Simultaneous precision spectroscopy of *pp*, ⁷Be, and *pep* solar neutrinos with Borexino Phase-II

M. Agostini,[1] K. Altenmüller,[1] S. Appel,[1] V. Atroshchenko,[2] Z. Bagdasarian,[3] D. Basilico,[4] G. Bellini,[4] J. Benziger,[5] G. Bonfini,[6] D. Bravo,[4,†] B. Caccianiga,[4] F. Calaprice,[7] A. Caminata,[8] L. Cappelli,[6] S. Caprioli,[4] M. Carlini,[6] P. Cavalcante,[9,‡] F. Cavanna,[8] A. Chepurnov,[10] K. Choi,[11] L. Collica,[4] D. D'Angelo,[4] S. Davini,[8] A. Derbin,[12] X. F. Ding,[13,6] A. Di Ludovico,[7] L. Di Noto,[8] I. Drachnev,[12] K. Fomenko,[14] A. Formozov,[14,4,10] D. Franco,[15] F. Gabriele,[6] C. Galbiati,[7,13] M. Gschwender,[16] C. Ghiano,[6] M. Giammarchi,[4] A. Goretti,[7,‡] M. Gromov,[10,14] D. Guffanti,[13,6] T. Houdy,[15] E. Hungerford,[17] Aldo Ianni,[6] Andrea Ianni,[7] A. Jany,[18] D. Jeschke,[1] S. Kumaran,[3,19] V. Kobychev,[20] G. Korga,[17,§] T. Lachenmaier,[16] M. Laubenstein,[6] E. Litvinovich,[2,21] P. Lombardi,[4] L. Ludhova,[3,19] G. Lukyanchenko,[2] L. Lukyanchenko,[2] I. Machulin,[2,21] G. Manuzio,[8] S. Marcocci,[13,‖] J. Maricic,[11] J. Martyn,[22] E. Meroni,[4] M. Meyer,[23] L. Miramonti,[4] M. Misiaszek,[18] V. Muratova,[12] B. Neumair,[1] M. Nieslony,[22] L. Oberauer,[1] V. Orekhov,[2] F. Ortica,[24] M. Pallavicini,[8] L. Papp,[1] Ö. Penek,[3,19] L. Pietrofaccia,[7] N. Pilipenko,[12] A. Pocar,[25] A. Porcelli,[22] G. Raikov,[2] G. Ranucci,[4] A. Razeto,[6] A. Re,[4] M. Redchuk,[3,19] A. Romani,[24] N. Rossi,[6,¶] S. Rottenanger,[16] S. Schönert,[1] D. Semenov,[12] M. Skorokhvatov,[2,21] O. Smirnov,[14] A. Sotnikov,[14] L. F. F. Stokes,[6] Y. Suvorov,[6,2,**] R. Tartaglia,[6] G. Testera,[8] J. Thurn,[23] E. Unzhakov,[12] F. Villante,[6,26] A. Vishneva,[14] R. B. Vogelaar,[9] F. von Feilitzsch,[1] S. Weinz,[22] M. Wojcik,[18] M. Wurm,[22] O. Zaimidoroga,[14] S. Zavatarelli,[8] K. Zuber,[23] and G. Zuzel[18]

(Borexino Collaboration)[*]

[1]*Physik-Department and Excellence Cluster Universe, Technische Universität München, 85748 Garching, Germany*
[2]*National Research Centre Kurchatov Institute, 123182 Moscow, Russia*
[3]*Institut für Kernphysik, Forschungszentrum Jülich, 52425 Jülich, Germany*
[4]*Dipartimento di Fisica, Università degli Studi and INFN, 20133 Milano, Italy*
[5]*Chemical Engineering Department, Princeton University, Princeton, New Jersey 08544, USA*
[6]*INFN Laboratori Nazionali del Gran Sasso, 67010 Assergi (AQ), Italy*
[7]*Physics Department, Princeton University, Princeton, New Jersey 08544, USA*
[8]*Dipartimento di Fisica, Università degli Studi and INFN, 16146 Genova, Italy*
[9]*Physics Department, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, USA*
[10]*Lomonosov Moscow State University Skobeltsyn Institute of Nuclear Physics, 119234 Moscow, Russia*
[11]*Department of Physics and Astronomy, University of Hawaii, Honolulu, Hawaii 96822, USA*
[12]*St. Petersburg Nuclear Physics Institute NRC Kurchatov Institute, 188350 Gatchina, Russia*
[13]*Gran Sasso Science Institute, 67100 L'Aquila, Italy*
[14]*Joint Institute for Nuclear Research, 141980 Dubna, Russia*
[15]*AstroParticule et Cosmologie, Université Paris Diderot, CNRS/IN2P3, CEA/IRFU, Observatoire de Paris, Sorbonne Paris Cité, 75205 Paris Cedex 13, France*
[16]*Kepler Center for Astro and Particle Physics, Universität Tübingen, 72076 Tübingen, Germany*
[17]*Department of Physics, University of Houston, Houston, Texas 77204, USA*
[18]*M. Smoluchowski Institute of Physics, Jagiellonian University, 30348 Krakow, Poland*
[19]*RWTH Aachen University, 52062 Aachen, Germany*
[20]*Kiev Institute for Nuclear Research, 03680 Kiev, Ukraine*
[21]*National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), 115409 Moscow, Russia*
[22]*Institute of Physics and Excellence Cluster PRISMA, Johannes Gutenberg-Universität Mainz, 55099 Mainz, Germany*
[23]*Department of Physics, Technische Universität Dresden, 01062 Dresden, Germany*
[24]*Dipartimento di Chimica, Biologia e Biotecnologie, Università degli Studi e INFN, 06123 Perugia, Italy*
[25]*Amherst Center for Fundamental Interactions and Physics Department, University of Massachusetts, Amherst, Massachusetts 01003, USA*

[26]*Dipartimento di Scienze Fisiche e Chimiche, Università dell'Aquila, 67100 L'Aquila, Italy*

We present the simultaneous measurement of the interaction rates R_{pp}, R_{Be}, R_{pep} of pp, ^{7}Be, and pep solar neutrinos performed with a global fit to the Borexino data in an extended energy range (0.19–2.93) MeV with particular attention to details of the analysis methods. This result was obtained by analyzing 1291.51 days of Borexino Phase-II data, collected after an extensive scintillator purification campaign. Using counts per day (cpd)/100 ton as unit, we find $R_{pp} = 134 \pm 10(\text{stat})^{+6}_{-10}(\text{sys})$, $R_{Be} = 48.3 \pm 1.1(\text{stat})^{+0.4}_{-0.7}(\text{sys})$; and $R_{pep}^{\text{HZ}} = 2.43 \pm 0.36(\text{stat})^{+0.15}_{-0.22}(\text{sys})$ assuming the interaction rate R_{CNO} of CNO-cycle (Carbon, Nitrogen, Oxigen) solar neutrinos according to the prediction of the high metallicity standard solar model, and $R_{pep}^{\text{LZ}} = 2.65 \pm 0.36(\text{stat})^{+0.15}_{-0.24}(\text{sys})$ according to that of the low metallicity model. An upper limit $R_{\text{CNO}} < 8.1$ cpd/100 ton (95% C.L.) is obtained by setting in the fit a constraint on the ratio R_{pp}/R_{pep} (47.7 ± 0.8 cpd/100 ton or 47.5 ± 0.8 cpd/100 ton according to the high or low metallicity hypothesis).

I. INTRODUCTION

Solar neutrinos produced in electron flavor (ν_e)in fusion reactions occurring in the Sun provide a unique and direct way to study the interior of our star. The main contribution to the solar luminosity (∼99%) comes from reactions belonging to the pp chain, while the Carbon–Nitrogen–Oxigen (CNO) cycle is expected to play a subdominant role [1].

The solar neutrino (ν) spectrum, as predicted by the standard solar model (SSM) [2,3], is dominated by the low-energy neutrinos produced in the primary pp reaction ($E_\nu < 0.42$ MeV) and it extends up to ∼18.8 MeV (maximum energy of the hep νs). It also features two monoenergetic lines from ^{7}Be νs ($E_\nu = 0.384$ MeV and 0.862 MeV) and one monoenergetic line from pep νs ($E_\nu = 1.44$ MeV). Neutrinos from the CNO cycle are expected to have a continuous energy spectrum extending

up to 1.74 MeV. The spectrum of ^{8}B νs is also continuous and it ends up at about 16.5 MeV.

The 50-year-long experimental effort to study solar neutrinos [4,5] has been extremely rewarding both in terms of solar physics, by confirming the SSM predictions [3], and in terms of particle physics, by giving a substantial contribution to the discovery of neutrino flavor oscillations [6,7]. The present-day precision spectroscopy of solar neutrinos aims at studying the details of their energy spectrum by disentangling the contributions from the different reactions (pp chain νs, namely pp, ^{7}Be, pep, ^{8}B, and hep νs, and CNO cycle νs).

On the one hand, if the SSM predictions of solar fluxes ϕ are assumed, measuring the solar neutrino interaction rates R for different reactions helps to pin down the electron-flavor neutrino survival probability P_{ee} for different energies (that is the probability that ν_es do not undergo flavor oscillations while traveling from their Sun production point to the detector). Consequently, it probes the predictions of the MSW-LMA model [8] and can set constraints on possible deviations, e.g., due to nonstandard interactions (NSI) [9].

On the other hand, if the neutrino oscillation parameters are assumed, the study of specific components of the solar neutrino spectrum can cross-check the SSM predictions. In particular, the experimental determination of the fluxes ϕ of ^{7}Be, ^{8}B, or CNO neutrinos, which are the most sensitive ones to the solar metallicity (the abundance of the elements heavier than He in the Sun), can help to settle the question of high (HZ) versus low (LZ) metallicity [3].

The Borexino experiment has recently reported a comprehensive measurement of the solar neutrino spectrum from the whole pp nuclear fusion chain in the energy range of (0.19–16) MeV. These results are presented in [10] together with their physical implications. They include the updated values of the neutrino survival probability P_{ee} as a

*Corresponding author.
spokeperson-borex@lngs.infn.it
[†]Present address: Universidad Autnoma de Madrid, Ciudad Universitaria de Cantoblanco, 28049 Madrid, Spain.
[‡]Present address: INFN Laboratori Nazionali del Gran Sasso, 67010 Assergi (AQ), Italy.
[§]Also at MTA-Wigner Research Centre for Physics, Department of Space Physics and Space Technology, 1121 Budapest, Hungary.
[∥]Present address: Fermi National Accelerato Laboratory (FNAL), Batavia, Illinois 60510, USA.
[¶]Present address: Dipartimento di Fisica, Sapienza Università di Roma e INFN, 00185 Roma, Italy.
[**]Present address: Dipartimento di Fisica, Università degli Studi Federico II e INFN, 80126 Napoli, Italy.

function of the neutrino energy, the first direct measurement of the ratio $\mathcal{R}$ between the $^3\mathrm{He} + {}^4\mathrm{He}$ (pp-II) and the $^3\mathrm{He} + {}^3\mathrm{He}$ (pp-I) branches of the pp chain obtained by combining our results on the $^7\mathrm{Be}$ and pp νs, and finally a preference for the HZ-metallicity choice in the SSM.

