

RECEIVED: August 9, 2024
REVISED: January 21, 2025
ACCEPTED: February 18, 2025
PUBLISHED: March 14, 2025

Investigating the effect of hadronic models on IACT images

Benedetta Bruno ,^a Rodrigo Guedes Lang ,^a Luan Bonneau Arbeletche ,^b
Vitor de Souza ,^b and Stefan Funk ^a

^a*Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen Centre for Astroparticle Physics,
Nikolaus-Fiebiger-Str. 2, 91058 Erlangen, Germany*

^b*Instituto de Física de São Carlos, Universidade de São Paulo,
Av. Trabalhador São Carlense 400, São Carlos-SP, 13566-590, Brazil
E-mail: benedetta.bruno@fau.de, rodrigo.lang@fau.de,
luan.arbeletche@gmail.com, vitor@ifsc.usp.br, s.funk@fau.de*

ABSTRACT: The predictions of hadronic interaction models for cosmic-ray induced air showers contain inherent uncertainties due to limitations of available accelerator data. This leads to differences in shower simulations using each of those models. Many studies have been carried out to track those differences by investigating the shower development or the particle content. In this work, we propose a new approach to search for discrepancies and similarities between the models, via the IACT images resulting from the observations of hadronic air showers. We use simulations of H.E.S.S. as a show-case scenario and, by investigating variables of the camera images, we find potential indicators to highlight differences between models. Number of pixels, Hillas image size, and density showed the largest difference between the models. We then further explore the (in)compatibility of the models by combining all the variables and using Boosted Decision Trees. For protons, a significant difference in the classifier output is found for **EPOS-LHC** when compared to both **QGSJET-II04** and **Sybill 2.3d**. For helium and nitrogen, **QGSJET-II04** is shown to be the outlier case. No significant differences are found for silicon and iron. The distribution of (in)compatibility between the models in the phase space of reconstructed shower parameters shows that a targeted search can be fruitful, with showers with reconstructed energies of a few TeV and reconstructed core closer to the large telescope presenting the largest power of separation. An investigation of the distribution of first interaction parameters has shown that **EPOS-LHC** and **QGSJET-II04** result in significantly different distributions of multiplicity and height of first interaction for protons and elasticity and fraction of energy carried by neutral pions for helium and nitrogen.

KEYWORDS: gamma ray experiments, cosmic ray theory, Machine learning

ARXIV EPRINT: [2408.04504](https://arxiv.org/abs/2408.04504)

Contents

1	Introduction	1
2	Simulations	3
3	Results from low-level variables	5
3.1	Mono	5
3.2	Hybrid	8
4	Results from high-level variables	8
4.1	Mono	10
4.2	Hybrid	14
5	Correlations between shower and Hillas parameter distributions	16
6	Conclusions	17
A	Distributions of low-level variables for helium	19

1 Introduction

Imaging Atmospheric Cherenkov Telescopes (IACTs) detect cosmic rays and gamma rays through particle showers initiated by the interaction of these particles with atoms in the atmosphere. Sensitive to the Cherenkov light emitted by relativistic charged particles in the cascade, IACTs enable for the indirect characterization of the primary particle using the properties of images registered on fast cameras. A combination of image-treatment algorithms and posterior multi-parametric analyses result in estimates of the primary particle’s energy, direction, and type. Despite being mainly designed for optimal sensitivity to primary gamma rays with energies > 100 GeV, a dominant background of cosmic rays produces the largest number of detected images. This motivates the use of IACT observatories for purposes other than measuring gamma rays, such as the study of cosmic-ray showers.

Analyses targeting the measurement of gamma-ray fluxes imply a crucial step of de-contaminating event samples from showers initiated by cosmic rays (the gamma/hadron separation step). While gamma-ray-induced showers generate elliptical images on IACTs, the hadronic interactions driving the evolution of cosmic-ray showers imbue on images a morphology that is typically erratic (see e.g. [43]). Therefore, removal of almost 100% of the background through simple image-parametrization schemes (Hillas parameters [21]) is achievable [16]. More complex algorithms (such as machine-learning [17, 23, 28] and template techniques [25, 35]) and improvement of detector capabilities can optimize background removal. However, a natural limit exists [41] because the stochastic nature of hadronic interactions can result in the transference of a large fraction of the primary energy into the electromagnetic component, thus mimicking gamma-ray events [26, 40]. These gamma-like

images are retained as background and their separation from the gamma-ray signal can only be performed statistically. Some methods include measuring off-signal samples (in the case of point-like sources) [11], comparing with simulations and/or modeling of the cosmic-ray background (in the case of extended sources [1, 4, 30] and, beyond, in measuring the diffuse electron flux [5, 10]). Moreover, the need for simulations of cosmic ray events is indispensable when designing and estimating sensitivities of future observatories [3].

Importantly, in cases where the simulation of cosmic-ray showers is required, it has been found that modeling hadronic interactions is a major source of systematic uncertainties, mostly due to the difference in the predicted fractions of gamma-like events. In an analysis targeting the future Cherenkov Telescope Array Observatory [31] (CTAO), the estimated sensitivity to point-like gamma-ray sources has been shown to differ up to 30% when computed with different hadronic interaction models (most notably, in the range $1 \text{ TeV} < E < 30 \text{ TeV}$). Models favoring the production of hard neutral pions result in a larger fraction of gamma-like events, and that fraction is known to be model-dependent, even in the IACT energy range [26]. Moreover, the prediction of shower properties, such as the electromagnetic energy deposit at ground level and Cherenkov-light density at ground level, was compared between different hadronic interaction models in refs. [34, 36] for proton-induced showers, where discrepancies between models were corroborated and shown to be present even without gamma-like event selection. In this direction, it is clear that a better understanding of hadronic interactions can potentially improve gamma-ray detection using IACTs. In particular, because gamma-like events are in the tails of the parameter distribution of protons, this might generate different systematic uncertainties for the general cosmic-ray population.

