2012/09/25
Head Id: 148989
Archive Id: 148989P
Archive Date: 2012/09/25
Archive Tag: trunk

Search for supersymmetry in events with opposite-sign dileptons and missing energy using an ANN

The CMS Collaboration

Abstract

In this paper, a search for supersymmetry (SUSY) is presented in events with two opposite sign isolated leptons in the final state, accompanied by hadronic jets and missing transverse energy. Advanced multivariate techniques, and in particular Artifical Neural Networks, are employed to discriminate between possible SUSY signals from Standard Model backgrounds. The analysis uses a data sample collected with the CMS detector during the 2011 LHC run and corresponds to an integrated luminosity of 4.98 fb^{-1} at a proton-proton center of mass energy of 7 TeV.

1 Introduction

One of the most natural extensions to the Standard Model (SM) of particle physics is Supersymmetry (SUSY) [1–8]. For every particle in the standard model, SUSY introduces a super-partner, the sparticle with spin difference 1/2. There are good theoretical arguments that require sparticle masses to be less than $\sim$ 1 TeV [7, 8] making the LHC experiments an ideal place for their discovery. Supersymmetry allows for gauge coupling unification at the Planck scale (10^{16} GeV), provides a good dark matter candidate (Lightest Supersymmetric Particle, LSP) [9], provides the needed mathematical framework (generator of space translations) for a theory of Quantum Gravity, and automatically cancels the Higgs mass quadratic divergences.

With the very successful 2011 LHC run, over 4.98 fb^{-1} of pp collisions at 7 TeV center of mass energy have been collected with the CMS experiment. This dataset is used to search for the presence of SUSY in events with two opposite-sign leptons (electrons and/or muons) in the final state utilizing an Artificial Neural Network (ANN). There are many possible ways to produce two leptons in Supersymmetric events. For example, squarks or gluinos can be pair produced in pp collisions, leading either to dileptons in the cascade decay chain of one of the supersymmetric particle or one or more leptons in the cascade decay chains of both sparticles. We assume that an additional R-parity symmetry is conserved [10], leading to a stable lightest sparticle (LSP) and hence to a missing energy signature. Depending on the mass splitting between accessible sparticles and the LSP, the missing energy signature can be large or small. Up to now, the criteria used in CMS to discriminate SUSY signals from the very large Standard Model (SM) backgrounds has tended to require several high transverse momentum jets which is manifest as large values of the scalar sum over all jet transverse momenta, $H_{\mathrm{T}} > 600$ GeV, and large missing energy ($\not\!\!E_{\mathrm{T}} > 200$ GeV).

Compared with previous CMS searches [11, 12], this analyses uses relaxed criteria on missing energy or H_{T}. For SUSY models with high $\not\!\!E_{\mathrm{T}}$ its performance is comparable to the ones using large $\not\!\!E_{\mathrm{T}}$ and H_{T}; the additional variables in an ANN are not required to discriminate between new physics (NP) and the Standard Model (SM) backgrounds. However, in the low $\not\!\!E_{\mathrm{T}}$ or low H_{T} regions of phase space, the discriminating power of the ANN is required to suppress the large SM backgrounds. This search studies events with low $\not\!\!E_{\mathrm{T}}$ or low H_{T}, uncovered by the previous analyses, and provides an independent and complementary probe of this particularly challenging region of phase space. In particular, events are selected as signal-like with $\not\!\!E_{\mathrm{T}}$ as low as 40 GeV or H_{T} as low as 120 GeV.

2 CMS Detector

A detailed description of the CMS detector can be found elsewhere [13]. A right-handed coordinate system is used with the origin at the nominal interaction point (IP). The x-axis points to the center of the LHC ring, the y-axis is vertical and points upward, and the z-axis is parallel to the counterclockwise beam direction. The azimuthal angle ϕ is measured with respect to the x-axis in the xy-plane and the polar angle θ is defined with respect to the z-axis, while the pseudorapidity is defined as $\eta = -ln[\tan(\theta/2)]$. The central feature of the CMS apparatus is a superconducting solenoid, of 6 m internal diameter, that produces a magnetic field of 3.8 T. Within the field volume are the silicon pixel and strip tracker and the barrel and endcap calorimeters ($|\eta| < 3$), composed of a crystal electromagnetic calorimeter (ECAL) and a brass scintillator hadronic calorimeter (HCAL). Calorimetry provides energy and direction measurements of electrons and hadronic jets. The detector is nearly hermetic, allowing for energy balance measurements in the plane transverse to the beam directions. Outside the field volume,

in the forward region ($3 < |\eta| < 5$), there is an iron/quartz-fiber hadronic calorimeter. The steel return yoke outside the solenoid is instrumented with gaseous detectors used to identify muons.

3 Event Samples, Trigger and Event Selection

The CMS experiment collects data using a two-level trigger system, the Level-1 hardware trigger (L1) [14] and a high-level software trigger (HLT) [15]. Data events are selected using a set of dilepton triggers which require the presence of at least two leptons, either two electrons or two muons or an electron-muon pair. In the case of the double muon trigger, the selection is asymmetric with a transverse momentum p_{T} threshold of 13 GeV for the first (higher p_{T}) muon and 8 GeV for the second one. In the case of the double electron trigger, the selection is asymmetric with a threshold applied to the energy of a cluster in the electromagnetic calorimeter, ECAL, projected on the plane transverse to the nominal beam line. The thresholds are fixed to 17 GeV for the more energetic electron and to 8 GeV for the second one. For the electron-muon trigger, the thresholds are 17 GeV for the electron and 8 GeV for the muon. For all triggers, additional identification and isolation criteria are also applied.

Muon candidates are reconstructed [16] by combining the information from the inner tracking system, the calorimeters, and the muon system. Electron candidates are reconstructed [17] by combining the information from the Electromagnetic Calorimeter with the silicon tracker, using shower shape and track-ECAL cluster matching variables in order to increase the sample purity. Jets are reconstructed using the anti-k_{T} clustering algorithm [18] with a size parameter $R = 0.5$. The inputs to the jet clustering algorithm are the four-momentum vectors of reconstructed particles. Each such particle is reconstructed with the particle-flow technique [19] which combines information from several subdetectors. Jets are corrected using factors derived from simulation; to correct for any differences in the energy response between simulation and data, a residual correction factor derived from data is applied to jets in data [20].

In general, $\not\!\!E_{\mathrm{T}}$ is the negative vector sum of the transverse momenta of all final-state particles reconstructed in the CMS detector. The total transverse energy of the event is calculated as the scalar sum of the transverse energies of leptons and jets. The total hadronic transverse energy, (H_{T}), is computed as the scalar sum of the transverse energies of all reconstructed jets in the event satisfying the jet selection criteria.

Simulated pp collision events are produced with the PYTHIA 6.4.22 [21] (using underlying event tune Z2) and MADGRAPH 4.4.24 [22] generators, and processed with a simulation of the CMS detector response based on GEANT4 [23]. Simulated events are reconstructed and analyzed in the same way as data events. Simulated event samples are also used to train the ANN and for extrapolating data-driven background estimates from a background-enriched control region to the expected signal-enriched region. Finally, the simulation is used to perform quality control checks of the data samples collected in the "background" enriched regions, and for estimating systematic uncertainties.