In this paper we present the details of the analysis of the data belonging to the lowest part of the energy spectrum which extends from 0.19 to 2.93 MeV. This low energy region (LER) is used to extract the interaction rates R_{pp}, R_{Be}, R_{pep}, as well as to set the limit on R_{CNO}. The analysis of the data from the so-called high energy region (HER) from 3.2 to 16 MeV, where our sensitivity to $^8\mathrm{B}$ νs is maximized and from 11 to 20 MeV energy region, in which the first Borexino limit on hep νs is set, is discussed in [11].

While our previous measurements of the pp [12], $^7\mathrm{Be}$ [5], pep [13], and $^8\mathrm{B}$ [14] νs were obtained separately by analyzing data in restricted energy ranges, the results of [10] provide a unified analysis over the interval covering the LER and HER. The experience from the previous analyses in different energy intervals, each of them having specific difficulties, was fundamental in the process of building up the comprehensive understanding of our data and of the detector response across the combined energy interval as a whole. In addition, other important elements of the measurement are: an accurate calibration campaign [15] in the energy interval ranging from 0.15 to 9 MeV carried out by deploying several radioactive sources inside the detector, a detailed Monte Carlo (MC) simulation fine-tuned to reproduce the calibration data simultaneously at low and at high energies [16], and the use of data processing and data selection as well as background-rejection tools common to the whole energy range.

The unified analysis approach in the LER, described in this work, together with a larger exposure and a reduction of the most relevant backgrounds in the Phase-II lead to a significant improvement of the accuracy of our previous Phase-I results about the R_{Be} (from 4.8% to 2.7%) and R_{pep} (from 21.6% to 17.4/16.3%, depending on the HZ/LZ-SSM assumption, respectively). For R_{pp} the improvement is smaller, from the precision of 11.4% to 10.6%.

II. THE BOREXINO DETECTOR AND THE DATA SELECTION

The Borexino experiment is located at the Laboratori Nazionali del Gran Sasso in Italy. The core of the detector [17] is 278 ton of ultrapure organic liquid scintillator, namely PC (pseudocumene, 1,2,4-trimethylbenzene) as a solvent and 1.5 g/l of fluor PPO (2,5-diphenyloxazole) as a solute, contained in a 125 μm-thick nylon inner vessel (IV) of 4.25 m radius, surrounded by nominally 2212 8-inch ETL 9351 photomultipliers (PMTs). Since the beginning of the data taking, we observed a slow PMT failure rate over time. As a reference, the number of working channels was

1769 at the beginning of the data-taking period considered in this work while it was 1383 at its end.

Neutrinos of any flavor interact by elastic scattering with electrons, whose recoil produces scintillation light ($\sim$500 photoelectrons/MeV/2000 PMTs). The density of target electrons in the scintillator is $(3.307 \pm 0.003) \times 10^{31}$/100 ton. A nonscintillating buffer fills the space between the IV and a stainless-steel sphere (SSS) of 6.85 m radius, which supports the PMTs. The buffer liquid is further divided in two regions by another nylon vessel of radius 5.5 m which prevents radon emanating from the SSS and the PMTs to enter the core of the detector. The entire detector is enclosed in a cylindrical tank filled with ultra-pure water and instrumented with 208 PMTs, acting as an active Cherenkov muon veto and as a passive shield against external γs and neutrons.

The present analysis is based on the data collected between December 14, 2011 to May 21, 2016, which corresponds to an exposure of 1291.51 days $\times$ 71.3 t ($\sim$1.6 times the exposure used in [5]). This period belongs to the so-called Borexino Phase-II, which started after an extensive purification campaign of the scintillator with 6 cycles of closed-loop water extraction, which has significantly reduced the radioactive contaminants: $^{238}\mathrm{U} < 9.4 \times 10^{-20}$ g/g (95% C.L.), $^{232}\mathrm{Th} < 5.7 \times 10^{-19}$ g/g (95% C.L.), $^{85}\mathrm{Kr}$ and $^{210}\mathrm{Bi}$ reduced respectively by a factor $\sim$4.6 and $\sim$2.3 (see this work).

The expected solar νs interaction rate in Borexino ranges from few to $\sim$100 cpd/100 ton depending on the neutrino component. Together with the lack of directionality information from the scintillation light, this low rate demands a high detector radio-purity, a deep understanding of the backgrounds, and an accurate modeling of the detector response.

The position and pulse-shape of each event are reconstructed by exploiting the number of detected photons and their detection times. The information about the event energy is carried by the number of detected photoelectrons or just the number of hit PMTs, as in our energy range the PMTs mainly work in a single photoelectron regime. In detail, we define different energy estimators: N_p which is the total number of hit PMTs in the event or $N_p^{dt_{1(2)}}$, the number of hit PMTs happening within a fixed time interval of 230 (400) ns; N_h the number of detected hits, including multiple hits on the same PMT and, finally N_{pe}, the total charge collected by each PMT anode, that is the number of photoelectrons, p.e. As it will be detailed in Sec. V C, the energy is not reconstructed meaning that, during the analysis procedure, we do not convert the values of the energy estimator into the event energy. On the contrary, we build the prediction of the measured variables transforming the theoretical event energy into the corresponding value of a given energy estimator. As a reference, at 1 MeV, the energy and position reconstruction resolutions are $\sim$50 keV and $\sim$10 cm, respectively. The trigger threshold

is $N_p > 20$ in a 100 ns time window, which corresponds to ~50 keV.

To account for the variation in the number of working channels as a function of time, in the analysis and simulation procedures, all the energy estimators are normalized to a fixed number N_{tot} of PMTs (typically $N_{tot} = 2000$ PMTs) [18] through the relation $N_{p,h,pe} = N^m_{p,h,pe} \cdot N_{tot}/N'(t)$, with $N^m_{p,h,pe}$ being the measured value of the energy estimator and $N'(t)$ is the time-dependent number of working PMTs.

Events in the entire LER are selected using the same cuts described in [12]: we remove internal (external) muons [19] and we apply a 300 (2) ms veto to suppress cosmogenic backgrounds. The total dead-time introduced by these vetoes is 1.5%. We remove ^{214}Bi $- ^{214}$Po fast coincidences from the ^{238}U chain and unphysical noise events. The fraction of good events removed by these cuts, estimated using MC simulations [16] and calibration data [15], is ~0.1%. Background from sources external to the scintillator (nylon vessel, SSS, and PMTs) is reduced with a fiducial volume (FV) cut, which selects the innermost region of the scintillator (71.3 ton), contained within the radius $R < 2.8$ m and the vertical coordinate $-1.8 < z < 2.2$ m.

III. BACKGROUND

The residual background, after the application of the described selection cuts, is mainly due to radioactive isotopes contaminating the scintillator itself, such as ^{14}C (β^- decay, Q = 0.156 MeV, $\tau = 8270$ years), ^{210}Po (α decay, $E_\alpha = 5.3$ MeV, $\tau = 200$ days, originating a scintillation light signal quenched by a factor ~10), ^{85}Kr (β^- decay, Q = 0.687 MeV, $\tau = 15.4$ years), and ^{210}Bi (β^- decay, Q = 1.16 MeV, $\tau = 7.23$ days), a relatively short lived daughter of ^{210}Pb (β^- decay, Q = 0.063 MeV, $\tau = 32.2$ years). The lowest energy region (below 0.3 MeV), which is most sensitive to pp νs, contains an additional background due to the *pile-up* of uncorrelated events (mostly ^{14}C, external background primarily due to radioactive contaminants of the SSS and PMTs, and ^{210}Po [12,16]). The energy region sensitive to pep and CNO νs (between ~1.1 and ~1.7 MeV) is also affected by the cosmogenic isotope ^{11}C (β^+ decay, Q = 0.960 MeV, $\tau = 29.4$ min) and by residual external background, mainly as γs from the decay of ^{208}Tl (2.614 MeV), ^{214}Bi (<1.764 MeV), and ^{40}K (1.460 MeV).

The ^{11}C isotope is continuously produced in the liquid scintillator by muons through spallation on ^{12}C. In order to limit its effect on the sensitivity to pep νs, we exploit the so-called three-fold coincidence (TFC) method and e^+/e^- pulse-shape discrimination [13,18].

The TFC takes advantage of the fact that ^{11}C is often produced together with one or even a burst of neutrons. The principle of the method is thus to tag events correlated in

space and time with a muon and a neutron. We have improved the TFC technique already employed by us [13] by implementing a new algorithm, which evaluates the likelihood $\mathcal{L}_{TFC}$ that an event is a ^{11}C candidate, considering relevant observables such as the distance in space and time from the parent muon, the distance from the neutron, the neutron multiplicity, and muon dE/dx. Based on this probability, the data set is divided in two samples: one depleted (*TFC-subtracted*), obtained removing the ^{11}C tagged events, and one enriched (*TFC-tagged*) in ^{11}C. These two sets are separately fitted in the multivariate scheme (see later). The new TFC algorithm has $(92 \pm 4)\%$ ^{11}C-tagging efficiency, while preserving $(64.28 \pm 0.01)\%$ of the total exposure in the TFC-subtracted spectrum. Figure 1 shows the distribution of $\log(\mathcal{L}_{TFC})$ of the present data set as a function of the $N_p^{dt_1}$ energy estimator and it demonstrates how ^{11}C decays can be identified by cutting the events on the basis of the value of $\mathcal{L}_{TFC}$.

A. Pulse shape discrimination of β^+/β^- events

The residual amount of ^{11}C in the TFC-subtracted spectrum can be disentangled from the neutrino signal through variables with β^+/β^- pulse-shape discrimination capability [13,18]. We build these variables considering that the probability density function (PDF) of the time detection of the scintillation light is different for β^+ and β^- events for two reasons: (i) for β^+ events, in 50% of the cases, the e^+ annihilation is delayed by ortho-positronium formation, which survives in the liquid scintillator with a mean time $\tau \sim 3$ ns [20]; (ii) the topology e^+ energy deposit is not pointlike, due to the two back-to-back
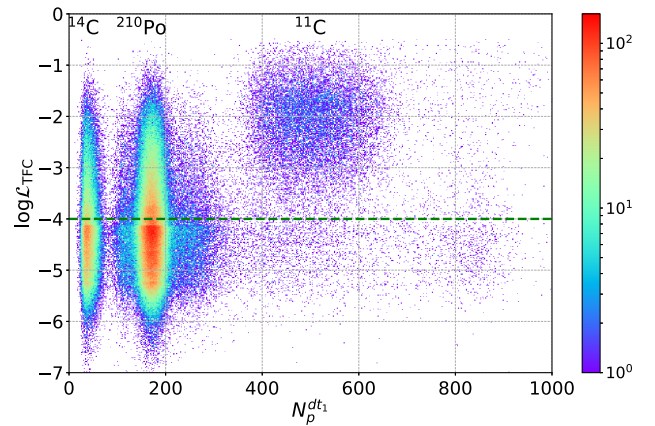

FIG. 1. Distribution of $\log(\mathcal{L}_{TFC})$ as a function of the $N_p^{dt_1}$ energy estimator. The plot is built using the entire set of data surviving the selection cuts described in Sec. II. The regions dominated by the abundant internal background of ^{14}C and ^{210}Po are indicated by the corresponding labels. The green-dashed horizontal line represents the $\mathcal{L}_{TFC}$-threshold, above/below which the events are assigned to the TFC-tagged/subtracted energy spectrum. It is clearly visible that the majority of the events of the ^{11}C energy decay spectrum lies above this threshold.