The existence of different approaches for describing hadronic interactions stems from the incompleteness of our current picture of strong interactions. Because the bulk of possible processes can not be computed in terms of first principles in perturbative Quantum Chromodynamics (QCD), phenomenological models have been elaborated to complement QCD and supply complete models for hadronic and nuclear processes. These models have free parameters that are constrained by collider data [15], and are extrapolated to unmeasured kinematic regions during air-shower simulations. Exhaustive comparisons exist [9, 13] showing that characteristics of particular models lead to uncertainties and discrepancies in the spectra of final-state particles, which are reflected on the simulation shower observables.

In the end, studies for astroparticle physics relying on shower simulations have to deal with the task of choosing one among many hadronic interaction models available as software packages. The most commonly applied are **EPOS-LHC** [37], **Sibyll 2.3d** [38], and **QGSJetII-04** [33], all of which are constrained by data from the LHC [15]. While for astrophysical studies the modeling of hadronic interactions enters as a source of uncertainty, it has been argued [29, 31] that since IACTs are sensitive to hadronic interactions, these detectors stand as a plausible scenario for probing the hadronic interactions themselves.

Therefore, in the direction of using IACT observatories to probe hadronic interactions, we propose in this study to explore the parameter space of detection (in terms of impact distance, primary energy, and image parameters) to find regions where discrepancies between the models are most relevant and can be used to constrain those models. For that, a simulation-based analysis will be performed in the context of the High Energy Stereoscopic System (H.E.S.S) experiment. H.E.S.S. is one of the currently operational IACT arrays and is located in the

Khomas Highlands of Namibia at 1800 m above sea level. It consists of five telescopes, four small telescopes with 108 m^2 mirror area and 15 m focal length, named CT1-4, arranged in a square with 120 m side length, and a single 614 m^2 mirror area telescope (CT5) of 36 m focal length placed in its center [18]. In this work, we use H.E.S.S. as a show-case scenario. Nevertheless, the obtained results are more general and the technique here developed can be applied to any IACT experiment.

This paper is organized as follows. In section 2 the techniques and details of our simulations are presented. In section 3 the difference in Hillas parameters for different models are investigated individually. In section 4, these variables are combined in a machine-learning approach leading to a single high-level variable that improves the power of separation. In section 5 the correlation between distributions of parameters of first shower interaction and Hillas parameters are presented for each model. Finally, in section 6, the work conclusions and outlooks are discussed.

2 Simulations

In order to investigate the effect of hadronic interaction models on IACT images, we simulated a large number of cosmic ray (CR) extensive air showers for different models using the software package CORSIKA (Cosmic Ray Simulation for Kascade, version 7.7402) [19]. Initially developed for the KASCADE experiment [8], this software is publicly available, open-source, and a standard tool for the broader astroparticle physics community. The showers were simulated using the QGSJET-II04, EPOS-LHC and Sibyll 2.3d models implemented in CORSIKA.

All events were simulated at 20° zenith, 180° azimuth, and with an opening angle of the view cone of 5° . The choice of these particular zenith and azimuth angles does not lead to a loss in generality, as it is the typical setting for studies evaluating IACTs performances [2, 7, 17, 22]. The simulations were performed for the so-called `phase2d3` configuration, which corresponds to the latest state of the H.E.S.S. array taking into account effects such as the optical degradation of the telescope mirrors and changes in electronics [39, 42]. Five representative primaries (proton, helium, nitrogen, silicon, and iron) were used. The energy range and total number of events were variable, depending on the primary particle (table 1), and a spectral index of -2 was used. Furthermore, we used the software `sim_telarray` (version 1.63) to fully simulate the detector response, ranging from the particles' ray tracing to the measurement with photomultiplier tubes (PMTs) and its digitization [12].

Once the simulations are recorded, a calibration and cleaning procedure is needed. Both were done using the standard H.E.S.S. Analysis Program (HAP); in particular, a dual threshold cleaning procedure is applied to exclude all pixels without a shower signal and to keep only pixels with shower signals in local proximity. In this paper, we used the so-called 5/10 tailcuts cleaning [6] for CT1-4. For CT5, we use a similar cleaning, but considering two neighboring pixels, 9/16NN2.

We further use pre-selection cuts, part of the standard analysis chain of H.E.S.S., selecting only camera images that are not truncated by the edge of the camera and contain enough light and pixels for a proper Hillas reconstruction. The cuts are listed in table 2 and ensure high-quality reconstructions even at low energies. The amount of remaining events resulting from the analysis through the different configurations is shown in table 1 and illustrated in figure 1.

Primary	Energy range [TeV]	# Triggered showers	# mono events	# hybrid events
Proton	0.03–150	$\sim 5.1 \cdot 10^7$	$\sim 3.3 \cdot 10^5$	$\sim 1.3 \cdot 10^5$
Helium	0.03–500	$\sim 3.0 \cdot 10^7$	$\sim 1.3 \cdot 10^5$	$\sim 5.0 \cdot 10^4$
Nitrogen	0.04–800	$\sim 3.0 \cdot 10^7$	$\sim 1.2 \cdot 10^5$	$\sim 5.0 \cdot 10^4$
Silicon	0.05–1000	$\sim 1.3 \cdot 10^7$	$\sim 5.3 \cdot 10^4$	$\sim 2.3 \cdot 10^4$
Iron	0.06–1200	$\sim 1.3 \cdot 10^7$	$\sim 5.2 \cdot 10^4$	$\sim 2.3 \cdot 10^4$

Table 1. Summary of the energy ranges and the number of triggered showers for each primary particle (which is roughly the same for each hadronic interactions model) together with the final number of events after cuts for the two different configurations considered in this work.

Name	Amplitude [p.e.]		Local distance [m]		# pixels		Multiplicity
	CT1-4	CT5	CT1-4	CT5	CT1-4	CT5	
Mono	-	≥ 80	-	≤ 0.8	-	≥ 5	= 1
Hybrid	≥ 60	≥ 250	≤ 0.525	≤ 0.72	≥ 1	≥ 10	= 5

Table 2. Preselection cuts used in this analysis. Image amplitude refers to the summed light of the pixels surviving the cleaning and local distance to the distance of the ellipse's center of gravity to the center of the camera. Multiplicity refers to the minimum number of telescopes surviving preselection cuts for the event to be considered.

