec{\nu}) = \sum_p M_p(\vec{\mu},\vec{\nu})\,\omega_{bp}(\vec{\nu}) + E_b(\vec{\mu},\vec{\nu}), \tag{11}$$

where the functions $\omega_{bp}(\vec{\nu})$ are the number of expected events for a process in a given bin b, $M_p(\vec{\mu},\vec{\nu})$ is the overall multiplicative factor representing the effects of the statistical model parameters on the total rate of a given process, and $E_b(\vec{\mu},\vec{\nu})$, discussed at the end of this section, accounts for the statistical uncertainties arising from limited simulated or collision data used to populate the histogram. The value of $\lambda_b(\vec{\mu},\vec{\nu})$ is restricted to be positive for any values of the parameters.

The functions $\omega_{bp}(\vec{\nu})$ are derived using the histograms provided by the user that represent the expected distribution for a given process, for the nominal model and for alternates in which a particular nuisance parameter is varied. Dropping the process labels henceforth, the nominal bin contents are denoted by ω_b^0. For each shape systematic uncertainty s, two additional histograms are specified, $\omega_b^{s,+}$ and $\omega_b^{s,-}$, typically corresponding to the distributions for the values $\nu_s = +1$ and $\nu_s = -1$, respectively. The datacard must include an additional line in the systematic uncertainties block with the form:

```
<name> shape[N] <effect_p_i>
```

where `effect_p_i` is a sequence of values that indicate the effect of the uncertainty in each process p and channel i in the datacard. Each value in the sequence can be "-" or 0 for no effect, or 1 to indicate that the process is affected. Values different from 1 can be specified to indicate that the histograms provided represent the expected distribution for other values of the associated nuisance parameter ν_s. For example, a value of 0.5 would indicate to COMBINE that the alternative histograms correspond to values of $\nu_s = \pm 2$ as is illustrated for the systematic uncertainty modeled by the nuisance parameter `sigma` in Datacard 2.

The alternative histograms can be different from the nominal histogram both in the fractional content of each bin, $f_b = \omega_b / \sum_{b'} \omega_{b'}$, and in their normalization. For the latter, the term $M_p(\vec{\mu}, \vec{\nu})$ in Eq. (11) contains an additional asymmetric log-normal multiplicative factor, as described in Table 1, where the values of κ^{Up} and κ^{Down} are defined by:

$$\kappa^{\text{Up/Down}} = \frac{\sum_b \omega_b^{+/-}}{\sum_b \omega_b^0}. \tag{12}$$

For the former, the user can choose between either parameterizing the variation due to each nuisance parameter directly in terms of the fractional bin contents, f_b, or their logarithms, $\ln(f_b)$, by specifying either `shape` or `shapeN` in the systematic line of the datacard, respectively. The functions for the fractions are given by:

$$f_b(\vec{\nu}) = f_b^0 + \sum_s F\left(\nu_s, \delta_b^{s,+}, \delta_b^{s,-}, \epsilon_s\right), \tag{13}$$

if using the `shape` algorithm or

$$f_b(\vec{\nu}) = \mathrm{e}^{\left(\ln\left(f_b^0\right) + \sum_s F\left(\nu_s, \Delta_b^{s,+}, \Delta_b^{s,-}, \epsilon_s\right)\right)}, \tag{14}$$

if using the `shapeN` algorithm, respectively, where $f_b^0 = \omega_b^0 / \sum_{b'} \omega_{b'}^0$, $\delta_b^{s,\pm} = f_b^{s,\pm} - f_b^0$, $\Delta_b^{s,\pm} = \ln\left(f_b^{s,\pm}\right) - \ln\left(f_b^0\right)$, and the subscript s iterates over each systematic uncertainty. Similarly to the nominal fractions, the values of $f_b^{s,\pm}$ are determined by $f_b^{s,\pm} = \omega_b^{s,\pm} / \sum_{b'} \omega_{b'}^{s,\pm}$. The `shape` algorithm is typically used when the variation due to a systematic uncertainty is relatively small compared to the contents in each bin, while the `shapeN` algorithm is used when a systematic uncertainties results in large variations. This is due to the fact that the latter yields a smooth function close to zero when the fractional bin contents are small.

The function F depends on a set of scaling factors, ϵ_s. These are assumed to be unity by default, but may be set to different values, e.g., if the values of $\omega_b^{s,\pm}$ correspond to $\nu_s = \pm X$, then $\epsilon_s = 1/X$. The function F is defined as:

$$F(\nu, \delta^+, \delta^-, \epsilon) = \begin{cases} \frac{1}{2}\nu'\left((\delta^+ - \delta^-) + \frac{1}{8}(\delta^+ + \delta^-)J(\overline{\nu})\right), & -q < \nu' < q; \\ \nu'\delta^+, & \nu' \geq q; \\ -\nu'\delta^-, & \nu' \leq -q; \end{cases} \tag{15}$$

where $\nu' = \nu\epsilon$, $\overline{\nu} = \nu'/q$, and $J(\overline{\nu}) = (3\overline{\nu}^5 - 10\overline{\nu}^3 + 15\overline{\nu})$. The minimum value of ϵ for a given process in a given channel is $q = \min(\epsilon_s)$. These functions ensure that the fractions and their first and second derivatives are continuous for all values of ν. A discussion of the different functions that are commonly used for parameterization in template-based analyses can be found in Ref. [24]. For each `shape[N]` systematic uncertainty line in the datacard corresponding to an uncertainty s, an auxiliary observable y_s and its probability distribution $p(y_s; \nu_s) = \mathcal{N}(y_s; \nu_s, 1)$ are included in the statistical model in Eq. (1). The default values are $\nu_s = y_s = 0$.

Datacard 2 is an example of a template-based analysis in which the observable x is the output of a multivariate analysis (MVA) discriminator. The histograms defining the distributions of all processes, and their variations due to systematic uncertainties, are contained in a single ROOT file named `template-analysis-datacard-input.root`. Datacard 2 contains two processes, `signal` and `background`.

Line 5 specifies the mapping between the histograms in the ROOT file (as shown in Table 2) and the processes defined in the datacard. Lines 17 and 18 provide the definition of systematic uncertainties that affect the probability distribution of the observable for the different processes. In the example, two such uncertainties are listed, one, `alpha`, that affects only the background process, and the other, `sigma`, that affects only the signal process. Any text beyond the number of columns expected in these lines is ignored by COMBINE and can be used to include descriptions of the systematic uncertainties in the datacard. Within the ROOT file, two histograms for each systematic uncertainty are provided, which represent the probability distributions and rates of the signal and background processes when the associated nuisance parameters are varied. In the ROOT file, these are `TH1` objects named `background_alphaUp` and `background_alphaDown`, and `signal_sigmaUp` and `signal_sigmaDown`. Table 2 shows the `TH1` inputs contained in the ROOT file used in Datacard 2.

Figure 1 shows the histograms used to determine the expected yield ω_b^0 and yields expected for the variations of the systematic uncertainties ω_b^+ and ω_b^- for the signal and background processes. The histogram for the observed data is also shown.

The term $E_b(\vec{\mu}, \vec{\nu})$ in Eq. (11) accounts for the statistical uncertainties in the histograms used to determine ω_{bp}. When the histograms are provided using `TH1` objects, these terms can be included using the following line at the end of the datacard:

```
<channel> autoMCStats <threshold>
```

The first string `channel` should give the name of the channels in the datacard for which these statistical uncertainties should be included. The wildcard "*" indicates that this should apply to all channels in the datacard. The value of `threshold` should be set to a value greater than or equal to zero to include the uncertainties. This value sets the threshold on the effective number of unweighted events above which the uncertainty is modeled with a single nuisance parameter for each bin following the Barlow–Beeston procedure

Datacard 2 Template analysis datacard - `datacard-2-template-analysis.txt`

```
imax 1
jmax 1
kmax 4
# ---------------
shapes * * template-analysis-datacard-input.root $PROCESS
  ↪ $PROCESS_$SYSTEMATIC
# ---------------
bin          ch1
observation 85
# -------------------------------
bin              ch1         ch1
process          signal      background
process          0           1
rate             24          100
# -------------------------------
lumi      lnN    1.1        1.0
bgnorm    lnN    -          1.3
alpha   shape    -           1    # uncertainty in the background template.
sigma   shape    0.5         -    # uncertainty in the signal template.
```

Table 2 `TH1` objects contained in the `template-analysis-datacard-input.root` file used in Datacard 2

Object name	Description
`data_obs`	Histogram containing the observed number of events in each bin of the analysis
`signal`	Histogram containing the expected yields ω_b^0 of the signal process in each bin of the analysis
`background`	Histogram containing the expected yields ω_b^0 of the background process in each bin of the analysis
`signal_sigmaUp`, `signal_sigmaDown`	Histograms containing the expected yields of the signal process for each bin in the analysis when the nuisance parameter `sigma` is set to -2 (Down) ω_b^-, and $+2$ (Up) ω_b^+, and all other nuisance parameters are set to their default values
`background_alphaUp`, `background_alphaDown`	Histograms containing the expected yields of the background process for each bin in the analysis when the nuisance parameter `alpha` is set to -1 (Down) ω_b^-, and $+1$ (Up) ω_b^+, and all other nuisance parameters are set to their default values

The histograms used to determine the effects of the `sigma` parameter on the signal distribution correspond to twice the variation resulting from the systematic uncertainty modeled by the parameter

outlined in Ref. [25], using the simplifying approximation introduced in Ref. [26]. This procedure is used to improve the computational performance without introducing significant impact on the accuracy of the results. Below the threshold, an individual nuisance parameter for each process is created. This threshold improves the computational stability of subsequent statistical routines by reducing the total number of parameters of the statistical model, while maintaining a reasonably accurate description of the statistical uncertainties in the histograms. For each nuisance parameter, a corresponding probability distribution for y is included when building the model in Eq. (1), the form of which is determined by this effective number.

When `threshold` is set to a number of effective unweighted events greater than or equal to zero, denoted $n^{\text{threshold}}$, the following algorithm is applied to each bin b:

1. Sum the yields ω_{bp}^0 and uncertainties e_{bp} of each background process p in the bin. $\omega_{\text{tot}} = \sum_{p \in \text{bkg}} \omega_{bp}^0$, and $e_{b,\text{tot}} = \sqrt{\sum_{p \in \text{bkg}} e_{bp}^2}$. The values of e_{bp} are obtained using the `TH1::GetBinError` method.
2. If $e_{b,\text{tot}} = 0$, the bin is skipped and no associated nuisance parameters are created.
3. The effective number of unweighted events is defined as $\omega_{b,\text{tot}}^{\text{eff}} = \omega_{b,\text{tot}}^2 / e_{b,\text{tot}}^2$, rounded to the nearest integer.

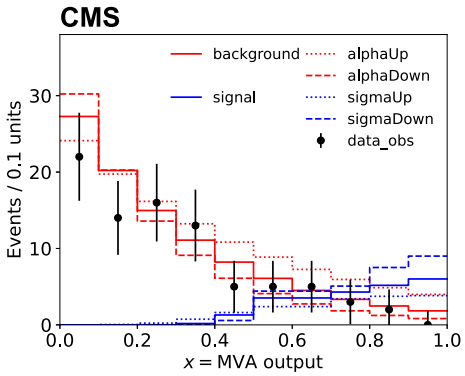

Fig. 1 Histograms used to define the pdfs for Datacard 2. The red and blue histograms show the nominal yields in each bin ω_b^0 for the background and signal processes, respectively. The dotted and dashed lines show the histograms that provide the values of ω_b^+ and ω_b^-, respectively, for each of the systematic uncertainties that modify the shape of the signal and background pdfs. The red dashed and dotted lines are associated with the effect of the nuisance parameter `alpha` on the background process, while the blue dashed and dotted lines are associated with the effect of the nuisance parameter `sigma` affecting the signal process. The black points show the observed number of events in data in each bin. The error bars indicate the statistical uncertainty

The value of $\omega_{b,\text{tot}}^{\text{eff}}$ determines the functional form of $E_b(\vec{\mu}, \nu)$. If $\omega_{b,\text{tot}}^{\text{eff}} > n^{\text{threshold}}$, a single auxiliary observable y is included in the statistical model that is normally distributed with $p(y; \nu) = \mathcal{N}(y; \nu, 1)$. The nuisance parameter ν determines the value of E_b:

$$E_b(\vec{\mu}, \vec{\nu}, \nu) = \nu \left(\sum_p e_{bp}^2 M_p^2(\vec{\mu}, \vec{\nu}) \right)^{\frac{1}{2}}, \tag{16}$$

where $\vec{\nu}$ here represents all of the other nuisance parameters in the statistical model. The default values in COMBINE are $\nu = y = 0$. If instead $\omega_{b,\text{tot}}^{\text{eff}} \leq n^{\text{threshold}}$, a vector of auxiliary observables $\vec{y}$ with one entry per process is included in the statistical model. For processes that have a number of effective events ω_{bp}^2/e_{pb}^2 less than $n^{\text{threshold}}$, the observable is Poisson distributed with $p(y_\alpha; \nu_\alpha) = \mathcal{P}(y_\alpha; \nu_\alpha)$. For processes with $w_{bp}^2/e_{pb}^2 \geq n^{\text{threshold}}$, the observable is normally distributed with $p(y_\beta; \nu_\beta) = \mathcal{N}(y_\beta; \nu_\beta, 1)$. The nuisance parameters determine the value of E_b:

$$\begin{aligned} E_b(\vec{\mu}, \vec{\nu}, \vec{\nu}_\alpha, \vec{\nu}_\beta) = &\sum_\alpha \left(\frac{\nu_\alpha}{y_\alpha} - 1 \right) \omega_{b\alpha} M_\alpha(\vec{\mu}, \vec{\nu}) \\ &+ \sum_\beta \nu_\beta e_{b\beta} M_\beta(\vec{\mu}, \vec{\nu}), \end{aligned} \tag{17}$$

where the indices α and β run over processes for which the auxiliary observables are Poisson and normally distributed, respectively. The default values of $\nu_\beta = y_\beta$ are zero, while the default values of y_α and ν_α are set to the effective number of unweighted events in the `TH1` histogram objects used to evaluate ω_{bp}^0.

4.2.2 *Parametric Shape Analysis*

A parametric shape analysis is one that uses analytic functions rather than histograms to describe the pdfs of continuous primary observables. In these cases, the primary observable x in each channel can be univariate or multivariate. For example, in the measurements of Higgs boson cross sections in the four-lepton decay mode, the primary observable is bivariate composed of the invariant mass of the four leptons and a kinematic discriminator designed to separate the signal and background processes [27]. The data in parametric shape analyses can be binned, as in the case of template-based analyses, or unbinned. Uncertainties affecting the expected distributions of the signal and background processes can be implemented directly as uncertainties in the parameters of those analytical functions.

Datacard 3 defines a parametric analysis with a single channel and two processes: one signal process and one background process. The datacard is similar to what would be used in a search for a narrow resonance over a smooth background, such as in the search for Higgs boson decays to muon pairs [28]. The primary observable is the invariant mass m of the decay products, and the signal distribution depends on the hypothesized mass of the resonance, m_H. The systematic uncertainties include uncertainties that affect the expected rates of the signal and background processes, and uncertainties in the parameters that describe the signal and background pdfs.

For datacards describing parametric-based analyses, the term $p(\vec{x}; \vec{\mu}, \vec{\nu})$ in Eq. (1) is constructed in COMBINE as:

$$p(\vec{x}; \vec{\mu}, \vec{\nu}) = \sum_p \frac{\lambda_p(\vec{\mu}, \vec{\nu}) f_p(x; \vec{\mu}, \vec{\nu})}{\sum_p \lambda_p(\vec{\mu}, \vec{\nu})}, \tag{18}$$

where $f_p(x; \vec{\mu}, \vec{\nu})$ are the pdfs for each process p for a given channel with primary observable x. These pdfs can depend on both the parameters of interest and the nuisance parameters. In parametric analyses, for a specific data set with n entries $\{x_d\}$ where d runs from 1 to n, a Poisson probability $\mathcal{P}\left(n; \sum_p \lambda_p(\vec{\mu}, \vec{\nu})\right)$ is included in the likelihood function in Eq. (2). The Poisson parameters λ_p depend only on the parameters of interest and any nuisance parameters affecting the rate of a given process. These parameters are products of multiplicative factors of the form given in Table 1 and any `RooAbsReal` object named `pdfname_norm` found in the input `RooWorkspace`, as described in the "Shape Analyses" section. When the data are provided as binned data sets with `RooDataHist` objects, the continuous observable x is replaced by a sequence of discrete the bin centers x_b, and the

pdfs are evaluated at these values. If the bins are relatively narrow, this approximation provides a good estimate of the probability density. When this approximation is not accurate, the COMBINE package provides a custom class named `RooParametricShapeBinPdf` that can wrap any univariate `RooAbsPdf` object such that

$$f_p(x_b; \vec{\mu}, \vec{\nu}) = \int_{\text{bin } b} f_p(x; \vec{\mu}, \vec{\nu})\, \mathrm{d}x. \tag{19}$$

The datacard lines for parametric shape analyses need two names to identify the ROOFIT object representing the pdf for a given process in each channel, separated by a colon in the following format:

```
shapes <process> <channel> <file> <
  ↪ workspace_name>:<pdf_name>
```

The label `workspace_name` identifies the input workspace, which is a `RooWorkspace` object containing the ROOFIT objects, while the second label `pdf_name` identifies the `RooAbsPdf` or `RooAbsData` contained therein. The pdfs f_p for each process p are defined by the objects identified with `pdf_name`.

Lines 5–7 indicate the name of the input ROOT file `parametric-analysis-datacard-input.root` that contains an input workspace (`w`), with `RooAbsPdf` objects named `sig` and `bkg` defining the pdfs for the signal and background processes, respectively. The contents of this workspace are summarized in Table 3.

There is a single `RooRealVar` object named `m` in the workspace, which represents the primary observable for the analysis. The `RooDataSet` object named `data_obs` provides the observed data. In this datacard the number of events observed in data, as indicated in line 10, is specified as 567, so COMBINE expects this data set to contain 567 entries, each with its own value of the `RooRealVar` object `m`.

Figure 2 shows the pdfs for the signal and background processes, and the distribution of m in observed data. The data are unbinned and treated as such in COMBINE; the binning is performed exclusively for visualization. The effects of varying the `sigma` and `alpha` nuisance parameters on the pdfs for the signal and background processes are also shown.

In this datacard, the signal process is parameterized as a normal distribution with a mean given by the hypothesized signal mass value `MH`. This variable is used in COMBINE when interpreting the command line argument value `--mass`. The value of the hypothesized signal mass will be fixed to the value specified by the `--mass` option unless `MH` is specifically listed as a parameter of interest in the physics model (see the "Physics Models" section). The background is an exponential distribution $p(m) \propto \mathrm{e}^{\alpha m}$, with a single nuisance parameter defined in the workspace as a `RooRealVar`

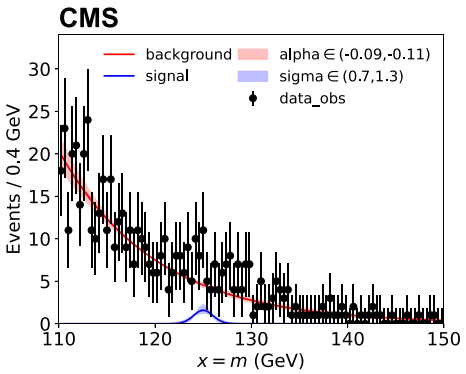

Fig. 2 Distributions of the invariant mass observable for the signal and background processes defined in Datacard 3. The red and blue curves show the parametric functions used to define the probability density for the invariant mass for the background and signal processes, respectively, at the default values of the nuisance parameters, normalized to their expected total yields. The blue shaded band shows the variation of the signal pdf when `sigma` is varied between 0.7 and 1.3. The red shaded region shows the variation of the background pdf when `alpha` is varied within 10% of its default value of −0.1. The black points show the distribution of the observed data. The binning and error bars are only for visualization and neither are used by COMBINE to build the likelihood function

object named `alpha`. The workspace also contains a nuisance parameter named `bkg_norm` that multiplies the background rate.

Parametric systematic uncertainties are included using datacard lines with the syntax:

```
<name> param <V> [<U>, <-UDown/+UUp>]
```

These datacard lines directly encode uncertainties in the parameters of the signal and background pdfs. For each of these lines, an additional term $\mathcal{N}(y; \nu, u)$, is included when constructing the statistical model in Eq. (1). The default values for y and ν are set to the value `V` and u is the value of `U` specified in the datacard line. Line 18 of Datacard 3 indicates that there is a `RooRealVar` object describing the parameter `sigma`, contained in the workspace, that is associated with a normal distribution with mean specified by `V`$= 1.0$ and standard deviation specified by `U`$= 0.1$. Asymmetric uncertainties in the parameter can be defined by using the syntax `-UDown/+UUp`$= -1/+1$ standard deviations in the relevant datacard line. The corresponding term $p(y; \nu)$ in the statistical model constructed by COMBINE is a dimidated Gaussian distribution [29]. It is possible to include linear correlations between the parameters by first diagonalizing the covariance among them and encoding the resulting linear combinations of parameters as the nuisance parameters that are declared in the datacard.

To specify that a parameter should be assigned a uniform distribution for $p(y; \nu)$, the datacard line should be:

```
<name> flatParam
```

Datacard 3 Parametric analysis datacard - `datacard-3-parametric-analysis.txt`

```
imax 1
jmax 1
kmax 2
# ---------------
shapes data_obs    bin1 parametric-analysis-datacard-input.root w:data_obs
shapes signal      bin1 parametric-analysis-datacard-input.root w:sig
shapes background bin1 parametric-analysis-datacard-input.root w:bkg
# ---------------
bin                  bin1
observation          567
# ---------------
bin                  bin1    bin1
process              signal  background
process              0       1
rate                 10      1
# ---------------
lumi          lnN    1.1     -
sigma         param  1.0     0.1
alpha         flatParam
bkg_norm      flatParam
```

Table 3 Contents of the `RooWorkspace` object contained in the `parametric-analysis-datacard-input.root` file providing inputs for the parametric analysis datacard

Object name	Type	Description
`m`	`RooRealVar`	The invariant mass observable
`data_obs`	`RooDataSet`	Invariant mass of each event in the observed data
`sig`	`RooGaussian`	Normal pdf describing the probability distribution of the invariant mass for the signal process
`bkg`	`RooExponential`	Exponential pdf describing the probability distribution of the invariant mass for the background process
`MH`	`RooRealVar`	Mean of the signal pdf
`sigma`	`RooRealVar`	Standard deviation of the signal pdf
`alpha`	`RooRealVar`	Slope parameter for the background pdf
`bkg_norm`	`RooRealVar`	Rate multiplier for the total background contribution

The range of allowed parameter values is determined from the `RooRealVar` methods `getMin` and `getMax`, and the default value of the parameter is determined from the `getVal` method. Lines 19 and 20 in Datacard 3 indicate that there are two such parameters. These lines do not count towards the value of `kmax` since for frequentist calculations they can be dropped from the datacard with no effect on the results. The same is not true for Bayesian calculations so it is recommended to include these lines in parametric shape analyses.