Non-collision backgrounds are removed by applying quality cuts ensuring the presence of a well-reconstructed primary vertex [24]. Events are required to have at least two leptons with $p_T > 20$ GeV and $|\eta| < 2.4$, and at least two jets with $p_T > 30$ GeV and $|\eta| < 2.4$. Jets are required to satisfy the quality criteria described in [25]. Leptons are required to be isolated from significant energy deposits and tracks in a cone of radius $\Delta R = 0.3$ around the direction of the lepton. The relative combined isolation, defined as $I_{rel.}^{comb.} = (\sum_{tracks} P_T + \sum_{ECAL} E_T + \sum_{HCAL} E_T)/P_T$, is required to be < 0.2 for muons and < 0.08 for electrons.

4 Methodology for event selection

Artificial Neural Networks (ANN) are used in order to separate "signal" (SUSY) from "background" (SM) events, accounting for any correlations between the discriminating variables, and thus providing improved results with respect to the use of sequential cuts.

Due to the natural presence of isolated leptons, the main SM background contributions to this analysis involve the production of $t\bar{t}$, and Z+jets. QCD multijet processes with two fake leptons and W+jets events with one fake lepton can also be part of the background, but are significantly reduced by the preselection criteria described below. Finally, two leptons in the final state are produced by WW, WZ, ZZ events but their contributions are found to be negligible compared to the above main ones.

The preselection criteria, which are imposed before ANN training, are the following : events are required to have $\not{E}_{\mathrm{T}} > 30$ GeV, the $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ between the lepton and the closest jet is required to be > 0.2, and the dilepton mass M_{ll} is required to be > 10 GeV. These criteria reject the vast majority of the background, whilst retaining most of the signal as shown in Table 1 for one characteristic CMSSM benchmark point: LM6 ($m_0 = 85$ GeV, $m_{1/2} = 400$ GeV, $\tan\beta = 10$, $A_0 = 0$ GeV). This greatly facilitates the ANN training and optimization by excluding a region heavily dominated by background in which no, or very little signal events are present. The signal region is defined by these preselection criteria with an additional cut on the fraction of lepton to total transverse energy, $\frac{E_T^{lepton}}{\sum E_T} < 0.4$.

Sample	Event Selection	Preselection
$t\bar{t}$	17395±57	8271±41
Z+Jets	507316±1204	4740±61
W+Jets	21094±739	416±35
WW	1204±11	15±1
WZ	1750±8	20±0.9
ZZ	1225±4	13±0.4
QCD	19578±7533	1313±263
Total SM Bkg.	569562±7665	14797±275
LM6	71±1.1	54±0.9

Table 1: Expected number of events for 4.98 fb^{-1} for signal and background after the preselection criteria are applied and for events in the signal region. For the CMSSM benchmark point the NLO cross sections have been used.

The ANN training samples are based on simulated events; the network architecture and an "early stopping" strategy are used to avoid overtraining. A mixture of $t\bar{t}$, Z+jets, W+jets, and QCD simulated samples, weighted according to the expected number of events passing the preselection cuts, are used as the SM background samples to train the ANN.

The ANN training is performed on a class of Simplified Models (SMS) [26], [27]. In the SMS considered here, gluinos are pair-produced, one of them decays as $\tilde{g} \to \chi_2^0 jj \to \chi_1^0 \ell^+\ell^- jj$, and the other decays as $\tilde{g} \to \chi_1^0 jj$. Here χ_2^0 is the second lightest neutralino, χ_1^0 is the lightest neutralino and the LSP, and $\ell = e, \mu, \tau$ with equal probability. This SMS thus always leads to a pair of opposite-sign leptons in the final state, in addition to the jets and $\not{E}_{\mathrm{T}}$. The SMS is fully described by three parameters, the masses of the gluino (m_{GL}), and the LSP (m_{LSP}), along with the neutralino mass in the gluino decay which is set using $M_{\chi_2^0} = 1/2 \cdot (M_{\tilde{g}} + M_{\chi_1^0})$.

For the ANN training, grid points close to the diagonal are used ($|m_{GL}$-$m_{LSP}|$ <400 GeV). These points are choosen since they exhibit low $\not{E}_{\mathrm{T}}$ and H_{T} characteristics : more than 90% of their events have $\not{E}_{\mathrm{T}} < 200$ GeV or $H_{\mathrm{T}} < 600$ GeV, the lowest $\not{E}_{\mathrm{T}}$ and H_{T} values from previously

published CMS results [11, 12]. The ANN performance is tested on a variety of different new physics models and similar sensitivity is obtained.

Several topological and kinematical variables are considered according to their potential to discriminate SM backgrounds versus possible SUSY signals, and the correlations among them. The variables studied are based on the general properties of Supersymmetric processes in many SUSY models, and are not tuned to a specific model.

Using different sets of input variables, taking into account their discriminating power, their correlations, and possible systematic uncertainties, several ANNs were constructed and compared in order to select the optimal configuration. The differences in performance are studied and quantified in terms of the efficiency as a function of background rejection with a figure of merit (FOM) defined as $\frac{S}{\sqrt{B}}$. A network having seven input variables, those with the smallest degree of correlation amongst themselves and with the highest discriminating power, showed the best performance. Table 2 lists the seven input ANN variables along with their description and their relative importance to the ANN after training.

Variable	Description	ANN Importance (%)
$\not{E}_{\mathrm{T}}$	Missing transverse energy	22
M_{ll}	Dilepton mass	20
$\frac{E_T^{lepton}}{\sum E_T}$	Ratio of energy of the dilepton system to the total transverse energy	18
N_{jets}	Number of jets	13
Jet2 p_T	Secondary jet p_T	12
Invariant Mass M_T	Invariant event mass	8
Jet1 p_T	Primary jet p_T	7

Table 2: Seven event, lepton and jet related variables used for the ANN construction.

The relative ANN importance is defined as the sum of the weights-squared of the connections between the variable's neuron in the input layer and the first hidden layer.

5 ANN Output for SM Background

In order to quantify the level of agreement and the significance of a possible excess between data and expectation, it is important to robustly estimate the ANN output distribution in the signal region under the SM-only hypothesis along with its systematic uncertainty.

The approach used to estimate the ANN prediction for the SM-only hypothesis from data is the following. A signal region (SR) is defined by the set of the preselection cuts and the additional criterion on the fraction of transverse energy carried by the dilepton system as described in section 3. A primary control region (CR) is defined by inverting two of the preselection cuts, the total missing transverse energy and the selection cut on the fraction of transverse energy carried by the dilepton system. This region is chosen so that it is dominated by SM processes. The contamination of signal in the primary control region is small and less than 0.03% for LM6, and less than 0.4% for the following three SMS points close to the diagonal ($m_{GL} = m_{LSP}$): $m_{GL} = 500$ GeV and $m_{LSP} = 325$ GeV, $m_{GL} = 600$ GeV and $m_{LSP} = 425$ GeV, $m_{GL} = 700$ GeV and $m_{LSP} = 500$ GeV). The ANN output distribution in the primary control region is then obtained using data, $ANN(SM)_{CR}^{Data}$.