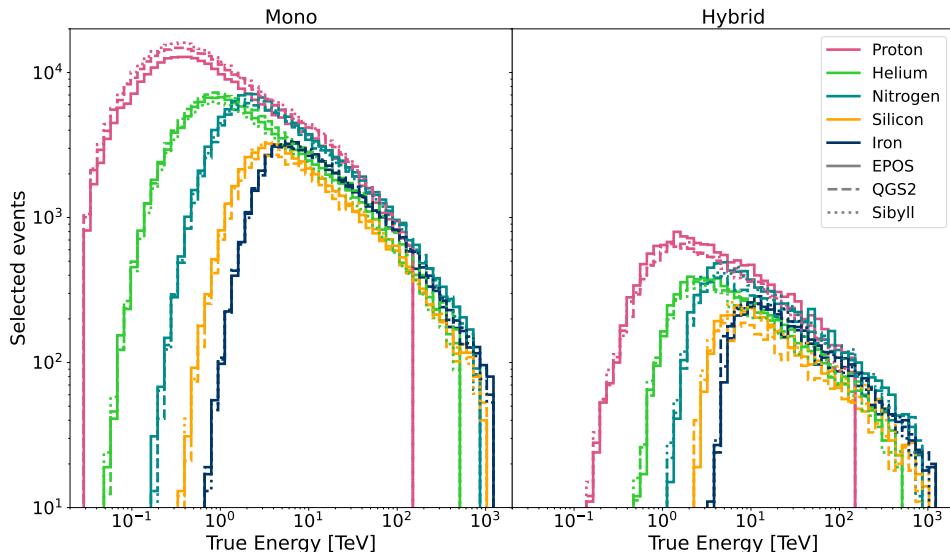


Figure 1. Number of selected events as a function of Monte Carlo true energy for each primary and hadronic model. Selected events are those that triggered the system and survived the preselection cuts. Each color represents a primary, while each line style represents a different hadronic model. Left and right panels are for different preselection cuts: mono, which only considers CT5, and hybrid, which considers CT1-5.

In H.E.S.S., the reconstruction can be performed with different configurations of the five telescopes enabling studies at different energy ranges. These configurations are motivated by either the start of operations of individual telescopes — for example — CT1-4 have been operational since 2003, and CT5 was added to the array in 2012 or by different science goals — lower energies (from tens to a few hundred of GeV) are only accessible to the CT5 due to its larger size, while higher energy can be detected with the whole array resulting in a higher quality reconstruction. At the time of writing, there are three main configurations: *mono*, *stereo*, and *hybrid*. “*Mono*”, corresponds to reconstructions that solely include CT5 — most effective at low energies. The “*stereo*” and “*hybrid*” configurations correspond to any combination of at least two telescopes from CT1-4 and CT1-5, respectively, having good quality data, i.e., satisfying the required selection. The analyses presented here are done with the mono and hybrid configurations, where the latter has an additional cut applied in order to keep only events detected in all telescopes.

3 Results from low-level variables

To search for differences in the IACT images resulting from the different hadronic models, first the effects on low-level variables were investigated. With those, we indicate some parameters directly related to the detected images in the cameras of the IACTs. In particular, in this work, we considered the Hillas parameters [20, 44] such as width, length, size (i.e. the total amplitude of the image), kurtosis, and skewness.

Furthermore, we also consider the number of pixels that survive the cleaning process and the so-called “density”, D , and “length over size”, L/S , defined as follows:

$$D = \frac{\text{Size}}{\text{Width} \cdot \text{Length}}, \quad L/S = \frac{\text{Length}}{\log(\text{Size})}.$$

We also explore the parameter space of true energy, i.e. the simulated energy of the primary particle initiating the shower, and true impact parameter, which here refers to the distance of the shower impact core from the center of the array. As the aim of this study is to investigate the differences between the models themselves, here we only focus on true Monte Carlo parameters.

In the following sections, 3.1 and 3.2, only results for proton are shown, while the results for helium regarding the low-level variables can be found in figures 14–17 in appendix A.

3.1 Mono

For each variable, the quantity $\frac{\overline{V_a} - \overline{V_b}}{\overline{V_b}}$ was calculated, where \overline{V} stands for the mean of the corresponding variable in a region of the parameter space (e.g., in a bin of true energy and true impact parameter) and a and b are two of the three models to be compared.

Figure 2 shows the 2D histogram for each variable (the rows of the grid-plot) and for each combination between the three models (the columns of the grid-plot) when having a proton initiating the air shower. When looking at the differences between the models QGSJET-II04 and EPOS-LHC or Sibyll 2.3d and EPOS-LHC, for all the variables in the plot, it is possible to note differences, in particular along the solid and dotted lines, which indicate the 68% and 95% containment radius for the Monte Carlo true impact parameter, i.e. where most of