4.3 Rate Parameters

Additional multiplicative scale factors can be introduced in the statistical model that directly modify the rate of a given process, in a given channel, by including additional lines in the datacard for any type of analysis, using the following syntax:

```
<name> rateParam <channel> <process> <
 ↪ initial_value> [<min>,<max>]
```

A nuisance parameter is included in the statistical model that multiplies the rate of that particular `process` in the given `channel` by its value. The default value for this parameter is set to the value indicated by `initial_value`. The values of `min` and `max` can be used to set a range for this parameter. The same `rateParam` nuisance parameter can be attached to multiple channels/processes by using a wild card. For example, "`*`" matches any process, while "`QCD_*`" matches any process whose name begins with "`QCD_`". Repeating the same datacard line with different channel/process values is also supported. A uniform probability distribution within the

range given is included in the statistical model if an additional `flatParam` datacard line is included for that parameter.

In addition to direct rate modifiers, modifiers that are functions of other parameters can be included using the following syntax:

```
<name> rateParam <channel> <process> <
  ↪ formula> <args>
```

where `formula` is a string with the syntax used by the ROOT package's `TFormula`, and `args` is a comma separated list of the arguments for the formula. Any nuisance parameter can be included in the `formula`.

Datacard 4 is an example datacard that uses the `rateParam` directive to implement an ABCD background estimation method. In the ABCD method, three samples of the data, labeled B, C, and D, that are depleted in signal contributions, are defined using two independent selection variables to estimate one or more background contributions to the signal-enriched data sample labeled A. In Ref. [30], the ABCD method is used to determine the contribution of the QCD multijet background using the missing transverse momentum and an isolation variable based on the measured energy around the electron in each event. The expected contribution from the QCD multijet background in sample A is estimated using the observed yields in samples B, C, and D by assuming the ratio of yields between samples A and B is the same as that between C and D. An example ABCD analysis can be constructed as a four-channel counting analysis in COMBINE, as described in Datacard 4: The parameters `beta`, `gamma`, and `delta` described by lines 16–18 are simple rate modifiers, β, γ, and δ, that directly scale the yields of the `bkg` process in channels B, C, and D, respectively. The parameter `alpha` is determined by the formula $\alpha = \beta\gamma/\delta$, as defined on line 15 of the datacard. The yield of the `bkg` process in channel A is scaled accordingly by this formula.

Finally, any pre-existing `RooAbsReal` object inside a ROOT file containing a `RooWorkspace` can be imported into the statistical model using the following syntax:

```
<name> rateParam <channel> <process> <
  ↪ rootfile>:<workspacename>
```

The value of `name` should correspond to the name of the `RooAbsReal` object inside the `RooWorkspace`. This allows for arbitrary functions of the statistical model parameters to be used to determine the rate of a particular process in a given channel.

5 Physics Models

The COMBINE package supports the construction and association of parameters of interest to the different signal processes declared in the datacard. This is achieved by defining the parameters of interest and how they affect the signal processes in a Python file: the physics model. Different physics models can be used with the same datacard, which facilitates the reinterpretation of a statistical analysis in terms of different parameters of interest [31]. To specify the physics model to use in the statistical model construction, the option `-P` in `text2workspace.py` should indicate the Python file and the model defined therein:

```
$ text2workspace.py <datacard.txt> -P
  ↪ HiggsAnalysis.CombinedLimit.<
  ↪ PythonFile>:<modelInstance> [--PO <
  ↪ options>]
```

The `PythonFile` should be contained in the `python/` subdirectory of the COMBINE package.

The default physics model is one for which the rate of every signal process is multiplied by a common factor r, which is the only parameter of interest. In this model $\vec{\mu} = r$. The default model is used if the `-P` option is not specified.

With the physics model defined, it is now possible to fully determine the statistical model for an input datacard. The statistical model created by COMBINE for Datacard 1 when using the default physics model is defined as:

$$\begin{aligned} p(n, \vec{y}; r, \vec{\nu}) = {} & \frac{\lambda(r,\vec{\nu})^n}{n!} e^{-\lambda(r,\vec{\nu})} \\ & \times \frac{1}{2\pi} e^{-(\nu_{\mathtt{lumi}} - y_{\mathtt{lumi}})^2} e^{-(\nu_{\mathtt{xs}} - y_{\mathtt{xs}})^2} \\ & \times \frac{(\nu_{\mathtt{nWW}})^{y_{\mathtt{nWW}}}}{y_{\mathtt{nWW}}!} e^{-\nu_{\mathtt{nWW}}}, \end{aligned} \tag{20}$$

where the function $\lambda(r, \vec{\nu})$ is given by:

$$\begin{aligned} \lambda(r, \vec{\nu}) = {} & r\ 1.47\ (1.11)^{\nu_{\mathtt{lumi}}} (1.2)^{\nu_{\mathtt{xs}}} \\ & + 0.22\ (1.11)^{\nu_{\mathtt{lumi}}} \\ & + 0.64\ (1.11)^{\nu_{\mathtt{lumi}}} \frac{\nu_{\mathtt{nWW}}}{4}. \end{aligned} \tag{21}$$

The observable values for the data are set to $n = 0$, $y_{\mathtt{lumi}} = y_{\mathtt{xs}} = 0$, and $y_{\mathtt{nWW}} = 4$.

Generic physics models can be implemented by writing a Python class that defines the parameters of interest and defines how the signal (and background) yields depend on these parameters. There are numerous example physics models provided in the COMBINE package in `python/PhysicsModel.py` and other Python files within the same directory. In the `PhysicsModel:floatingXSHiggs` physics model, the signal processes expected in the datacard correspond to the four dominant Higgs boson production modes at the LHC: gluon fusion, vector boson fusion, and Higgs boson production associated with a vector boson or a pair of top quarks. Their rates are modified by separate scaling parameters; `r_ggH`, `r_qqH`, `r_VH` (or

Datacard 4 Example ABCD datacard - `datacard-4-abcd.txt`

```
imax 4
jmax 1
kmax *
# -------
bin          B       C       D       A
observation 50     100      500     16
# -------
bin          B       C       D       A      A
process      bkg     bkg     bkg     bkg    sig
process      1       2       3       4      0
rate         1       1       1       1      3
# -------
lumi    lnN  -       -       -       -      1.02
eff     lnN  -       -       -       -      1.01
alpha rateParam A bkg (@0*@1/@2) beta,gamma,delta
beta  rateParam B bkg 50
gamma rateParam C bkg 100
delta rateParam D bkg 500
```

`r_WH` and `r_ZH`), and `r_ttH` as defined in the following block of code:

```
def doParametersOfInterest(self):
  """Create parameters of interest (
 ↪ POIs) and other parameters, and
 ↪ define the POI set."""
  # --- Signal Strength as only POI ---
  if "ggH" in self.modes: self.
 ↪ modelBuilder.doVar("r_ggH[1,%s,%s]"
 ↪  % (self.ggHRange[0], self.ggHRange
 ↪ [1]))
  if "qqH" in self.modes: self.
 ↪ modelBuilder.doVar("r_qqH[1,%s,%s]"
 ↪  % (self.qqHRange[0], self.qqHRange
 ↪ [1]))
  if "VH"  in self.modes: self.
 ↪ modelBuilder.doVar("r_VH[1,%s,%s]"
 ↪  % (self.VHRange [0], self.VHRange
 ↪ [1]))
  if "WH"  in self.modes: self.
 ↪ modelBuilder.doVar("r_WH[1,%s,%s]"
 ↪  % (self.WHRange [0], self.WHRange
 ↪ [1]))
  if "ZH"  in self.modes: self.
 ↪ modelBuilder.doVar("r_ZH[1,%s,%s]"
 ↪  % (self.ZHRange [0], self.ZHRange
 ↪ [1]))
  if "ttH" in self.modes: self.
 ↪ modelBuilder.doVar("r_ttH[1,%s,%s]"
 ↪  % (self.ttHRange[0], self.ttHRange
 ↪ [1]))
  poi = ",".join(["r_"+m for m in self.
 ↪ modes])
  if self.pois: poi = self.pois
  ...
```

Each of these is a parameter of interest in the statistical model that COMBINE constructs. The arrays `ggHRange`, `qqHRange`, `VHRange`, `WHRange`, `ZHRange` and `ttHRange` specify the range of each parameter of interest and are defined in the same Python class. The association of each parameter of interest with each production process is defined in the following function:

```
def getHiggsSignalYieldScale(self,
 ↪ production,decay, energy):
  if production == "ggH": return ("
 ↪ r_ggH" if "ggH" in self.modes else
 ↪ 1)
  if production == "qqH": return ("
 ↪ r_qqH" if "qqH" in self.modes else
 ↪ 1)
  if production in [ "WH", "ZH", "VH"
 ↪ ]: return ("r_VH" if "VH" in self.
 ↪ modes else 1)
  if production == "ttH": return ("
 ↪ r_ttH" if "ttH" in self.modes else
 ↪ ("r_ggH" if self.ttHasggH else 1))
  raise RuntimeError, "Unknown
 ↪ production mode '
```

An example datacard with two signal processes and two channels for use with `PhysicsModel:floatingXSHiggs` is shown in Datacard 5. In this datacard, there are two signal processes, `ggH_hgg` and `qqH_hgg`, that correspond to Higgs boson production in the gluon fusion and vector boson fusion modes, respectively, decaying to two photons. The background process is estimated by fitting the data outside the signal peak. The channels correspond to events with an additional pair of jets reconstructed (`dijet`) or otherwise, leading to a more inclusive channel (`incl`). The `FAKE` directive in lines 5 and 6 are used to indicate that each channel of the counting analysis datacard represents a single bin in a histogram. This is required in counting analysis datacards to run some of the diagnostic methods described in the "Goodness-of-Fit Tests and Diagnostics" section, and does not change the statistical model constructed by COMBINE.

Datacard 5 Multi signal datacard - `datacard-5-multi-signal.txt`

```
imax 2
jmax 2
kmax *
------------
shapes *  dijet   FAKE
shapes *  incl    FAKE
------------
bin             incl  dijet
observation     166   8
------------
bin                 incl     incl     incl     dijet    dijet    dijet
process             ggH_hgg  qqH_hgg  bkg      ggH_hgg  qqH_hgg  bkg
process              -1      0        1        -1       0         1
rate                21       1.6      140      0.4      0.95      3.2
------------
QCDscale_ggH lnN    1.12     -        -        1.12     -         -
pdf_gg       lnN    1.08     -        -        1.08     -         -
pdf_qqbar    lnN     -       1.025    -        -        1.025     -
bg_incl      lnN     -       -        1.05     -        -         -
```

It is possible to include generic constraints on the parameters of the physics model. These can be included in the datacard with lines having the following syntax:

```
<name> constr <formula> <args> <delta>
```

where `name` should be a unique identifier for the constraint and `formula` and `args` follow the ROOT `TFormula` syntax. The result is to multiply the probability term $p(\vec{x};\vec{\mu},\vec{\nu})$ in Eq. (1) by the product of constraint terms:

$$p(\vec{x};\vec{\mu},\vec{\nu}) \to p(\vec{x};\vec{\mu},\vec{\nu}) \prod_l \frac{1}{\delta_l\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{g_l(\vec{\mu})}{\delta_l}\right)^2}, \qquad (22)$$

where l runs over the `constr` lines in the datacard. This feature can be used to include additional restrictions on the parameters of the model imposed by external theoretical or experimental constraints. This feature has been used to perform regularization in measurements of unfolded differential Higgs boson cross sections in the $\mathrm{H} \to \tau\tau$ decay mode [32]. As an example, the following datacard line produces a single constraint term in the statistical model with $g(\vec{\mu}) = r_{\mathrm{ggH}} - 2r_{\mathrm{VH}} + r_{\tau\tau\mathrm{H}}$ and $\delta = 0.03$, when used with the `PhysicsModel:floatingXSHiggs` physics model:

```
constraint_higgs constr @0-2*@2+@1
 ↪ r_ggH,r_VH,r_ttH 0.03
```

The order of the list of parameters in the fourth column of the datacard line defines which of the parameters is assigned to each term in $g(\vec{\mu})$.

Throughout this paper, μ generically denotes the first parameter of interest defined in the physics model, while r specifically refers to the single parameter of interest in the default physics model. For any physics model, it is possible to redefine the list of parameters of interest, or their order within the list, using the COMBINE command line option `--redefineSignalPOIs`. Parameters of interest not included in this list are demoted to nuisance parameters. This command may include nuisance parameters, which results in the removal of the probability density in the statistical model for the associated auxiliary observable. This can be used to test how well any parameter of the model can be measured using only the primary observables, and any remaining auxiliary observables of a given data set.

6 How to Run COMBINE

This section gives an overview of the command line executable `combine` provided by the package, which is used to perform a number of different statistical routines using the statistical model constructed by COMBINE. The executable runs using the command:

```
$ combine <datacard.[txt|root]> -M <
 ↪ Method>
```

where the `Method` specifies the statistical calculation to be performed. A list of available options for the executable is displayed by adding the command line option `--help`.

6.1 Generic Minimizer Options

A number of methods available in COMBINE make use of numerical optimization of the likelihood function given in Eq. (2). Typically, these methods make use of the profile negative-log-likelihood function, $-\ln \mathcal{L}(\vec{\mu}, \hat{\hat{\vec{\nu}}}(\vec{\mu}))$, in which

the nuisance parameters are profiled; $\hat{\hat{\nu}}(\vec{\mu})$ are the values of the nuisance parameters ν that maximize the likelihood function at a fixed set of values of the parameters of interest $\vec{\mu}$. The class `CascadeMinimizer` is used to steer these routines and allows for a sequential minimization of $-\ln \mathcal{L}(\vec{\Phi})$ using different algorithms. The class also allows for COMBINE to perform minimization over any discrete nuisance parameters like those needed for the implementation of the discrete profiling method described in Ref. [33]. The combinations of minimizers and algorithms supported in COMBINE are given in Table 4. Details of these algorithms can be found in Refs. [14] and [34].

6.2 Output from COMBINE

Most of the methods available in COMBINE output the results of the computation to the terminal. In addition, these results are also saved in a ROOT file containing a `TTree` called `limit`. The name of this file has the following format:

```
higgsCombine$NAME.MethodName.mH$MASS.[
 ↪ $WORD$VALUE].root
```

where `NAME` is set to the value passed to the option `-n`, which defaults to `Test`, and `$WORD$VALUE` is any user-defined keyword `WORD` in the datacard that has been set to a particular value `VALUE` using the command line option `--keyword-value WORD=VALUE`. The option can be repeated multiple times for multiple keywords. The keyword-value pairs are also stored in the output ROOT file. The option `-m` sets the value of `$MASS` and the parameter `MH` if it is included in the statistical model. Its value is written to the branch `mh` in the output `TTree` object.

The structure of the `TTree` contained in the output ROOT file is given in Table 5. The branch names `limit` and `limitErr` always refer to the main result of any statistical routine and an estimate of its uncertainty, whether or not that routine calculates a frequentist limit.

6.3 Pseudo-Data Generation

By default, COMBINE performs the calculation using the observed data. For example, in frequentist methods the observed data are automatically used to construct the likelihood function in Eq. (2). It is possible to run these routines instead using pseudo-data sets to determine the distributions of various statistical quantities such as maximum likelihood estimates, or perform optimization studies that are blind to the observed data.

The option `--toys` is used to instruct COMBINE to first generate one or more pseudo-data sets, which are used in place of the observed data. There are two variants of this procedure available in COMBINE. In the first, specified by `--toys <N>` with $N > 0$, COMBINE generates N pseudo-data sets from the statistical model and runs the specified statistical routine once per data set. The pseudo-data set is constructed by generating random values of the observables $\vec{x}$. The random number seed for the generation can be modified with the option `--seed <value>`, which allows the user to ensure each run of the COMBINE command produces different pseudo-data sets, or identical pseudo-data sets. This allows certain calculations to be split into parallel tasks in the former case and for performing diagnostic studies in particular pseudo-data sets in the latter. The output `TTree` contains one entry for each of these data sets when generated with $N > 0$.

In the second variant, specifying `--toys -1` produces an Asimov data set [35]. An Asimov data set is defined as that in which the maximum likelihood estimates for all of the model parameters are equal to the values used to generate the data set. Asimov data sets are used for deriving the expected outcome of frequentist calculations such as in the determination of upper limits and confidence intervals. Where valid, their use makes these calculations much more computationally efficient than using the first variant of pseudo-data generation.

In COMBINE, Asimov data sets are constructed using the expectation value for the probability $p(\vec{x}; \vec{\mu}, \vec{\nu})$ in counting analyses and shape analyses for which the data are binned, or by using a large sample of weighted events, which are generated according to $p(\vec{x}; \vec{\mu}, \vec{\nu})$. In the former case, the values of n_b for the Poisson probabilities in Eq. (10) are allowed to be non-integer, since the terms $n_b!$ are ignored when constructing a likelihood for the statistical routines available in COMBINE. The event weights in the latter case are identical for every event in the same channel, and are accounted for when estimating parameter uncertainties from Asimov data sets [36]. The default values of the parameters of interest are used when generating pseudo-data sets in both variants. The command line option `--setParameters <x>=<value_x>,<y>=<value_y>,...` can be used to specify other values of the parameters to be used for the generation of the pseudo-data sets.

By default, pseudo-data sets in COMBINE make use of marginalization where the value of each nuisance parameter ν_k is randomly sampled from its probability distribution $p(\nu_k|y_k)$ in Eq. (3) before generating values for the primary observables in each pseudo-data set. The auxiliary observables $\vec{y}$ are set to their default values. This can be modified by specifying the option `--toysFrequentist`. With this option and $N > 0$, a parametric bootstrap [37,38] is instead performed: each y is generated according to its probability distribution $p(y; \nu)$, where the value of the nuisance parameter ν is set to the maximum likelihood estimate obtained using the observed data and fixing the parameters of interest to the values specified in the `--setParameters` option. If instead $N = -1$, the value of y is this maximum likelihood

Table 4 Available combinations of minimizer and algorithms in COMBINE

Minimizer	References	Available algorithms
`Minuit`	[34]	`Migrad`, `Simplex`, `Combined`, `Scan`
`Minuit2`	[34]	`Migrad`, `Simplex`, `Combined`, `Scan`
`GSLMultiMin`	[14]	`ConjugateFR`, `ConjugatePR`, `BFGS`, `BFGS2`, `SteepestDescent`

Table 5 `TTree` branches contained in the output ROOT file from COMBINE

Branch name	Type	Description
`limit`	`Double_t`	Main result of the statistical routine being performed
`limitErr`	`Double_t`	Estimated uncertainty in the result
`mh`	`Double_t`	Value specified with `--mass` command line option. The default value is 120
`iToy`	`Int_t`	Pseudo-data set identifier if running with `--toys`
`iSeed`	`Int_t`	Random seed specified with `-s`
`t_cpu`	`Float_t`	Estimated processing time
`t_real`	`Float_t`	Elapsed wall-clock time for routine
`quantileExpected`	`Float_t`	Quantile identifier for methods that calculate expected and observed results. The meaning is method-dependent. Negative values are reserved for entries that are not related to quantiles of a calculation. The default is set to -1 and specifies that the entry corresponds to the result obtained from the observed data

estimate for the corresponding nuisance parameter ν. When specifying the option `--bypassFrequentistFit`, the default values of the nuisance parameters instead of the maximum likelihood estimates are used.

It is possible to separate the tasks of generating the pseudo-data sets from running statistical routines, by first saving the pseudo-data sets to a ROOT file on disk, and then passing them to any COMBINE method later. The following sections describe the most commonly used statistical methods available in COMBINE. These represent only part of the functionality of the tool and users are recommended to consult the online documentation for a full description of its capabilities.

6.4 Frequentist Limits and Confidence Intervals

The `HybridNew` method can be used for calculating upper limits with statistical models created with the default physics model, and for calculating confidence intervals for models with one or more parameters of interest using the Feldman–Cousins procedure [39]. Each of these methods utilizes a test statistic based on the likelihood function given in Eq. (2).

In the case of upper limits, the single parameter of interest μ corresponds to the parameter r in the default physics model. A number of prescriptions using different test statistics are supported in COMBINE as follows:

- LEP-style: `--testStat LEP --generateNuisances=1 --fitNuisances=0`. The test statistic is defined using the ratio of likelihoods:

$$q_{\text{LEP}}(\mu) = -2\ln\left(\frac{\mathcal{L}(\mu=0,\vec{\nu}_0)}{\mathcal{L}(\mu,\vec{\nu}_0)}\right), \tag{23}$$

  where $\vec{\nu}_0$ are the default values of the nuisance parameters. This test statistic was used in the searches for the Higgs boson at the LEP [40].
- TEV-style: `--testStat TEV --generateNuisances=0 --generateExternalMeasurements=1 --fitNuisances=1`. The test statistic is defined using a ratio of profile likelihoods:

$$q_{\text{TEV}}(\mu) = -2\ln\left(\frac{\mathcal{L}(\mu=0,\hat{\hat{\vec{\nu}}}(0))}{\mathcal{L}(\mu,\hat{\hat{\vec{\nu}}}(\mu))}\right), \tag{24}$$

  where $\hat{\hat{\vec{\nu}}}(0)$, and $\hat{\hat{\vec{\nu}}}(\mu)$ are the values of the nuisance parameters that maximize the likelihood function at $\mu = 0$ and μ, respectively. For the purposes of pseudo-data generation, the nuisance parameters are set to the values of $\hat{\hat{\vec{\nu}}}(\mu)$ obtained using the observed data, while the values of $\vec{y}$ are randomly sampled according to their probability densities. This test statistic was used in the context of Higgs boson searches at the Tevatron [41].