511 keV annihilation γs. These two features originate a pattern of the energy deposit of β^+ with a larger time and spatial spread than the corresponding one generated by β^-. Based on this fact, a pulse-shape (PS) discrimination algorithm has been constructed using the neural network of a boosted decision tree (BDT) and used for previous analysis as detailed in [18]. In the present analysis, we have introduced a novel discrimination parameter, called PS-$\mathcal{L}_{PR}$, defined as the maximum value of the likelihood function $\mathcal{L}_{PR}$ used in the position reconstruction (PR), divided by the value of the energy estimator. The latter normalization removes the $\mathcal{L}_{PR}$ energy-dependence, since it is calculated as the summation over the collected hits [18]. The PR-algorithm is based on the expected distribution of the arrival times of optical photons on the PMTs. For all events, the algorithm uses the scintillation light emission PDF of point-like β^- events. For this reason the distribution of the maximum likelihood value shows some discrimination capability for different types of particles, if they originate photon time patterns distinct from that of β^-.

The study of the performances of the PS-$\mathcal{L}_{PR}$ variable demands, from one side, the identification of samples of true β^- and β^+ events and, from another side, it requires to properly account for the variable number of working channels that influences its value.

A pure, high-statistics β^- sample can only be obtained from a limited time period of the water-extraction phase of the scintillator purification campaign. During this time, a temporary ^{222}Rn contamination entered the detector. Using the space-and-time correlation of the fast coinciding ^{214}Bi$(\beta^-) - {}^{214}$Po(α) decays, we have tagged about 10^4 ^{214}Bi events. The ability of the MC to reproduce the PS-L_{PR} parameter of these events and the comparison to data is shown in Fig. 2. The agreement between data and simulation demonstrates that the MC can accurately construct
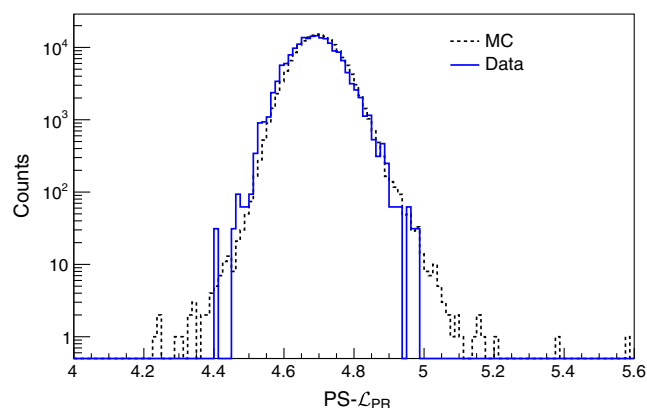
the PDF of this parameter for the entire set of data thus accounting for the variable number of working channels.

Our best β^+ sample is obtained from the TFC tagged events with hard cuts on the energy and on the time correlation with the neutron and muon tracks. These events are selected from the whole data set and thus they naturally follow the live-channels distribution. The discrimination capability of PS-L_{PR} is demonstrated comparing them with a MC sample of pure electrons with a flat energy distribution in the energy interval of the ^{11}C events, while also following the realistic live channel distribution over the whole data set. The PS-L_{PR} for these MC generated electrons was used as β^- sample in a further analysis (analytical multivariate fit) described in Sec. IV). Figure 3 shows the distribution of the PS-L_{PR} parameter for the MC generated electrons compared with that of β^+ events obtained from ^{11}C data. The difference between the two distributions at high values of PS-$\mathcal{L}_{PR}$ is the key element allowing the discrimination between β^- and β^+. Note that we do not need to build a position reconstruction algorithm based on the time profile of the scintillation light of β^+ events. Figure 4 shows the PS-$\mathcal{L}_{PR}$ pulse-shape discriminator as a function of $N_p^{dt_1}$ energy estimator for events selected with the cuts described in Sec. II and used in the present analysis.

It is interesting to note that the comparison between the BDT and PS-$\mathcal{L}_{PR}$ parameters, using the samples of true β^- and β^+ events, shows that they have similar discrimination power and they similarly help in reducing the systematic uncertainty of the $pep\ \nu$s result. However, the use of PS-$\mathcal{L}_{PR}$ offers some advantages like its simplicity, the fact that it can be calculated without the training procedure necessary for BDT (that suffers the limited size of the available β^- training sample), and finally, the possibility to easily reproduce it through the MC.
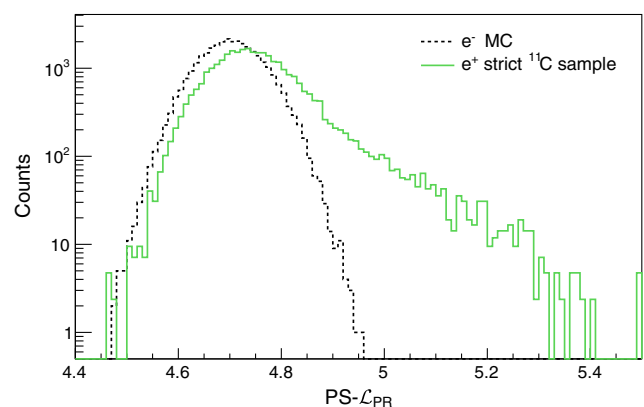

FIG. 2.   Comparison of the distributions of the PS-$\mathcal{L}_{PR}$ parameter for ^{214}Bi events extracted from data (blue, continuous line) and generated using MC (black, dashed line). The MC sample of ^{214}Bi was generated using the same spatial distribution of the ^{214}Bi events of the data. The simulation also takes into account the proper values of the working channels $N'(t)$.

FIG. 3.   Comparison of the distributions of the PS-$\mathcal{L}_{PR}$ variable for MC generated e^- events (black, dashed line) and for e^+ events selected from the data (green, continuous line). The latter events are a high-purity ^{11}C sample, obtained with the optimized TFC method, using very strict cuts on the energy and on the time correlation with the neutron and muon tracks.
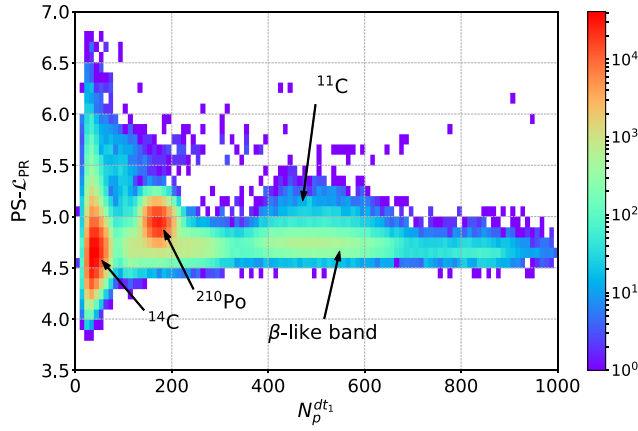
FIG. 4.   Distribution of PS-$\mathcal{L}_{\text{PR}}$ pulse-shape discriminator as a function of $N_p^{dt_1}$ energy estimator. The plot is built using the entire set of data surviving the selection cuts described in Sec. II. The comparison with Fig. 3 allows to identify the range of values belonging to the β^--like band indicated by the arrow.

IV. MULTIVARIATE FIT

The most powerful signatures for the detection of solar neutrinos in Borexino are the shapes of the energy spectra from electrons that underwent elastic scattering interactions with neutrinos. However, the recognition of these shapes is somewhat hindered by the contribution of various types of background events. In addition, the spectral details are also masked through the finite energy resolution of the detector and eventually distorted by nonlinear effects linking the energy deposit in the scintillator and the observed energy estimator.

Signal and background can be disentangled through an accurate fit. In order to enhance our sensitivity to the neutrino signal, we have adopted in the entire LER the multivariate fit approach already exploited in [13]. We maximize a binned likelihood function containing the information from the TFC-tagged and TFC-subtracted energy spectra. Additional information from the PS-$\mathcal{L}_{\text{PR}}$ parameter and the radial distributions of the events in the optimized energy regions are included in the fit. The radial information is important to accurately measure the background rates due to external γs produced by the contamination of the PMTs and the supporting SSS. The pulse shape parameter PS-$\mathcal{L}_{\text{PR}}$ helps in the separation of the residual $^{11}\text{C}(e^+)$ background from the e^-–like components, and this is relevant for the determination of R_{pep} and R_{CNO}.

Several ingredients are necessary to perform the fit. The first one is a background model, that is a list of possible radioactive contaminants that we assume give a contribution to the measured signal. The second one is the detector response function, i.e., a full model of the distributions of all the physical variables that we measure. The knowledge of the detector response function allows the prediction of the probability density functions of all the quantities entering the fit procedure.

As done in previous Borexino analyses, we have adopted two complementary methods to build the detector response function: an analytical approach and a MC based procedure. The only free parameters of the fit in the MC approach are the interaction rates of neutrino and background species, while in the analytical method (see later), in addition, some of the parameters related to the response function and to the energy scale are also free and determined by the fit procedure. These two methods share the same background model.

Fitting tools based on the use of graphical processing units (GPU) have been developed and used with the analytical fit method. They decrease the computation time by about 3 orders of magnitude compared with the standard CPU based algorithms previously used [21].

A. Multivariate likelihood function

The TFC-subtracted and TFC-tagged data sets are fitted simultaneously by maximizing a likelihood function $\mathcal{L}_{3D}(\vec{\theta}|\vec{k})$ defined as

$$\mathcal{L}_{3D}(\vec{k}|\vec{\theta}) = \mathcal{L}_{\text{sub}}^{\text{TFC}}(\vec{k}|\vec{\theta}) \cdot \mathcal{L}_{\text{tag}}^{\text{TFC}}(\vec{k}|\vec{\theta}). \tag{1}$$

The symbol $\vec{\theta}$ indicates the set of the arguments with respect to which the function is maximized and $\vec{k}$ generically indicates the set of the experimental data used to evaluate the likelihood. The two factors in Eq. (1) are the likelihood functions related to TFC-subtracted and TFC-tagged energy spectra, respectively.

$\mathcal{L}_{\text{sub}}^{\text{TFC}}(\vec{k}|\vec{\theta})$ is the standard Poisson binned likelihood function:

$$\mathcal{L}_{\text{sub}}^{\text{TFC}}(\vec{k}|\vec{\theta}) = \prod_{j,l,m=0}^{N_{E,R,P}} \frac{\lambda_{jlm}^{k_{jlm}}(\vec{\theta})}{k_{jlm}!} e^{-\lambda_{jlm}(\vec{\theta})}, \tag{2}$$

where $\vec{k}$ in this case is the ensemble of the data entries $k_{j,l,m}$ in the energy bin j, position bin l, and pulse shape parameter bin m; $\lambda_{j,l,m}(\vec{\theta})$ are the expected number of entries in the same bins, and $N_{E,R,P}$ are the total number of energy, radial, and pulse shape parameter bins. $\mathcal{L}_{\text{tag}}^{\text{TFC}}(\vec{k}|\vec{\theta})$ is constructed in a similar way but it does not include the pulse shape variable:

$$\mathcal{L}_{\text{tag}}^{\text{TFC}}(\vec{k}|\vec{\theta}) = \prod_{j,l=0}^{N_{E,R}} \frac{\lambda_{j,l}^{k_{j,l}}(\vec{\theta})}{k_{j,l}!} e^{-\lambda_{j,l}(\vec{\theta})}. \tag{3}$$

and $\vec{k}$ represents in this case the set of data entries $k_{j,l}$ in the energy and radial bins j, l integrated with respect to the pulse shape parameter. The signal of ^{11}C in the TFC-tagged spectrum is relatively strong compared to the other spectral components and the fit procedure extracts it very efficiently

thanks to its spectral shape. This is the reason driving the choice of using the two dimensional (2D) likelihood function of Eq. (3) for the TFC-tagged spectrum instead of a complete function of Eq. (2) that, for the TFC-tagged spectrum, only increases the computation time without bringing additional information.