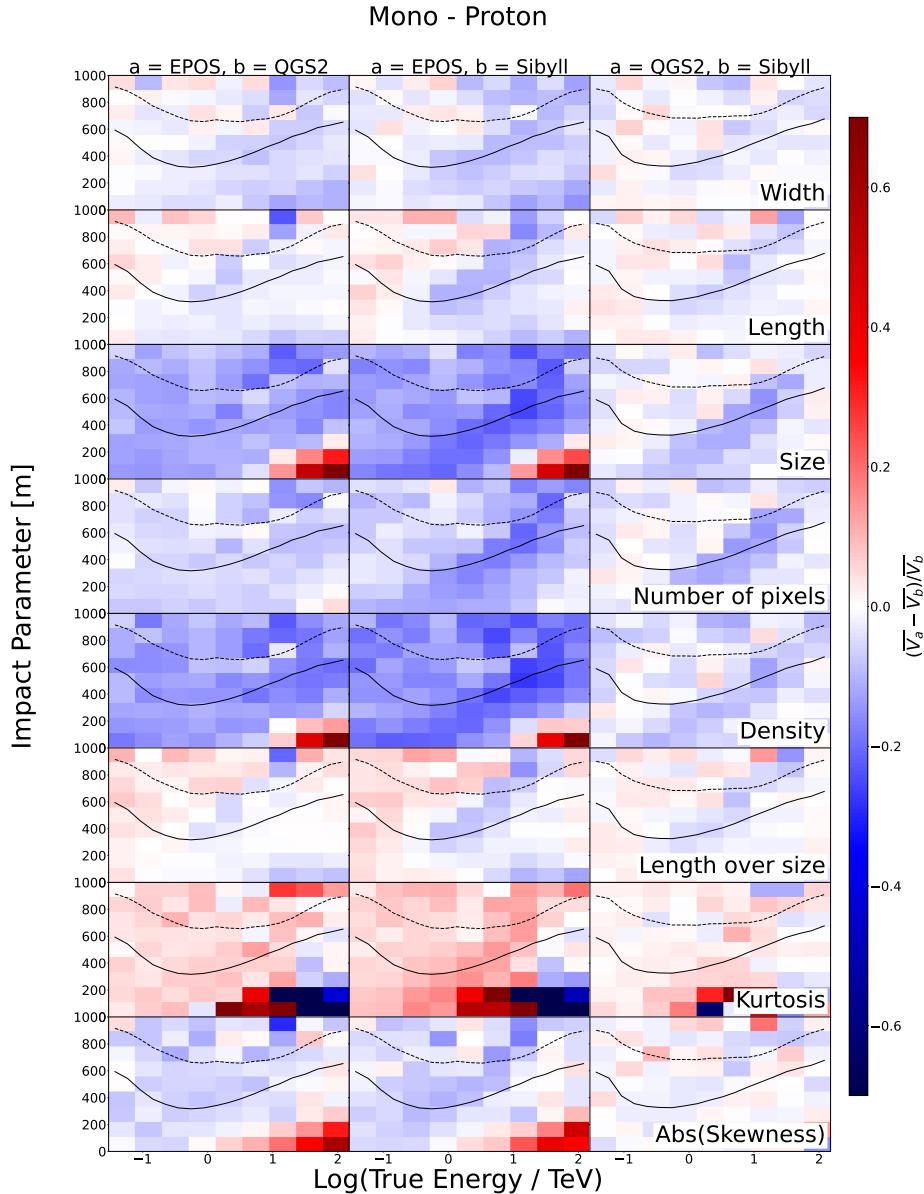


Figure 2. Given a pair of models, a and b, and indicating with V the mean of a given variable, this grid-plot shows the 2D histogram of the $\frac{V_a - V_b}{V_b}$ in impact parameter and true energy bins for the mono configuration. Each row corresponds to a low-level variable, while each column refers to a pair of compared models, specified at the top of the corresponding column. Solid and dashed lines indicate the 68% and 95% containment radius for the Monte Carlo true impact parameter. Only proton primaries are shown.

the events concentrate. The comparison between **QGSJET-II04** and **Sibyll1 2.3d** shows also some differences but not as strong as the other cases and only for some of the variables. It is clear that, for each subplot, the differences are concentrated in specific regions of the phase space. In particular, for the size, density, kurtosis and skewness parameters it is possible to spot an area of high energy and small impact parameter with large differences. Overall, the

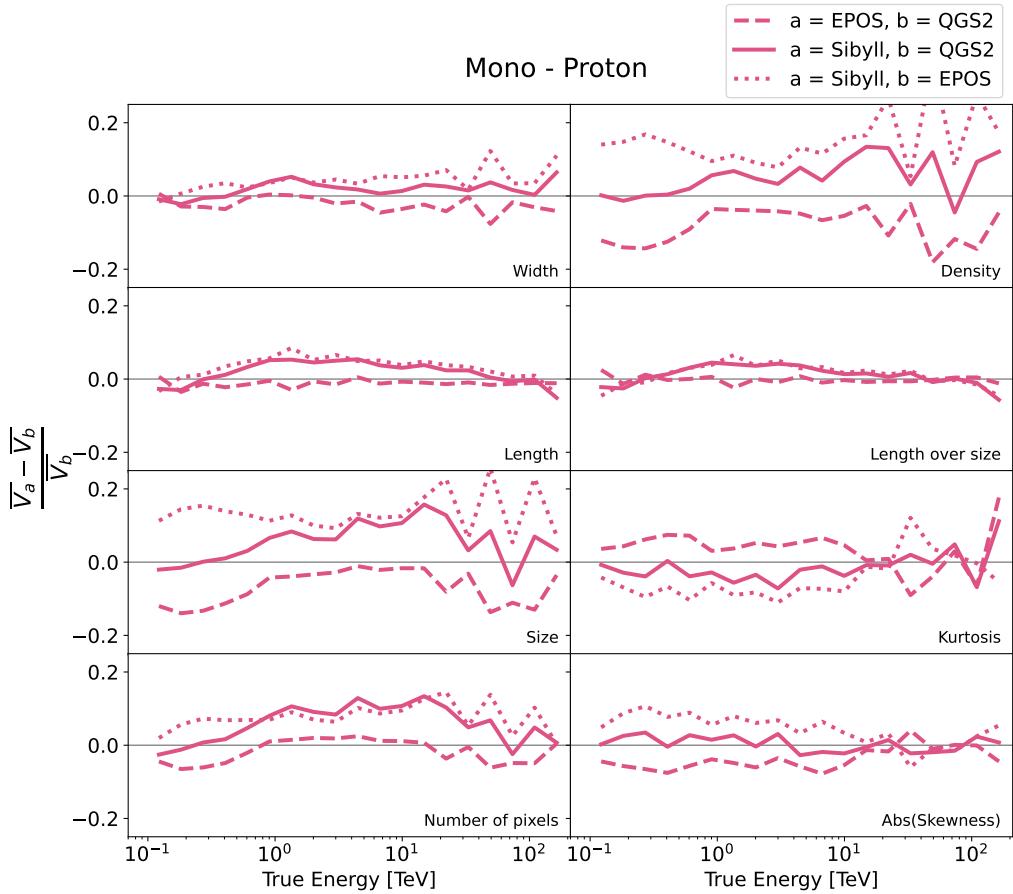


Figure 3. The same quantity for the same set of variables as in figure 2 is shown here only in energy bins. In each bin, the values are calculated by selecting events around 0.5σ of the mean of the impact parameter. Solid, dashed, and dotted lines indicate the pair of different models being compared, according to the legend on the top right of the plot. Only proton primaries are shown.

variables that show the strongest dependencies from the used hadronic interaction model are size, number of pixels, density, kurtosis, and skewness.