- LHC-style: `--LHCmode LHC-limits`. The test statistic is defined using a ratio of profile likelihoods:

$$\tilde{q}_{\mathrm{LHC}}(\mu) = \begin{cases} -2\ln\left(\dfrac{\mathcal{L}(\mu, \hat{\hat{\vec{\nu}}}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\vec{\nu}})}\right) & \text{if } 0 \le \hat{\mu} \le \mu, \\ -2\ln\left(\dfrac{\mathcal{L}(\mu, \hat{\hat{\vec{\nu}}}(\mu))}{\mathcal{L}(0, \hat{\hat{\vec{\nu}}}(0))}\right) & \text{if } \hat{\mu} < 0, \\ 0 & \text{if } \hat{\mu} > \mu, \end{cases} \tag{25}$$

  where $\hat{\mu}$ is the maximum likelihood estimator for μ. The same result is obtained with the option `--testStat LHC --generateNuisances=0 --generateExternalMeasurements=1 --fitNuisances=1`. The values of the nuisance parameters $\vec{\nu}$ that maximize the likelihood $\mathcal{L}$ assuming a specific value of μ and for $\mu = \hat{\mu}$ are denoted $\hat{\hat{\vec{\nu}}}(\mu)$ and $\hat{\vec{\nu}}$, respectively.

The test statistic in Eq. (25) is the most widely used for setting upper limits in searches for new physics at the LHC [3]. In COMBINE, these test statistics can be modified to perform two point hypothesis tests such as those performed to test different hypotheses for the spin and parity of the Higgs boson [42]. In the LEP-style prescription, the nuisance parameter values for each pseudo-data set are randomly sampled from their distributions $p(\nu_k|y_k)$ in Eq. (3). In the TEV- and LHC-style prescriptions, a parametric bootstrap is used where the nuisance parameters are set to the values of $\hat{\hat{\vec{\nu}}}(\mu)$ obtained using the observed data before generating the primary and auxiliary observables from their probability distributions. It is also possible to integrate out (marginalize) the nuisance parameters in a Bayesian-inspired procedure [43] and advocated for upper limits in high-energy physics analyses [44]. This amounts to calculating

$$\mathcal{L}_{\mathrm{int}}(\vec{\mu}) = \int \mathcal{L}(\vec{\Phi}) \prod_k \pi_k(\nu_k) d\vec{\nu}, \tag{26}$$

where $\mathcal{L}_{\mathrm{int}}(\vec{\mu})$ is the integrated likelihood [43]. In COMBINE, this is achieved by modifying the options to `--generateNuisances=1` and `--generateExternalMeasurements=0`. This is required to calculate upper limits in cases where there are few or no background events in channels dominated by signal processes such as in the search for lepton flavor violating tau lepton decays performed by the CMS Collaboration [45].

For a specific value of μ, the value of the test statistic using the observed data $q^{\mathrm{obs}}(\mu)$ is calculated, along with the two p-values p_μ and p_b defined as:

$$p_\mu = \begin{cases} \int_{q_{\mathrm{x}}^{\mathrm{obs}}(\mu)}^{\infty} f(q_{\mathrm{x}}(\mu)|\mu)\, \mathrm{d}q_{\mathrm{x}} & \text{if x=LHC,} \\ \int_{-\infty}^{q_{\mathrm{x}}^{\mathrm{obs}}(\mu)} f(q_{\mathrm{x}}(\mu)|\mu)\, \mathrm{d}q_{\mathrm{x}} & \text{if x=TEV or LEP,} \end{cases} \tag{27}$$

and

$$p_b = \begin{cases} \int_{0}^{q_{\mathrm{x}}^{\mathrm{obs}}(\mu)} f(q_{\mathrm{x}}(\mu)|0)\, \mathrm{d}q_{\mathrm{x}} & \text{if x=LHC,} \\ \int_{q_{\mathrm{x}}^{\mathrm{obs}}(\mu)}^{\infty} f(q_{\mathrm{x}}(\mu)|0)\, \mathrm{d}q_{\mathrm{x}} & \text{if x=TEV or LEP,} \end{cases} \tag{28}$$

where the distributions of the test statistics $f(q_{\mathrm{x}}(\mu)|\mu)$ and $f(q_{\mathrm{x}}(\mu)|0)$ are determined using pseudo-data sets, assuming the value of μ indicated and the values of $\vec{\nu}$ and $\vec{y}$ depending on the options used, as described above. From these p-values, the tool calculates the $\mathrm{CL_s}$ criterion [20,46,47] defined by $\mathrm{CL_s} = p_\mu/(1 - p_b)$. The tool then uses a bisection algorithm to find the value of μ for which $\mathrm{CL_s} = \alpha$ corresponding to the upper limit on μ at the $100(1-\alpha)\%$ confidence level (CL). The tool can also calculate the median, and 2.5, 16, 85, or 97.5% quantiles of the expected distribution of the upper limit assuming $\mu = 0$, by including the option `--expectedFromGrid=<X>`, where `X` is either 0.5, 0.025, 0.16, 0.84, or 0.975, respectively.

The 95% CL upper limit on μ in the example template analysis can be calculated using Datacard 2 with the command:

```
$ combine datacard-2-template-analysis.
  ↪ txt -M HybridNew --LHCmode LHC-
  ↪ limits --rMax 2.0 --clsAcc 0.01
```

The option `--rMax` specifies the maximum value of μ in the algorithm that performs the search for the upper limit. The results of the calculation are output to the terminal as:

```
>   -- Hybrid New --
> Limit: r < 0.346362 +/- 0.0134581 @
  ↪ 95
> Done in 0.31 min (cpu), 0.32 min (
  ↪ real)
```

The bisection algorithm for calculating the upper limit terminates when the estimate of the precision on the upper limit value is below a specified threshold, or when the precision cannot be improved further. The user can specify the following options to control this behavior:

- `--rAbsAcc` and `--rRelAcc`: Define the accuracy on the upper limit, μ_{up} at which the algorithm terminates. The default values are 0.1 and 0.05, respectively, meaning that the search terminates when the absolute accuracy $\Delta\mu_{\mathrm{up}} < 0.1$ or the relative accuracy $\Delta\mu_{\mathrm{up}}/\mu_{\mathrm{up}} < 0.05$, where $\Delta\mu_{\mathrm{up}}$ is the estimated uncertainty on the upper limit.

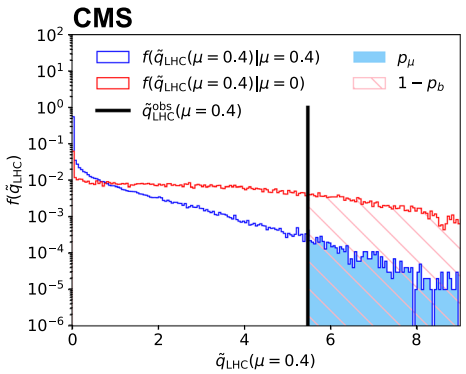

Fig. 3 Distributions of $\tilde{q}_{\mathrm{LHC}}(\mu = 0.4)$ from 100,000 pseudo-data sets for $\mu = 0$ (red histogram) and $\mu = 0.4$ (blue histogram) using the analysis described in Datacard 2. The observed value of the test statistic is indicated by the black vertical line and the regions used to determine $1 - p_b$ and p_μ are indicated by the pink hatched and light blue shaded regions, respectively

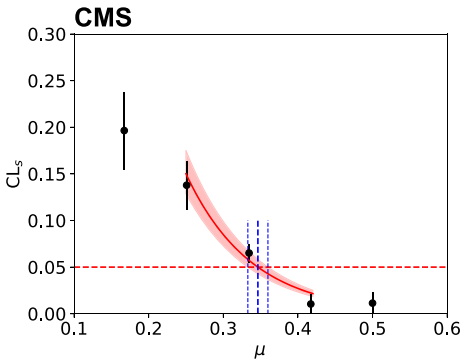

Fig. 4 Calculated CL_s as a function of μ, used to determine the 95% CL upper limit for Datacard 2. The solid red line is used to interpolate the CL_s values to find the crossing at 0.05, and the shaded band indicates the uncertainty in the interpolation that is used to estimate an uncertainty in the upper limit. The vertical dashed blue lines show the derived upper limit and the estimated uncertainty due to the number of pseudo-data sets used in the calculation

- `--clsAcc`: Determines the absolute accuracy up to which the value of CL_s (or p_μ) values are computed when searching for the upper limit. The default is 0.5%.
- `-T` or `--toysH`: Determines the minimum number of pseudo-data sets that are generated for each value of μ with a default value of 500.

The distributions of the test statistic for the pseudo-data sets generated assuming each value of μ that is tested during the bisection algorithm and $\mu = 0$ can be saved in the output file by specifying the option `--saveHybridResult`. Figure 3 shows the distributions of the test statistic $\tilde{q}_{\mathrm{LHC}}(\mu = 0.4)$ for $\mu = 0$ and 0.4 in pseudo-data sets obtained with Datacard 2.

To further improve the accuracy when searching for the upper limit, COMBINE interpolates across several results to estimate μ_{up}. The interpolation uses an exponential function that is fit to the set of results that are closest to the chosen CL and the range in μ used for the fit is determined by the accuracy specified in the command line. A plot of the calculated CL_s value as a function of μ can be produced using the option `--plot=name.png`. Figure 4 shows the calculation of the upper limit at the 95% CL using the CL_s criterion with Datacard 2.

The `AsymptoticLimits` method can be used to calculate the upper limits in statistical models that use the default physics model with the LHC-style prescription. This is the default method that will be run if the command line option `-M` is not specified. In this method, the limit calculation relies on asymptotic approximations for the distributions of the $\tilde{q}_{\mathrm{LHC}}(\mu)$, following the prescription described in Ref. [35]. The tool also calculates the median and 2.5, 16, 84 and 97.5% quantiles of the expected distribution of the upper limit assuming $\mu = 0$, using an Asimov data set. The output `TTree` contains an entry for each of these results, which can be identified by the `quantileExpected` branch.

By default upper limits are calculated at the 95% CL ($\alpha = 0.05$). This can be modified using the option `--cl=<X>` where `X` is $(1 - \alpha)$. Upper limits are calculated using the CL_s criterion by default. Alternatively, it is possible to only use p_μ by specifying the option `--rule Pmu` in the command line. It is also possible to calculate the values of CL_s and p_μ for a single value of μ, bypassing the bisection algorithm, by specifying `--singlePoint <r>`, where `r` is the desired value of μ.

The `HybridNew` method can also be used to compute Feldman–Cousins intervals by specifying the option `--LHCmode LHC-feldman-cousins`. This method allows for calculating confidence intervals with accurate coverage both in scenarios with low event counts or where physical boundaries are placed on the parameters of interest μ. For example, this method has been used in Higgs boson property measurements at CMS [48]. The following procedure can be used to produce one-dimensional confidence intervals or multidimensional confidence regions for physics models with multiple parameters of interest:

- For each parameter point, run the `HybridNew` method with the option `--LHCmode LHC-feldman-cousins --singlePoint <mu1>=<v1>,<mu2>=<v2>,<mu3>=<v3>,... --saveHybridResult` to generate the distributions of the test statistic:

$$q_{\mathrm{FC}}(\vec{\mu}) = -2\ln\left(\frac{\mathcal{L}(\vec{\mu}, \hat{\hat{\vec{\nu}}}(\vec{\mu}))}{\mathcal{L}_\Omega(\hat{\vec{\mu}}, \hat{\vec{\nu}})}\right), \tag{29}$$

where $\vec{\mu} = \mu_1, \mu_2, \ldots$ are the parameters of interest and $\mathcal{L}_\Omega(\hat{\vec{\mu}}, \hat{\vec{\nu}})$ indicates the maximum of the likelihood func-

tion within the bounded region Ω of its parameters. This region can be defined using the command line option `--setParameterRanges <mu1>=<mu1min>,<mu1max>:<mu2>=<mu2min>,<mu2max>:...` This step also calculates the value of the test statistic for the observed data $q_{\mathrm{FC}}^{\mathrm{obs}}(\vec{\mu})$.

- Collect the resulting output files into a single ROOT file and find the set of points for which

$$\int_{q_{\mathrm{FC}}^{\mathrm{obs}}(\vec{\mu})}^{\infty} f(q_{\mathrm{FC}}(\vec{\mu})|\vec{\mu})\,\mathrm{d}q_{\mathrm{FC}} > \alpha, \tag{30}$$

  to form the $100(1-\alpha)\%$ CL allowed region.
- The output ROOT file contains the test statistic value for each pseudo-data set as ROOFIT `RooStats::HybridResult` objects. These can be used to determine confidence intervals or contours at different values of α.

6.5 Significance Calculation

The `HybridNew` method is also used to calculate the significance of the observed data when considered against a null hypothesis. For statistical models constructed using the default physics model, this estimates the significance of the presence of a signal contribution in the data where the null hypothesis represents the absence of the signal. By specifying the options `--LHCmode LHC-significance`, COMBINE generates pseudo-data under the background-only hypothesis ($\mu = 0$) and evaluates the test statistic q_0 defined by:

$$q_0 = \begin{cases} -2\ln\left(\frac{\mathcal{L}(0,\hat{\hat{\nu}}(0))}{\mathcal{L}(\hat{\mu},\hat{\nu})}\right) & \text{if } \hat{\mu} > 0 \\ 0 & \text{otherwise,} \end{cases} \tag{31}$$

for each pseudo-data set. The value of the test statistic for the observed data q_0^{obs} is also calculated in order to determine the p-value

$$p_0 = \int_{q_0^{\mathrm{obs}}}^{\infty} f(q_0|0)\,\mathrm{d}q_0, \tag{32}$$

where $f(q_0|0)$ is the distribution of the test statistic determined using the pseudo-data sets. Figure 5 shows the observed value of q_0 and the distribution of q_0 in pseudo-data sets assuming $\mu = 0$ for Datacard 3.

The value of p_0 and corresponding significance of the signal in the example parametric analysis can be calculated using 100,000 pseudo-data sets with Datacard 3 using the command line below:

```
$ combine datacard-3-parametric-
 ↪ analysis.txt -M HybridNew --LHCmode
 ↪  LHC-significance -T 100000 --mass
 ↪ 125
```

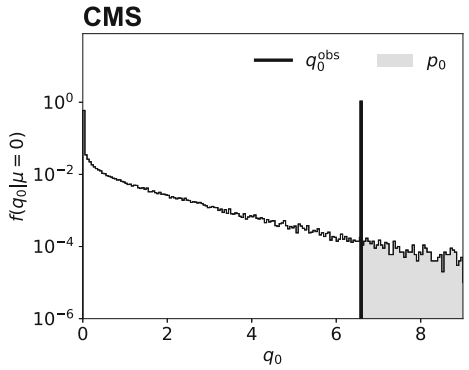

Fig. 5 Distribution of q_0 in 100,000 pseudo-data sets from Datacard 3. The observed value of the test statistic is indicated by the black vertical line and the region used to determine p_0 is indicated by the light gray shaded region

The results are output to the terminal as:

```
>   -- Hybrid New --
> Significance: 2.54397
 ↪  -0.0146063/+0.0151701
> Null p-value: 0.00548 +/- 0.000233452
> Done in 5.95 min (cpu), 7.57 min (
 ↪ real)
```

The value of p_0 is converted into a significance using a standard normal distribution [20]. The method `Significance` can be used in order to speed up the calculation using the asymptotic approximation for the distribution $f(q_0|\mu = 0)$ given in Ref. [35] thereby avoiding the need to generate pseudo-data in cases where the number of events in data is large. For the same datacard, the asymptotic approximation for the significance can be calculated with:

```
$ combine datacard-3-parametric-
 ↪ analysis.txt -M Significance --mass
 ↪  125
```

The result of the calculation using the asymptotic approximation is output to the terminal as:

```
>   -- Significance --
> Significance: 2.56729
> Done in 0.00 min (cpu), 0.00 min (
 ↪ real)
```

The significance result and its estimated uncertainty are stored in the output ROOT file in the `limit` and `limitErr` branches, respectively. Using the `--pval` option, the p-value is stored instead of the significance.

6.6 Bayesian Upper Limits and Credible Regions

Bayesian calculations in COMBINE are based on the posterior probability $p(\vec{\mu})$ defined by:

$$p(\vec{\mu}) = \frac{1}{p(\{\vec{x}\}_d)} \int \mathcal{L}\left(\vec{\Phi}\right) \prod_k \pi_k(\nu_k)\rho(\vec{\mu})\,\mathrm{d}\vec{\nu}, \tag{33}$$

where the index d runs over the events in the observed data set $\{\vec{x}\}_d$, and $p(\{\vec{x}\}_d)$ is defined such that $\int p(\vec{\mu})\,\mathrm{d}\vec{\mu} = 1$. The prior term for the parameters of interest $\rho(\vec{\mu})$ must be specified by the user. By default this prior is assumed to be uniform over the ranges (a_i, b_i) specified for each parameter of interest μ_i:

$$\rho(\vec{\mu}) \propto \prod_i \mathcal{U}(\mu_i; a_i, b_i). \tag{34}$$

Bayesian upper limits are calculated in COMBINE using either the `BayesianSimple` method for relatively simple statistical models or the `MarkovChainMC` method for models with multiple parameters of interest or nuisance parameters, which perform the marginalization over the nuisance parameters. For statistical models with a single parameter of interest μ, the prior can be modified via the command line using the option `--prior <prior>` with the following options available:

- `flat`: the default uniform prior.
- `1/sqrt(r)`: inverse square root prior. This is the Jeffreys prior for the mean of a Poisson distribution [49].
- Any valid ROOT `TMath::Formula` expression with `@0` as the parameter of interest.
- Any string that names a ROOFIT `RooAbsPdf` object contained in one of the input workspaces. This option can also be used for statistical models with multiple parameters of interest.

Both methods compute the $100(1-\alpha)\%$ credible upper limit on the parameter of interest μ_{up} as:

$$\int_{-\infty}^{\mu_{\mathrm{up}}} p(\mu)\,\mathrm{d}\mu = 1 - \alpha. \tag{35}$$

The value of α can be modified by specifying the option `--cl=`$1-\alpha$.

The `BayesianSimple` method computes μ_{up} using numerical integration, while the `MarkovChainMC` method uses Markov chain integration [50].

The number of steps in the Markov chain and number of chains to compute can be specified via the command line, as well as the number of steps to ignore from the start of the chain. The user can specify a proposal algorithm by which the Markov chain evolves using the option `--proposal <algorithm>` with the following options:

- `uniform`: Selects the next parameter point in the chain at random.
- `gaus`: Uses a product of independent normal distributions, one for each nuisance parameter where the standard deviation of the distribution for each variable is set to some fraction of the range of the parameter defined by the option `--propHelperWidthRangeDivisor`.
- `ortho`: This is the default proposal and is similar to the `gaus` proposal except that at each point in the chain, only a single parameter is varied.
- `fit`: With this proposal, COMBINE computes the Hessian matrix of $-2\ln p(\vec{x}; \mu, \vec{\nu})$ with respect to the nuisance parameters $\vec{\nu}$ to construct the proposal function. The accuracy can be improved by including the option `--runMinos`.

The value of μ_{up} and an estimate of its uncertainty are obtained from the average over N independent Markov chains, specified by the command line option `--tries <N>`. The 95% Bayesian upper limit on the default physics model parameter r for Datacard 1 using 100 Markov chains can be calculated using the following:

```
$ combine datacard-1-counting-
  ↪ experiment.txt -M MarkovChainMC --
  ↪ tries 100
```

These values are saved in the output ROOT file and output to the terminal as below:

```
>  -- MarkovChainMC --
> Limit: r < 2.21031 +/- 0.0133576 @ 95
> Done in 0.05 min (cpu), 0.05 min (
  ↪ real)
```

If using the `MarkovChainMC` method, it is also possible to store the resulting Markov chains in the output file from COMBINE using the option `--saveChain`. This allows for estimating the posterior distribution $p(\vec{\mu})$ for one or more parameters of the model and deriving credible intervals or regions.

6.7 Maximum Likelihood Estimates and Scans

Likelihood-based parameter estimation is performed in COMBINE using the `MultiDimFit` method. This method can be used to calculate maximum likelihood estimates for the parameter values and their uncertainties through several different approaches. This method has been used by the CMS Collaboration to provide measurements in a number of different scenarios, including Higgs boson production and decay rates, and measurements of the Higgs boson couplings [51]. The `MultiDimFit` method is used to evaluate the negative-log-likelihood function, $-\ln\mathcal{L}(\vec{\Phi})$, and obtain maximum likelihood estimates and estimates of confidence intervals for the parameters of interest $\vec{\mu}$ using the profile likelihood ratio:

$$q(\vec{\mu}) = -\ln\left(\frac{\mathcal{L}(\vec{\mu}, \hat{\hat{\vec{\nu}}}(\vec{\mu}))}{\mathcal{L}(\hat{\hat{\vec{\mu}}}, \hat{\hat{\vec{\nu}}})}\right), \tag{36}$$

where $\hat{\vec{\mu}}$ are the maximum likelihood estimates for the parameters of interest, and $\hat{\hat{\vec{\nu}}}(\vec{\mu})$ and $\hat{\vec{\nu}}$ are the values of the nuisance parameters for which $\mathcal{L}$ is maximized for a specific set of parameter values $\vec{\mu}$ and for the maximum likelihood estimates $\hat{\vec{\mu}}$, respectively. The process of finding the parameter values that maximize the likelihood function is typically referred to as a "fit". In Eq. (36) there are two such fits. The one in the denominator is often referred to as the "overall best fit". The default parameter values are commonly referred to as "pre-fit", while the maximum likelihood estimates are commonly referred to as "post-fit". Throughout this and following sections, the process of determining the maximum likelihood estimates is referred to as maximum likelihood optimization so as not to confuse this procedure with the more general "goodness of fit" methods described in the "Goodness-of-Fit Tests and Diagnostics" section.