Next, an extrapolation ratio, $R_{Ext.} = \frac{ANN(SM)_{SR}^{MC}}{ANN(SM)_{CR}^{MC}}$ obtained from simulated events, is defined for each bin in the ANN output distribution as the ANN output for the SM-only hypothesis in

the signal region (SR) divided by the ANN output for the SM-only hypothesis in the control region. The extrapolation factor, $R_{Ext.}$, is shown in Figure 1; it is a smooth, monotonic function, and varies between 10^{-2} at low ANN values to 10 at high ANN values.

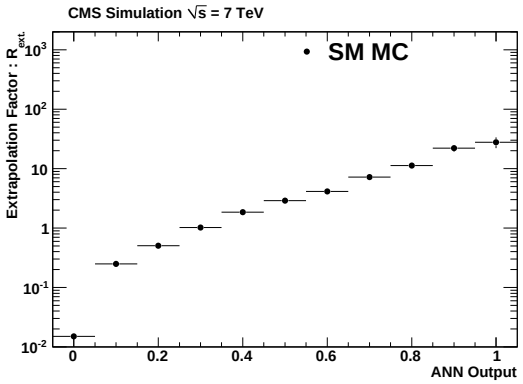

Figure 1: Extrapolation factor $R_{Ext.}$, as obtained from simulated events.

Finally, the ANN output from data in the control region, where only SM physics is assumed to be present, is multiplied by the extrapolation factor, $R_{Ext.}$, in order to predict the ANN output SM in the signal region, $ANN(SM)_{SR}^{Prediction}$.

$$ANN(SM)_{SR}^{Prediction} = ANN(SM)_{CR}^{Data} \times \frac{ANN(SM)_{SR}^{MC}}{ANN(SM)_{CR}^{MC}} \quad (1)$$

The primary control region is further subdivided into a $t\bar{t}$ enriched one with $\not\!\!E_{\mathrm{T}} > 30$ GeV and $M_{ll} > 105$ GeV OR $M_{ll} < 75$ GeV, denoted as "Control Region A", and separately into a Z+jets enriched one with $\not\!\!E_{\mathrm{T}} < 30$ GeV or 75 GeV $< M_{ll} <$ 105 GeV, denoted as "Control Region B". These are not used in the analysis, however they provide quality control tests (level of agreement between data and simulation) for the main two backgrounds that affect this analysis.

6 Systematic Uncertainties

Systematic uncertainties of the ANN output prediction for the SM-only hypothesis, obtained as described in section 5, are estimated with simulated data using the following procedure. A systematic effect is introduced into the simulated data for all events in the sample before any preselection is applied. The nominal SM extrapolation factor $R_{Ext.}$ is then used to obtain a new ANN output prediction for the signal region corresponding to the systematic effect under study. Next, the ANN output prediction, corresponding to the systematic alteration, is compared against the ANN output for the original sample, without any systematic effects introduced. The relative difference in ANN outputs for each bin, is assigned as a bin-by-bin systematic uncertainty. Finally, for each bin, the relative differences for all systematic effects studied are added in quadrature. This results in a bin-by-bin total systematic uncertainty in the ANN output prediction. The overall systematic uncertainty above some ANN value is calculated as the quadratic sum of the integrated uncertainties from the ANN cut value to infinity. With an ANN output of greater than 0.95, the overall systematic uncertainties, corresponding to the seven input variables used for the ANN construction, as well as the uncertainties in the cross sections of the SM backgrounds, are shown in Table 3.

The magnitude of the systematic alterations for the jet energy scale is taken from dedicated CMS measurements [28]. While the clustered energy scale of $\not\!\!E_{\mathrm{T}}$ is known to the 3% level

Missing Transverse Energy $\not\!\!E_{\mathrm{T}}$	$\pm$10%	26
Number of Jets N_{jets}	$\pm$0.5%	$<$1
$\frac{E_T^{lepton}}{\sum E_T}$	$\pm$2%	9
Primary and Secondary jet p_T	$\pm$3%	10
$t\bar{t}$ cross section	$\pm$10%	1
Dilepton Mass M_{ll}	$\pm$1%	1
Invariant Mass M_T	$\pm$5%	6
QCD and W+jets cross sections	$\pm$20%	$<$1
Total		30

Table 3: Systematic uncertainties considered in the predicted background, along with their magnitude, and the impact they have on the final ANN output prediction when the signal selection cut at 0.95 is applied.

in CMS and the unclustered energy scale for $\not\!\!E_{\mathrm{T}}$ is known to within 10% [29], this analysis uses a conservative 10% for the overall $\not\!\!E_{\mathrm{T}}$ systematic uncertainty. The individual component uncertainties on jets and leptons are used to estimate the uncertainty for the M_T scale. For the input ANN variables for which there is no dedicated CMS measurement, the level of agreement between data and simulation in the control region is used to obtain an estimate of the systematic uncertainty. Therefore, the control region is used to constrain the systematic uncertainties in these cases. Given the above, the difference between data and simulation for the migration of events from the one-jet to the two-jet bin is estimated to be 0.5%. Similarly, the systematic uncertainty on the ratio of the lepton to the total transverse energy is estimated to be 2%. The dilepton mass scale uncertainty of 1% is taken from the CMS measurements of the Z peak [30].

Finally, the relative fraction of $t\bar{t}$ and Z+jets backgrounds are observed to vary both as a function of the ANN cut, as well as across the signal and control regions. In order to account for any remaining differences, not already accounted for in the extrapolation factor $R_{Ext.}$, the cross sections of all background components are left to vary within their uncertainties, taken from the recent CMS measurements for the $t\bar{t}$ [31], and QCD [32] events. The Z+jet cross section uncertainty ($<$ 3%) [30], and the W+jet cross section uncertainty ($<$ 3%) [30] produce a negligible systematic uncertainty effect on the ANN output, and hence are neglected.

7 Performance of the ANN

The ANN output after the training is shown in Figure 2 for the signal (blue) and SM background (red) populations; the efficiency and purity of the selected samples is also shown as a function of the ANN output cut.

When statistical and systematic uncertainties are taken into account, the ANN output cut yielding the best exclusion limit in the SMS plane is ANN$>$ 0.95. The expected number of SM and signal events for CMSSM benchmark point LM6 after imposing the ANN output cut are shown in Table 4. The remaining backgrounds are dominated by $t\bar{t}$ to dileptons, and Z+jets at a much smaller level.

Figure 3 compares the ANN output distributions of data and simulated events in the control regions as defined in section 5. Agreement between data and simulation is observed both in the primary control region used to define the ANN output, as well as in the $t\bar{t}$ and Z+jets dominated control regions "A" and "B".