Both the TFC-subtracted and TFC-tagged spectra are fitted keeping the rates of the majority of the components in common, except ^{11}C itself, ^{6}He and ^{10}C (which have cosmogenic origin), and ^{210}Po, that is not distributed homogeneoulsy through the detector volume.

Constraints on the values of the multivariate fit parameters are implemented (if not specified otherwise) as multiplicative Gaussian terms in the likelihood function.

The likelihood function of Eq. (2) and Eq. (3) are exactly the ones which are maximized using our most recent version of the MC-based fit procedure (see Sec. V A). Precisely, we generate with the MC every signal and background component and we build and properly normalize 3D (or 2D) histograms of the simulated number of events as a function of the energy estimator, PS-$\mathcal{L}_{PR}$ parameter, and radius (or of the energy estimator and radius only). The quantities λ_{jlm} and λ_{jl} of Eq. (2) and Eq. (3) represent the sum of the bin content of the histograms, each one weighted by the rate of the specific component ($\vec{\theta}$).

Earlier versions of the MC fit and the present analytical fit maximize an approximated version of the likelihood $\mathcal{L}_{3D}(\vec{k}|\vec{\theta})$, as already described in [18]. This function, called $\mathcal{L}(\vec{k}|\vec{\theta})$, is written as a product of four factors coming from the TFC-subtracted and TFC-tagged energy spectra ($\mathcal{L}_{E,\text{sub}}^{\text{TFC}}$ and $\mathcal{L}_{E,\text{tag}}^{\text{TFC}}$) and from the PS-$\mathcal{L}_{PR}$ ($\mathcal{L}_P$) and radial ($\mathcal{L}_R$) distributions of events in the ^{11}C-energy-range of the TFC-subtracted spectrum:

$$\mathcal{L}(\vec{k}|\vec{\theta}) = \mathcal{L}_{E,\text{sub}}^{\text{TFC}}(\vec{k}|\vec{\theta}) \cdot \mathcal{L}_{E,\text{tag}}^{\text{TFC}}(\vec{k}|\vec{\theta}) \cdot \mathcal{L}_P(\vec{k}|\vec{\theta}) \cdot \mathcal{L}_R(\vec{k}|\vec{\theta}). \quad (4)$$

The first two terms, $\mathcal{L}_{E,\text{sub}}^{\text{TFC}}(\vec{k}|\vec{\theta})$ and $\mathcal{L}_{E,\text{tag}}^{\text{TFC}}(\vec{k}|\vec{\theta})$, are Poisson likelihoods [like Eq. (2) and (3)] with $\vec{k}$ being the data entries k_j in the energy bin j integrated with respect to the other variables.

The other two terms in Eq. (4) have been built considering that in the framework of the analytical approach, there is no model able to produce precise multidimensional PDFs. Thus we have projected the events, from the optimized energy intervals of the TFC-subtracted spectrum and integrated over energy ranges larger than the binning of the energy spectrum, into 1D histograms of the pulse-shape and radial distributions. $\mathcal{L}_P(\vec{k}|\vec{\theta})$ and $\mathcal{L}_R(\vec{k}|\vec{\theta})$ of Eq. (4) are then built fitting these 1D distributions using PDFs obtained either from the data (high purity ^{11}C sample for β^+ pulse shape) or based on the MC simulation (β^- pulse shape, radial distributions). In the calculation of the

corresponding likelihoods, we introduce a correlation between the number of counts in different histograms, as events that are in the energy spectrum will also be entries in the projections. To handle this issue, we normalize the functions to the total number of entries N in the projected data histograms. Consequently, we define the likelihood of the PS-$\mathcal{L}_{PR}$ parameter as we did in [18] for the previously used PS-BDT parameter:

$$\mathcal{L}_P(\vec{k}|\vec{\theta}) = \prod_{m=1}^{N_P^{1D}} \frac{a\lambda_m^{k_m}(\vec{\theta})e^{-a\lambda_m(\vec{\theta})}}{k_m!}, \quad (5)$$

where the scaling parameter a enforces the normalization and is set such

$$N = a \sum_{m=1}^{N_P^{1D}} \lambda_m(\vec{\theta}), \quad (6)$$

where N is the total number of entries in the projected histogram and a is a scaling factor. Here, k_m is the actual number of entries of bin m of the 1D projection of the PS-$\mathcal{L}_{PR}$ distribution in a fixed energy interval, N_P^{1D} is the total number of bins of this histogram, and $\lambda_m(\vec{\theta})$ represents the expected content in bin m. $\mathcal{L}_R(\vec{k}|\vec{\theta})$ is defined in a way similar to $\mathcal{L}_P(\vec{k}|\vec{\theta})$.

The results of the MC-based fit, which is either performed using $\mathcal{L}_{3D}(\vec{k}|\vec{\theta})$ or $\mathcal{L}(\vec{k}|\vec{\theta})$, are consistent, confirming that no systematic uncertainty is introduced when using the approximated likelihood function.

V. DETECTOR RESPONSE FUNCTION

A. The Monte Carlo method

The MC code developed for Borexino [16] is a customized Geant4-based simulation package [22], which can simulate all processes following the interaction of a particle in the detector (energy loss including ionisation quenching in the scintillator; scintillation and Cherenkov light production; optical photon propagation and interaction in the scintillator modelling absorption and re-emission, Rayleigh scattering, interaction of the optical photons with the surface of the materials; photon detection on the PMTs, and response of the electronics chain) including all known characteristics of the apparatus (geometry, properties of the materials, variable number of the working channels over the duration of the experiment as in the real data) and their evolution in time. The code thus produces a fully simulated detector response function because it provides a simulated version of all the measured physical variables.

All the MC input parameters have been chosen or optimized using samples of data independent from the ones used in the present analysis (laboratory measurements and Borexino calibrations with radioactive sources [15])

and the simulation of the variables relevant for the present analysis has reached sub-percent precision [16].

Once the MC input parameters have been tuned, the PDFs of all the needed variables related to each of the ν and background components are built simulating events according to the specific energy spectrum. In order to properly reproduce the spatial dependence of the energy response, events are simulated in the detector following their expected spatial distribution: while the ν and most of background events are expected to be uniformly distributed in the detector, ^{210}Po decays are simulated according to their actual spatial and time distribution obtained from experimental data. Note that data events due to the α decay of ^{210}Po are efficiently identified by tagging ^{210}Po with a pulse–shape discrimination method based on the multilayer perceptron (MLP) algorithm [23] (a particular class of neural network algorithms). Similarly, γs from external background are generated on the SSS and PMTs surfaces so that the radial distribution of the interactions inside the scintillator volume shows a clear decrease from the outer region of the detector toward the center.

Events generated according to the theoretical signal and background energy spectra are then processed as real data. As already anticipated, for every species, 3D or 2D histograms are built for the energy estimators, the reconstructed radius, and the PS-$\mathcal{L}_{PR}$ variable. When properly binned and normalized, these histograms represent the PDFs to be used in the fit and they provide the values $\lambda_{jlm}(\vec{\theta})$ in Eq. (2) and $\lambda_{jl}(\vec{\theta})$ in Eq. (3). In the MC approach there are no free fit parameters other than the interaction rates of all species. The goodness of the fit simultaneously demonstrates the accuracy of the MC simulation, as well as the stability of the detector response over the period of five years.

In the wide energy range covered by this analysis, there is a huge difference between the number of measured counts per bin in the lower and in the higher energy regions. In the construction of the 3D PDFs, the need to simulate large numbers of events becomes really important, since they are scattered over a larger number of bins. To mitigate the consequences due to low populated bins and to have a good approximation to a χ^2, we have replaced the energy estimator and the radius R with some transformed variables. We choose to use R^3 instead of R, thus using bins of 5 m^3 each and still achieving a very effective separation of the external background from the bulk components. Similarly, we introduced a transformed variable $N_{h'}$ based on the N_h energy estimator: this change of variable is equivalent to adopting a variable binning size that scales with energy proportionally to the width of the N_h distribution obtained simulating monoenergetic electrons. This approach allows to reduce the statistical fluctuations without losing any physical information. As a by-product, this efficient binning significantly reduced the computing time needed to perform a single fit, speeding up the analysis of

the MC pseudoexperiments used to estimate the statistical and systematic uncertainties of the measurement described in Sec. VI.

The multivariate analysis was not applied on the whole energy range: the radial information was considered only for $N_h > 290$ to exclude from the analysis the spatial distribution of ^{210}Po, while the PS-$\mathcal{L}_{PR}$ was used where ^{11}C is present ($409 < N_h < 645$). The shape of the probability density function of the PS-$\mathcal{L}_{PR}$ variable for β^+ was obtained from an empirical parametrization of the distribution generated by the MC, with an additional small shift to compensate differences between the MC simulation outcome and a sample of strictly selected ^{11}C events.

B. The analytical model of the energy response function

In the analytical approach, we introduce a PDF for the energy estimator under consideration and analytical expressions for its mean value and variance. This PDF describes the detector's energy response function to monoenergetic events and, in brief, it is mainly influenced by the number of scintillation and Cherenkov photons and effects due to the nonuniformity of the light collection. As already anticipated, we then transform the energy spectra of each species into the corresponding distributions of the energy estimators. Effects like the ionization quenching in the scintillator, the contribution of the Cherenkov light, the spatial dependence of the reconstructed energy and its resolution are accounted for through some parameters, part of which are fixed, while others are free to vary in the final fit.

We describe here the present model for N_p which is derived from [18], with several improvements to extend the energy range of the fit to the entire LER. The same model describes the variables $N_p^{dt_{1(2)}}$. All used energy estimators are obtained after normalising the corresponding measured values to a reference configuration of $N_{tot} = 2000$ channels (defined in Sec. II).

As energy response function for the entire LER, we use the scaled Poisson function $f(N_p)$ [and similarly $f(N_p^{dt_{1(2)}})$] already introduced for analyzing events in the lowest region of the energy spectrum and detailed in [24] and in [25]

$$f(N_p) = \frac{m^{(N_p s)}}{(N_p s)!}\exp^{-m}. \qquad (7)$$

The two free parameters of this function, m and s, are fixed using the expressions for the mean value $\bar{N}_p(E)$ and variance σ_p^2 developed in the context of our model and described below:

$$m = \frac{\bar{N}_p^2(E)}{\sigma_p^2} \qquad (8)$$

and

$$s = \frac{\sigma_p^2}{\bar{N}_p(E)}. \tag{9}$$

In order to obtain $\bar{N}_p(E)$ we first consider that the mean number of photoelectrons $\bar{N}_{pe}(E)$ for each event of energy E takes its main contribution from the scintillation photons with a subdominant correction from the Cherenkov light, and it can be written as follows

$$\bar{N}_{pe}(E) = Y_0^{pe} \cdot [Q(E) \cdot E + f_{Ch} \cdot F_{Ch}(E)] \tag{10}$$

where Y_0^{pe} is the photoelectron yield expressed in photoelectrons/MeV for events in the detector center; the quenching term, $Q(E)$, accounts for nonlinearity of the scintillator response; $F_{Ch}(E)$, an analytical parametrization of Cherenkov light dependence on energy valid for electrons, provides the smooth transition between linear

dependency at the energies above 1–2 MeV and zero contribution for electrons below the Cherenkov threshold $E_0 = 0.165$ MeV; f_{Ch} is a parameter allowing to adjust the relative weight of the scintillation and Cherenkov light. Table I reports details of the analytical expressions.