The following choices for the `--algo` option are supported in COMBINE:

- `none`: this is the default algorithm. The algorithm finds the parameter values $\hat{\hat{\Phi}}$ that maximize $\mathcal{L}(\vec{\Phi})$ and reports the maximum likelihood estimates of the parameters of interest. For a model with N parameters of interest, the output `TTree` contains N branches, one for each parameter of interest with the maximum likelihood estimates. The output of this algorithm can be used as the starting point for the other algorithms to reduce their evaluation time.
- `singles`: the algorithm determines the maximum likelihood estimates for each parameter of interest μ and sequentially determines 68% confidence intervals for each parameter of interest. The output `TTree` contains one branch for each parameter of interest. One of the entries contains the maximum likelihood estimates with `quantileExpected` set to -1. Two additional entries for each parameter of interest provide the upper μ^+ and lower μ^- bounds of the 68% CL interval for that parameter determined as the range of that parameter for which $q(\mu) < 1/2$.
- `cross`: the algorithm determines the maximum likelihood estimates for each parameter of interest and a set of intervals for each parameter μ_l for which $q(\mu_l) < \frac{c}{2}$, where c is determined by $\int_c^\infty \chi^2(x;n)\,\mathrm{d}x = \alpha$ and $\chi^2(x;n)$ is a chi-squared distribution with the number of degrees of freedom equal to the number of parameters of interest n. The output `TTree` has one entry with the maximum likelihood estimates for the parameters of interest $\hat{\mu}$, and two entries for each parameter of interest, corresponding to the points that define the interval.
- `contour2d`: for statistical models with two parameters of interest, this algorithm constructs two dimensional contours bounding the regions for which $q(\vec{\mu}) < \frac{c}{2}$, where c is computed from $\int_c^\infty \chi^2(x;2)\,\mathrm{d}x = \alpha$. The output contains values corresponding to the maximum likelihood estimates with `quantileExpected` set to -1, and additional entries that define the contour. For $\alpha = 0.32$ the contour produced corresponds to $c = 2.3$ [20].
- `random`: in this algorithm, the value of $q(\vec{\mu})$ is evaluated for N uniformly distributed random points in the space spanned by the parameters of interest. The number of points is set by the option `--points=<N>`.
- `fixed`: in this algorithm, the value of $q(\vec{\mu})$ is calculated at a specified point in the space spanned by the parameters of interest. The fixed point is specified using the `--fixedPointPOIs` option.
- `grid`: the value of $q(\vec{\mu})$ is evaluated on a grid of parameter of interest points with N points in total. The number of points N is specified with the option `--points=<N>`. It is possible to partition the scan into multiple runs of COMBINE using the options `--firstPoint <n>` and `--lastPoint <m>`, where $0 \leq$`n`$<$`m`$\leq N$.

The `fixed` and `grid` algorithms can be used to evaluate the likelihood function and $q(\vec{\mu})$ for any value of the statistical model parameters. In each of the above options the value of α can be defined using the option `--cl=`$1-\alpha$. The output `TTree` branch named `deltaNLL` stores the value of $q(\vec{\mu})$. Additional branches can be included to store the values of any nuisance parameter, any `RooCategory` object, or any `RooFormulaVar` object contained in the workspace produced by COMBINE by adding the options `--saveSpecifiedNuis`, `--saveSpecifiedIndex` or `--saveSpecifiedFunc`, respectively.

Confidence intervals calculated with COMBINE using these algorithms will yield results with approximately correct coverage in the absence of large non-Gaussian uncertainties [52–54]. If the overall best fit values lie on physical boundaries in the parameter space, the Feldman–Cousins procedure described in the "Frequentist Limits and Confidence Intervals" section should be used to obtain intervals with improved coverage properties.

The maximum likelihood estimates for the parameters r_{ggH} and r_{qqH} and estimates of their uncertainties can be obtained using Datacard 5 with the `PhysicsModel:floatingXSHiggs` physics model using the commands below,

```
$ text2workspace.py datacard-5-multi-
  ↪ signal.txt -P HiggsAnalysis.
  ↪ CombinedLimit.PhysicsModel:
  ↪ floatingXSHiggs --PO modes=ggH,qqH
  ↪ -o datacard-5-multi-signal.root --
  ↪ mass 125
$ combine datacard-5-multi-signal.root
  ↪ -M MultiDimFit --algo singles --
  ↪ mass 125
```

The output from COMBINE is given below:

```
>   --- MultiDimFit ---
> best fit parameter values and profile
 ↪ -likelihood uncertainties:
>     r_ggH :     +0.882    -0.749/+0.795
 ↪ (68%)
>     r_qqH :     +4.683    -2.746/+3.464
 ↪ (68%)
> Done in 0.00 min (cpu), 0.04 min (
 ↪ real)
```

For each parameter, the result is presented as a measurement $\hat{\mu} - \Delta^{-}\mu / + \Delta^{+}\mu$, where the values of $\Delta^{\pm}\mu$ are determined from the 68% confidence intervals as:

$$\Delta^{\pm}\mu = \left|\mu^{\pm} - \hat{\mu}\right|. \tag{37}$$

The relative precision on r_{qqH} is better than for r_{ggH} due to the larger signal to background ratio in the `dijet` channel of this datacard. The parameters are correlated due to the contribution of each process across both channels. This can be seen by considering the shape of the function $q(r_{\mathrm{ggH}}, r_{\mathrm{qqH}})$. Figure 6 shows $q(r_{\mathrm{ggH}}, r_{\mathrm{qqH}})$ using the same datacard and physics model. The points indicate the output from COMBINE using the `grid` and `contour2d` algorithms. The box shown in the figure is constructed from the set of intervals calculated using the `cross` algorithm with $(1-\alpha) = 0.68$.

A subset of parameters for any of the algorithms can be specified using multiple instances of the option `--parameters`. In this case, all other parameters of interest are set equal to their default values throughout the calculation. This behavior can be adjusted by including the option `--floatOtherPOI=1`, which instructs COMBINE to include the remaining parameters of interest in the list of profiled parameters for the purposes of the calculation of profile likelihoods. This is not the same as using the `--redefineSignalPOIs` option as selecting subsets of parameters does not remove the associated probability distribution $p(y; \nu)$ when nuisance parameters are included. Figure 7 shows the values of $q(r_{\mathrm{ggH}}, \hat{r}_{\mathrm{qqH}})$ and $q(r_{\mathrm{qqH}}, \hat{r}_{\mathrm{ggH}})$ obtained from COMBINE using the `grid` algorithm, and the 68% CL intervals on each parameter obtained using the `singles` algorithm. In this case, the options `--parameters=<X> --floatOtherPOI=1` are included where `X` is either `r_ggH` or `r_qqH`.

6.8 Goodness-of-Fit Tests and Diagnostics

The `GoodnessOfFit` method can be used to perform a test of compatibility between the observed data and the statistical model. While the statistical model constructed by COMBINE is used as the null distribution, goodness-of-fit (GoF) tests do not have a well-defined alternative hypothesis, unlike conventional hypothesis tests.

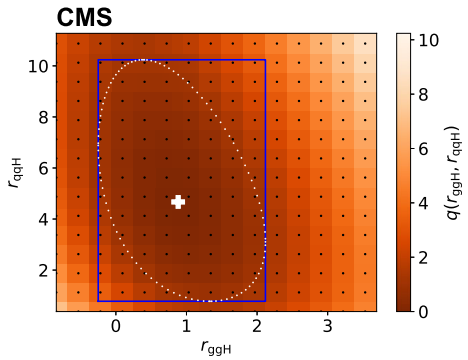

Fig. 6 Values of $q(r_{\mathrm{ggH}}, r_{\mathrm{qqH}})$ for Datacard 5 in a model with two parameters of interest r_{ggH} and r_{qqH}. The orange scale shows the values obtained in COMBINE at the set of points indicated by the black dots, using the `grid` algorithm. The blue box is constructed using the `cross` algorithm with $(1-\alpha) = 0.68$. The white cross and white dots indicate, respectively, the maximum likelihood estimates for r_{ggH} and r_{qqH} from the best fit, and the 68% CL confidence region obtained using the `contour2d` algorithm defined as the values of $(r_{\mathrm{ggH}}, r_{\mathrm{qqH}})$ for which $q(r_{\mathrm{ggH}}, r_{\mathrm{qqH}}) = 2.3$

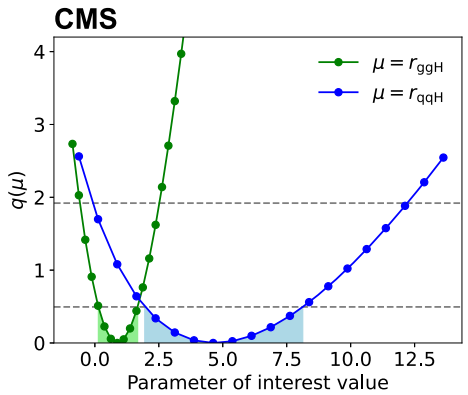

Fig. 7 Example of $q(r_{\mathrm{ggH}}, \hat{r}_{\mathrm{qqH}})$ and $q(r_{\mathrm{qqH}}, \hat{r}_{\mathrm{ggH}})$ obtained from COMBINE with Datacard 5. The points indicate the values at which the functions are evaluated using the `grid` algorithm, and the shaded region indicates the 68% CL intervals on each parameter obtained using the `singles` algorithm. The horizontal dashed lines indicate the values of $q(\mu)$ used to define 68% and 95% CL intervals

The `GoodnessOfFit` method evaluates a test statistic for the observed data and generates the expected distribution of the test statistic under the null distribution using pseudo-data. In COMBINE this is done in two separate commands:

```
$ combine -M GoodnessOfFit <datacard.
 ↪ txt> --algo=<test-statistic>

$ combine -M GoodnessOfFit <datacard.
 ↪ txt> --algo=<test-statistic> --toys
 ↪   <N>
```

The first command evaluates the observed test statistic and the second determines the distribution of the test statistic under the null hypothesis using pseudo-data sets. The statis-

tical model parameters used to specify the null hypothesis can be set using the `--setParameters` option.

All of the supported test statistics are based on binned data. For statistical models that include parametric shape analyses, an automatic binning is applied to the data to evaluate the probability density and calculate the test statistic, unless a specific binning is specified for the observables using the `RooAbsReal::setBinning` method. If no binning is specified, 100 uniform bins will be used for each observable. The following test statistics for the `--algo` option are supported in COMBINE:

- saturated: evaluates a GoF test statistic defined as:

$$t = -2 \ln \frac{\mathcal{L}(\hat{\hat{\Phi}})}{\mathcal{L}_S}, \tag{38}$$

  where $\mathcal{L}_S$ represents the likelihood for the saturated model [55].
- KS: evaluates the Kolmogorov–Smirnov test statistic [19, 56,57], based on the largest difference between the cumulative distribution function and the empirical distribution function across all bins in all channels.
- AD: evaluates the Anderson–Darling test statistic based on the integral of the difference between the cumulative distribution function and the empirical distribution function across all bins in all channels. This test statistic gives more importance to the tails of the distribution in data [58,59].

The output `TTree` contains a branch called `limit` that contains the value of the test statistic in each pseudo-data set if running with the option `--toys`, or from the observed data. These can be used to determine a p-value under the null hypothesis. Figure 8 shows the distribution of the saturated test statistic t in pseudo-data sets using the statistical model constructed by COMBINE from Datacard 2 with default nuisance parameter values. The p-value is 0.73 and is computed from the distribution and the observed value of t. In this example, the template analysis in Datacard 2 was used to generate pseudo-data sets and calculate the corresponding p-value. The distribution of the test statistic has a peak close to the number of bins for the observable in the datacard, which is expected for this test statistic. The method of pseudo-data generation can be modified using the options described in the "Pseudo-Data Generation" section.

The `ChannelCompatibilityCheck` method is used to evaluate the compatibility between measurements of the signal rate in the N_c separate channels defined in the statistical model. The method can be used with the default physics

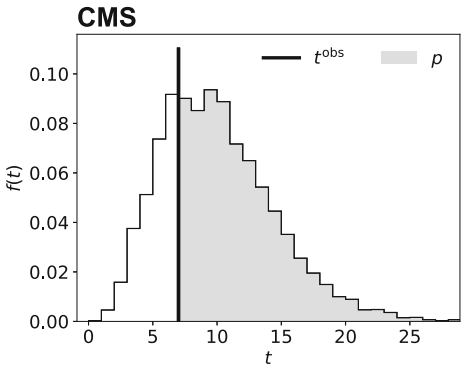

Fig. 8 Distribution of the saturated test statistic t in 10,000 pseudo-data sets using Datacard 2. The observed value of the test statistic is indicated by the black vertical line and the region used to determine p is indicated by the light gray shaded region

model and calculates the value of

$$q = -2 \ln \left(\frac{\mathcal{L}(\hat{r})}{\mathcal{L}(\hat{r}_1, \hat{r}_2, ..., \hat{r}_{N_c})} \right), \tag{39}$$

where each r_i, for $i = 1, 2, \ldots, N$, multiplies the rate of all signal processes in a specific channel i, and $\hat{r}$ denotes the maximum likelihood estimate.

The channel compatibility of the multi-signal Higgs boson analysis from Datacard 5 can be calculated using the command line below:

```
$ combine datacard-5-multi-signal.txt -
 ↪ M ChannelCompatibilityCheck --mass
 ↪ 125
```

The output from these commands is given below:

```
>  --- ChannelCompatibilityCheck ---
> Nominal fit  : r =  1.4812
 ↪ -0.5987/+0.6644
> Alternate fit: r =  3.5554
 ↪ -1.8576/+2.3619   in channel dijet
> Alternate fit: r =  1.1506
 ↪ -0.6393/+0.6939   in channel incl
> Chi2-like compatibility variable:
 ↪ 1.5417
> Done in 0.00 min (cpu), 0.02 min (
 ↪ real)
```

These results are not expected to match those from the maximum likelihood estimates for the parameters r_{ggH} and r_{qqH} as the parameters in that model are related to the process signal multipliers instead of multipliers for the total signal in each channel. The compatibility variable in the output is the value of q. Assuming the values of $\hat{r}_i$ are normally distributed, this quantity can be converted to a p-value $p = \int_q^\infty \chi^2(x; N_c - 1)\, \mathrm{d}x$ where χ^2 is a chi-squared distribution with $N_c - 1$ degrees of freedom.

The output `TTree` contains the value of q in the `limit` branch. If the option `--saveFitResult`is specified,

Table 6 Additional output file contents from the `FitDiagnostics` method

Object	Description
`nuisances_prefit`	`RooArgSet` containing the default values of $\vec{\nu}$, and their uncertainties estimated from their probability distributions
`fit_b`	`RooFitResult` object containing the outcome of the maximum likelihood optimization with r fixed to zero
`fit_s`	`RooFitResult` object containing the outcome of the maximum likelihood optimization with r allowed to vary
`tree_prefit`	`TTree` of default values of $\vec{\nu}$ and $\vec{y}$
`tree_fit_b`	`TTree` of maximum likelihood estimates $\hat{\hat{\vec{\nu}}}(0)$ and values of $\vec{y}$ from the maximum likelihood optimization fixing r to zero
`tree_fit_sb`	`TTree` of maximum likelihood estimates $\hat{\vec{\nu}}$ and values of $\vec{y}$ from the maximum likelihood optimization allowing r to vary
Objects below are present only if the option `--plots` is included	
`covariance_fit_b`	`TH2D` covariance matrix of the parameters from the maximum likelihood optimization fixing r to zero
`covariance_fit_s`	`TH2D` covariance matrix of the parameters from the maximum likelihood optimization allowing r to vary
`channel_observable_prefit`	`RooPlot` plot of the probability density, normalized to the yield in data, projected onto each observable in each channel along with a histogram of the data. The statistical model parameters are set to their default values
`channel_observable_fit_b`	Same as `channel_observable_prefit`, except the statistical model parameters are set to their maximum likelihood estimates from the optimization fixing r to zero
`channel_observable_fit_s`	Same as `channel_observable_prefit`, except the statistical model parameters are set to their maximum likelihood estimates from the optimization allowing r to vary

the output ROOT file also contains two `RooFitResult` objects `fit_nominal`, and `fit_alternate` with the results of the two maximum likelihood optimizations used to calculate the numerator and denominator in q, respectively.

The `FitDiagnostics` method provides a number of diagnostic routines to investigate the statistical model. The method can be used with the default physics model and performs two optimizations of the parameters $\vec{\Phi}$ that maximize the function $\mathcal{L}(\vec{\Phi})$, the first with the parameter of interest r allowed to vary and the second with r set at a constant value of zero. The output `TTree` contains the maximum likelihood estimator for r and an estimate of its uncertainty. In addition to the usual output file, an additional ROOT file is produced with the name `fitDiagnostics$NAME.root`. This file contains additional details about the optimizations performed that can be used to investigate the statistical model and the optimization procedure, the details of which are given in Table 6.

In addition to maximum likelihood estimates for all of the model parameters, these results contain estimates of their uncertainties. By default, the uncertainties for all nuisance parameters $\Delta\nu$ are estimated as $\Delta\nu_k = H_{kk}^{-1}$, where H_{ij} is the Hessian matrix of $q(\hat{\hat{\mu}})$:

$$H_{ij} = \left.\frac{\partial^2 q}{\partial\nu_i \partial\nu_j}\right|_{\vec{\Phi}=\hat{\hat{\vec{\Phi}}}}, \tag{40}$$

where $\hat{\hat{\vec{\Phi}}}$ are the maximum likelihood estimates of the statistical model parameters. The accuracy can be improved by including the option `--minos all`, which determines the uncertainty for each nuisance parameter as:

$$\Delta^{\pm}\nu_k = \left|\nu^{\pm} - \hat{\nu}\right|, \tag{41}$$

where the range $\left(\nu^{-}, \nu^{+}\right)$ is determined as the region for which $q(\nu) < 1/2$, similar to the `singles` algorithm described in the "Maximum Likelihood Estimates and Scans" section.

The results contained in this file can be used to study the additional constraints imposed by the observed data by considering the difference between the maximum likelihood estimates of the nuisance parameters ("post-fit") and their uncertainties compared to their default ("pre-fit") values.

The `FitDiagnostics` method can also be used to calculate nuisance parameter impacts [60] for statistical models

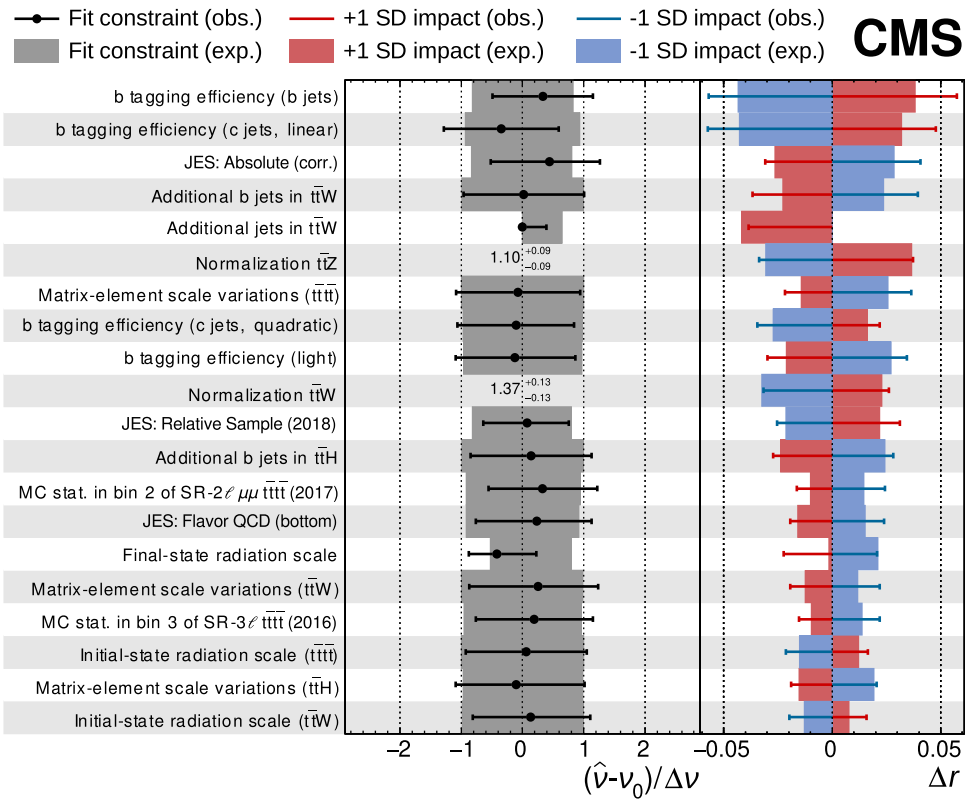

Fig. 9 Example of nuisance parameter uncertainties and impacts calculated in COMBINE for the observation of four top quark production. Each row gives the name of the nuisance parameter, the difference in its maximum likelihood estimate $\hat{\nu}$ with respect to its default value ν_0 relative to its uncertainty $\Delta\nu$, and the impact with respect to the default physics model parameter Δr. The nuisance parameter constraints and impacts are calculated using the observed data set (obs.) and an Asimov dataset constructed assuming standard model production of four top quarks (exp.). The red and blue lines in each row represent the positive impact Δr^{+} and negative impact Δr^{-}, respectively, for the observed data. Similarly, the red and blue shaded boxes represent the same quantities for the Asimov dataset. The error bars on the fit constraint values indicate the ratio of $\Delta^{-}\nu$ or $\Delta^{+}\nu$, to their default values. The two numerical values displayed in the figure give the value of $\hat{\nu}^{+\Delta^{+}\nu}_{-\Delta^{-}\nu}$ for two rate parameters, which do not have well-defined default uncertainty values. Figure adapted from Ref. [22]

with one or more parameters of interest, defined with respect to any particular parameter of interest μ for each nuisance parameter ν_k as:

$$\Delta\mu^{\pm} = \hat{\mu}\left(\hat{\nu}_k \pm \Delta^{\pm}\nu_k\right) - \hat{\mu}, \tag{42}$$

where $\hat{\mu}(\nu_k \pm \Delta_k^{\pm})$ is the value of μ that maximizes the likelihood function when the nuisance parameter is shifted from its maximum likelihood estimator value by its uncertainty. Large impacts typically result from nuisance parameters that contribute significantly to the total uncertainty in μ and the sign of the impact indicates the sign of the (anti-)correlation between the parameter of interest and that nuisance parameter. These impacts provide diagnostic information that encapsulates both the constraints imposed on the nuisance parameter by the data and the correlation of that nuisance parameter with a particular parameter of interest. Figure 9 shows the impacts for the default physics model parameter r calculated in the statistical analysis leading to the observation of four top quark production by the CMS Collaboration [22]. These impacts are calculated using both the observed data set ("obs.") and an Asimov data set constructed assuming standard model production of four top quarks to obtain the expected ("exp.") impacts.