Similar agreement between data and simulation for the ANN input variables in the control region is observed as well. This helps validate that the simulation is sufficiently good to train

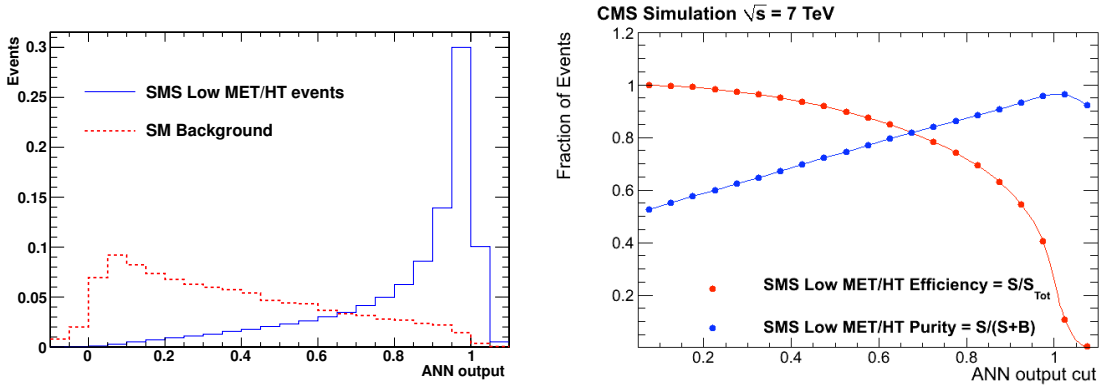

Figure 2: Left : ANN output for the SM background (red line) and SMS low $\not\!\!E_T$/low H_T events (blue line) Right : Efficiency (red) and purity (blue) vs ANN output cut for SMS low $\not\!\!E_T$/low H_T events in the signal region.

Sample	ANN $>$ 0.95
$t\bar{t}$	125.1 ± 37.8
Z+Jets	13.9 ± 4.2
W+Jets	$< 2.8 \pm 2.8$
WW	0.60 ± 0.2
WZ	0.20 ± 0.1
ZZ	0.25 ± 0.05
QCD	0.005 ± 0.002
Total SM Bkg.	140.0 ± 41.7
LM6	40.3 ± 0.7

Table 4: Expected number of events for 4.98 fb^{-1} for signal and SM backgrounds and for the ANN output greater than $>$ 0.95 which yields maximum Figure of Merit. The NLO cross section is used for CMSSM benchmark point LM6.

the ANN, and adequate to be used for the estimation of systematic uncertainties. It is important to note that this analysis does not use the SM-only ANN output from simulation to search for a possible signal in the data. Rather, as previously described, it uses the ANN output distribution obtained from data in the control region, and extrapolates that distribution, using the bin-by-bin factors $R_{Ext.}$ from the simulation, into the signal region to obtain an ANN output prediction for the SM-only hypothesis.

Figure 4 shows the $\not\!\!E_T$ and H_T distributions for data and simulated events in the final signal-like (ANN $>$ 0.95) region. This analysis is sensitive to signal-like events with $\not\!\!E_T$ as low as 40 GeV or H_T as low as 120 GeV, regions not explored by other CMS analyses.

8 Results

The seven input ANN variables are shown in Figures 5-6 for simulated and data events, after all preselection cuts are applied and for events in the signal region. Data and simulation are consistent with each other, given the statistical and systematic uncertainties.

Figure 7 shows the comparison between the SM ANN prediction and the data in in the signal region including statistical and systematic uncertainties. With the optimal ANN cut at 0.95, 140 ± 42 events are expected from SM processes and 171 are observed, showing agreement at the 68% confidence level. The error on the expectation includes both statistical and systematic uncertainties.

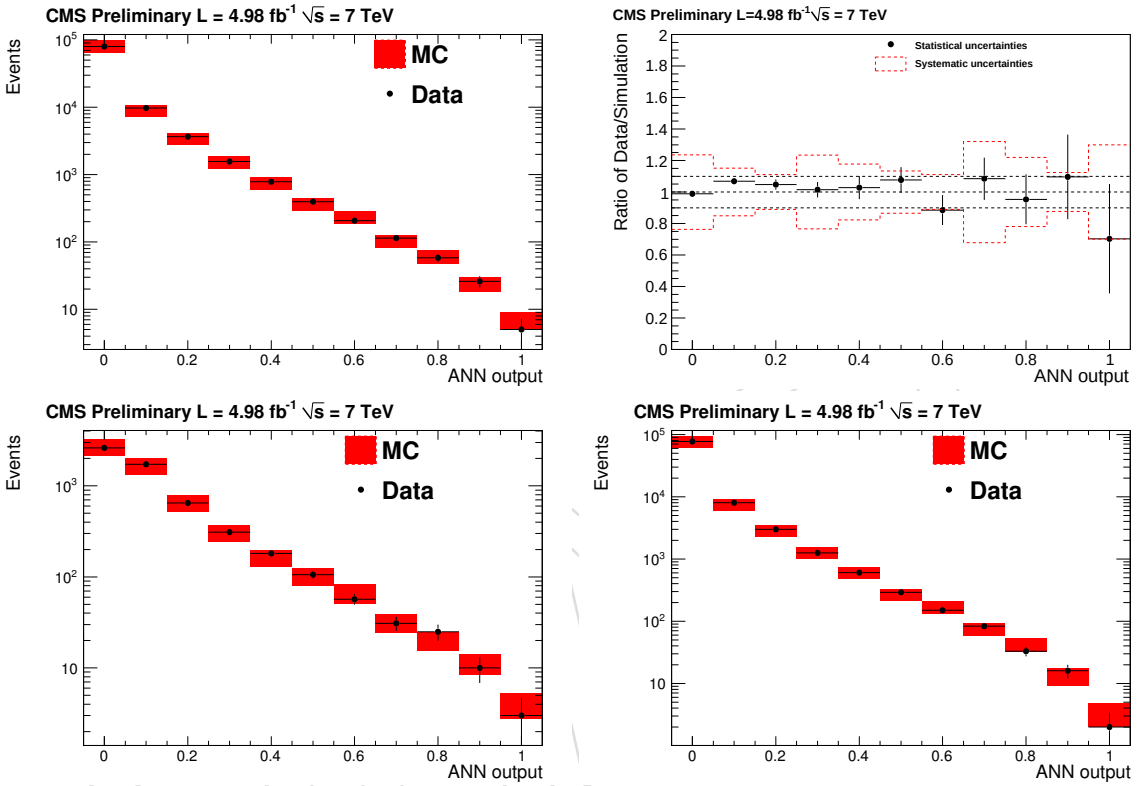

Figure 3: Data (black points) vs simulated events (red line) comparisons of the ANN output distributions in the various control regions. Top : The ANN output in the control region used in order to perform the extrapolation with 20% systematic uncertainties included (left), and the ratio between data and simulated events (right) with both statistical (black error bars) and systematic (red bands) uncertainties shown. Bottom : Two additional control regions used to perform sanity checks : $t\bar{t}$ enriched control region "A" (left), Z+jet enriched control region "B" (right) with 20% systematic uncertainties included. The 20% systematic uncertainty is the smallest bin-by-bin systematic error used for illustration purposes.