Similarly to that described in [18], $\bar{N}_p(E)$, is linked to $\bar{N}_{pe}(E)$ through

$$\bar{N}_p(E) = N_{\text{tot}}[1 - e^{-\mu}(1 + p_t\mu)](1 - g_C\mu) \tag{11}$$

where $\mu = \frac{\bar{N}_{pe}(E)}{N_{\text{tot}}}$, g_C is a geometric correction factor, calculated for the given fiducial volume, and p_t is the fraction of a single photoelectron signal below the electronics threshold. These expressions extend the ones previously used in [18] with the introduction of the f_{Ch} and p_t parameters.

The second ingredient of the analytical model is the variance σ_p^2 of the N_p energy estimator. It is described by

TABLE I. Summary of the parameters used in the analytical model of the energy response function. In case of the free parameters, the values given in the table are obtained with the multivariate analytical fit of the data discussed in this paper and reported in Sec. VII. Only statistical errors are shown here. Energy is expressed in MeV; v_T^0, v_T^α (v_T^q) must be multiplied by 10^{-6} (10^{-4}).

Parameter	Fix./Free	Meaning/Approach to fixing	Value
Y_0^{pe}	Free	Photoelectron yield [p.e./MeV] for events in the detector center and with N_{tot} = 2000 PMTs	551 ± 1
g_C	Fixed	Fit $N_p^{dt_{1(2)}}$ vs true N_{pe} of MC with Eq. (11) using MC mono-energetic electron samples at 4 energies, simulated along the whole data-set.	0.101
p_t	Fixed	Fraction of a single photoelectron charge spectrum below the electronics threshold; fixed from the earlier calibration measurements and calculations.	0.12
f_{Ch}	Fixed	Relative weight of the scintillation and Cherenkov light; fixed by performing many analytical fits on data with it as a free/fixed parameter.	1.0
$F_{Ch}(E)$	Fixed	$F_{Ch}(E) = (C_0 + C_1 \cdot x + C_2 \cdot x^2 + C_3 \cdot x^3)(1 + C_4 \cdot E)$ $x = \ln(1 + E/E_0);$ $E_0 = 0.165$ MeV	$C_0 = 1.415;$ $C_1 = -3.397;$ $C_2 = 1.107;$ $C_3 = 0.072;$ $C_4 = 1.337$
$Q(E)$	Fixed	Quenching term summarizing the effects related to nonlinearity of the scintillator response according to Birk's quenching model [18]: $Q(E, k_B) = \frac{1}{E} \int_0^E \frac{dE'}{1 + k_B dE'/dx}$, where k_B is the Birk's constant, and Q(E) can be parametrized as: $Q(E, k_B) = \frac{A_1 + A_2 \ln E + A_3 \ln E^2}{1 + A_4 \ln E + A_5 \ln E^2}$; fixed from the fit of N_{pe} vs E with MC simulation of γ calibration data.	$k_B[\text{cm/MeV}] = 0.0109;$ $A_1 = 0.972;$ $A_2 = 0.201;$ $A_3 = 0.0105;$ $A_4 = 0.195;$ $A_5 = 0.014$
v_1	Fixed	Relative variance of the probability that a PMT triggers for events uniformly distributed in the detector volume, calculated using dedicated MC studies. It has some energy dependence and then we are using a value averaged over the LER.	0.16
v_T^0	Free	Spatial nonuniformity of the number of triggered PMTs.	0.50 ± 0.37
v_N	Free	Scintillator intrinsic resolution parameter for βs (caused by δ-electrons) that also effectively takes into account other contributions at low energies.	11.5 ± 1.0
v_T^q	Fixed	Nonuniformity of the light collection, calculated from MC events uniformly distributed in FV.	7.0
v_T^α	Free	Spatial nonuniformity resolution, corresponding to the width of ^{210}Po-α peak.	4.73 ± 0.21
σ_d	Fixed	PMT dark noise contribution	$0.23 N_p^{dt_1}$, $0.4 N_p^{dt_2}$

the following expression which extends the model already described in [18] in particular with the modification of the term linear with $\bar{N}_p(E)$ and the addition of a quadratic one

$$\sigma_p^2 = f_{eq}[1 - (1 + v_1)p_1]\bar{N}_p(E) + v_T^0\bar{N}_p^3(E)$$
$$+ v_T^q\left(\mu\frac{p_0}{p_1}\right)^2\bar{N}_p^2(E) + v_N\bar{N}_p(E) + \sigma_d^2 \qquad (12)$$

where v_1 is the relative variance of the PMT triggering probability for events uniformly distributed in the detector volume, $p_1 = 1 - e^{-\mu}$ is the probability of having a signal at any PMT, $p_0 = e^{-\mu}$ is the probability of absence of the signal, v_T^0 accounts for the spatial nonuniformity of the number of triggered PMTs, v_T^q accounts for the nonuniformity of the light collection, v_N is the intrinsic resolution parameter of the scintillator for βs that effectively includes other contributions at low energies, and the last term σ_d describes the effects of the dark noise an of PMTs. The channel equalization factor f_{eq} is the ratio between N_{tot} and the actual number of working PMTs and it changes during the data taking period.

In summary, the cubic term takes into account the variance of the number of the triggered PMTs for the events with fixed collected charge in the IV. The quadratic term takes into account the variance of the light collection function over the detector, and is generally weaker compared to the cubic term (and was neglected in previous analyses with more uniform PMTs distribution).

Formula (12) was derived analytically and verified against the MC simulations. For α particles we are using a simplified form with only the first and cubic term of relation 15 since we need to model a single energy point (^{210}Po). It is thus not necessary to follow the energy dependence of the variance. The coefficient of the cubic term is called v_T^α and it corresponds to the width of the ^{210}Po-α peak. As anticipated, we use the previous relations also for describing the mean value and variance of the estimators $N_p^{dt_{1(2)}}$.

Most of the above listed parameters are tuned using data independent from the ones used in the solar neutrino fit, calibrations or MC and are fixed in the fit ($Q(E)$, f_{Ch}, p_t, g_C, v_1, v_T^q), while other parameters (Y^{pe}, v_T^α, v_T^0, v_N) are left free to vary in the fit, together with the neutrino and background interaction rates. The two parameters p_t, g_C could be in principle free fit parameters, however they are fixed because the fit results have a low sensitivity to them.

In summary, the model has one free parameter describing the yield and three free parameters describing the energy resolution. Leaving the above listed parameters free gives the analytical fit the freedom to account for unexpected effects or unforeseen variations of the detector response in time. Table I reports all the parameters, free or fixed, appearing in the analytical fit with a short explanation about how they are obtained. In case of parameters kept free in the

fit, we report in the table the values obtained fitting the present data set as described in this paper. The corresponding values of the ν interaction rates and background are reported in Sec. VII.

C. Handling of the energy variables in the fit

We perform the fit of the energy spectra with the experimental data binned as a function of the energy estimators instead of transforming that distribution into the energy scale. Among the reasons driving this choice we remark that the analytical approach does not assume *a priori* knowledge of the prec energy transformation rules and the energy scale is automatically adjusted while fitting the experimental data. The use of the transformed experimental spectra would significantly slow down the fitting procedure, as the data reprocessing will be needed each time the energy scale parameters are changed in the fit. In addition, the presence of the contributions from ^{14}C and ^{210}Po with very high statistics makes the fit sensitive to tiny details of the energy response function (response to the monoenergetic event with a fixed energy distributed uniformly in the detector's volume). The shape of the energy response for the detected number of p.e. (or the number of the triggered PMTs) in the sub-MeV energy region is defined mainly by the statistical factor, with small additional smearing due to the nonuniformity of the amount of the collected light throughout the detector. The study performed using MC model showed that the shape of the charge response can be approximated by the generalized gamma function, and the shape of the N_p response can be approximated by the scaled Poisson function. But the energy response function in the energy scale does not allow a simple description with an analytical function, and thus complex calculations would be necessary if the transformed energy is used. In the MC approach the transformation to the energy scale is in principle feasible, because the energy scale and energy response in this approach are fixed from the calibrations, but it was not applied to keep internal consistency with the analytical approach.

Moreover, the amount of light emitted for a given energy deposit in the scintillator differs for the electrons, γs and α particles and then the energy scale calibrated for electrons is not valid for αs and γs. The experimental spectrum contains contributions from all these types of particle and the event-by-event identification of the type of interaction is not possible while the different contributions are statistically identified using the fit procedure. The binning of the data in the physical energy scale (as shown in the figures reporting the fit results) is performed only after the fit is completed.

VI. SENSITIVITY STUDIES

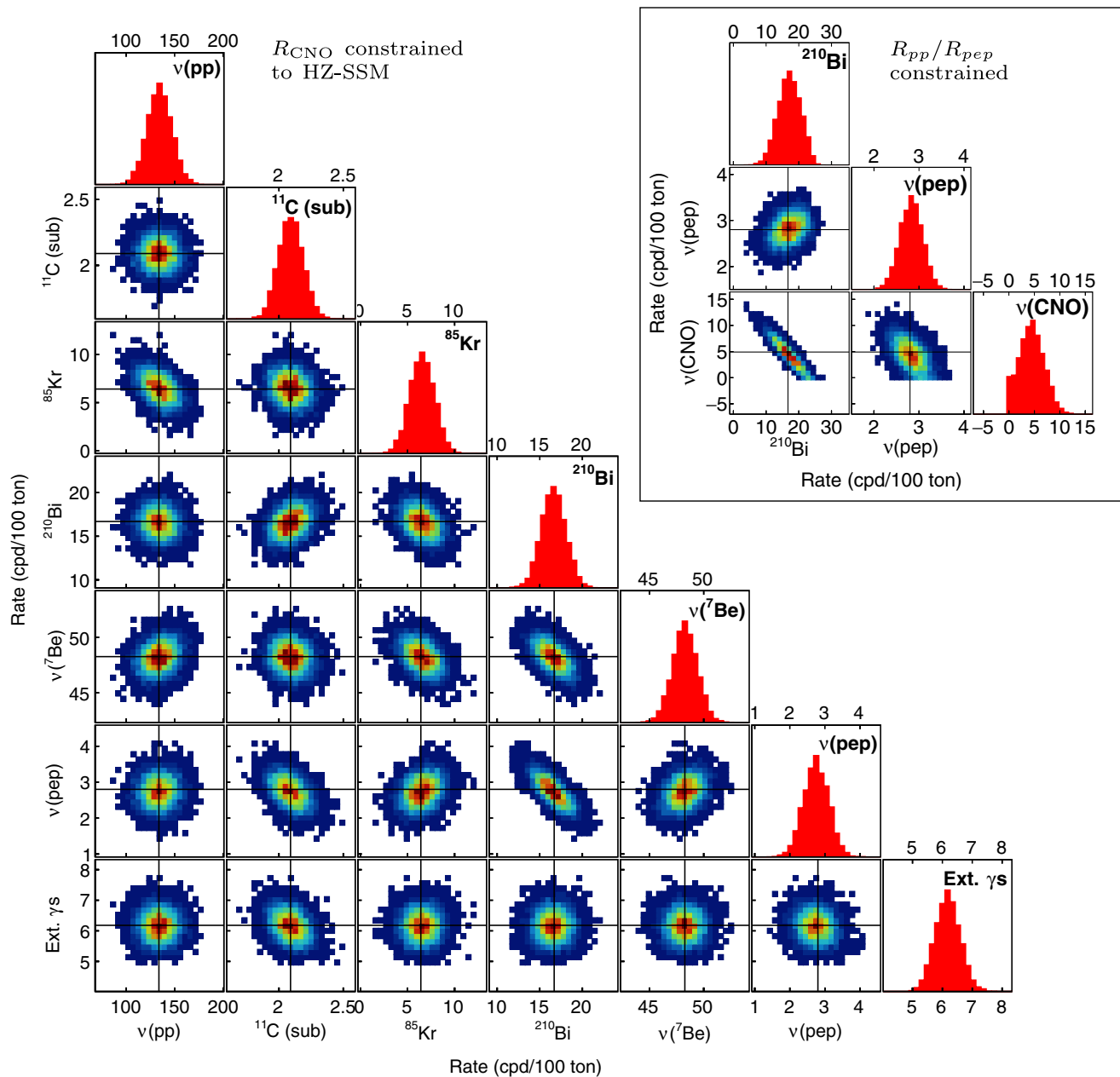Sensitivity studies have been performed by generating many pseudoexperiments with the MC and fitting this