By including the option `--saveShapes` with this method, COMBINE saves additional ROOT `TH1F` histogram objects in the output file. These represent the contributions from each process in each channel of the statistical model evaluated at the default values of the parameters, and at the maximum likelihood estimates from the two optimizations. This allows a visualization of the pdfs of the primary observables at the pre-fit and post-fit values of the statistical model parameters individually for each process and each channel. For statistical models that include parametric shape analyses, an automatic binning is used to construct the histograms unless a particular binning is specified for the observables using the `RooAbsReal::setBinning`

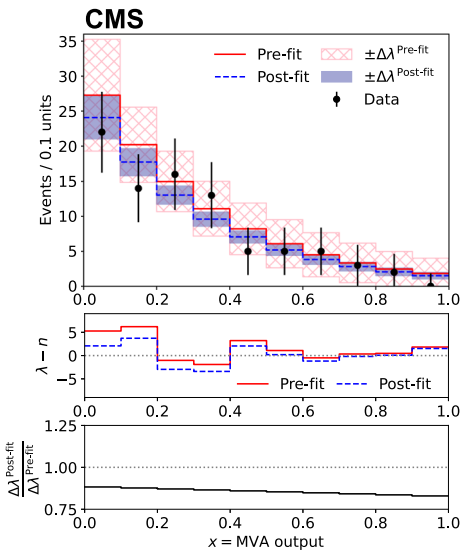

Fig. 10 Distributions of the observable x for the data and background process in Datacard 2 and their uncertainties. The upper panel shows the distribution for the default values of the nuisance parameters (red solid line, pre-fit) and for the maximum likelihood estimates assuming no signal (blue dashed line, post-fit). The pink hatched and blue shaded bands show the estimate of the uncertainty in each bin for the pre-fit and post-fit distributions, respectively. The middle panel shows the difference between the expected number of events in the background processes (λ) and the data (n) in the pre-fit (red solid line) and post-fit (blue dashed line) cases, and the lower panel shows the ratios of the estimated uncertainties of the post-fit distribution $\Delta\lambda^{\text{Post-fit}}$ to the pre-fit $\Delta\lambda^{\text{Pre-fit}}$ in each bin

method. If no binning is specified, 100 uniform bins will be used for each observable. The total signal, total background, and total contributions in each channel are also saved as separate `TH1F` histogram objects. By adding the option `--saveWithUncertainties`, the output also includes estimates of the covariance between each of the bin yields, accounting for any correlations between the parameters of the statistical model. This is achieved using the `RooFitResult::randomizePars()` method from the results of the optimization. Figure 10 shows the distribution of the observable x for the data and the background process in Datacard 2 using the default physics model. The uncertainties are estimated by sampling from the distributions $p(\nu|y)$ in the pre-fit case and from the covariance matrix of the model parameters obtained from a fit to the data shown assuming $r = 0$ (post-fit). The change in the expected number of events in each bin λ and their uncertainties $\Delta\lambda$ visually indicate the additional constraint imposed by the data on the statistical model parameters.

The `FitDiagnostics` method can be run with the `--toys` option such that the process normalization and bin yields in each channel resulting from each of the two optimizations performed using pseudo-data sets are stored, allowing for more detailed studies of the statistical model.

7 Summary

After a decade of development, the COMBINE package has become the main tool used for statistical analysis of data by the CMS Collaboration. The tool is based on the ROOT [1], ROOFIT [2], and ROOSTATS [2] software packages to provide a command-line interface to several common statistical workflows used in high-energy physics. The statistical model is constructed from a text file provided by the user and a configurable physics model that encodes the parameters of interest and the nuisance parameters that model systematic uncertainties. The COMBINE package can perform a variety of statistical procedures including calculating confidence or credible intervals, evaluating profile likelihoods, and performing goodness of fit tests. The online documentation [12] contains comprehensive information on the capabilities and instructions for running the COMBINE package, as well as detailed instructions for its installation.

Acknowledgements We congratulate our colleagues in the CERN accelerator departments for the excellent performance of the LHC and thank the technical and administrative staffs at CERN and at other CMS institutes for their contributions to the success of the CMS effort. In addition, we gratefully acknowledge the computing centers and personnel of the Worldwide LHC Computing Grid and other centers for delivering so effectively the computing infrastructure essential to our analyses. Finally, we acknowledge the enduring support for the construction and operation of the LHC, the CMS detector, and the supporting computing infrastructure provided by the following funding agencies: SC (Armenia), BMBWF and FWF (Austria); FNRS and FWO (Belgium); CNPq, CAPES, FAPERJ, FAPERGS, and FAPESP (Brazil); MES and BNSF (Bulgaria); CERN; CAS, MoST, and NSFC (China); MINCIENCIAS (Colombia); MSES and CSF (Croatia); RIF (Cyprus); SENESCYT (Ecuador); ERC PRG, RVTT3 and MoER TK202 (Estonia); Academy of Finland, MEC, and HIP (Finland); CEA and CNRS/IN2P3 (France); SRNSF (Georgia); BMBF, DFG, and HGF (Germany); GSRI (Greece); NKFIH (Hungary); DAE and DST (India); IPM (Iran); SFI (Ireland); INFN (Italy); MSIP and NRF (Republic of Korea); MES (Latvia); LMTLT (Lithuania); MOE and UM (Malaysia); BUAP, CINVESTAV, CONACYT, LNS, SEP, and UASLP-FAI (Mexico); MOS (Montenegro); MBIE (New Zealand); PAEC (Pakistan); MES and NSC (Poland); FCT (Portugal); MESTD (Serbia); MCIN/AEI and PCTI (Spain); MOSTR (Sri Lanka); Swiss Funding Agencies (Switzerland); MST (Taipei); MHESI and NSTDA (Thailand); TUBITAK and TENMAK (Turkey); NASU (Ukraine); STFC (United Kingdom); DOE and NSF (USA).
Individuals have received support from the Marie-Curie programme and the European Research Council and Horizon 2020 Grant, contract Nos. 675440, 724704, 752730, 758316, 765710, 824093, 101115353, 101002207, and COST Action CA16108 (European Union); the Leventis Foundation; the Alfred P. Sloan Foundation; the Alexander von Humboldt Foundation; the Science Committee, project no. 22rl-037 (Armenia); the Belgian Federal Science Policy Office; the Fonds pour la Formation à la Recherche dans l'Industrie et dans l'Agriculture (FRIA-Belgium); the Agentschap voor Innovatie door Wetenschap en Technologie (IWT-Belgium); the F.R.S.-FNRS and FWO (Belgium) under the "Excellence of Science—EOS"—be.h project n. 30820817; the Beijing Municipal Science & Technology Commission, No. Z191100007219010 and Fundamental Research Funds for the Central Universities (China); the Ministry of Education, Youth and Sports (MEYS) of the Czech Republic; the Shota Rustaveli National Sci-

ence Foundation, grant FR-22-985 (Georgia); the Deutsche Forschungsgemeinschaft (DFG), under Germany's Excellence Strategy— EXC 2121 "Quantum Universe"—390833306, and under project number 400140256-GRK2497; the Hellenic Foundation for Research and Innovation (HFRI), Project Number 2288 (Greece); the Hungarian Academy of Sciences, the New National Excellence Program—ÚNKP, the NKFIH research grants K 131991, K 133046, K 138136, K 143460, K 143477, K 146913, K 146914, K 147048, 2020-2.2.1-ED-2021-00181, and TKP2021-NKTA-64 (Hungary); the Council of Science and Industrial Research, India; ICSC—National Research Centre for High Performance Computing, Big Data and Quantum Computing, funded by the EU NexGeneration program (Italy); the Latvian Council of Science; the Ministry of Education and Science, project no. 2022/WK/14, and the National Science Center, contracts Opus 2021/41/B/ST2/01369 and 2021/43/B/ST2/01552 (Poland); the Fundação para a Ciência e a Tecnologia, grant CEECIND/01334/2018 (Portugal); the National Priorities Research Program by Qatar National Research Fund; MCIN/AEI/10.13039/501100011033, ERDF "a way of making Europe", and the Programa Estatal de Fomento de la Investigación Científica y Técnica de Excelencia María de Maeztu, grant MDM-2017-0765 and Programa Severo Ochoa del Principado de Asturias (Spain); the Chulalongkorn Academic into Its 2nd Century Project Advancement Project, and the National Science, Research and Innovation Fund via the Program Management Unit for Human Resources & Institutional Development, Research and Innovation, grant B37G660013 (Thailand); the Kavli Foundation; the Nvidia Corporation; the SuperMicro Corporation; the Welch Foundation, contract C-1845; and the Weston Havens Foundation (USA).

Author contributions We do not provide individual contributions.

Funding Open access funding provided by CERN (European Organization for Nuclear Research).

Declarations

Competing Interests The authors declare no competing interests.

References

1. Brun R, Rademakers F (1997) ROOT: An object oriented data analysis framework. Nucl. Instrum. Meth. A 389:81. https://doi.org/10.1016/S0168-9002(97)00048-X
2. Verkerke W, Kirkby DP (2003) "The RooFit toolkit for data modeling". In: Proceedings of the 13th International Conference for Computing in High-Energy and Nuclear Physics (CHEP03). arXiv:physics/0306116. [eConf C0303241, MOLT007]. http://inspirehep.net/record/634021
3. ATLAS and CMS Collaborations, and LHC Higgs Combination Group (2011) "Procedure for the LHC Higgs boson search combination in Summer 2011". Technical Report CMS-NOTE-2011-005, ATL-PHYS-PUB-2011-11. https://cds.cern.ch/record/1379837
4. CMS Collaboration (2012) Combined results of searches for the standard model Higgs boson in pp collisions at $\sqrt{s} = 7$ TeV. Phys. Lett. B 710:26. https://doi.org/10.1016/j.physletb.2012.02.064. arXiv:1202.1488
5. CMS Collaboration (2012) Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. Phys. Lett. B 716:30. https://doi.org/10.1016/j.physletb.2012.08.021. arXiv:1207.7235
6. CMS Collaboration (2013) Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV. JHEP 06:081. https://doi.org/10.1007/JHEP06(2013)081. arXiv:1303.4571
7. CMS Collaboration (2019) Combined measurements of Higgs boson couplings in proton-proton collisions at $\sqrt{s} = 13$ TeV. Eur. Phys. J. C 79:421. https://doi.org/10.1140/epjc/s10052-019-6909-y. arXiv:1809.10733
8. CMS Collaboration (2021) Combined searches for the production of supersymmetric top quark partners in proton-proton collisions at $\sqrt{s} = 13$ TeV. Eur. Phys. J. C 81:970. https://doi.org/10.1140/epjc/s10052-021-09721-5. arXiv:2107.10892
9. CMS Collaboration (2023) Measurement of the top quark pole mass using $\mathrm{t\bar{t}}$+jet events in the dilepton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV. JHEP 07:077. https://doi.org/10.1007/JHEP07(2023)077. arXiv:2207.02270
10. ATLAS and CMS Collaborations (2015) Combined measurement of the Higgs boson mass in pp collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS experiments. Phys. Rev. Lett. 114:191803. https://doi.org/10.1103/PhysRevLett.114.191803. arXiv:1503.07589
11. ATLAS and CMS Collaborations (2016) Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV. JHEP 08:45. https://doi.org/10.1007/JHEP08(2016)045. arXiv:1606.02266
12. CMS Collaboration. "Combine". Accessed 2024. http://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/
13. Piparo D, Innocente V, Hauth T (2014) Speeding up HEP experiment software with a library of fast and auto-vectorisable mathematical functions. J. Phys. Conf. Ser 513:052027. https://doi.org/10.1088/1742-6596/513/5/052027
14. Galassi M, et al (2009) "GNU Scientific Library Reference Manual". 3rd edition. ISBN 978-0-9546120-7-8
15. Jacob B, Guennebaud G, et al. "Eigen". Accessed 2024. https://eigen.tuxfamily.org
16. The Boost C++ community. "Boost C++ libraries". http://www.boost.org/. Accessed 2024
17. Free Software Foundation. "GNU Gzip". https://www.gnu.org/software/gzip/
18. Merkel D (2014) "Docker: Lightweight Linux containers for consistent development and deployment". Linux J. **2014**. https://www.linuxjournal.com/content/docker-lightweight-linux-containers-consistent-development-and-deployment
19. James FE (2006) Statistical methods in experimental physics, 2nd edn. World Scientific, Singapore. https://doi.org/10.1142/6096 (**ISBN 978-981-270-527-3**)
20. Particle Data Group, Workman RL, et al (2022) "Review of particle physics". Prog. Theor. Exp. Phys. **2022**, 083C01, https://doi.org/10.1093/ptep/ptac097
21. CMS Collaboration (2018) Observation of Higgs boson decay to bottom quarks. Phys. Rev. Lett. 121:121801. https://doi.org/10.1103/PhysRevLett.121.121801. arXiv:1808.08242
22. CMS Collaboration (2023) Observation of four top quark production in proton-proton collisions at $\sqrt{s} = 13$ TeV. Phys. Lett. B 847:138290. https://doi.org/10.1016/j.physletb.2023.138290. arXiv:2305.13439

23. CMS Collaboration (2012) Search for the standard model Higgs boson decaying to W^+W^- in the fully leptonic final state in pp collisions at $\sqrt{s} = 7$ TeV. Phys. Lett. B 710:91. https://doi.org/10.1016/j.physletb.2012.02.076. arXiv:1202.1489
24. Cranmer K (2014) "Practical statistics for the LHC". In: Proc. 2011 European School of High-Energy Physics, p. 267. https://doi.org/10.5170/CERN-2014-003.267. arXiv:1503.07622
25. Barlow R, Beeston C (1993) Fitting using finite Monte Carlo samples. Comput. Phys. Commun. 77:219. https://doi.org/10.1016/0010-4655(93)90005-W
26. Conway JS (2011) "Incorporating nuisance parameters in likelihoods for multisource spectra". In: PHYSTAT (2011): Workshop on statistical issues related to discovery claims in search experiments and unfolding, p. 115. https://doi.org/10.5170/CERN-2011-006.115. arXiv:1103.0354
27. CMS Collaboration (2021) Measurements of production cross sections of the Higgs boson in the four-lepton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV. Eur. Phys. J. C 81:488. https://doi.org/10.1140/epjc/s10052-021-09200-x. arXiv:2103.04956
28. CMS Collaboration (2021) Evidence for Higgs boson decay to a pair of muons. JHEP 01:148. https://doi.org/10.1007/JHEP01(2021)148. arXiv:2009.04363
29. Barlow R (2003) "Asymmetric errors". In: PHYSTAT (2003): Statistical problems in particle physics, astrophysics, and cosmology. arXiv:physics/0401042. [eConf C030908, WEMT002]
30. CDF Collaboration (1991) A measurement of $\sigma B(W \to e\nu)$ and $\sigma B(Z^0 \to e^+e^-)$ in $\bar{p}p$ collisions at $\sqrt{s} = 1800$ GeV. Phys. Rev. D 44:29. https://doi.org/10.1103/PhysRevD.44.29
31. Cranmer K et al (2022) Publishing statistical models: Getting the most out of particle physics experiments. SciPost Phys. 12:037. https://doi.org/10.21468/SciPostPhys.12.1.037. arXiv:2109.04981
32. CMS Collaboration (2022) Measurement of the inclusive and differential Higgs boson production cross sections in the decay mode to a pair of τ leptons in pp collisions at $\sqrt{s} = 13$ TeV. Phys. Rev. Lett. 128:081805. https://doi.org/10.1103/PhysRevLett.128.081805. arXiv:2107.11486
33. Dauncey PD, Kenzie M, Wardle N, Davies GJ (2015) Handling uncertainties in background shapes: the discrete profiling method. JINST 10:P04015. https://doi.org/10.1088/1748-0221/10/04/P04015. arXiv:1408.6865
34. James F, Roos M (1975) Minuit: A system for function minimization and analysis of the parameter errors and correlations. Comput. Phys. Commun. 10:343. https://doi.org/10.1016/0010-4655(75)90039-9
35. Cowan G, Cranmer K, Gross E, Vitells O (2011) "Asymptotic formulae for likelihood-based tests of new physics". Eur. Phys. J. C **71**, 1554, https://doi.org/10.1140/epjc/s10052-011-1554-0. arXiv:1007.1727. [Erratum: https://doi.org/10.1140/epjc/s10052-013-2501-z]
36. Langenbruch C (2022) Parameter uncertainties in weighted unbinned maximum likelihood fits. Eur. Phys. J. C 82:393. https://doi.org/10.1140/epjc/s10052-022-10254-8. arXiv:1911.01303
37. Efron B (1979) Bootstrap methods: Another look at the jackknife. Ann. Statist. 7:1. https://doi.org/10.1214/aos/1176344552. See "Remark K"
38. Lee SMS, Young GA (2005) Parametric bootstrapping with nuisance parameters. Stat. Probab. Lett. 71:143. https://doi.org/10.1016/j.spl.2004.10.026
39. Feldman GJ, Cousins RD (1998) Unified approach to the classical statistical analysis of small signals. Phys. Rev. D 57:3873. https://doi.org/10.1103/PhysRevD.57.3873. arXiv:physics/9711021
40. LEP Working Group for Higgs boson searches, ALEPH, DELPHI, L3 and OPAL Collaborations (2003) Search for the standard model Higgs boson at LEP. Phys. Lett. B **565**, 61. https://doi.org/10.1016/S0370-2693(03)00614-2. arXiv:hep-ex/0306033
41. CDF and D0 Collaborations, (2013) Higgs boson studies at the Tevatron. Phys. Rev. D **88**(5), 052014. https://doi.org/10.1103/PhysRevD.88.052014. arXiv:1303.6346
42. CMS Collaboration (2015) Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV. Phys. Rev. D 92:012004. https://doi.org/10.1103/PhysRevD.92.012004. arXiv:1411.3441
43. Berger JO, Liseo B, Wolpert RL (1999) Integrated likelihood methods for eliminating nuisance parameters. Stat. Scien. 14:1. https://doi.org/10.1214/ss/1009211804
44. Cousins RD, Highland VL (1992) Incorporating systematic uncertainties into an upper limit. Nucl. Instrum. Meth. A 320:331. https://doi.org/10.1016/0168-9002(92)90794-5
45. CMS Collaboration (2023) Search for the lepton flavor violating $\tau \to 3\mu$ decay in proton-proton collisions at $\sqrt{s} = 13$ TeV. Phys. Lett. B **853**:138633. arXiv:2312.02371
46. Junk T (1999) Confidence level computation for combining searches with small statistics. Nucl. Instrum. Meth. A 434:435. https://doi.org/10.1016/S0168-9002(99)00498-2. arXiv:hep-ex/9902006
47. Read AL (2002) Presentation of search results: The CL_s technique. J. Phys. G 28:2693. https://doi.org/10.1088/0954-3899/28/10/313
48. CMS Collaboration (2014) Measurement of the properties of a Higgs boson in the four-lepton final state. Phys. Rev. D 89:092007. https://doi.org/10.1103/PhysRevD.89.092007. arXiv:1312.5353
49. Jeffreys H (1961) Theory of Probability, 3rd edn. Oxford University Press, Oxford (**ISBN 9780198503682**)
50. Moneta L, et al (2010) "The RooStats project". In 13th International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT2010). SISSA. arXiv:1009.1003. http://pos.sissa.it/archive/conferences/093/057/ACAT2010_057.pdf
51. CMS Collaboration (2022) A portrait of the Higgs boson by the CMS experiment ten years after the discovery. Nature 607:60. https://doi.org/10.1038/s41586-022-04892-x. arXiv:2207.00043
52. Wilks SS (1938) The large-sample distribution of the likelihood ratio for testing composite hypotheses. Ann. Math. Statist. 9:60. https://doi.org/10.1214/aoms/1177732360
53. Wald A (1943) Tests of statistical hypotheses concerning several parameters when the number of observations is large. Trans. Amer. Math. Soc. 54:426. https://doi.org/10.1090/S0002-9947-1943-0012401-3
54. Engle RF (1984) "Chapter 13 Wald, likelihood ratio, and Lagrange multiplier tests in econometrics". In: Handbook of Econometrics, volume 2, p. 775. Elsevier. https://doi.org/10.1016/S1573-4412(84)02005-5
55. Cousins RD (2018) "Lectures on statistics in theory: Prelude to statistics in practice". arXiv:1807.05996
56. Kolmogorov AN (1933) Sulla determinazione empirica di una legge di distribuzione. Giorn. Ist. Ital. Attuari 4:83
57. Smirnov N (1948) Table for estimating the goodness of fit of empirical distributions. Ann. Math. Statist. 19:279. https://doi.org/10.1214/aoms/1177730256
58. Anderson TW, Darling DA (1952) Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes. Ann. Math. Statist. 23:193. https://doi.org/10.1214/aoms/1177729437
59. Stephens MA (1974) EDF statistics for goodness of fit and some comparisons. J. Am. Stat. Assoc. 69:730. https://doi.org/10.2307/2286009
60. ATLAS Collaboration (2015) Search for the $b\bar{b}$ decay of the standard model Higgs boson in associated $(W/Z)H$ production with the ATLAS detector. JHEP **01**, 069. https://doi.org/10.1007/JHEP01(2015)069. arXiv:1409.6212

CMS Collaboration

Yerevan Physics Institute, Yerevan, Armenia
A. Hayrapetyan, A. Tumasyan[1]

Institut für Hochenergiephysik, Vienna, Austria
W. Adam, J. W. Andrejkovic, T. Bergauer, S. Chatterjee, K. Damanakis, M. Dragicevic, P. S. Hussain, M. Jeitler[2], N. Krammer, A. Li, D. Liko, I. Mikulec, J. Schieck[2], R. Schöfbeck, D. Schwarz, M. Sonawane, S. Templ, W. Waltenberger, C.-E. Wulz[2]

Universiteit Antwerpen, Antwerpen, Belgium
M. R. Darwish[3], T. Janssen, P. Van Mechelen

Vrije Universiteit Brussel, Brussel, Belgium
N. Breugelmans, J. D'Hondt, S. Dansana, A. De Moor, M. Delcourt, F. Heyen, S. Lowette, I. Makarenko, D. Müller, S. Tavernier, M. Tytgat[4], G. P. Van Onsem, S. Van Putte, D. Vannerom

Université Libre de Bruxelles, Bruxelles, Belgium
B. Clerbaux, A. K. Das, G. De Lentdecker, H. Evard, L. Favart, P. Gianneios, D. Hohov, J. Jaramillo, A. Khalilzadeh, F. A. Khan, K. Lee, M. Mahdavikhorrami, A. Malara, S. Paredes, M. A. Shahzad, L. Thomas, M. Vanden Bemden, C. Vander Velde, P. Vanlaer

Ghent University, Ghent, Belgium
M. De Coen, D. Dobur, G. Gokbulut, Y. Hong, J. Knolle, L. Lambrecht, D. Marckx, G. Mestdach, K. Mota Amarilo, A. Samalan, K. Skovpen, N. Van Den Bossche, J. van der Linden, L. Wezenbeek