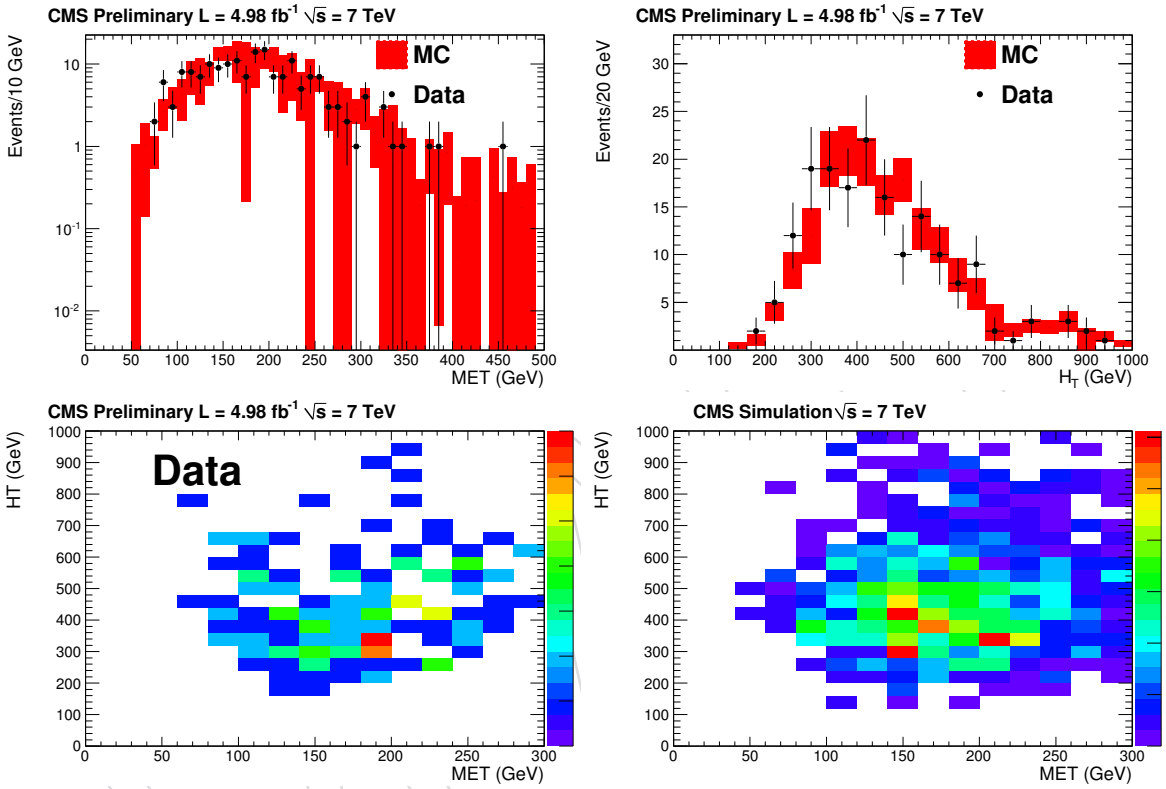

Figure 4: $\not\!\!E_T$ (top left) and H_T (top right) distributions for signal-like events (ANN > 0.95) in data (black points) and simulation (red band) including both statistical and systematic uncertainties. $\not\!\!E_T$ vs H_T distributions for signal-like events (ANN > 0.95) in data (bottom left) and simulation (bottom right).

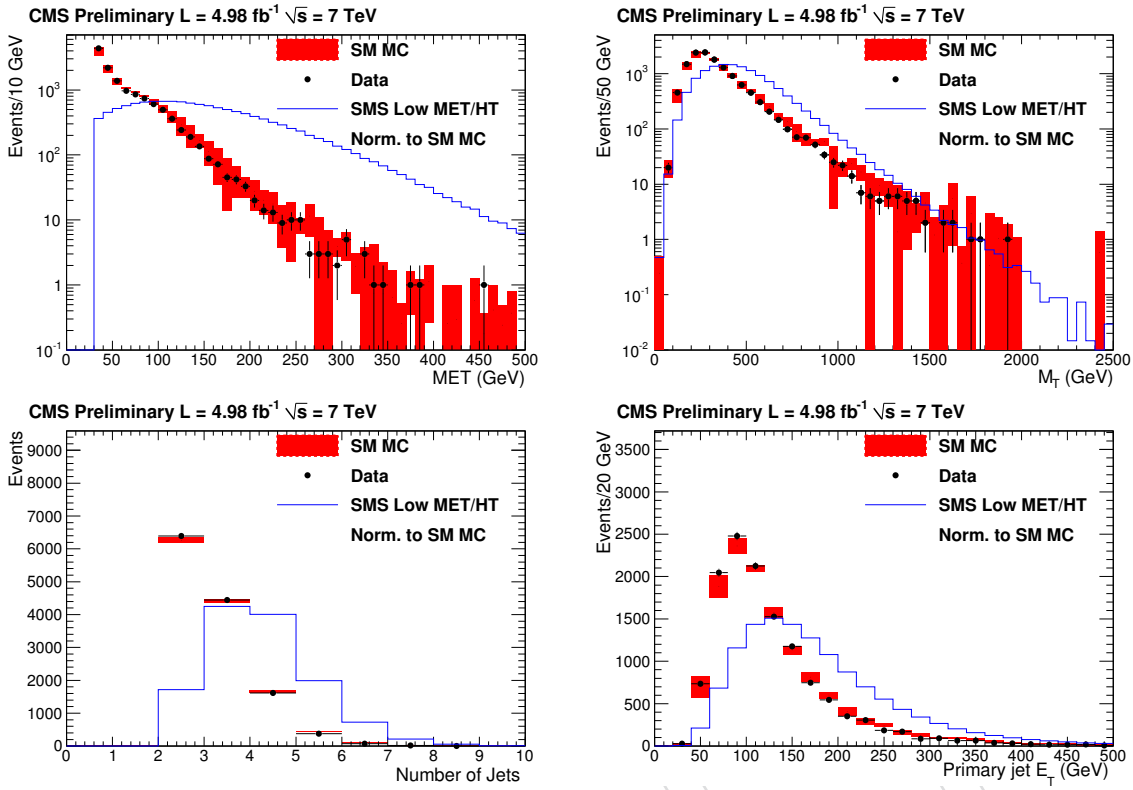

Figure 5: The seven input ANN variables for simulated SM background events (red) and SMS low $\not\!\!E_{\mathrm{T}}$/ low H_{T} events (blue), normalized to the same number of events. The data are also shown for comparison (black points). Statistical and systematic errors on the data and simulated events are shown.

Finally, the observed and expected number of events are translated into limits on the SMS models, which is depicted in Fig. 8, and on the $(m_0, m_{1/2})$ plane in the Constrained Minimal Supersymmetric extension of the SM (CMSSM), as shown in Fig. 9. The systematic uncertainties associated with the signal selection efficiency, along with their magnitude, are summarized in Table 5. The uncertainty on the lepton triggers and the lepton isolation are the same as the

Systematic	Magnitude
Lepton Triggers ($p_T > 20$ GeV)	3%
Lepton Isolation	5%
Luminosity	2.2%
ANN selection	17%
Total	18%

Table 5: Signal selection efficiency systematic uncertainties

ones estimated in [33]. The relative ANN uncertainty for the signal is lower than the corresponding uncertainty for the background, due mainly to the different ANN shapes for these two populations (signal and background).