While the different behaviors in different regions of the phase space may allow for a more targeted search, certain regions may suffer from a lack of statistics. In order to inspect the differences that could still be statistically significant, we have calculated the quantity $\frac{\bar{V}_a - \bar{V}_b}{\bar{V}_b}$ for all variables, but binned in true energy and by selecting events that are in the range of $\pm 0.5\sigma$ from the mean of the impact parameter in that energy bin. This is shown in figure 3.

Here, it is easier to spot which variables are more important in distinguishing the models and which models are more consistent with each other, by looking at how far the lines are from zero. The largest effect is found for density, size, and number of pixels. For the first two, the solid line related to the comparison between Sibyll 2.3d and QGSJET-II04 does not deviate much from zero. This means that the two models are more in agreement with each other than when these same models are compared to EPOS-LHC.

As seen in figure 16, size and number of pixels are also the most significant variables for helium. Nevertheless, the model that differs the most from the other two is **QGSJET-II04**. The same behavior is found for nitrogen and silicon, while no significant difference is found for iron.

3.2 Hybrid

Hadronic showers may contain several sub-electromagnetic showers, which result in asymmetric camera images for the different telescopes. This effect, however, is not encapsulated by computing the mean of the variables for each telescope and for each event. Instead, to fully exploit the hybrid configuration, for each variable, the standard deviation among CT1-5 was computed for each event ($\sigma(V)$), then the average among all the events was taken to compare one model to the other. The resulting 2D- and 1D-histograms as before are shown in figure 4 and figure 5.

The 2D-histograms show fewer structures with the hybrid configuration in comparison to the mono configuration, but it is still possible to spot some differences along the 68% and 95% containment radius for some variables like density, number of pixels, and size. Again, those differences are more visible only in the first two columns (**EPOS-LHC** vs **QGSJET-II04** and **EPOS-LHC** vs **Sibyll 2.3d**), highlighting once more **EPOS-LHC** as the model that differs the most from the others.

This is also confirmed by the hybrid 1D-histogram, where the solid line for the same three mentioned parameters (size, density, and number of pixels) is the closest to zero, meaning that **QGSJET-II04** and **Sibyll 2.3d** are harder to distinguish.

As for the mono configuration, helium results are shown in appendix A. It is harder to draw clear conclusions from these plots, but a hint is given from the width parameter for which it seems that the **QGSJET-II04** model is the one differing the most from the others, as already shown for the mono configuration when considering helium, silicon and nitrogen. Iron remains the primary for which it is not possible to assert a particular outcome.

4 Results from high-level variables

Once the effects of each individual variable are better understood, a more sensitive investigation of the (in)compatibility of the models can be explored by combining all the variables. In particular, we use the Boosted Decision Trees (BDT) algorithm, a widely used method in astrophysics [14, 24, 27, 32] and proven efficient for classification tasks. We have trained a BDT for each combination of: configuration (mono and hybrid), primary (H, He, N, Si, and Fe), and pair of models (**EPOS-LHC** versus **QGSJET-II04**, **EPOS-LHC** versus **Sibyll 2.3d**, and **QGSJET-II04** vs **Sibyll 2.3d**), resulting in $2 \times 5 \times 3 = 30$ different BDT trainings. The data was split, with 90% of it used for training the model and 10% used for evaluating it. These are called *train* and *test* datasets. The eight variables discussed in section 3 were used as an input. Density and size were modified to log density and log size given the vast range of magnitudes covered by these variables. To make it easier to check the stability of the BDT for *a* versus *b* and the differences between the evaluated datasets, we convolved the BDT output distribution with a function such that the *train* distribution for model *a* is flat between 0 and 1. In such a way, a dataset that agrees with model *a* will present a flat