FIG. 5. The figure shows the distributions of the interaction rates (cpd/100 ton) of solar ν and of the background species as they result from the MC fit of pseudo-experiments simulated with the same exposure as the experimental data discussed in this paper. The fit is performed in the entire LER region and, as in the real data analysis, penalty terms are added in the likelihood to constrain the values of the ^{14}C and pileup rates within the measured ones. It is interesting to note the correlation between the pp and ^{85}Kr rates, physically driven by the fact that a not negligible portion of the ^{85}Kr spectrum lies in the energy region around about 200 keV where we are sensitive to the pp νs signal. In the left plot, 6700 pseudoexperiments have been generated assuming the R_{CNO} according to HZ-SSM and fitted imposing a constraint on R_{CNO} to the same value. The same MC PDFs have been used to simulate and fit data, so these plots show only uncertainties due to statistical fluctuations and the effects of the correlations among the various components. The top right inset represents the results of the fit of 10000 pseudo-experiments fitted with the MC method while keeping the R_{CNO} free but constraining the R_{pp}/R_{pep} ratio to (47.7 ± 0.8) (HZ-SSM [3,26]). Constraining R_{pp}/R_{pep} to the LZ-SSM prediction, 47.5 ± 0.8, gives consistent results. The study included all the background and neutrino species: here we only show those components that mostly influence the sensitivity to CNO neutrinos.

simulated data using the same response functions adopted for fitting the real experimental data, using both analytical and MC procedures. The simulated data of the pseudoexperiments are obtained from a random sampling

of PDFs produced with the full Borexino MC, including solar neutrino interaction rates as predicted by the HZ/LZ–SSM and with the rates of the different background components compatible with the final results presented

in this work. As an example, Fig. 5 shows the distribution of the results of the MC fit of 6700 pseudoexperiments each one with the same exposure as the real data. In this particular example, by construction, the fit model perfectly matches the simulated data. The 1D distributions of the fit results, i.e., the rates R of different solar neutrino and background species, are Gaussian and do not show any significant biases with respect to the rates used as simulation inputs. The widths of these distributions show the expected statistical precision of the measurement of the corresponding component. The shapes of the analogous 2D distributions visualize the correlations among the different components. In particular, we underline that since the energy spectrum of the CNO neutrinos is quite similar to that of the ^{210}Bi internal contamination and the fit procedure cannot separate them, the sensitivity studies for all the pp-cycle neutrino and background components are performed by constraining the CNO rate. These results are depicted in the left portion of Fig. 5 with R_{CNO} generated and constrained assuming, as an example, the HZ–SSM. The same constraint on R_{CNO} is used in fitting the real data, as it will be reported below. Some additional significant correlations are present among some of the

various species, as the figure is showing. This is one of the reasons why the best accuracy in the determination of the interaction rates of solar neutrinos is obtained by fitting the entire energy spectrum, as in the present analysis, thus best using all the available information about details of the entire spectral shapes, instead of choosing partial energy regions.

The top right inset in Fig. 5 demonstrates the sensitivity of the present data set to CNO neutrinos. In this case, no constraint on R_{CNO} is applied, but, to decrease the effect of the degeneracy of the spectral shapes, a constraint on the ratio between R_{pp} and R_{pep}, as expected from the SSM, is applied. It is interesting to note the strong anticorrelation between the ^{210}Bi and CNO components which is originated by the above discussed similarities of their energy spectra.

Finally, Fig. 6 is obtained removing all the constraints on the CNO and pep components and clearly shows that the strong correlations (and anticorrelations) among R_{CNO}, R_{pep}, and the ^{210}Bi decay rate significantly limit the possibility to determine all the three species at the same time.

Similar MC studies have been performed to quantify the systematic uncertainty associated to the fit models, by generating MC data with a response function modified with respect to the one used in the fit (see next section). Finally, pseudo-experiments MC data have been used to obtain the distribution of the likelihood functions and thus evaluate the p-values of our results.

VII. RESULTS

The interaction rates R_{pp}, R_{Be}, R_{pep} are obtained from the fit together with the decay rates of ^{85}Kr, ^{210}Po, ^{210}Bi, ^{11}C internal backgrounds, and the external backgrounds rates (^{208}Tl, ^{214}Bi, and ^{40}K γ rays).

In the MC approach, the MC-based pile-up spectrum [16] is included in the fit with a constraint of $(137.5 \pm 2.8 \text{ cpd}/100 \text{ ton})$ on the ^{14}C $- ^{14}$C contribution based on an independent measurement of the ^{14}C rate [12]. In the analytical approach, pile-up is taken into account with the convolution of each spectral component with the solicited-trigger spectrum [12]. Alternatively, the analytical fit uses a synthetic pile-up spectrum [12] built directly from data. The differences between these methods are quoted in the systematic error (see Table IV).

In order to break the degeneracy between the ^{210}Bi and the CNO ν spectral shapes, we constrain the CNO ν interaction rate to the HZ-SSM predictions, including MSW-LMA oscillations $(4.92 \pm 0.56 \text{ cpd}/100 \text{ ton})$ [3,26] as anticipated in Sec. VI. The analysis is repeated constraining the CNO ν rate to the LZ-SSM predictions $(3.52 \pm 0.37 \text{ cpd}/100 \text{ ton})$ and in case of difference, the two results are quoted separately. The contribution of ^{8}B νs is small and its rate was constrained to the value obtained from the HER analysis [11].
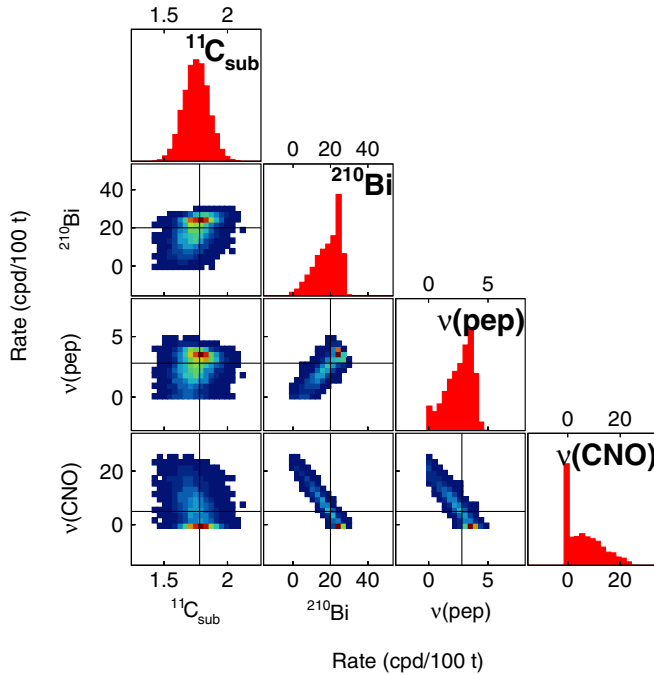

FIG. 6. This figure shows the results of the fit of MC simulated experiments obtained in the same conditions of Fig. 5 but, this time, removing all the constraints on R_{CNO} and R_{pep}. We only show here the correlation between pep, CNO, ^{11}C and ^{210}Bi, but the study included all the spectral components. The significant correlations among these species forbid to measure at the same time R_{CNO} and R_{pep} and to determine the ^{210}Bi decay rate. As described in the text, we have constrained the CNO rate to get the pep one and set a constraint on the ratio R_{pp}/R_{pep} to obtain a limit on the CNO flux.

TABLE II. Borexino Phase-II results on pp, ^{7}Be ($862 + 384$ keV), pep, and CNO solar νs: interaction rates and fluxes inferred assuming the MSW-LMA oscillation parameters [26]. The first error is the statistical derived by profiling the likelihood under Wilks' approximation. The interval extracted is consistent with the expectation from the MC sensitivity study. The second error is the systematic uncertainty. Different contributions to the systematic error are detailed in Table IV. The result on pep νs depends on whether we assume HZ-SSM or LZ-SSM metallicity for CNO νs. The remaining columns show the theoretical interaction rates and fluxes predicted by the standard solar model under the high and low metallicity assumptions [3].

Solar ν	Borexino experimental results		B16(GS98)-HZ		B16(AGSS09)-LZ	
	Rate [cpd/100 ton]	Flux [cm^{-2} s^{-1}]	Rate [cpd/100 ton]	Flux [cm^{-2} s^{-1}]	Rate [cpd/100 ton]	Flux [cm^{-2} s^{-1}]
pp	$134 \pm 10^{+6}_{-10}$	$(6.1 \pm 0.5^{+0.3}_{-0.5}) \times 10^{10}$	131.1 ± 1.4	$5.98(1 \pm 0.006) \times 10^{10}$	132.2 ± 1.4	$6.03(1 \pm 0.005) \times 10^{10}$
^{7}Be	$48.3 \pm 1.1^{+0.4}_{-0.7}$	$(4.99 \pm 0.11^{+0.06}_{-0.08}) \times 10^9$	47.9 ± 2.8	$4.93(1 \pm 0.06) \times 10^9$	43.7 ± 2.5	$4.50(1 \pm 0.06) \times 10^9$
pep (HZ)	$2.43 \pm 0.36^{+0.15}_{-0.22}$	$(1.27 \pm 0.19^{+0.08}_{-0.12}) \times 10^8$	2.74 ± 0.04	$1.44(1 \pm 0.009) \times 10^8$	2.78 ± 0.04	$1.46(1 \pm 0.009) \times 10^8$
pep (LZ)	$2.65 \pm 0.36^{+0.15}_{-0.24}$	$(1.39 \pm 0.19^{+0.08}_{-0.13}) \times 10^8$	2.74 ± 0.04	$1.44(1 \pm 0.009) \times 10^8$	2.78 ± 0.04	$1.46(1 \pm 0.009) \times 10^8$
CNO	$<8.1(95\% \text{C.L.})$	$<7.9 \times 10^8 (95\% \text{C.L.})$	4.92 ± 0.55	$4.88(1 \pm 0.11) \times 10^8$	3.52 ± 0.37	$3.51(1 \pm 0.10) \times 10^8$

The interaction rates of solar neutrinos and the decay rates of background species, obtained by averaging the results of the analytical and MC approaches, are summarized in Tables II and III, respectively.