The 95% C.L. upper limits are computed using a frequentist CL_s method with profile likelihood test statistics, and lognormal distributions for the background expectation [34], [35]. The uncertainties in the NLO+NLL cross sections from the parton distribution functions [36], the choice of the factorization and renormalization scale, and α_S are taken into account for each point, and are evaluated according to the PDF4LHC recommendation [37]. A uniform signal acceptance systematic is assumed for each point. As described previously, the contamination

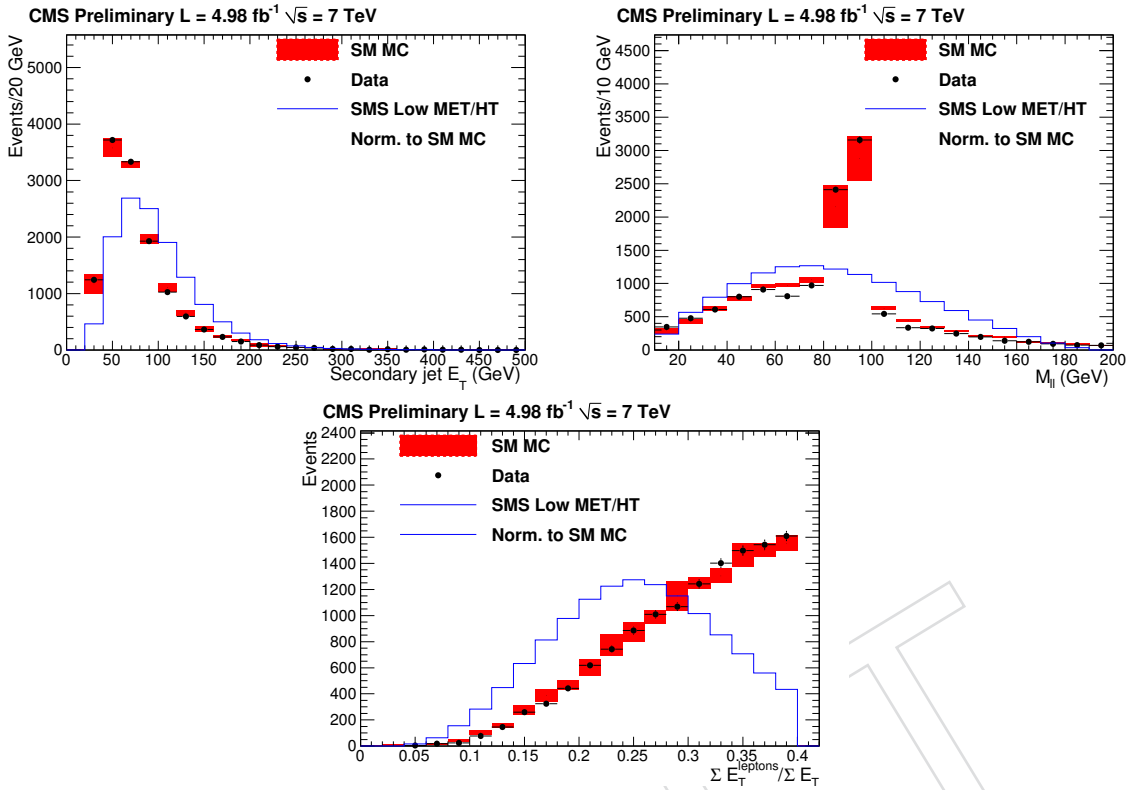

Figure 6: The seven input ANN variables for the 7 TeV simulated events and for SM backgrounds (red) and SMS low $\not\!\!E_T$/ low H_T events (blue), normalized to the same number of events. The 7 TeV data are also shown for comparison (black points). Statistical and systematic errors on the data and simulated events are shown.

Selection	Expected	Observed	95% C.L. Upper Limit on number of events
ANN > 0.95	140 ± 42	171	95

Table 6: Number of predicted and observed events for 4.98 fb^{-1} for ANN > 0.95.

of the signal in the control region is negligible and hence not taken into account in the limit extraction.

In case of the CMSSM limits and for this choice of parameter values, squark masses below ~ 700 GeV are excluded at 95% C.L., as are gluino masses below the same value for the region $m_0 <$ 600 GeV. In the region $1000 < m_0 <$3000 GeV, gluino masses below 500 GeV are excluded, while the squark mass in the excluded models varies in the range 1000 GeV to 2500 GeV, depending on the value m_0. For the SMS limits, for gluino masses below $\sim$ 800 GeV, LSP masses below $\sim$ 400 GeV are excluded. For gluino masses greater than $\sim$ 800 GeV no limits on the mass of LSP can be set.

9 Conclusions

A search for supersymmetry in events with two opposite sign leptons in the final state and with the use of Artificial Neural Networks has been presented, using the 2011 dataset collected with the CMS experiment, corresponding to an integrated luminosity of 4.98 fb^{-1} at a proton-proton center of mass energy of 7 TeV. This search is complementary to the ones already published [11, 12] and yields comparable exclusion limits for high $\not\!\!E_T$, high H_T SUSY models. In addition,

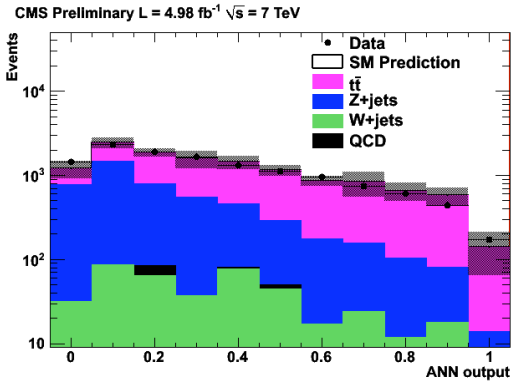

Figure 7: ANN output for the data (black points) and the SM data-driven prediction (gray dashed area) in the signal region. The error on the SM prediction (gray dashed area) includes both statistical and systematic uncertainties.

the significantly relaxed criteria on $\not{E}_{\mathrm{T}}$ ($\not{E}_{\mathrm{T}} > 40$ GeV) or H_{T} ($H_{\mathrm{T}} > 120$ GeV) with respect the previously published analyses [11, 12] ($\not{E}_{\mathrm{T}} > 200$ and $H_{\mathrm{T}} > 600$ GeV) allows for the study of events uncovered by previous analyses, and provides an independent and complementary probe of this particularly challenging region of phase space. Agreement is observed between the expectation from the SM and the data, with no significant excess. This results in setting limits in the CMSSM ($m_0, m_{1/2}$) and SMS (m_{GL}, m_{LSP}) planes. In case of the CMSSM limits and for a specific choice of parameter values, squark masses below $\sim$ 700 GeV are excluded at 95%CL, as are gluino masses below the same value for the region $m_0 < 1000$ GeV. In the region $1000 < m_0 <$3000 GeV, gluino masses below 500 GeV are excluded, while the squark mass in the excluded models varies in the range 1000 GeV to 2500 GeV, depending on the value m_0. In case of the the SMS limits, for gluino masses below $\sim$ 800 GeV, LSP masses below $\sim$ 400 GeV are excluded. For gluino masses greater than $\sim$ 800 GeV, no limits on the mass of LSP can be set.