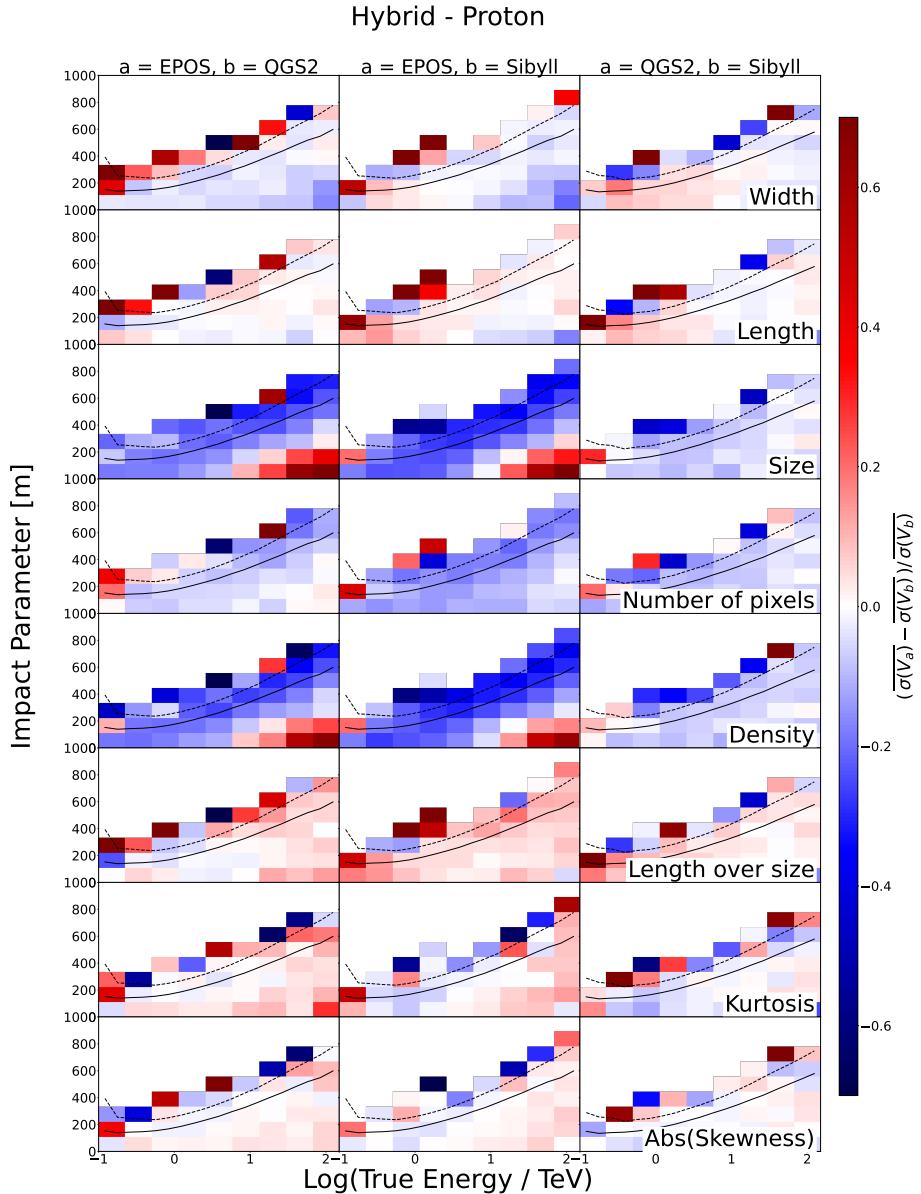


Figure 4. Given a pair of models, a and b , and indicating with V the mean of the standard deviation among CT1-5 of a given variable, this grid-plot shows the 2D histogram of the $\frac{\sigma(V_a) - \sigma(V_b)}{\sigma(V_b)}$ in impact parameter and true energy bins for the hybrid configuration. Each row corresponds to a low-level variable, while each column refers to a pair of compared models, specified at the top of the corresponding column. Solid and dashed lines indicate the 68% and 95% containment radius for the Monte Carlo true impact parameter. Only proton primaries are shown.

BDT output distribution, while a dataset that agrees with model b will present a non-flat distribution peaked around 1. After the training, the BDT was also evaluated with the full dataset of the third model not involved in the training.

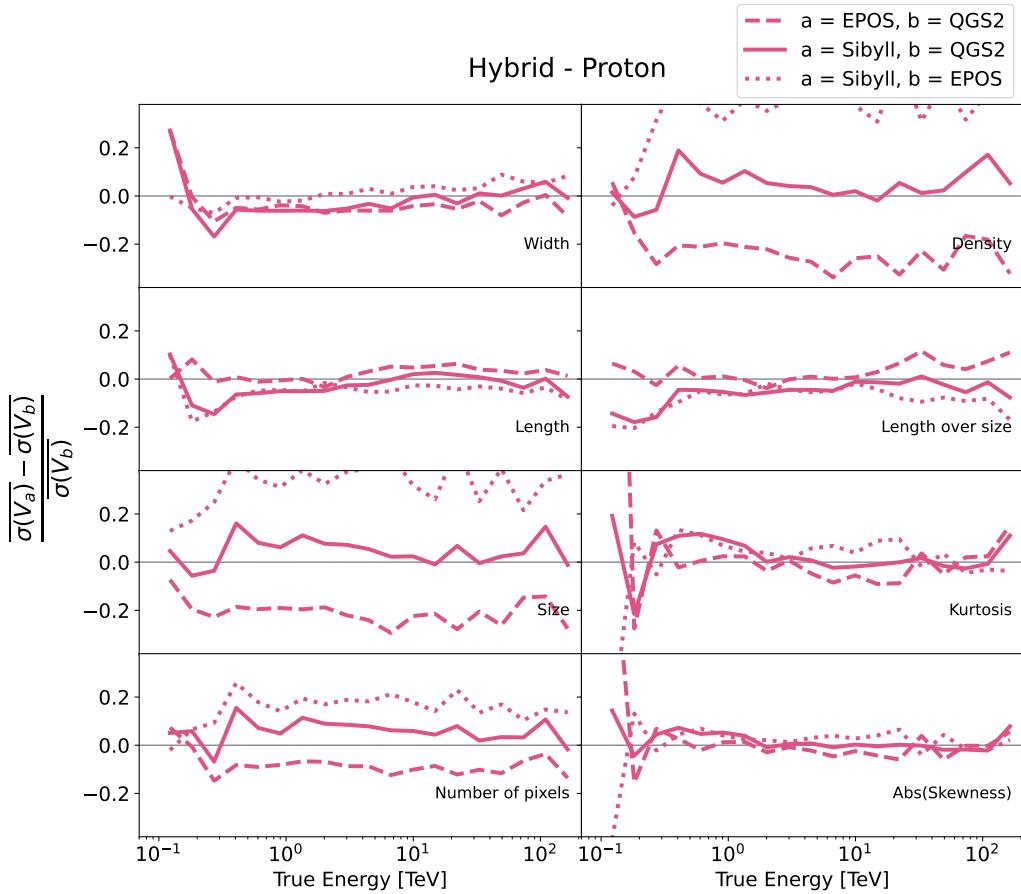


Figure 5. The same ratio for the same set of variables as in figure 4 is shown here only in energy bins. In each bin, the values are calculated by selecting events around 1σ of the mean of the impact parameter. Solid, dashed, and dotted lines indicate the pair of different models being compared, according to the legend on the top right of the plot. Only proton primaries are shown.