Université Catholique de Louvain, Louvain-la-Neuve, Belgium
A. Benecke, A. Bethani, G. Bruno, C. Caputo, J. De Favereau De Jeneret, C. Delaere, I. S. Donertas, A. Giammanco, A. O. Guzel, Sa Jain, V. Lemaitre, J. Lidrych, P. Mastrapasqua, T. T. Tran, S. Wertz

Centro Brasileiro de Pesquisas Fisicas, Rio de Janeiro, Brazil
G. A. Alves, M. Alves Gallo Pereira, E. Coelho, G. Correia Silva, C. Hensel, T. Menezes De Oliveira, A. Moraes, P. Rebello Teles, M. Soeiro, A. Vilela Pereira[5]

Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil
W. L. Aldá Júnior, M. Barroso Ferreira Filho, H. Brandao Malbouisson, W. Carvalho, J. Chinellato[6], E. M. Da Costa, G. G. Da Silveira[7], D. De Jesus Damiao, S. Fonseca De Souza, R. Gomes De Souza, M. Macedo, J. Martins[8], C. Mora Herrera, L. Mundim, H. Nogima, J. P. Pinheiro, A. Santoro, A. Sznajder, M. Thiel

Universidade Estadual Paulista, Universidade Federal do ABC, São Paulo, Brazil
C. A. Bernardes[7], L. Calligaris, T. R. Fernandez Perez Tomei, E. M. Gregores, I. Maietto Silverio, P. G. Mercadante, S. F. Novaes, B. Orzari, Sandra S. Padula

Institute for Nuclear Research and Nuclear Energy, Bulgarian Academy of Sciences, Sofia, Bulgaria
A. Aleksandrov, G. Antchev, R. Hadjiiska, P. Iaydjiev, M. Misheva, M. Shopova, G. Sultanov

University of Sofia, Sofia, Bulgaria
A. Dimitrov, L. Litov, B. Pavlov, P. Petkov, A. Petrov, E. Shumka

Instituto De Alta Investigación, Universidad de Tarapacá, Casilla 7 D, Arica, Chile
S. Keshri, S. Thakur

Beihang University, Beijing, China
T. Cheng, T. Javaid, L. Yuan

Department of Physics, Tsinghua University, Beijing, China
Z. Hu, Z. Liang, J. Liu, K. Yi[9,10]

Institute of High Energy Physics, Beijing, China
G. M. Chen[11], H. S. Chen[11], M. Chen[11], F. Iemmi, C. H. Jiang, A. Kapoor[12], H. Liao, Z.-A. Liu[13], R. Sharma[14], J. N. Song[13], J. Tao, C. Wang[11], J. Wang, Z. Wang[11], H. Zhang, J. Zhao

State Key Laboratory of Nuclear Physics and Technology, Peking University, Beijing, China
A. Agapitos, Y. Ban, S. Deng, B. Guo, C. Jiang, A. Levin, C. Li, Q. Li, Y. Mao, S. Qian, S. J. Qian, X. Qin, X. Sun, D. Wang, H. Yang, L. Zhang, Y. Zhao, C. Zhou

Guangdong Provincial Key Laboratory of Nuclear Science and Guangdong-Hong Kong Joint Laboratory of Quantum Matter, South China Normal University, Guangzhou, China
S. Yang

Sun Yat-Sen University, Guangzhou, China
Z. You

University of Science and Technology of China, Hefei, China
K. Jaffel, N. Lu

Nanjing Normal University, Nanjing, China
G. Bauer[15], B. Li, J. Zhang

Key Laboratory of Nuclear Physics and Ion-beam Application (MOE), Institute of Modern Physics, Fudan University, Shanghai, China
X. Gao[16]

Zhejiang University, Hangzhou, Zhejiang, China
Z. Lin, C. Lu, M. Xiao

Universidad de Los Andes, Bogota, Colombia
C. Avila, D. A. Barbosa Trujillo, A. Cabrera, C. Florez, J. Fraga, J. A. Reyes Vega

Universidad de Antioquia, Medellin, Colombia
F. Ramirez, C. Rendón, M. Rodriguez, A. A. Ruales Barbosa, J. D. Ruiz Alvarez

Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split, Split, Croatia
D. Giljanovic, N. Godinovic, D. Lelas, A. Sculac

Faculty of Science, University of Split, Split, Croatia
M. Kovac, A. Petkovic, T. Sculac

Institute Rudjer Boskovic, Zagreb, Croatia
P. Bargassa, V. Brigljevic, B. K. Chitroda, D. Ferencek, K. Jakovcic, S. Mishra, A. Starodumov[17], T. Susa

University of Cyprus, Nicosia, Cyprus
A. Attikis, K. Christoforou, A. Hadjiagapiou, C. Leonidou, J. Mousa, C. Nicolaou, L. Paizanos, F. Ptochos, P. A. Razis, H. Rykaczewski, H. Saka, A. Stepennov

Charles University, Prague, Czech Republic
M. Finger, M. Finger Jr., A. Kveton

Universidad San Francisco de Quito, Quito, Ecuador
E. Carrera Jarrin

Egyptian Network of High Energy Physics, Academy of Scientific Research and Technology of the Arab Republic of Egypt, Cairo, Egypt
Y. Assran[18,19], B. El-mahdy, S. Elgammal[19]

Center for High Energy Physics (CHEP-FU), Fayoum University, El-Fayoum, Egypt
M. A. Mahmoud, Y. Mohammed

National Institute of Chemical Physics and Biophysics, Tallinn, Estonia
K. Ehataht, M. Kadastik, T. Lange, S. Nandan, C. Nielsen, J. Pata, M. Raidal, L. Tani, C. Veelken

Department of Physics, University of Helsinki, Helsinki, Finland
H. Kirschenmann, K. Osterberg, M. Voutilainen

Helsinki Institute of Physics, Helsinki, Finland
S. Bharthuar, N. Bin Norjoharuddeen, E. Brücken, F. Garcia, P. Inkaew, K. T. S. Kallonen, T. Lampén, K. Lassila-Perini, S. Lehti, T. Lindén, L. Martikainen, M. Myllymäki, M. M. Rantanen, H. Siikonen, J. Tuominiemi

Lappeenranta-Lahti University of Technology, Lappeenranta, Finland
P. Luukka, H. Petrow

IRFU, CEA, Université Paris-Saclay, Gif-sur-Yvette, France
M. Besancon, F. Couderc, M. Dejardin, D. Denegri, J. L. Faure, F. Ferri, S. Ganjour, P. Gras, G. Hamel de Monchenault, V. Lohezic, J. Malcles, F. Orlandi, L. Portales, A. Rosowsky, M.Ö. Sahin, A. Savoy-Navarro[20], P. Simkina, M. Titov, M. Tornago

Laboratoire Leprince-Ringuet, CNRS/IN2P3, Ecole Polytechnique, Institut Polytechnique de Paris, Palaiseau, France
F. Beaudette, P. Busson, A. Cappati, C. Charlot, M. Chiusi, F. Damas, O. Davignon, A. De Wit, I. T. Ehle, B. A. Fontana Santos Alves, S. Ghosh, A. Gilbert, R. Granier de Cassagnac, A. Hakimi, B. Harikrishnan, L. Kalipoliti, G. Liu, M. Nguyen, C. Ochando, R. Salerno, J. B. Sauvan, Y. Sirois, L. Urda Gómez, E. Vernazza, A. Zabi, A. Zghiche

CNRS, IPHC UMR 7178, Université de Strasbourg, Strasbourg, France
J.-L. Agram[21], J. Andrea, D. Apparu, D. Bloch, J.-M. Brom, E. C. Chabert, C. Collard, S. Falke, U. Goerlach, R. Haeberle, A.-C. Le Bihan, M. Meena, O. Poncet, G. Saha, M. A. Sessini, P. Van Hove, P. Vaucelle

CNRS/IN2P3, Centre de Calcul de l'Institut National de Physique Nucleaire et de Physique des Particules, Villeurbanne, France
A. Di Florio

Institut de Physique des 2 Infinis de Lyon (IP2I), Villeurbanne, France
D. Amram, S. Beauceron, B. Blancon, G. Boudoul, N. Chanon, D. Contardo, P. Depasse, C. Dozen[22], H. El Mamouni, J. Fay, S. Gascon, M. Gouzevitch, C. Greenberg, G. Grenier, B. Ille, E. Jourd'huy, I. B. Laktineh, M. Lethuillier, L. Mirabito, S. Perries, A. Purohit, M. Vander Donckt, P. Verdier, J. Xiao

Georgian Technical University, Tbilisi, Georgia
I. Lomidze, T. Toriashvili[23], Z. Tsamalaidze[17]

I. Physikalisches Institut, RWTH Aachen University, Aachen, Germany
V. Botta, L. Feld, K. Klein, M. Lipinski, D. Meuser, A. Pauls, D. Pérez Adán, N. Röwert, M. Teroerde

III. Physikalisches Institut A, RWTH Aachen University, Aachen, Germany
S. Diekmann, A. Dodonova, N. Eich, D. Eliseev, F. Engelke, J. Erdmann, M. Erdmann, P. Fackeldey, B. Fischer, T. Hebbeker, K. Hoepfner, F. Ivone, A. Jung, M. Y. Lee, F. Mausolf, M. Merschmeyer, A. Meyer, S. Mukherjee, D. Noll, F. Nowotny, A. Pozdnyakov, Y. Rath, W. Redjeb, F. Rehm, H. Reithler, V. Sarkisovi, A. Schmidt, A. Sharma, J. L. Spah, A. Stein, F. Torres Da Silva De Araujo[24], S. Wiedenbeck, S. Zaleski

III. Physikalisches Institut B, RWTH Aachen University, Aachen, Germany
C. Dziwok, G. Flügge, T. Kress, A. Nowack, O. Pooth, A. Stahl, T. Ziemons, A. Zotz

Deutsches Elektronen-Synchrotron, Hamburg, Germany
H. Aarup Petersen, M. Aldaya Martin, J. Alimena, S. Amoroso, Y. An, J. Bach, S. Baxter, M. Bayatmakou, H. Becerril Gonzalez, O. Behnke, A. Belvedere, S. Bhattacharya, F. Blekman[25], K. Borras[26], A. Campbell, A. Cardini, C. Cheng, F. Colombina, S. Consuegra Rodríguez, M. De Silva, G. Eckerlin, D. Eckstein, L. I. Estevez Banos, O. Filatov, E. Gallo[25], A. Geiser, V. Guglielmi, M. Guthoff,

A. Hinzmann, L. Jeppe, B. Kaech, M. Kasemann, C. Kleinwort, R. Kogler, M. Komm, D. Krücker, W. Lange, D. Leyva Pernia, K. Lipka[27], W. Lohmann[28], F. Lorkowski, R. Mankel, I.-A. Melzer-Pellmann, M. Mendizabal Morentin, A. B. Meyer, G. Milella, K. Moral Figueroa, A. Mussgiller, L. P. Nair, J. Niedziela, A. Nürnberg, Y. Otarid, J. Park, E. Ranken, A. Raspereza, D. Rastorguev, J. Rübenach, L. Rygaard, A. Saggio, M. Scham[26,29], S. Schnake[26], P. Schütze, C. Schwanenberger[25], D. Selivanova, K. Sharko, M. Shchedrolosiev, D. Stafford, F. Vazzoler, A. Ventura Barroso, R. Walsh, D. Wang, Q. Wang, Y. Wen, K. Wichmann, L. Wiens[26], C. Wissing, Y. Yang, A. Zimermmane Castro Santos

University of Hamburg, Hamburg, Germany
A. Albrecht, S. Albrecht, M. Antonello, S. Bein, L. Benato, S. Bollweg, M. Bonanomi, P. Connor, K. El Morabit, Y. Fischer, E. Garutti, A. Grohsjean, J. Haller, H. R. Jabusch, G. Kasieczka, P. Keicher, R. Klanner, W. Korcari, T. Kramer, C. C. Kuo, V. Kutzner, F. Labe, J. Lange, A. Lobanov, C. Matthies, L. Moureaux, M. Mrowietz, A. Nigamova, Y. Nissan, A. Paasch, K. J. Pena Rodriguez, T. Quadfasel, B. Raciti, M. Rieger, D. Savoiu, J. Schindler, P. Schleper, M. Schröder, J. Schwandt, M. Sommerhalder, H. Stadie, G. Steinbrück, A. Tews, M. Wolf

Karlsruher Institut fuer Technologie, Karlsruhe, Germany
S. Brommer, M. Burkart, E. Butz, T. Chwalek, A. Dierlamm, A. Droll, N. Faltermann, M. Giffels, A. Gottmann, F. Hartmann[30], R. Hofsaess, M. Horzela, U. Husemann, J. Kieseler, M. Klute, R. Koppenhöfer, J. M. Lawhorn, M. Link, A. Lintuluoto, B. Maier, S. Maier, S. Mitra, M. Mormile, Th. Müller, M. Neukum, M. Oh, E. Pfeffer, M. Presilla, G. Quast, K. Rabbertz, B. Regnery, N. Shadskiy, I. Shvetsov, H. J. Simonis, L. Sowa, L. Stockmeier, K. Tauqeer, M. Toms, N. Trevisani, R. F. Von Cube, M. Wassmer, S. Wieland, F. Wittig, R. Wolf, X. Zuo

Institute of Nuclear and Particle Physics (INPP), NCSR Demokritos, Aghia Paraskevi, Greece
G. Anagnostou, G. Daskalakis, A. Kyriakis, A. Papadopoulos[30], A. Stakia

National and Kapodistrian University of Athens, Athens, Greece
P. Kontaxakis, G. Melachroinos, Z. Painesis, I. Papavergou, I. Paraskevas, N. Saoulidou, K. Theofilatos, E. Tziaferi, K. Vellidis, I. Zisopoulos

National Technical University of Athens, Athens, Greece
G. Bakas, T. Chatzistavrou, G. Karapostoli, K. Kousouris, I. Papakrivopoulos, E. Siamarkou, G. Tsipolitis, A. Zacharopoulou

University of Ioánnina, Ioánnina, Greece
K. Adamidis, I. Bestintzanos, I. Evangelou, C. Foudas, C. Kamtsikis, P. Katsoulis, P. Kokkas, P. G. Kosmoglou Kioseoglou, N. Manthos, I. Papadopoulos, J. Strologas

HUN-REN Wigner Research Centre for Physics, Budapest, Hungary
C. Hajdu, D. Horvath[31,32], K. Márton, A. J. Rádl[33], F. Sikler, V. Veszpremi

MTA-ELTE Lendület CMS Particle and Nuclear Physics Group, Eötvös Loránd University, Budapest, Hungary
M. Csanád, K. Farkas, A. Fehérkuti[34], M. M. A. Gadallah[35], Á. Kadlecsik, P. Major, G. Pásztor, G. I. Veres

Faculty of Informatics, University of Debrecen, Debrecen, Hungary
B. Ujvari, G. Zilizi

Institute of Nuclear Research ATOMKI, Debrecen, Hungary
G. Bencze, S. Czellar, J. Molnar, Z. Szillasi

Karoly Robert Campus, MATE Institute of Technology, Gyongyos, Hungary
T. Csorgo[34], T. Novak

Panjab University, Chandigarh, India
J. Babbar, S. Bansal, S. B. Beri, V. Bhatnagar, G. Chaudhary, S. Chauhan, N. Dhingra[36], A. Kaur, A. Kaur, H. Kaur, M. Kaur, S. Kumar, K. Sandeep, T. Sheokand, J. B. Singh, A. Singla

University of Delhi, Delhi, India
A. Ahmed, A. Bhardwaj, A. Chhetri, B. C. Choudhary, A. Kumar, A. Kumar, M. Naimuddin, K. Ranjan, M. K. Saini, S. Saumya

Saha Institute of Nuclear Physics, HBNI, Kolkata, India
S. Baradia, S. Barman[37], S. Bhattacharya, S. Das Gupta, S. Dutta, S. Dutta, S. Sarkar

Indian Institute of Technology Madras, Madras, India
M. M. Ameen, P. K. Behera, S. C. Behera, S. Chatterjee, G. Dash, P. Jana, P. Kalbhor, S. Kamble, J. R. Komaragiri[38], D. Kumar[38], P. R. Pujahari, N. R. Saha, A. Sharma, A. K. Sikdar, R. K. Singh, P. Verma, S. Verma, A. Vijay

Tata Institute of Fundamental Research-A, Mumbai, India
S. Dugad, M. Kumar, G. B. Mohanty, B. Parida, M. Shelake, P. Suryadevara

Tata Institute of Fundamental Research-B, Mumbai, India
A. Bala, S. Banerjee, R. M. Chatterjee, M. Guchait, Sh. Jain, A. Jaiswal, S. Kumar, G. Majumder, K. Mazumdar, S. Parolia, A. Thachayath

National Institute of Science Education and Research, An OCC of Homi Bhabha National Institute, Bhubaneswar, India
S. Bahinipati[39], C. Kar, D. Maity[40], P. Mal, T. Mishra, V. K. Muraleedharan Nair Bindhu[40], K. Naskar[40], A. Nayak[40], S. Nayak, K. Pal, P. Sadangi, S. K. Swain, S. Varghese[40], D. Vats[40]

Indian Institute of Science Education and Research (IISER), Pune, India
S. Acharya[41], A. Alpana, S. Dube, B. Gomber[41], P. Hazarika, B. Kansal, A. Laha, B. Sahu[41], S. Sharma, K. Y. Vaish

Isfahan University of Technology, Isfahan, Iran
H. Bakhshiansohi[42], A. Jafari[43], M. Zeinali[44]

Institute for Research in Fundamental Sciences (IPM), Tehran, Iran
S. Bashiri, S. Chenarani[45], S. M. Etesami, Y. Hosseini, M. Khakzad, E. Khazaie[46], M. Mohammadi Najafabadi, S. Tizchang[47]

University College Dublin, Dublin, Ireland
M. Felcini, M. Grunewald

INFN Sezione di Bari[a], Università di Bari[b], Politecnico di Bari[c], Bari, Italy
M. Abbrescia[a,b], A. Colaleo[a,b], D. Creanza[a,c], B. D'Anzi[a,b], N. De Filippis[a,c], M. De Palma[a,b], L. Fiore[a], G. Iaselli[a,c], L. Longo[a], M. Louka[a,b], G. Maggi[a,c], M. Maggi[a], I. Margjeka[a], V. Mastrapasqua[a,b], S. My[a,b], S. Nuzzo[a,b], A. Pellecchia[a,b], A. Pompili[a,b], G. Pugliese[a,c], R. Radogna[a,b], D. Ramos[a], A. Ranieri[a], L. Silvestris[a], F. M. Simone[a,c], Ü. Sözbilir[a], A. Stamerra[a,b], D. Troiano[a,b], R. Venditti[a,b], P. Verwilligen[a], A. Zaza[a,b]

INFN Sezione di Bologna[a], Università di Bologna[b], Bologna, Italy
G. Abbiendi[a], C. Battilana[a,b], D. Bonacorsi[a,b], L. Borgonovi[a], P. Capiluppi[a,b], A. Castro[a,b,†], F. R. Cavallo[a], M. Cuffiani[a,b], G. M. Dallavalle[a], T. Diotalevi[a,b], F. Fabbri[a], A. Fanfani[a,b], D. Fasanella[a], P. Giacomelli[a], L. Giommi[a,b], C. Grandi[a], L. Guiducci[a,b], S. Lo Meo[a,48], M. Lorusso[a,b], L. Lunerti[a], S. Marcellini[a], G. Masetti[a], F. L. Navarria[a,b], G. Paggi[a,b], A. Perrotta[a], F. Primavera[a,b], A. M. Rossi[a,b], S. Rossi Tisbeni[a,b], T. Rovelli[a,b], G. P. Siroli[a,b]

INFN Sezione di Catania[a], Università di Catania[b], Catania, Italy
S. Costa[a,b,49], A. Di Mattia[a], A. Lapertosa[a], R. Potenza[a,b], A. Tricomi[a,b,49], C. Tuve[a,b]

INFN Sezione di Firenze[a], Università di Firenze[b], Firenze, Italy
P. Assiouras[a], G. Barbagli[a], G. Bardelli[a,b], B. Camaiani[a,b], A. Cassese[a], R. Ceccarelli[a], V. Ciulli[a,b], C. Civinini[a], R. D'Alessandro[a,b], E. Focardi[a,b], T. Kello[a], G. Latino[a,b], P. Lenzi[a,b], M. Lizzo[a], M. Meschini[a], S. Paoletti[a], A. Papanastassiou[a,b], G. Sguazzoni[a], L. Viliani[a]

INFN Laboratori Nazionali di Frascati, Frascati, Italy
L. Benussi, S. Bianco, S. Meola[50], D. Piccolo

INFN Sezione di Genova[a], Università di Genova[b], Genova, Italy
P. Chatagnon[a], F. Ferro[a], E. Robutti[a], S. Tosi[a,b]

INFN Sezione di Milano-Bicocca[a], Università di Milano-Bicocca[b], Milano, Italy
A. Benaglia[a], G. Boldrini[a,b], F. Brivio[a], F. Cetorelli[a,b], F. De Guio[a,b], M. E. Dinardo[a,b], P. Dini[a], S. Gennai[a], R. Gerosa[a,b], A. Ghezzi[a,b], P. Govoni[a,b], L. Guzzi[a], M. T. Lucchini[a,b], M. Malberti[a], S. Malvezzi[a], A. Massironi[a], D. Menasce[a], L. Moroni[a], M. Paganoni[a,b], S. Palluotto[a,b], D. Pedrini[a], A. Perego[a,b], B. S. Pinolini[a], G. Pizzati[a,b], S. Ragazzi[a,b], T. Tabarelli de Fatis[a,b]

INFN Sezione di Napoli[a], Università di Napoli 'Federico II'[b], Napoli, Italy; Università della Basilicata[c], Potenza, Italy; Scuola Superiore Meridionale (SSM)[d], Napoli, Italy
S. Buontempo[a], A. Cagnotta[a,b], F. Carnevali[a,b], N. Cavallo[a,c], F. Fabozzi[a,c], A. O. M. Iorio[a,b], L. Lista[a,b,51], P. Paolucci[a,30], B. Rossi[a]