An example of the multivariate fit (with the MC approach) is shown in Fig. 7 (TFC-subtracted and TFC-tagged energy spectra) and in Fig. 8 (radial distribution and PS-$\mathcal{L}_{\text{PR}}$ pulse-shape distribution). The details of the fit at low energies (between $\sim$230 and 830 keV) can be appreciated in Fig. 9. In this example, obtained with the analytical fit procedure, the pile-up is not present as a separate fit component, since it is taken into account with the convolution method mentioned above.

To recognize the pep ν contribution to the measured electron-recoil spectrum, the TFC-subtracted spectrum, zoomed into the highest energy region (between 800 and 2700 keV), is shown after applying stringent selection cuts on the radial distribution (R < 2.4 m) and on the pulse-shape variable distribution (PS-$\mathcal{L}_{\text{PR}}$ < 4.8) (see Fig. 10): the CNO and pep neutrino interactions are clearly visible between 1250 and 1500 keV, and the spectrum is

TABLE III. Best estimates for the total rates of the background species included in the fit with statistical and systematic uncertainties added in quadrature. These numbers are obtained by averaging the results of the fits with the HZ and LZ hypothesis.

Background	Rate [cpd/100 ton]
^{14}C [Bq/100 t]	40.0 ± 2.0
^{85}Kr	6.8 ± 1.8
^{210}Bi	17.5 ± 1.9
^{11}C	26.8 ± 0.2
^{210}Po	260.0 ± 3.0
Ext. ^{40}K	1.0 ± 0.6
Ext. ^{214}Bi	1.9 ± 0.3
Ext. ^{208}Tl	3.3 ± 0.1

consistent with the Compton-like shoulder expected from the pep line.

An extensive study of the systematic errors has been performed and the results are summarized in Table IV.

Differences between the results of the analytical and the MC fits are quoted as systematic errors. Further systematic uncertainties associated with the fitting procedure were studied by performing the fit in many different configurations by generating simulated data using a family of response functions whose parameters has been varied within calibration accuracy with respect to the nominal response function and by varying the energy estimator, the number and width of the bins, as well as the fit range).

Systematic uncertainties related to the fit models were evaluated using the method described in Sec. VI. Ensembles of pseudoexperiments were generated from a family of PDFs based on the full MC simulations and fitted using both the MC and analytical methods. PDFs including deformations due to possible inaccuracies in the modeling of the detector response (energy scale, uniformity of the energy response, shape of PS-$\mathcal{L}_{\text{PR}}$) and uncertainties in the theoretical energy spectra (^{210}Bi) were considered. The magnitude of the deformation was chosen to be within the range allowed by the available calibration data.

In an additional systematic study, the fit was repeated taking into account the upper limit on the ^{85}Kr decay rate following the procedure described in [18], which exploits the ^{85}Kr $-$ ^{85m}Rb delayed coincidences (^{85}Kr rate < 7.5 cpd/100 ton at 95% C.L.).

The last three lines of Table IV list the uncertainties associated with the determination of the exposure. The one about the fiducial volume is one of the dominant. Its value is the same as quoted in [5] and it is estimated using calibration sources of known positions.

Fully consistent results are obtained when adopting a larger fiducial volume (R < 3.02 m, $|z|$ < 1.67 m). This FV contains more external background (critical for the
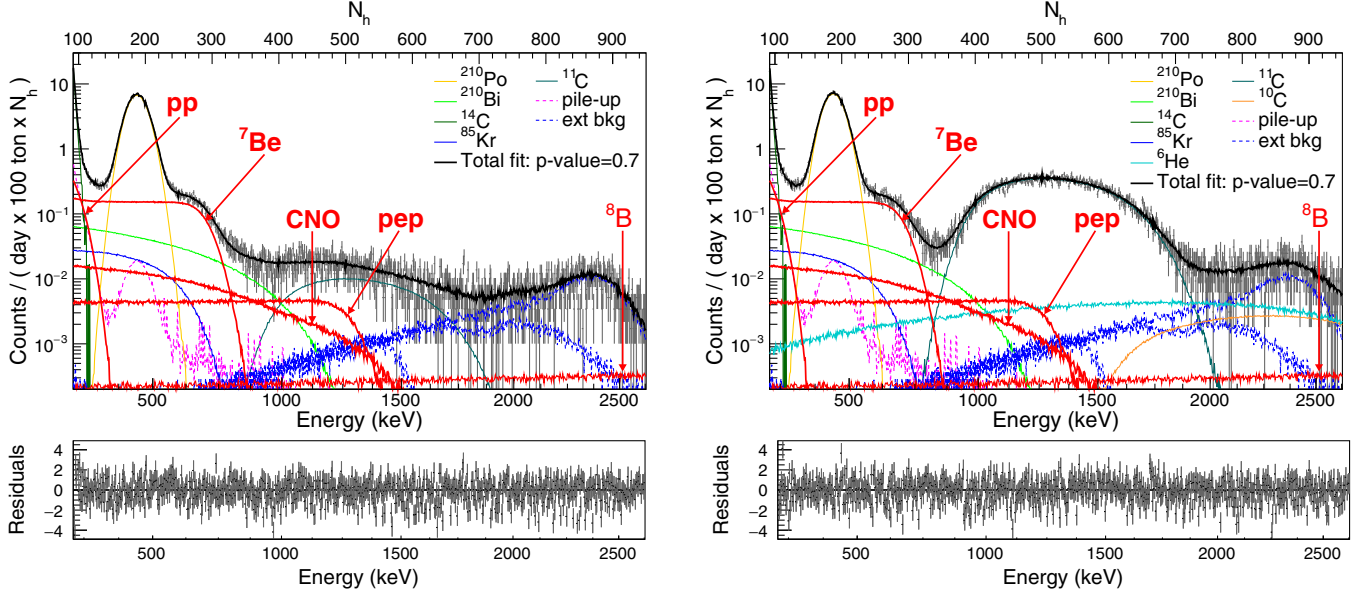
FIG. 7. Multivariate fit results (an example obtained with the MC method) for the TFC-subtracted (left) and the TFC-tagged (right) energy spectra, with residuals. The sum of the individual components from the fit (black lines) is superimposed on the data (grey points). The analysis has been performed using N_h as energy estimator and the transformation to keV-energy scale was performed only for the plotting purposes. The residuals are calculated in every bin as the difference between the data counts and the fit result, divided by the square root of the data counts.
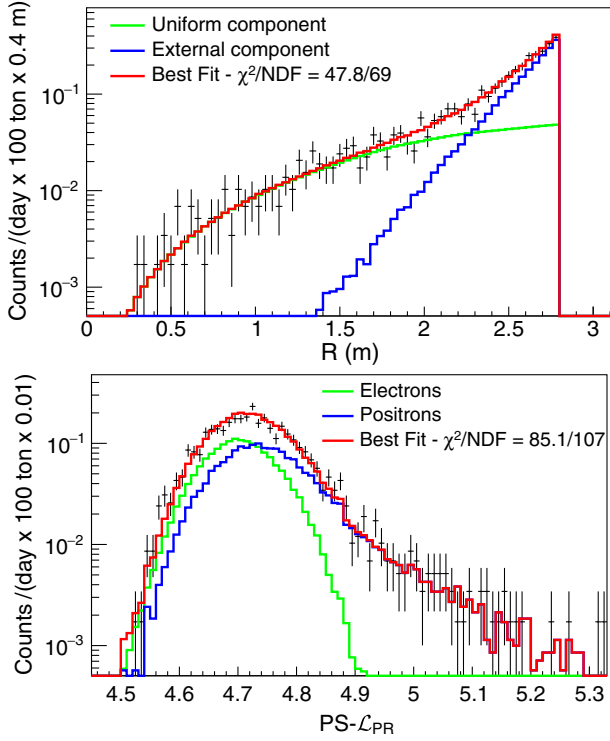
FIG. 8. An example of the multivariate fit showing the radial (top) and the PS-$\mathcal{L}_{PR}$ (bottom) distributions of the events (black crosses) from the TFC-subtracted spectrum, in the energy intervals $N_h > 290$ and $409 < N_h < 645$, respectively.

pep νs) which is, however, properly disentangled by the multivariate fit thanks to its energy shape and radial distribution. The previously published Borexino results regarding pp νs [12] and ^{7}Be νs [5] were obtained in this enlarged fiducial volume.
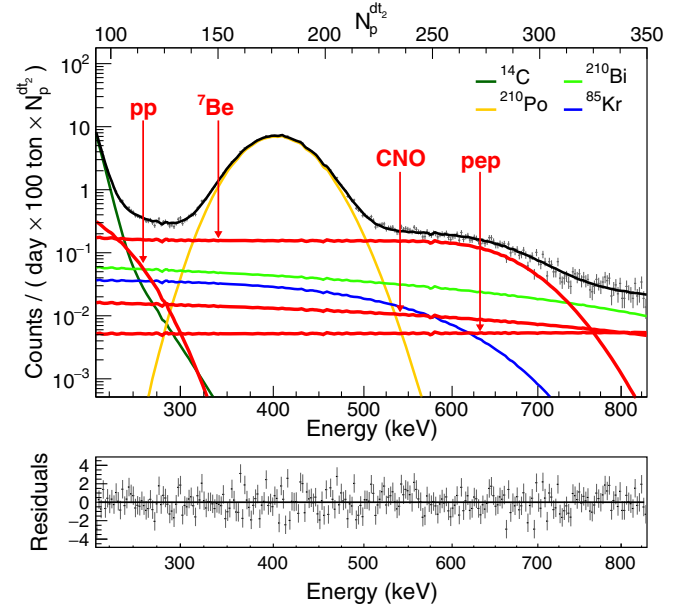

FIG. 9. Results of the fit for TFC-subtracted energy spectrum zoomed in to the lowest energy region (an example obtained with the analytical method) and residuals. The residual are calculated as in Fig. 7.
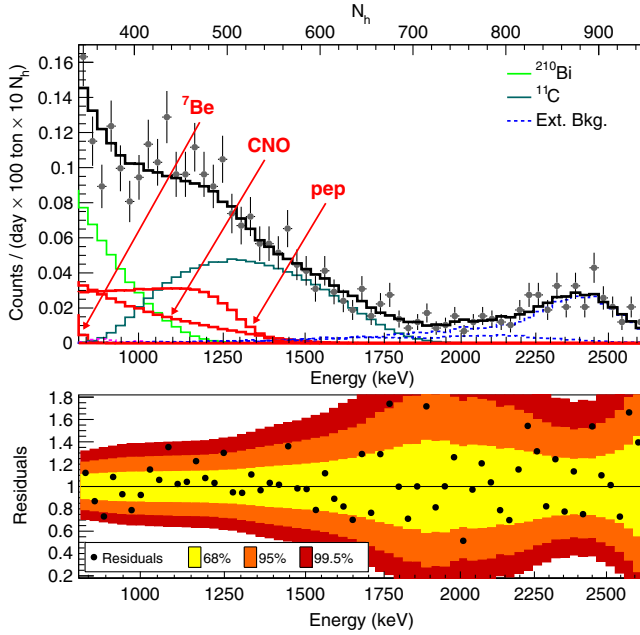
FIG. 10. TFC-subtracted energy spectrum zoomed between 800 keV and 2700 keV after applying stringent selection cuts on the radial distribution ($R < 2.4$ m) and on the pulse-shape variable distribution (PS-$\mathcal{L}_{PR} < 4.8$) to better see features due to pep νs interactions. The residuals (bottom plot) are the ratio between the data and the fit model.