4.1 Mono

Figure 6 shows the results of the BDT for each pair of models for mono and protons. By choice, the distribution for the *train* dataset of one of the models will always be flat. The distribution for the *train* dataset of the other model involved in the training is expected to be significantly different from flat and peak at 1. If that's the case, a successful separation between these models is achieved. Another important indication of the stability of the training is an agreement between the *test* and *train* datasets for each model. The dataset for the third model not involved in the training has also been evaluated in each case. This reveals similarities/differences between the models which are unbiased with respect to any overfitting of the BDT.

With that, it can be noted that the training was stable and successful for EPOS-LHC vs QGSJET-II04 and EPOS-LHC vs Sibyll 2.3d. Density, size, and number of pixels are shown to be the most relevant variables, as expected.

For QGSJETII-04 versus Sibyll 2.3d, the BDT output distributions are very similar and a bad agreement is found between *test* and *train*. This suggests that the BDT could not find

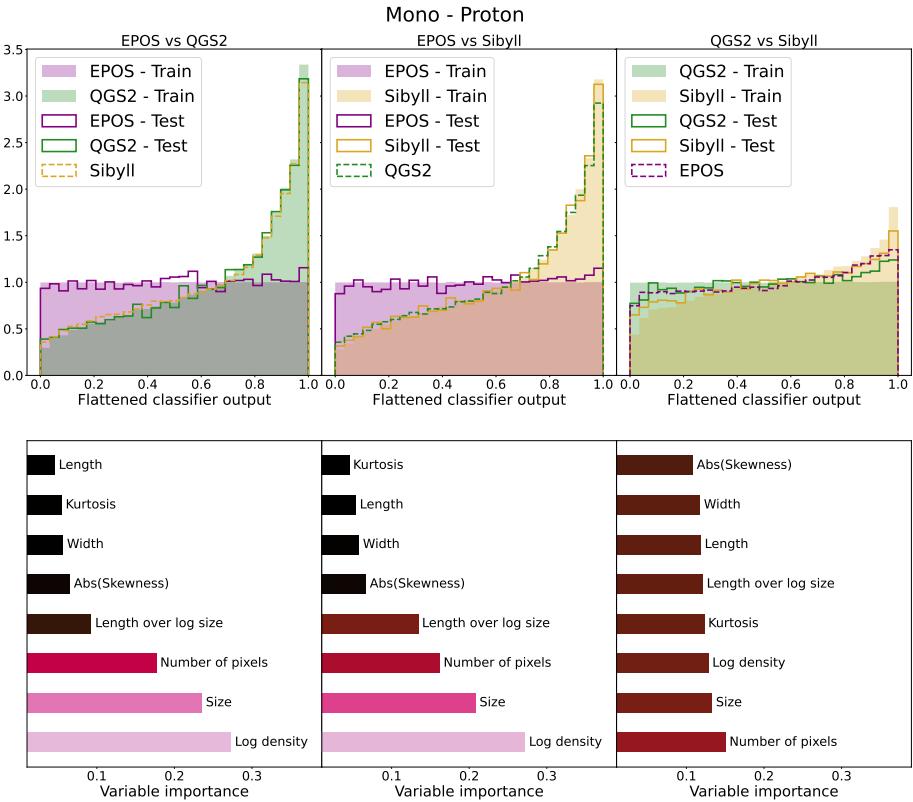


Figure 6. Results of the mono BDT trained for separating each pair of hadronic models. Only primary protons are considered. The top panels show the flattened classifier output distribution for the *train* (shaded areas) and *test* (lines) data as well as for the third model not involved in the training. The bottom panels show the variable importance for each case.

significant differences between these models. A further exploration of the hyperparameters of the BDT, in particular, number of estimators, maximum depth and learning rate, has proven not enough to improve the separation. This demonstrates even further the similarity between **QGSJETII-04** versus **Sibyll 2.3d** models for protons.

Figures 7 and 8 show the results for helium and nitrogen. For these, the situation changes and **QGSJET-II04** is found to be the most different model. For **EPOS-LHC** versus **Sibyll 2.3d**, a significant difference between the BDT distribution for the *test* and *train* datasets can be seen. Both *test* distributions fall somewhere in the middle between the *train* distribution. This is also true for the distribution coming for the third model (**QGSJET-II04** in this case). This indicates a strong similarity between these models, in agreement with the findings of section 3. No significant separation is found for silicon and iron.

No large differences between the distributions are found such that an event-by-event classification could be possible. This is expected since the effects of the different models are known to not be of leading order in showers for these energies. Nevertheless, if the shapes of the BDT distributions are significantly different and stable (i.e., with the *test* distribution agreeing with the *train* distribution), this effect could be investigated in data by looking at the distribution of events.