10 Aknowledgements

We congratulate our colleagues in the CERN accelerator departments for the excellent performance of the LHC machine during 2011. We thank the technical and administrative staff at CERN and other CMS institutes, and acknowledge support from: FMSR (Austria); FNRS and FWO (Belgium); CNPq, CAPES, FAPERJ, and FAPESP (Brazil); MES (Bulgaria); CERN; CAS, MoST, and NSFC (China); COLCIENCIAS (Colombia); MSES (Croatia); RPF (Cyprus); Academy of Sciences and NICPB (Estonia); Academy of Finland, ME, and HIP (Finland); CEA and CNRS/IN2P3 (France); BMBF, DFG, and HGF (Germany); GSRT (Greece); OTKA and NKTH (Hungary); DAE and DST (India); IPM (Iran); SFI (Ireland); INFN (Italy); NRF andWCU (Korea); LAS (Lithuania); CINVESTAV, CONACYT, SEP, and UASLP-FAI (Mexico); PAEC (Pakistan); SCSR (Poland); FCT (Portugal); JINR (Armenia, Belarus, Georgia, Ukraine, Uzbekistan); MST and MAE (Russia); MSTD (Serbia); MICINN and CPAN (Spain); Swiss Funding Agencies (Switzerland); NSC (Taipei); TUBITAK and TAEK (Turkey); STFC (United Kingdom); DOE and NSF (USA). Individuals have received support from the Marie- Curie programme and the European Research Council (European Union); the Leventis Foun dation; the A. P. Sloan Foundation; the Alexander von Humboldt Foundation; the Belgian Federal Science Policy Office; the Fonds pour la Formation a‘ la Recherche dans lIndustrie et dans lAgriculture (FRIA-Belgium); the Agentschap voor Innovatie doorWetenschap en Technologie (IWT-Belgium); the Council of

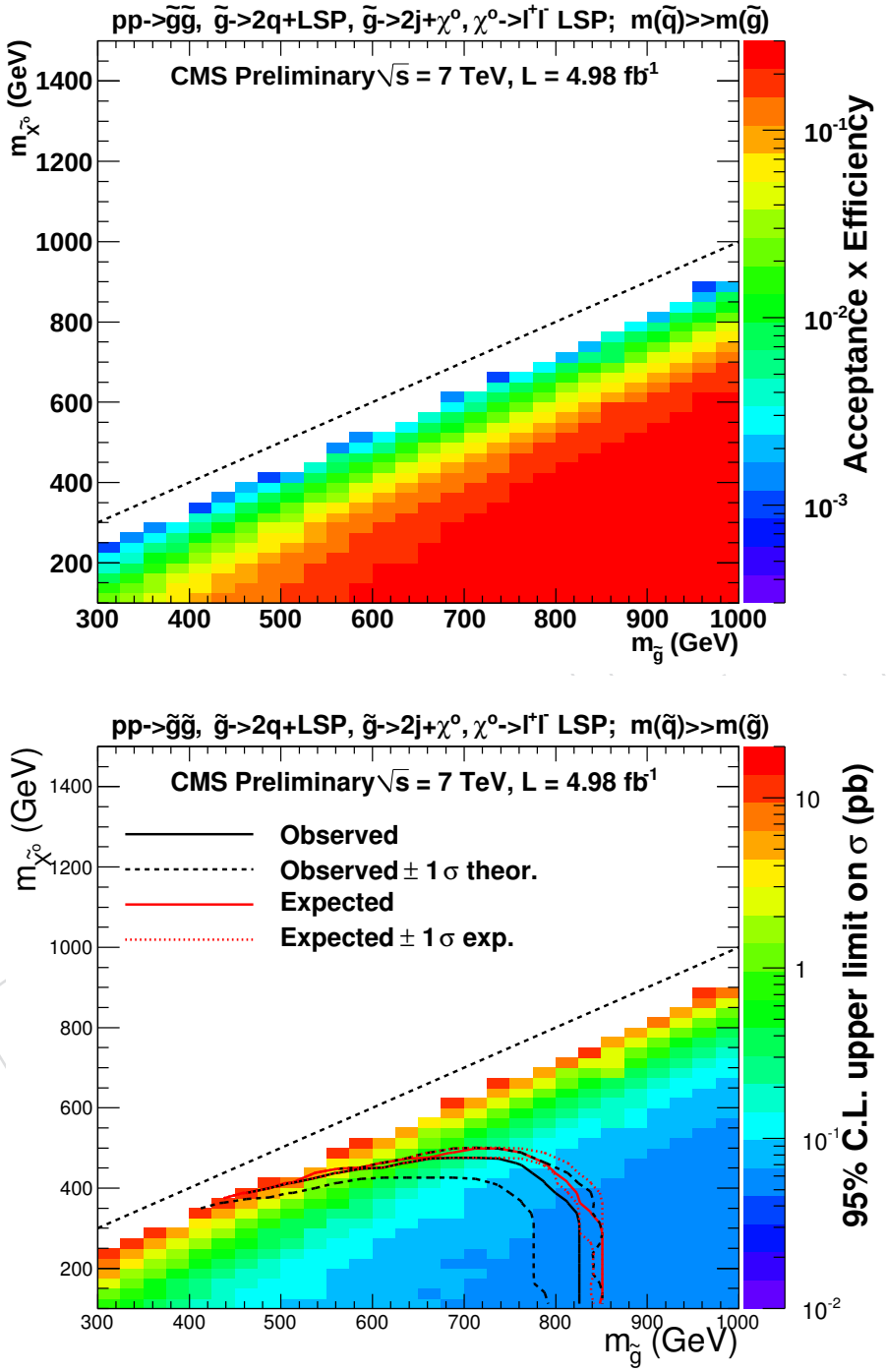

Figure 8: SMS 95% C.L. exclusion limits on Simplified Models with the ANN analysis. The acceptance$\times$ efficiency (top) and 95% C.L. upper cross section limit (bottom) are shown for different gluino and neutralino masses. The region just below the diagonal is not considered due to inadequate initial state radiation modelling.

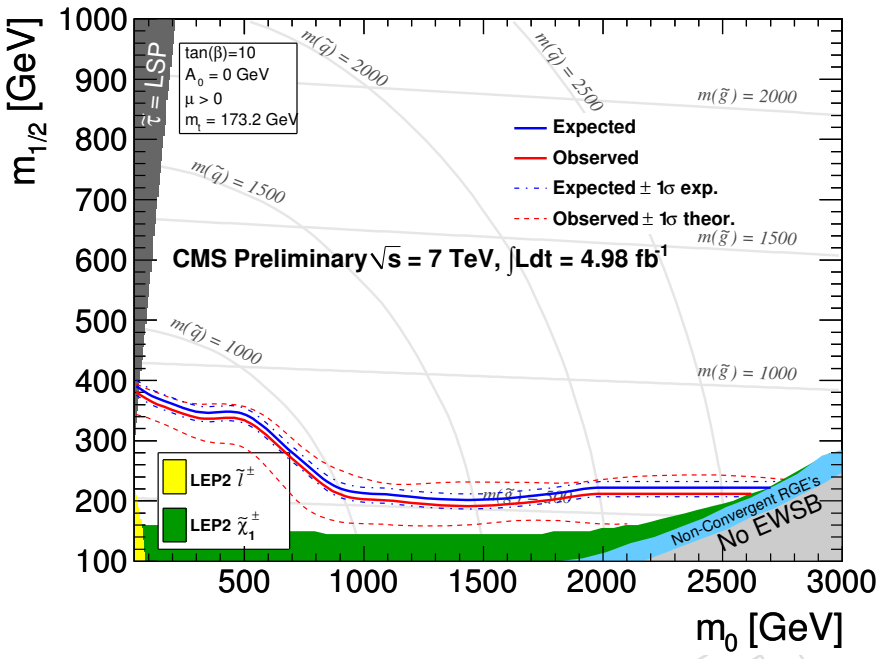

Figure 9: Expected (blue) and observed (red) 95% C.L. exclusion limit for the ANN analysis (ANN$>$ 0.95)

296 Science and Industrial Research, India; and the HOMING PLUS programme of Foundation for
297 Polish Science, cofinanced from European Union, Regional Development Fund.