Finally, the analytical fit performed on a restricted energy range (not sensitive to pp neutrinos) using the N_{pe} energy estimator gives consistent results (within 2σ) for R_{Be} and R_{pep}.

The ^{7}Be solar ν flux listed in Table II is the sum of the two monoenergetic lines at 384 and 862 keV. It corresponds to a rate for the 862 keV line of $46.3 \pm 1.1^{+0.4}_{-0.7}$ cpd/100 ton, fully compatible with the Borexino Phase-I measurement [5]. The ^{7}Be solar ν flux is determined with a total uncertainty of 2.7%, which

represents a factor of 1.8 improvement with respect to our previous result [5] and is two times smaller than the theoretical uncertainty.

The present value of R_{pp} is consistent with our previous result and the uncertainty is reduced by about 7.9%.

The correlation between the CNO and pep ν is broken by constraining the R_{CNO} in the fit. The values of R_{Be} and R_{pp} are not affected by the hypothesis on CNO νs within our sensitivity. However, R_{pep} depends on it, being 0.22 cpd/100 ton higher if the LZ hypothesis is assumed (see Table II).

The $\Delta\chi^2$ profile obtained by marginalizing the pep rate is shown in Fig. 11 (left) for both the HZ and LZ assumptions on CNO ν rate. Both curves are symmetric and allow us to establish, for the first time, that the absence of the pep reaction in the Sun is rejected at more than 5σ.

As anticipated, the similarity between the e^- recoil spectrum induced by the CNO neutrinos and the ^{210}Bi spectrum makes it impossible to disentangle the two contributions with the spectral fit without an external constraint on the ^{210}Bi rate. For this reason, we can only provide an upper limit on the CNO neutrinos interaction rate R_{CNO}. In order to do so, we need further to break the correlation between the CNO and pep contributions. In Phase-I, this was achieved by fixing the pep ν rate to the theoretical value [13]. In the current analysis, where pp νs are included in the extended energy range of the fit, we place an indirect constraint on pep νs by exploiting the theoretically well-known pp and pep flux ratio. The interaction rate ratio R_{pp}/R_{pep}, is constrained to (47.7 ± 0.8) (HZ) [3,26]. Constraining R_{pp}/R_{pep} to the LZ hypothesis value 47.5 ± 0.8 gives identical results.

We carried out a sensitivity study by performing the analysis on thousands of data sets simulated with the MC sensitivity tool: this study shows that under the current

TABLE IV. Relevant sources of systematic uncertainty and their contributions to the measured neutrino interaction rates. More details are in the text.

|	pp		^{7}Be		pep	
Source of uncertainty	$-\%$	$+\%$	$-\%$	$+\%$	$-\%$	$+\%$
Fit method (analytical/MC)	-1.2	1.2	-0.2	0.2	-4.0	4.0
Choice of energy estimator	-2.5	2.5	-0.1	0.1	-2.4	2.4
Pile-up modeling	-2.5	0.5	0	0	0	0
Fit range and binning	-3.0	3.0	-0.1	0.1	1.0	1.0
Fit models (see text)	-4.5	0.5	-1.0	0.2	-6.8	2.8
Inclusion of ^{85}Kr constraint	-2.2	2.2	0	0.4	-3.2	0
Live Time	-0.05	0.05	-0.05	0.05	-0.05	0.05
Scintillator density	-0.05	0.05	-0.05	0.05	-0.05	0.05
Fiducial volume	-1.1	0.6	-1.1	0.6	-1.1	0.6
Total systematics (%)	-7.1	4.7	-1.5	0.8	-9.0	5.6
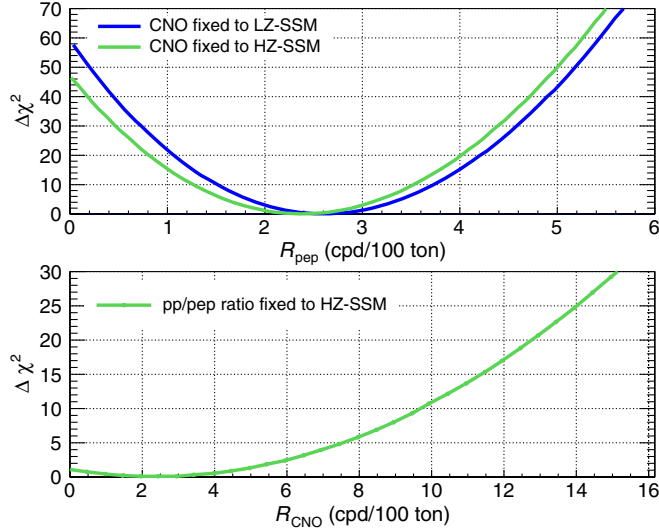
FIG. 11. $\Delta\chi^2$ profile for the *pep* (top plot) and CNO (bottom plot) ν interaction rates.

experimental conditions the total expected uncertainty (statistical plus systematical) is 3.4 cpd/100 ton. With this error, we expect the median 95% C.L. upper limit for R_{CNO} to be $\sim$9 cpd/100 ton and 10 cpd/100 ton, for low and high metallicity, respectively. On data, we obtain the upper limit on $R_{CNO} = 8.1$ cpd/100 ton (95% C.L.) (see Table II), which is slightly stronger than the median limit expected from the MC based sensitivity study. The $\Delta\chi^2$ profile for the CNO rate is shown in Fig. 11 (bottom). This result, using a weaker hypothesis on *pep* ν, confirms the current best limit on the flux of CNO νs previously obtained with Borexino Phase-I data [13].

VIII. CONCLUSIONS

In summary, we have reported the details of the analysis and the results of the first simultaneous measurement of the *pp*, ^{7}Be, and *pep* components of the solar neutrino

spectrum providing a comprehensive investigation of the main *pp* chain in the Sun [10]. These results are in agreement with and improve the precision of our previous measurements. In particular, R_{Be} is measured with an unprecedented precision of 2.7%. The absence of *pep* neutrinos is rejected for the first time at more than 5σ. These data, together with our measurement about ^{8}B ν flux in the HER [11], provide a unique measurement of the interaction rates and thus of the fluxes of the different components of the solar neutrinos from the *pp* chain with a single detector and a unified analysis approach.

The upper limit on R_{CNO} has the same significance as that of Borexino Phase-I and currently is providing the tightest bound on this component.

Several analysis methods and details here reported and discussed have a general interest which is going beyond the understanding of the Borexino results: as example the ^{11}C suppression, the multivariate fit, the analytical model of the energy response, the full MC description of the detector and the fitting procedures can be easily adapted to large volume liquid scintillator based detectors similar to Borexino [27,28].

[1] W. A. Fowler, Astrophys. J. **127,** 551 (1958); A. W. G. Cameron, Annu. Rev. Nucl. Sci. **8,** 299 (1958).

[2] J. Bahcall, http://www.sns.ias.edu/~jnb/SNdata (2005).

[3] N. Vinyoles, A. M. Serenelli, F. L. Villante, S. Basu, J. Bergström, M. C. Gonzalez-Garcia, M. Maltoni, C. Peña-Garay, and N. Song, Astrophys. J. **835,** 202 (2017).

[4] P. Anselmann *et al.* (GALLEX Collaboration), Phys. Lett. B **285,** 376 (1992); L. M. Krauss, Nature (London) **357,** 437 (1992); J. N. Abdurashitov *et al.*, Phys. Lett. B **328,** 234 (1994); B. T. Cleveland, T. Daily, R. Davis, Jr., J. R. Distel, K. Lande, C. K. Lee, P. S. Wildenhain, and J. Ullman, Astrophys. J. **496,** 505 (1998); Y. Fukuda *et al.*, Phys. Rev.

Lett. **77,** 1683 (1996); Y. Abe *et al.* (SuperKamiokande Collaboration), Phys. Rev. D **83,** 052010 (2011); Q. R. Ahmad *et al.* (SNO Collaboration), Phys. Rev. Lett. **87,** 071301 (2001).

[5] G. Bellini *et al.* (Borexino Collaboration), Phys. Rev. Lett. **107,** 141302 (2011).

[6] T. Kajita and A. B. McDonald, The Nobel prize in physics 2015, http://www.nobelprize.org/nobel_prizes/physics/laureates/2015 (2015).

[7] R. Davis, Jr. and M. Koshiba, The Nobel prize in physics 2002, http://www.nobelprize.org/nobel_prizes/physics/laureates/2002 (2002).

[8] L. Wolfenstein, Phys. Rev. D **17,** 2369 (1978); S. P. Mikheev and A. Y. Smirnov, Sov. J. Nucl. Phys. **42,** 913 (1985).

[9] A. Friedland, C. Lunardini, and C. Peña-Garay, Phys. Lett. B **594,** 347 (2004); S. Davidson, C. Peña-Garay, N. Rius, and A. Santamaria, J. High Energy Phys. 03 (2003); P. C. de Holanda and A. Y. Smirnov, Phys. Rev. D **69,** 113002 (2004); A. Palazzo and J. W. F. Valle, Phys. Rev. D **80,** 091301 (2009).

[10] M. Agostini *et al.* (Borexino Collaboration), Nature (London) **562,** 505 (2018).

[11] M. Agostini *et al.* (Borexino Collaboration), arXiv:1709 .00756.

[12] G. Bellini *et al.* (Borexino Collaboration), Nature (London) **512,** 383 (2014).

[13] G. Bellini *et al.* (Borexino Collaboration), Phys. Rev. Lett. **108,** 051302 (2012).

[14] G. Bellini *et al.* (Borexino Collaboration), Phys. Rev. D **82,** 033006 (2010).

[15] H. Back *et al.* (Borexino Collaboration), J. Instrum. **7,** P10018 (2012).

[16] M. Agostini *et al.* (Borexino Collaboration), Astropart. Phys. **97,** 136 (2018).

[17] G. Alimonti *et al.* (Borexino Collaboration), Nucl. Instrum. Methods Phys. Res., Sect. A **600,** 568 (2009).

[18] G. Bellini *et al.* (Borexino Collaboration), Phys. Rev. D **89,** 112007 (2014).

[19] G. Bellini *et al.* (Borexino Collaboration), J. Instrum. **6,** P05005 (2011).

[20] D. Franco, G. Consolati, and D. Trezzi, Phys. Rev. C **83,** 015504 (2011).

[21] X. F. Ding *et al.* (Borexino Collaboration), arXiv:1805 .11125.

[22] S. Agostinelli *et al.*, Nucl. Instrum. Methods Phys. Res., Sect. A **506,** 250 (2003).

[23] M. Agostini *et al.* (Borexino Collaboration), Astropart. Phys. **92,** 21 (2017).

[24] M. Agostini *et al.* (Borexino Collaboration), Phys. Rev. Lett. **115,** 231802 (2015).

[25] O. Yu. Smirnov *et al.* (Borexino Collaboration), Phys. Part. Nucl. **47,** 995 (2016).

[26] I. Esteban, M. C. Gonzalez-Garcia, M. Maltoni, I. Martinez-Soler, and T. Schwetz, J. High Energy Phys. 01 (2017) 87.

[27] E. Ciuffoli, J. Evslin, and X. Zhang, Phys. Rev. D **88,** 033017 (2013).

[28] C. Kraus, Prog. Part. Nucl. Phys. **57,** 150 (2006).