INFN Sezione di Padova[a], Università di Padova[b], Padova, Italy; Università di Trento[c], Trento, Italy
R. Ardino[a], P. Azzi[a], N. Bacchetta[a,52], D. Bisello[a,b], P. Bortignon[a], G. Bortolato[a,b], A. Bragagnolo[a,b], A. C. M. Bulla[a], R. Carlin[a,b], P. Checchia[a], T. Dorigo[a], U. Gasparini[a,b], A. Gozzelino[a], M. Gulmini[a,53], E. Lusiani[a], M. Margoni[a,b], A. T. Meneguzzo[a,b], M. Migliorini[a,b], J. Pazzini[a,b], P. Ronchese[a,b], R. Rossin[a,b], F. Simonetto[a,b], G. Strong[a], M. Tosi[a,b], A. Triossi[a,b], M. Zanetti[a,b], P. Zotto[a,b], A. Zucchetta[a,b], G. Zumerle[a,b]

INFN Sezione di Pavia[a], Università di Pavia[b], Pavia, Italy
C. Aimè[a], A. Braghieri[a], S. Calzaferri[a], D. Fiorina[a], P. Montagna[a,b], V. Re[a], C. Riccardi[a,b], P. Salvini[a], I. Vai[a,b], P. Vitulo[a,b]

INFN Sezione di Perugia[a], Università di Perugia[b], Perugia, Italy
S. Ajmal[a,b], M. E. Ascioti[a,b], G. M. Bilei[a], C. Carrivale[a,b], D. Ciangottini[a,b], L. Fanò[a,b], M. Magherini[a,b], V. Mariani[a,b], M. Menichelli[a], F. Moscatelli[a,54], A. Rossi[a,b], A. Santocchia[a,b], D. Spiga[a], T. Tedeschi[a,b]

INFN Sezione di Pisa[a], Università di Pisa[b], Scuola Normale Superiore di Pisa[c], Pisa, Italy; Università di Siena[d], Siena, Italy
C. A. Alexe[a,c], P. Asenov[a,b], P. Azzurri[a], G. Bagliesi[a], R. Bhattacharya[a], L. Bianchini[a,b], T. Boccali[a], E. Bossini[a], D. Bruschini[a,c], R. Castaldi[a], M. A. Ciocci[a,b], M. Cipriani[a,b], V. D'Amante[a,d], R. Dell'Orso[a], S. Donato[a], A. Giassi[a], F. Ligabue[a,c], D. Matos Figueiredo[a], A. Messineo[a,b], M. Musich[a,b], F. Palla[a], A. Rizzi[a,b], G. Rolandi[a,c], S. Roy Chowdhury[a], T. Sarkar[a], A. Scribano[a], P. Spagnolo[a], R. Tenchini[a], G. Tonelli[a,b], N. Turini[a,d], F. Vaselli[a,c], A. Venturi[a], P. G. Verdini[a]

INFN Sezione di Roma[a], Sapienza Università di Roma[b], Roma, Italy
C. Baldenegro Barrera[a,b], P. Barria[a], C. Basile[a,b], M. Campana[a,b], F. Cavallari[a], L. Cunqueiro Mendez[a,b], D. Del Re[a,b], E. Di Marco[a], M. Diemoz[a], F. Errico[a,b], E. Longo[a,b], J. Mijuskovic[a,b], G. Organtini[a,b], F. Pandolfi[a], R. Paramatti[a,b], C. Quaranta[a,b], S. Rahatlou[a,b], C. Rovelli[a], F. Santanastasio[a,b], L. Soffi[a]

INFN Sezione di Torino[a], Università di Torino[b], Torino, Italy; Università del Piemonte Orientale[c], Novara, Italy
N. Amapane[a,b], R. Arcidiacono[a,c], S. Argiro[a,b], M. Arneodo[a,c], N. Bartosik[a], R. Bellan[a,b], A. Bellora[a,b], C. Biino[a], C. Borca[a,b], N. Cartiglia[a], M. Costa[a,b], R. Covarelli[a,b], N. Demaria[a], L. Finco[a], M. Grippo[a,b], B. Kiani[a,b], F. Legger[a], F. Luongo[a,b], C. Mariotti[a], L. Markovic[a,b], S. Maselli[a], A. Mecca[a,b], L. Menzio[a,b], P. Meridiani[a], E. Migliore[a,b], M. Monteno[a], R. Mulargia[a], M. M. Obertino[a,b], G. Ortona[a], L. Pacher[a,b], N. Pastrone[a], M. Pelliccioni[a], M. Ruspa[a,c], F. Siviero[a,b], V. Sola[a,b], A. Solano[a,b], A. Staiano[a], C. Tarricone[a,b], D. Trocino[a], G. Umoret[a,b], R. White[a,b]

INFN Sezione di Trieste[a], Università di Trieste[b], Trieste, Italy
S. Belforte[a], V. Candelise[a,b], M. Casarsa[a], F. Cossutti[a], K. De Leo[a], G. Della Ricca[a,b]

Kyungpook National University, Daegu, Korea
S. Dogra, J. Hong, C. Huh, B. Kim, J. Kim, D. Lee, H. Lee, S. W. Lee, C. S. Moon, Y. D. Oh, M. S. Ryu, S. Sekmen, B. Tae, Y. C. Yang

Department of Mathematics and Physics-GWNU, Gangneung, Korea
M. S. Kim

Institute for Universe and Elementary Particles, Chonnam National University, Kwangju, Korea
G. Bak, P. Gwak, H. Kim, D. H. Moon

Hanyang University, Seoul, Korea
E. Asilar, J. Choi, D. Kim, T. J. Kim, J. A. Merlin, Y. Ryou

Korea University, Seoul, Korea
S. Choi, S. Han, B. Hong, K. Lee, K. S. Lee, S. Lee, J. Yoo

Kyung Hee University, Department of Physics, Seoul, Korea
J. Goh, S. Yang

Sejong University, Seoul, Korea
H. S. Kim, Y. Kim, S. Lee

Seoul National University, Seoul, Korea
J. Almond, J. H. Bhyun, J. Choi, J. Choi, W. Jun, J. Kim, S. Ko, H. Kwon, H. Lee, J. Lee, J. Lee, B. H. Oh, S. B. Oh, H. Seo, U. K. Yang, I. Yoon

University of Seoul, Seoul, Korea
W. Jang, D. Y. Kang, Y. Kang, S. Kim, B. Ko, J. S. H. Lee, Y. Lee, I. C. Park, Y. Roh, I. J. Watson

Department of Physics, Yonsei University, Seoul, Korea
S. Ha, H. D. Yoo

Sungkyunkwan University, Suwon, Korea
M. Choi, M. R. Kim, H. Lee, Y. Lee, I. Yu

College of Engineering and Technology, American University of the Middle East (AUM), Dasman, Kuwait
T. Beyrouthy

Riga Technical University, Riga, Latvia
K. Dreimanis, A. Gaile, G. Pikurs, A. Potrebko, M. Seidel, D. Sidiropoulos Kontos

University of Latvia (LU), Riga, Latvia
N. R. Strautnieks

Vilnius University, Vilnius, Lithuania
M. Ambrozas, A. Juodagalvis, A. Rinkevicius, G. Tamulaitis

National Centre for Particle Physics, Universiti Malaya, Kuala Lumpur, Malaysia
I. Yusuff[55], Z. Zolkapli

Universidad de Sonora (UNISON), Hermosillo, Mexico
J. F. Benitez, A. Castaneda Hernandez, H. A. Encinas Acosta, L. G. Gallegos Maríñez, M. León Coello, J. A. Murillo Quijada, A. Sehrawat, L. Valencia Palomo

Centro de Investigacion y de Estudios Avanzados del IPN, Mexico City, Mexico
G. Ayala, H. Castilla-Valdez, H. Crotte Ledesma, E. De La Cruz-Burelo, I. Heredia-De La Cruz[56], R. Lopez-Fernandez, J. Mejia Guisao, C. A. Mondragon Herrera, A. Sánchez Hernández

Universidad Iberoamericana, Mexico City, Mexico
C. Oropeza Barrera, D. L. Ramirez Guadarrama, M. Ramírez García

Benemerita Universidad Autonoma de Puebla, Puebla, Mexico
I. Bautista, I. Pedraza, H. A. Salazar Ibarguen, C. Uribe Estrada

University of Montenegro, Podgorica, Montenegro
I. Bubanja, N. Raicevic

University of Canterbury, Christchurch, New Zealand
P. H. Butler

National Centre for Physics, Quaid-I-Azam University, Islamabad, Pakistan
A. Ahmad, M. I. Asghar, A. Awais, M. I. M. Awan, H. R. Hoorani, W. A. Khan

Faculty of Computer Science, Electronics and Telecommunications, AGH University of Krakow, Krakow, Poland
V. Avati, L. Grzanka, M. Malawski

National Centre for Nuclear Research, Swierk, Poland
H. Bialkowska, M. Bluj, M. Górski, M. Kazana, M. Szleper, P. Zalewski

Faculty of Physics, Institute of Experimental Physics, University of Warsaw, Warsaw, Poland
K. Bunkowski, K. Doroba, A. Kalinowski, M. Konecki, J. Krolikowski, A. Muhammad

Warsaw University of Technology, Warsaw, Poland
K. Pozniak, W. Zabolotny

Laboratório de Instrumentação e Física Experimental de Partículas, Lisboa, Portugal
M. Araujo, D. Bastos, C. Beirão Da Cruz E Silva, A. Boletti, M. Bozzo, T. Camporesi, G. Da Molin, P. Faccioli, M. Gallinaro, J. Hollar, N. Leonardo, G. B. Marozzo, T. Niknejad, A. Petrilli, M. Pisano, J. Seixas, J. Varela, J. W. Wulff

Faculty of Physics, University of Belgrade, Belgrade, Serbia
P. Adzic, P. Milenovic

VINCA Institute of Nuclear Sciences, University of Belgrade, Belgrade, Serbia
M. Dordevic, J. Milosevic, L. Nadderd, V. Rekovic

Centro de Investigaciones Energéticas Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain
J. Alcaraz Maestre, Cristina F. Bedoya, Oliver M. Carretero, M. Cepeda, M. Cerrada, N. Colino, B. De La Cruz, A. Delgado Peris, A. Escalante Del Valle, D. Fernández Del Val, J. P. Fernández Ramos, J. Flix, M. C. Fouz, O. Gonzalez Lopez, S. Goy Lopez, J. M. Hernandez, M. I. Josa, E. Martin Viscasillas, D. Moran, C. M. Morcillo Perez, Á. Navarro Tobar, C. Perez Dengra, A. Pérez-Calero Yzquierdo, J. Puerta Pelayo, I. Redondo, S. Sánchez Navas, J. Sastre, J. Vazquez Escobar

Universidad Autónoma de Madrid, Madrid, Spain
J. F. de Trocóniz

Instituto Universitario de Ciencias y Tecnologías Espaciales de Asturias (ICTEA), Universidad de Oviedo, Oviedo, Spain
B. Alvarez Gonzalez, J. Cuevas, J. Fernandez Menendez, S. Folgueras, I. Gonzalez Caballero, J. R. González Fernández, P. Leguina, E. Palencia Cortezon, C. Ramón Álvarez, V. Rodríguez Bouza, A. Soto Rodríguez, A. Trapote, C. Vico Villalba, P. Vischia

Instituto de Física de Cantabria (IFCA), CSIC-Universidad de Cantabria, Santander, Spain
S. Bhowmik, S. Blanco Fernández, J. A. Brochero Cifuentes, I. J. Cabrillo, A. Calderon, J. Duarte Campderros, M. Fernandez, G. Gomez, C. Lasaosa García, R. Lopez Ruiz, C. Martinez Rivero, P. Martinez Ruiz del Arbol, F. Matorras, P. Matorras Cuevas, E. Navarrete Ramos, J. Piedra Gomez, L. Scodellaro, I. Vila, J. M. Vizan Garcia

University of Colombo, Colombo, Sri Lanka
B. Kailasapathy[57], D. D. C. Wickramarathna

Department of Physics, University of Ruhuna, Matara, Sri Lanka
W. G. D. Dharmaratna[58], K. Liyanage, N. Perera

CERN, European Organization for Nuclear Research, Geneva, Switzerland
D. Abbaneo, C. Amendola, E. Auffray, G. Auzinger, J. Baechler, D. Barney, A. Bermúdez Martínez, M. Bianco, B. Bilin, A. A. Bin Anuar, A. Bocci, C. Botta, E. Brondolin, C. Caillol, G. Cerminara, N. Chernyavskaya, D. d'Enterria, A. Dabrowski, A. David, A. De Roeck, M. M. Defranchis, M. Deile, M. Dobson, G. Franzoni, W. Funk, S. Giani, D. Gigi, K. Gill, F. Glege, J. Hegeman, J. K. Heikkilä, B. Huber, V. Innocente, T. James, P. Janot, O. Kaluzinska, S. Laurila, P. Lecoq, E. Leutgeb, C. Lourenço, L. Malgeri, M. Mannelli, A. C. Marini, M. Matthewman, A. Mehta, F. Meijers, S. Mersi, E. Meschi, V. Milosevic, F. Monti, F. Moortgat, M. Mulders, I. Neutelings, S. Orfanelli, F. Pantaleo, G. Petrucciani, A. Pfeiffer, M. Pierini, H. Qu, D. Rabady, B. Ribeiro Lopes, M. Rovere, H. Sakulin, S. Sanchez Cruz, S. Scarfi, C. Schwick, M. Selvaggi, A. Sharma, K. Shchelina, P. Silva, P. Sphicas[59], A. G. Stahl Leiton, A. Steen, S. Summers, D. Treille, P. Tropea, D. Walter, J. Wanczyk[60], J. Wang, S. Wuchterl, P. Zehetner, P. Zejdl, W. D. Zeuner

Paul Scherrer Institut, Villigen, Switzerland
T. Bevilacqua[61], L. Caminada[61], A. Ebrahimi, W. Erdmann, R. Horisberger, Q. Ingram, H. C. Kaestli, D. Kotlinski, C. Lange, M. Missiroli[61], L. Noehte[61], T. Rohe

ETH Zurich-Institute for Particle Physics and Astrophysics (IPA), Zurich, Switzerland
T. K. Aarrestad, K. Androsov[60], M. Backhaus, G. Bonomelli, A. Calandri, C. Cazzaniga, K. Datta, P. De Bryas Dexmiers D'archiac[60], A. De Cosa, G. Dissertori, M. Dittmar, M. Donegà, F. Eble, M. Galli, K. Gedia, F. Glessgen, C. Grab, N. Härringer, T. G. Harte, D. Hits, W. Lustermann, A.-M. Lyon, R. A. Manzoni, M. Marchegiani, L. Marchese, C. Martin Perez, A. Mascellani[60], F. Nessi-Tedaldi, F. Pauss, V. Perovic, S. Pigazzini, C. Reissel, T. Reitenspiess, B. Ristic, F. Riti, R. Seidita, J. Steggemann[60], A. Tarabini, D. Valsecchi, R. Wallny

Universität Zürich, Zurich, Switzerland
C. Amsler[62], P. Bärtschi, M. F. Canelli, K. Cormier, M. Huwiler, W. Jin, A. Jofrehei, B. Kilminster, S. Leontsinis, S. P. Liechti, A. Macchiolo, P. Meiring, F. Meng, U. Molinatti, J. Motta, A. Reimers, P. Robmann, M. Senger, E. Shokr, F. Stäger, R. Tramontano

National Central University, Chung-Li, Taiwan
C. Adloff[63], D. Bhowmik, C. M. Kuo, W. Lin, P. K. Rout, P. C. Tiwari[38], S. S. Yu

National Taiwan University (NTU), Taipei, Taiwan
L. Ceard, K. F. Chen, P. S. Chen, Z. G. Chen, A. De Iorio, W.-S. Hou, T. H. Hsu, Y. W. Kao, S. Karmakar, G. Kole, Y. Y. Li, R.-S. Lu, E. Paganis, X. F. Su, J. Thomas-Wilsker, L. S. Tsai, H. Y. Wu, E. Yazgan

High Energy Physics Research Unit, Department of Physics, Faculty of Science, Chulalongkorn University, Bangkok, Thailand
C. Asawatangtrakuldee, N. Srimanobhas, V. Wachirapusitanand

Physics Department, Science and Art Faculty, Çukurova University, Adana, Turkey
D. Agyel, F. Boran, F. Dolek, I. Dumanoglu[64], E. Eskut, Y. Guler[65], E. Gurpinar Guler[65], C. Isik, O. Kara, A. Kayis Topaksu, U. Kiminsu, G. Onengut, K. Ozdemir[66], A. Polatoz, B. Tali[67], U. G. Tok, S. Turkcapar, E. Uslan, I. S. Zorbakir

Physics Department, Middle East Technical University, Ankara, Turkey
G. Sokmen, M. Yalvac[68]

Bogazici University, Istanbul, Turkey
B. Akgun, I. O. Atakisi, E. Gülmez, M. Kaya[69], O. Kaya[70], S. Tekten[71]

Istanbul Technical University, Istanbul, Turkey
A. Cakir, K. Cankocak[64,72], G. G. Dincer[64], Y. Komurcu, S. Sen[73]

Istanbul University, Istanbul, Turkey
O. Aydilek[74], B. Hacisahinoglu, I. Hos[75], B. Kaynak, S. Ozkorucuklu, O. Potok, H. Sert, C. Simsek, C. Zorbilmez

Yildiz Technical University, Istanbul, Turkey
S. Cerci[67], B. Isildak[76], D. Sunar Cerci, T. Yetkin

Institute for Scintillation Materials of National Academy of Science of Ukraine, Kharkiv, Ukraine
A. Boyaryntsev, B. Grynyov

National Science Centre, Kharkiv Institute of Physics and Technology, Kharkiv, Ukraine
L. Levchuk

University of Bristol, Bristol, UK
D. Anthony, J. J. Brooke, A. Bundock, F. Bury, E. Clement, D. Cussans, H. Flacher, M. Glowacki, J. Goldstein, H. F. Heath, M.-L. Holmberg, L. Kreczko, S. Paramesvaran, L. Robertshaw, S. Seif El Nasr-Storey, V. J. Smith, N. Stylianou[77], K. Walkingshaw Pass

Rutherford Appleton Laboratory, Didcot, UK
A. H. Ball, K. W. Bell, A. Belyaev[78], C. Brew, R. M. Brown, D. J. A. Cockerill, C. Cooke, A. Elliot, K. V. Ellis, K. Harder, S. Harper, J. Linacre, K. Manolopoulos, D. M. Newbold, E. Olaiya, D. Petyt, T. Reis, A. R. Sahasransu, G. Salvi, T. Schuh, C. H. Shepherd-Themistocleous, I. R. Tomalin, K. C. Whalen, T. Williams

Imperial College, London, UK
I. Andreou, R. Bainbridge, P. Bloch, C. E. Brown, O. Buchmuller, V. Cacchio, C. A. Carrillo Montoya, G. S. Chahal[79], D. Colling, J. S. Dancu, I. Das, P. Dauncey, G. Davies, J. Davies, M. Della Negra, S. Fayer, G. Fedi, G. Hall, M. H. Hassanshahi, A. Howard, G. Iles, M. Knight, J. Langford, J. León Holgado, L. Lyons, A.-M. Magnan, S. Mallios, M. Mieskolainen, J. Nash[80], M. Pesaresi, P. B. Pradeep, B. C. Radburn-Smith, A. Richards, A. Rose, K. Savva, C. Seez, R. Shukla, A. Tapper, K. Uchida, G. P. Uttley, L. H. Vage, T. Virdee[30], M. Vojinovic, N. Wardle, D. Winterbottom

Brunel University, Uxbridge, UK
K. Coldham, J. E. Cole, A. Khan, P. Kyberd, I. D. Reid

Baylor University, Waco, TX, USA
S. Abdullin, A. Brinkerhoff, B. Caraway, E. Collins, J. Dittmann, K. Hatakeyama, J. Hiltbrand, B. McMaster, J. Samudio, S. Sawant, C. Sutantawibul, J. Wilson

Catholic University of America, Washington, DC, USA
R. Bartek, A. Dominguez, C. Huerta Escamilla, A. E. Simsek, R. Uniyal, A. M. Vargas Hernandez

The University of Alabama, Tuscaloosa, AL, USA
B. Bam, A. Buchot Perraguin, R. Chudasama, S. I. Cooper, C. Crovella, S. V. Gleyzer, E. Pearson, C. U. Perez, P. Rumerio[81], E. Usai, R. Yi

Boston University, Boston, MA, USA
A. Akpinar, C. Cosby, G. De Castro, Z. Demiragli, C. Erice, C. Fangmeier, C. Fernandez Madrazo, E. Fontanesi, D. Gastler, F. Golf, S. Jeon, J. O'cain, I. Reed, J. Rohlf, K. Salyer, D. Sperka, D. Spitzbart, I. Suarez, A. Tsatsos, A. G. Zecchinelli

Brown University, Providence, RI, USA
G. Benelli, X. Coubez[26], D. Cutts, L. Gouskos, M. Hadley, U. Heintz, J. M. Hogan[82], T. Kwon, G. Landsberg, K. T. Lau, D. Li, J. Luo, S. Mondal, M. Narain[†], N. Pervan, T. Russell, S. Sagir[83], F. Simpson, M. Stamenkovic, N. Venkatasubramanian, X. Yan, W. Zhang

University of California, Davis, Davis, CA, USA
S. Abbott, C. Brainerd, R. Breedon, H. Cai, M. Calderon De La Barca Sanchez, M. Chertok, M. Citron, J. Conway, P. T. Cox, R. Erbacher, F. Jensen, O. Kukral, G. Mocellin, M. Mulhearn, S. Ostrom, W. Wei, Y. Yao, S. Yoo, F. Zhang

University of California, Los Angeles, CA, USA
M. Bachtis, R. Cousins, A. Datta, G. Flores Avila, J. Hauser, M. Ignatenko, M. A. Iqbal, T. Lam, E. Manca, A. Nunez Del Prado, D. Saltzberg, V. Valuev

University of California, Riverside, Riverside, CA, USA
R. Clare, J. W. Gary, M. Gordon, G. Hanson, W. Si, S. Wimpenny †

University of California, San Diego, La Jolla, CA, USA
A. Aportela, A. Arora, J. G. Branson, S. Cittolin, S. Cooperstein, D. Diaz, J. Duarte, L. Giannini, Y. Gu, J. Guiang, R. Kansal, V. Krutelyov, R. Lee, J. Letts, M. Masciovecchio, F. Mokhtar, S. Mukherjee, M. Pieri, M. Quinnan, B. V. Sathia Narayanan, V. Sharma, M. Tadel, E. Vourliotis, F. Würthwein, Y. Xiang, A. Yagil