References

[1] Y. A. Golfand and E. P. Likhtman, "Extension of the Algebra of Poincaré Group Generators and Violation of p Invariance", *JETP Lett.* **13** (1971) 323.

[2] J. Wess and B. Zumino, "Supergauge transformations in four dimensions", *Nucl. Phys. B* **70** (1974) 39, doi:10.1016/0550-3213(74)90355-1.

[3] H. P. Nilles, "Supersymmetry, Supergravity and Particle Physics", *Phys. Reports* **110** (1984) 1, doi:10.1016/0370-1573(84)90008-5.

[4] H. Haber and G. Kane, "The Search for Supersymmetry: Probing Physics Beyond the Standard Model", *Phys. Reports* **117** (1987) 75, doi:10.1016/0370-1573(85)90051-1.

[5] R. Barbieri, S. Ferrara, and C. A. Savoy, "Gauge Models with Spontaneously Broken Local Supersymmetry", *Phys. Lett. B* **119** (1982) 343, doi:10.1016/0370-2693(82)90685-2.

[6] S. Dawson, E. Eichten, and C. Quigg, "Search for Supersymmetric Particles in Hadron - Hadron Collisions", *Phys. Rev. D* **31** (1985) 1581, doi:10.1103/PhysRevD.31.1581.

[7] E. Witten, "Dynamical Breaking of Supersymmetry", *Nucl. Phys. B* **188** (1981) 513, doi:10.1016/0550-3213(81)90006-7.

[8] S. Dimopoulos and H. Georgi, "Softly Broken Supersymmetry and SU(5)", *Nucl. Phys. B* **193** (1981) 150, doi:10.1016/0550-3213(81)90522-8.

[9] G. Jungman and M. Kamionkowski, "Supersymmetric dark Matter", *Phys. Report* **267** (1996) 195, doi:10.1016/0370-1573(95)00058-5.

[10] G. R. Farrar and P. Fayet, "Phenomenology of the Production, Decay, and Detection of New Hadronic States Associated with Supersymmetry", *Phys. Lett. B* **76** (1978) 575, doi:10.1016/0370-2693(78)90858-4.

[11] CMS Collaboration, "Search for new physics in events with opposite-sign leptons, jets, and missing transverse energy in pp collisions at $\sqrt{s}$=7 TeV", *CMS Physics Analysis Summary* **CMS-PAS-SUS-11-011** (2012).

[12] CMS Collaboration, "Search for physics beyond the standard model in events with a Z boson, at $\sqrt{s}$=7 TeV", *CMS Physics Analysis Summary* **CMS-PAS-SUS-11-021** (2012).

[13] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **03** (2008) S08003, doi:doi:10.1088/1748-0221/3/08/S08004.

[14] CMS Collaboration, "CMS TRIDAS Project Technical Design Report, Volume 1 The Trigger Systems", *CERN/LHCC 2000-38 CMS TDR 6.1.* (2000).

[15] CMS Collaboration, "The CMS High Level Trigger", *Eur. Phys. J.* **C46** (2006) 605.

[16] CMS Collaboration, "Performance of muon identification in pp collisions at $\sqrt{s}$=7 TeV", *CMS Physics Analysis Summary* **CMS-PAS-MUO-10-002** (2010).

[17] CMS Collaboration, "Electron reconstruction and identification at $\sqrt{s}$=7 TeV", *CMS Physics Analysis Summary* **CMS-PAS-EGM-10-004** (2010).

[18] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-kt jet clustering algorithm", *JHEP* **04** (2008) 063, doi:doi:10.1088/1126-6708/2008/04/063.

[19] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", *CMS Physics Analysis Summary* **CMS-PAS-PFT-09-001** (2009).

[20] CMS Collaboration, "Jet Energy Corrections determination at 7 TeV", *CMS Physics Analysis Summary* **CMS-PAS-JME-10-010** (2010).

[21] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual", *JHEP* **05** (2006) 026, doi:doi:10.1088/1126-6708/2006/05/026.

[22] J. Alwall et al., "MadGraph/MadEvent v4: The NewWeb Generation", *JHEP* **09** (2007) 028, doi:doi:10.1088/1126-6708/2007/09/028.

[23] S. Agostinelli et al., "GEANT4: A simulation toolkit", *Nucl. Instrum. Meth.* **A506** (2003) 250, doi:doi:10.1016/S0168-9002(03)01368-8.

[24] CMS Collaboration, "Tracking and Primary Vertex Results in First 7 TeV Collisions", *CMS Physics Analysis Summary* **CMS-PAS-TRK-10-005** (2010).

[25] CMS Collaboration, "Calorimeter Jet Quality Criteria for the First CMS Collision Data", *CMS Physics Analysis Summary* **CMS-PAS-JME-09-008** (2010).

[26] B. Knuteson and S. Mrenna, "BARD: Interpreting new frontier energy collider physics", arXiv:0602101.

[27] N.Arkani-Hamed et al., "MARMOSET: The Path from LHC Data to the New Standard Model via On-Shell Effective Theories", arXiv:0703088.

[28] CMS Collaboration, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS", *J. Instrum.* **6** (2011) P11002, doi:10.1088/1748-0221/6/11/P11002. arXiv:1107.4277.

[29] CMS Collaboration, "Missing transverse energy performance of the CMS detector", *J. Instrum.* **6** (2011) P09001, doi:10.1088/1748-0221/6/09/P09001. arXiv:1106.5048.

[30] CMS Collaboration, "Measurement of the Inclusive W and Z Production Cross Sections in pp Collisions at *sqrts* = 7 TeV", *J. High Energy Phys.* **10** (2011) 132, doi:10.1007/JHEP10(2011)132. arXiv:1107.4789.

[31] CMS Collaboration, "Measurement of the $t\bar{t}$ production cross section in the dilepton channel in pp collisions at *sqrts* = 7 TeV", *J. High Energy Phys.* **07** (2011) 049, doi:10.1007/JHEP07(2011)049. arXiv:1105.5661.

[32] CMS Collaboration, "Measurement of the differential dijet production cross section in proton-proton collisions at $\sqrt{s}$=7 TeV", *Phys. Lett. B* **700** (2011) 187206, doi:10.1016/j.physletb.2011.05.027. arXiv:1104.1693.

[33] CMS Collaboration, "Search for new physics with same-sign isolated dilepton events with jets and missing transverse energy", *CMS Physics Analysis Summary* **CMS-PAS-SUS-11-011** (2012).

[34] A. Read, "Presentation of search results: the CLs technique", *J. Phys. G: Nucl. Part. Phys.* **28** (2002) 2693.

[35] T. Junk, "Confidence level computation for combining searches with small statistics", *Nucl. Instrum. Meth. A* **434** (1999) 435.

[36] M. Kramer et al., "Supersymmetry production cross sections in pp collisions at sqrts = 7 TeV", arXiv:1206.2892.

[37] S. Alekhin et al., "The PDF4LHC Working Group Interim Report", arXiv:1101.0536v1.