Department of Physics, University of California, Santa Barbara, Santa Barbara, CA, USA
A. Barzdukas, L. Brennan, C. Campagnari, K. Downham, C. Grieco, J. Incandela, J. Kim, A. J. Li, P. Masterson, H. Mei, J. Richman, S. N. Santpur, U. Sarica, R. Schmitz, F. Setti, J. Sheplock, D. Stuart, T.Á. Vámi, S. Wang, D. Zhang

California Institute of Technology, Pasadena, CA, USA
A. Bornheim, O. Cerri, A. Latorre, J. Mao, H. B. Newman, G. Reales Gutiérrez, M. Spiropulu, J. R. Vlimant, C. Wang, S. Xie, R. Y. Zhu

Carnegie Mellon University, Pittsburgh, PA, USA
J. Alison, S. An, M. B. Andrews, P. Bryant, M. Cremonesi, V. Dutta, T. Ferguson, A. Harilal, A. Kallil Tharayil, C. Liu, T. Mudholkar, S. Murthy, P. Palit, K. Park, M. Paulini, A. Roberts, A. Sanchez, W. Terrill

University of Colorado Boulder, Boulder, CO, USA
J. P. Cumalat, W. T. Ford, A. Hart, A. Hassani, G. Karathanasis, N. Manganelli, A. Perloff, C. Savard, N. Schonbeck, K. Stenson, K. A. Ulmer, S. R. Wagner, N. Zipper, D. Zuolo

Cornell University, Ithaca, NY, USA
J. Alexander, S. Bright-Thonney, X. Chen, D. J. Cranshaw, J. Fan, X. Fan, S. Hogan, P. Kotamnives, J. Monroy, M. Oshiro, J. R. Patterson, M. Reid, A. Ryd, J. Thom, P. Wittich, R. Zou

Fermi National Accelerator Laboratory, Batavia, IL, USA
M. Albrow, M. Alyari, O. Amram, G. Apollinari, A. Apresyan, L. A. T. Bauerdick, D. Berry, J. Berryhill, P. C. Bhat, K. Burkett, J. N. Butler, A. Canepa, G. B. Cerati, H. W. K. Cheung, F. Chlebana, G. Cummings, J. Dickinson, I. Dutta, V. D. Elvira, Y. Feng, J. Freeman, A. Gandrakota, Z. Gecse, L. Gray, D. Green, A. Grummer, S. Grünendahl, D. Guerrero, O. Gutsche, R. M. Harris, R. Heller, T. C. Herwig, J. Hirschauer, B. Jayatilaka, S. Jindariani, M. Johnson, U. Joshi, T. Klijnsma, B. Klima, K. H. M. Kwok, S. Lammel, D. Lincoln, R. Lipton, T. Liu, C. Madrid, K. Maeshima, C. Mantilla, D. Mason, P. McBride, P. Merkel, S. Mrenna, S. Nahn, J. Ngadiuba, D. Noonan, S. Norberg, V. Papadimitriou, N. Pastika, K. Pedro, C. Pena [84], F. Ravera, A. Reinsvold Hall [85], L. Ristori, M. Safdari, E. Sexton-Kennedy, N. Smith, A. Soha, L. Spiegel, S. Stoynev, J. Strait, L. Taylor, S. Tkaczyk, N. V. Tran, L. Uplegger, E. W. Vaandering, I. Zoi

University of Florida, Gainesville, FL, USA
C. Aruta, P. Avery, D. Bourilkov, P. Chang, V. Cherepanov, R. D. Field, E. Koenig, M. Kolosova, J. Konigsberg, A. Korytov, K. Matchev, N. Menendez, G. Mitselmakher, K. Mohrman, A. Muthirakalayil Madhu, N. Rawal, S. Rosenzweig, Y. Takahashi, J. Wang

Florida State University, Tallahassee, FL, USA
T. Adams, A. Al Kadhim, A. Askew, S. Bower, R. Habibullah, V. Hagopian, R. Hashmi, R. S. Kim, S. Kim, T. Kolberg, G. Martinez, H. Prosper, P. R. Prova, M. Wulansatiti, R. Yohay, J. Zhang

Florida Institute of Technology, Melbourne, FL, USA
B. Alsufyani, M. M. Baarmand, S. Butalla, S. Das, T. Elkafrawy [86], M. Hohlmann, M. Rahmani, E. Yanes

University of Illinois Chicago, Chicago, USA
M. R. Adams, A. Baty, C. Bennett, R. Cavanaugh, R. Escobar Franco, O. Evdokimov, C. E. Gerber, M. Hawksworth, A. Hingrajiya, D. J. Hofman, J. h. Lee, D. S. Lemos, A. H. Merrit, C. Mills, S. Nanda, G. Oh, B. Ozek, D. Pilipovic, R. Pradhan, E. Prifti, T. Roy, S. Rudrabhatla, M. B. Tonjes, N. Varelas, M. A. Wadud, Z. Ye, J. Yoo

The University of Iowa, Iowa City, IA, USA
M. Alhusseini, D. Blend, K. Dilsiz[87], L. Emediato, G. Karaman, O. K. Köseyan, J.-P. Merlo, A. Mestvirishvili[88], O. Neogi, H. Ogul[89], Y. Onel, A. Penzo, C. Snyder, E. Tiras[90]

Johns Hopkins University, Baltimore, MD, USA
B. Blumenfeld, L. Corcodilos, J. Davis, A. V. Gritsan, L. Kang, S. Kyriacou, P. Maksimovic, M. Roguljic, J. Roskes, S. Sekhar, M. Swartz

The University of Kansas, Lawrence, KS, USA
A. Abreu, L. F. Alcerro Alcerro, J. Anguiano, S. Arteaga Escatel, P. Baringer, A. Bean, Z. Flowers, D. Grove, J. King, G. Krintiras, M. Lazarovits, C. Le Mahieu, J. Marquez, N. Minafra, M. Murray, M. Nickel, M. Pitt, S. Popescu[91], C. Rogan, C. Royon, R. Salvatico, S. Sanders, C. Smith, G. Wilson

Kansas State University, Manhattan, KS, USA
B. Allmond, R. Gujju Gurunadha, A. Ivanov, K. Kaadze, Y. Maravin, J. Natoli, D. Roy, G. Sorrentino

University of Maryland, College Park, MD, USA
A. Baden, A. Belloni, J. Bistany-riebman, Y. M. Chen, S. C. Eno, N. J. Hadley, S. Jabeen, R. G. Kellogg, T. Koeth, B. Kronheim, Y. Lai, S. Lascio, A. C. Mignerey, S. Nabili, C. Palmer, C. Papageorgakis, M. M. Paranjpe, L. Wang

Massachusetts Institute of Technology, Cambridge, MA, USA
J. Bendavid, I. A. Cali, P. C. Chou, M. D'Alfonso, J. Eysermans, C. Freer, G. Gomez-Ceballos, M. Goncharov, G. Grosso, P. Harris, D. Hoang, D. Kovalskyi, J. Krupa, L. Lavezzo, Y.-J. Lee, K. Long, C. Mcginn, A. Novak, C. Paus, D. Rankin, C. Roland, G. Roland, S. Rothman, G. S. F. Stephans, Z. Wang, B. Wyslouch, T. J. Yang

University of Minnesota, Minneapolis, MN, USA
B. Crossman, B. M. Joshi, C. Kapsiak, M. Krohn, D. Mahon, J. Mans, B. Marzocchi, M. Revering, R. Rusack, R. Saradhy, N. Strobbe

University of Nebraska-Lincoln, Lincoln, NE, USA
K. Bloom, D. R. Claes, G. Haza, J. Hossain, C. Joo, I. Kravchenko, J. E. Siado, W. Tabb, A. Vagnerini, A. Wightman, F. Yan, D. Yu

State University of New York at Buffalo, Buffalo, NY, USA
H. Bandyopadhyay, L. Hay, H. W. Hsia, I. Iashvili, A. Kalogeropoulos, A. Kharchilava, M. Morris, D. Nguyen, S. Rappoccio, H. Rejeb Sfar, A. Williams, P. Young

Northeastern University, Boston, MA, USA
G. Alverson, E. Barberis, J. Bonilla, J. Dervan, Y. Haddad, Y. Han, A. Krishna, J. Li, M. Lu, G. Madigan, R. Mccarthy, D. M. Morse, V. Nguyen, T. Orimoto, A. Parker, L. Skinnari, D. Wood

Northwestern University, Evanston, IL, USA
J. Bueghly, S. Dittmer, K. A. Hahn, Y. Liu, Y. Miao, D. G. Monk, M. H. Schmitt, A. Taliercio, M. Velasco

University of Notre Dame, Notre Dame, IN, USA
G. Agarwal, R. Band, R. Bucci, S. Castells, A. Das, R. Goldouzian, M. Hildreth, K. W. Ho, K. Hurtado Anampa, T. Ivanov, C. Jessop, K. Lannon, J. Lawrence, N. Loukas, L. Lutton, J. Mariano, N. Marinelli, I. Mcalister, T. McCauley, C. Mcgrady, C. Moore, Y. Musienko[17], H. Nelson, M. Osherson, A. Piccinelli, R. Ruchti, A. Townsend, Y. Wan, M. Wayne, H. Yockey, M. Zarucki, L. Zygala

The Ohio State University, Columbus, OH, USA
A. Basnet, B. Bylsma, M. Carrigan, L. S. Durkin, C. Hill, M. Joyce, M. Nunez Ornelas, K. Wei, B. L. Winer, B. R. Yates

Princeton University, Princeton, NJ, USA
H. Bouchamaoui, P. Das, G. Dezoort, P. Elmer, A. Frankenthal, B. Greenberg, N. Haubrich, K. Kennedy, G. Kopp, S. Kwan, D. Lange, A. Loeliger, D. Marlow, I. Ojalvo, J. Olsen, A. Shevelev, D. Stickland, C. Tully

University of Puerto Rico, Mayaguez, PR, USA
S. Malik

Purdue University, West Lafayette, IN, USA
A. S. Bakshi, S. Chandra, R. Chawla, A. Gu, L. Gutay, M. Jones, A. W. Jung, A. M. Koshy, M. Liu, G. Negro, N. Neumeister, G. Paspalaki, S. Piperov, V. Scheurer, J. F. Schulte, M. Stojanovic, J. Thieman, A. K. Virdi, F. Wang, W. Xie

Purdue University Northwest, Hammond, IN, USA
J. Dolen, N. Parashar, A. Pathak

Rice University, Houston, TX, USA
D. Acosta, T. Carnahan, K. M. Ecklund, P. J. Fernández Manteca, S. Freed, P. Gardner, F. J. M. Geurts, W. Li, J. Lin, O. Miguel Colin, B. P. Padley, R. Redjimi, J. Rotter, E. Yigitbasi, Y. Zhang

University of Rochester, Rochester, NY, USA
A. Bodek, P. de Barbaro, R. Demina, J. L. Dulemba, A. Garcia-Bellido, O. Hindrichs, A. Khukhunaishvili, N. Parmar, P. Parygin[92], E. Popova[92], R. Taus

Rutgers, The State University of New Jersey, Piscataway, NJ, USA
B. Chiarito, J. P. Chou, S. V. Clark, D. Gadkari, Y. Gershtein, E. Halkiadakis, M. Heindl, C. Houghton, D. Jaroslawski, O. Karacheban[28], S. Konstantinou, I. Laflotte, A. Lath, R. Montalvo, K. Nash, J. Reichert, H. Routray, P. Saha, S. Salur, S. Schnetzer, S. Somalwar, R. Stone, S. A. Thayil, S. Thomas, J. Vora, H. Wang

University of Tennessee, Knoxville, TN, USA
H. Acharya, D. Ally, A. G. Delannoy, S. Fiorendi, S. Higginbotham, T. Holmes, A. R. Kanuganti, N. Karunarathna, L. Lee, E. Nibigira, S. Spanier

Texas A&M University, College Station, Texas, USA
D. Aebi, M. Ahmad, T. Akhter, O. Bouhali[93], R. Eusebi, J. Gilmore, T. Huang, T. Kamon[94], H. Kim, S. Luo, R. Mueller, D. Overton, D. Rathjens, A. Safonov

Texas Tech University, Lubbock, TX, USA
N. Akchurin, J. Damgov, N. Gogate, V. Hegde, A. Hussain, Y. Kazhykarim, K. Lamichhane, S. W. Lee, A. Mankel, T. Peltola, I. Volobouev

Vanderbilt University, Nashville, TN, USA
E. Appelt, Y. Chen, S. Greene, A. Gurrola, W. Johns, R. Kunnawalkam Elayavalli, A. Melo, F. Romeo, P. Sheldon, S. Tuo, J. Velkovska, J. Viinikainen

University of Virginia, Charlottesville, VA, USA
B. Cardwell, B. Cox, J. Hakala, R. Hirosky, A. Ledovskoy, C. Neu

Wayne State University, Detroit, MI, USA
S. Bhattacharya, P. E. Karchin

University of Wisconsin-Madison, Madison, WI, USA
A. Aravind, S. Banerjee, K. Black, T. Bose, S. Dasu, I. De Bruyn, P. Everaerts, C. Galloni, H. He, M. Herndon, A. Herve, C. K. Koraka, A. Lanaro, R. Loveless, J. Madhusudanan Sreekala, A. Mallampalli, A. Mohammadi, S. Mondal, G. Parida, L. Pétré, D. Pinna, A. Savin, V. Shang, V. Sharma, W. H. Smith, D. Teague, H. F. Tsoi, W. Vetens, A. Warden

Authors Affiliated with an Institute or an International Laboratory Covered by a Cooperation Agreement with CERN, Geneva, Switzerland
S. Afanasiev, V. Alexakhin, V. Andreev, Yu. Andreev, T. Aushev, M. Azarkin, A. Babaev, V. Blinov[95], E. Boos, V. Borshch, D. Budkouski, V. Bunichev, M. Chadeeva[95], V. Chekhovsky, R. Chistov[95], A. Dermenev, T. Dimova[95], D. Druzhkin[96], M. Dubinin[84], L. Dudko, G. Gavrilov, V. Gavrilov, S. Gninenko, V. Golovtcov, N. Golubev, I. Golutvin, I. Gorbunov, A. Gribushin, Y. Ivanov, V. Kachanov, V. Karjavine, A. Karneyeu, V. Kim[95], M. Kirakosyan, D. Kirpichnikov, M. Kirsanov, V. Klyukhin, O. Kodolova[97], D. Konstantinov, V. Korenkov, A. Kozyrev[95], N. Krasnikov, A. Lanev, P. Levchenko[98], N. Lychkovskaya, V. Makarenko, A. Malakhov, V. Matveev[95], V. Murzin, A. Nikitenko[97,99], S. Obraztsov, V. Oreshkin, V. Palichik, V. Perelygin, M. Perfilov, S. Polikarpov[95], V. Popov, O. Radchenko[95], M. Savina, V. Savrin, V. Shalaev, S. Shmatov, S. Shulha, Y. Skovpen[95], S. Slabospitskii, V. Smirnov, D. Sosnov, V. Sulimov, E. Tcherniaev, A. Terkulov, O. Teryaev, I. Tlisova, A. Toropin, L. Uvarov, A. Uzunian, P. Volkov, A. Vorobyev†, G. Vorotnikov, N. Voytishin, B. S. Yuldashev[100], A. Zarubin, I. Zhizhin, A. Zhokin

† **Deceased**

1: Also at Yerevan State University, Yerevan, Armenia
2: Also at TU Wien, Vienna, Austria
3: Also at Faculty of Engineering, Institute of Basic and Applied Sciences, Arab Academy for Science, Technology and Maritime Transport, Alexandria, Egypt
4: Also at Ghent University, Ghent, Belgium
5: Also at Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil
6: Also at Universidade Estadual de Campinas, Campinas, Brazil
7: Also at Federal University of Rio Grande do Sul, Porto Alegre, Brazil
8: Also at UFMS, Nova Andradina, Brazil
9: Also at Nanjing Normal University, Nanjing, China
10: Now at The University of Iowa, Iowa City, IA, USA
11: Also at University of Chinese Academy of Sciences, Beijing, China
12: Also at China Center of Advanced Science and Technology, Beijing, China
13: Also at University of Chinese Academy of Sciences, Beijing, China
14: Also at China Spallation Neutron Source, Guangdong, China
15: Now at Henan Normal University, Xinxiang, China
16: Also at Université Libre de Bruxelles, Bruxelles, Belgium
17: Also at an Institute or an International Laboratory Covered by a Cooperation Agreement with CERN, Geneva, Switzerland
18: Also at Suez University, Suez, Egypt
19: Now at British University in Egypt, Cairo, Egypt
20: Also at Purdue University, West Lafayette, Indiana, USA
21: Also at Université de Haute Alsace, Mulhouse, France
22: Also at Department of Physics, Tsinghua University, Beijing, China
23: Also at Tbilisi State University, Tbilisi, Georgia
24: Also at The University of the State of Amazonas, Manaus, Brazil
25: Also at University of Hamburg, Hamburg, Germany
26: Also at III. Physikalisches Institut A, RWTH Aachen University, Aachen, Germany
27: Also at Bergische University Wuppertal (BUW), Wuppertal, Germany
28: Also at Brandenburg University of Technology, Cottbus, Germany
29: Also at Forschungszentrum Jülich, Juelich, Germany
30: Also at CERN, European Organization for Nuclear Research, Geneva, Switzerland

31: Also at Institute of Nuclear Research ATOMKI, Debrecen, Hungary
32: Now at Facultatea de Fizica, Universitatea Babes-Bolyai, Cluj-Napoca, Romania
33: Also at MTA-ELTE Lendület CMS Particle and Nuclear Physics Group, Eötvös Loránd University, Budapest, Hungary
34: Also at HUN-REN Wigner Research Centre for Physics, Budapest, Hungary
35: Also at Physics Department, Faculty of Science, Assiut University, Assiut, Egypt
36: Also at Punjab Agricultural University, Ludhiana, India
37: Also at University of Visva-Bharati, Santiniketan, India
38: Also at Indian Institute of Science (IISc), Bangalore, India
39: Also at IIT Bhubaneswar, Bhubaneswar, India
40: Also at Institute of Physics, Bhubaneswar, India
41: Also at University of Hyderabad, Hyderabad, India
42: Also at Deutsches Elektronen-Synchrotron, Hamburg, Germany
43: Also at Isfahan University of Technology, Isfahan, Iran
44: Also at Sharif University of Technology, Tehran, Iran
45: Also at Department of Physics, University of Science and Technology of Mazandaran, Behshahr, Iran
46: Also at Department of Physics, Isfahan University of Technology, Isfahan, Iran
47: Also at Department of Physics, Faculty of Science, Arak University, Arak, Iran
48: Also at Italian National Agency for New Technologies, Energy and Sustainable Economic Development, Bologna, Italy
49: Also at Centro Siciliano di Fisica Nucleare e di Struttura Della Materia, Catania, Italy
50: Also at Università degli Studi Guglielmo Marconi, Roma, Italy
51: Also at Scuola Superiore Meridionale, Università di Napoli 'Federico II', Napoli, Italy
52: Also at Fermi National Accelerator Laboratory, Batavia, IL, USA
53: Also at Laboratori Nazionali di Legnaro dell'INFN, Legnaro, Italy
54: Also at Consiglio Nazionale delle Ricerche - Istituto Officina dei Materiali, Perugia, Italy
55: Also at Department of Applied Physics, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Malaysia
56: Also at Consejo Nacional de Ciencia y Tecnología, Mexico City, Mexico
57: Also at Trincomalee Campus, Eastern University, Sri Lanka, Nilaveli, Sri Lanka
58: Also at Saegis Campus, Nugegoda, Sri Lanka
59: Also at National and Kapodistrian University of Athens, Athens, Greece
60: Also at Ecole Polytechnique Fédérale Lausanne, Lausanne, Switzerland
61: Also at Universität Zürich, Zurich, Switzerland
62: Also at Stefan Meyer Institute for Subatomic Physics, Vienna, Austria
63: Also at Laboratoire d'Annecy-le-Vieux de Physique des Particules, IN2P3-CNRS, Annecy-le-Vieux, France
64: Also at Research Center of Experimental Health Science, Near East University, Mersin, Turkey
65: Also at Konya Technical University, Konya, Turkey
66: Also at Izmir Bakircay University, Izmir, Turkey
67: Also at Adiyaman University, Adiyaman, Turkey
68: Also at Bozok Universitetesi Rektörlügü, Yozgat, Turkey
69: Also at Marmara University, Istanbul, Turkey
70: Also at Milli Savunma University, Istanbul, Turkey
71: Also at Kafkas University, Kars, Turkey
72: Now at stanbul Okan University, Istanbul, Turkey
73: Also at Hacettepe University, Ankara, Turkey
74: Also at Erzincan Binali Yildirim University, Erzincan, Turkey
75: Also at Faculty of Engineering, Istanbul University-Cerrahpasa, Istanbul, Turkey
76: Also at Yildiz Technical University, Istanbul, Turkey
77: Also at Vrije Universiteit Brussel, Brussel, Belgium
78: Also at School of Physics and Astronomy, University of Southampton, Southampton, UK
79: Also at IPPP Durham University, Durham, UK
80: Also at Faculty of Science, Monash University, Clayton, Australia
81: Also at Università di Torino, Torino, Italy
82: Also at Bethel University, St. Paul, Minnesota, USA

83: Also at Karamanoğlu Mehmetbey University, Karaman, Turkey
84: Also at California Institute of Technology, Pasadena, CA, USA
85: Also at United States Naval Academy, Annapolis, MD, USA
86: Also at Ain Shams University, Cairo, Egypt
87: Also at Bingol University, Bingol, Turkey
88: Also at Georgian Technical University, Tbilisi, Georgia
89: Also at Sinop University, Sinop, Turkey
90: Also at Erciyes University, Kayseri, Turkey
91: Also at Horia Hulubei National Institute of Physics and Nuclear Engineering (IFIN-HH), Bucharest, Romania
92: Now at an Institute or an International Laboratory Covered by a Cooperation Agreement with CERN, Geneva, Switzerland
93: Also at Texas A&M University at Qatar, Doha, Qatar
94: Also at Kyungpook National University, Daegu, Korea
95: Also at Another Institute or International Laboratory Covered by a Cooperation Agreement with CERN, Geneva, Switzerland
96: Also at Universiteit Antwerpen, Antwerpen, Belgium
97: Also at Yerevan Physics Institute, Yerevan, Armenia
98: Also at Northeastern University, Boston, MA, USA
99: Also at Imperial College, London, UK
100: Also at Institute of Nuclear Physics of the Uzbekistan Academy of Sciences, Tashkent, Uzbekistan