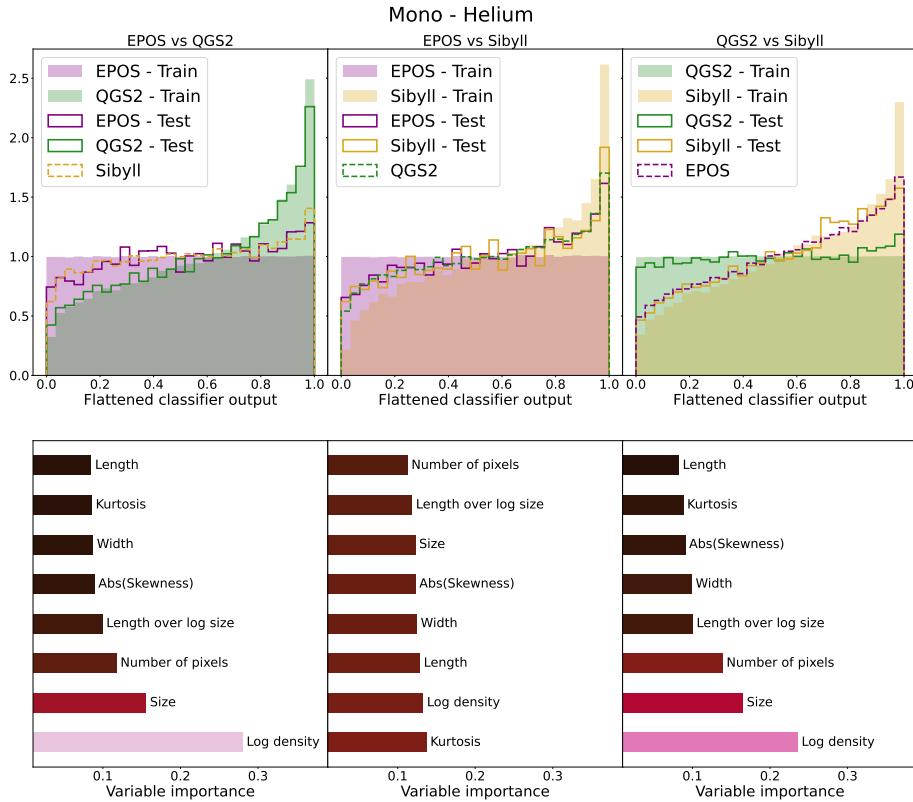


Figure 7. Results of the mono BDT trained for separating each pair of hadronic models. Only primary helium is considered. The top panels show the flattened classifier output distribution for the *train* (shaded areas) and *test* (lines) data as well as for the third model not involved in the training. The bottom panels show the variable importance for each case.

A more targeted search could be performed if the differences could be reduced to specific regions of the phase space. We further investigate this by choosing the outlier model for each case (EPOS-LHC for protons and QGSJET-II04 for helium and nitrogen) and dividing their *test* data set into two: low BDT, for flattened BDT output < 0.35 and high BDT, for flattened BDT output > 0.65 , i.e., the 35% of events with lowest and highest BDT outputs. Events with low/high BDT output are those in which a high/low separation was found and are, thus, more incompatible/compatible with the other two models. Finally, we again divide the data into different regions of the phase space by binning in true energy and true impact parameter as discussed before. For each bin, we calculate the value of $\xi = (n_{\text{lowBDT}} - n_{\text{highBDT}})/(n_{\text{lowBDT}} + n_{\text{highBDT}})$ as an indicator of incompatibility, with positive/negative values indicating higher/lower incompatibility.

Here, two different parameter phase spaces are considered. On the one hand, we use true Monte Carlo parameters, which are important for pointing out where are the differences coming from in the models themselves. In this case, we select a few regions of interest (marked in figure 9) to further investigate in the following section. On the other hand, the phase space of reconstructed parameters is explored. This is crucial as future investigations using data can only be performed with reconstructed parameters. It is, though, important

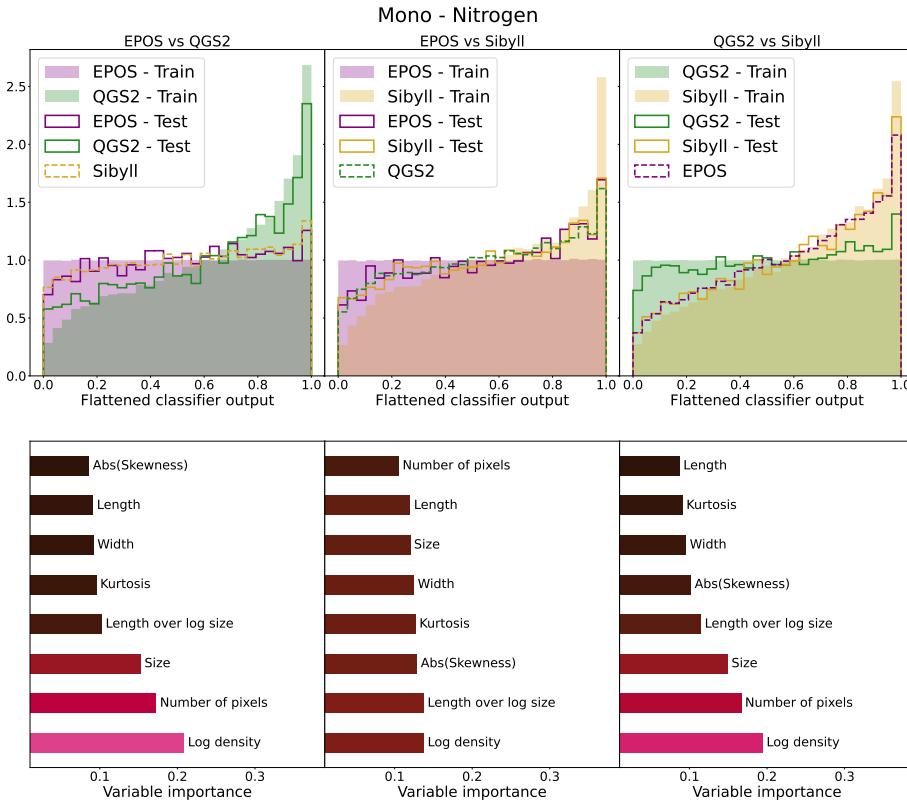


Figure 8. Results of the mono BDT trained for separating each pair of hadronic models. Only primary nitrogen is considered. The top panels show the flattened classifier output distribution for the *train* (shaded areas) and *test* (lines) data as well as for the third model not involved in the training. The bottom panels show the variable importance for each case.

to note that IACT reconstruction is designed for gamma-rays and, for that reason, it is not expected that the distributions of reconstructed and true parameters should agree. This disagreement is more significant for mono, which uses lower energy events and no stereo reconstruction. Nevertheless, this is not a drawback of the analysis, since the goal is to understand which region of the phase space for data is interesting, which will be consistent even if the reconstructed values do not reflect the real ones.

The distributions of ξ are shown in figure 9. A clear region of incompatibility can be seen in the phase space, with the most incompatible events being those for which the core of the air shower was located closer to the center of the array (correspondingly closer to CT5) in the true parameter case. These events are mostly within the 68% containment radius of the impact parameter, showing that they can be explored with the bulk of the statistics without the need to stretch the analysis for very uncommon events. The increase in impact parameter for higher energies probably comes from the fact that, for the highest energies, very low impact parameter events will be truncated and therefore discarded by the preselection cuts.

Six representative regions of the true parameters phase space are chosen to be further investigated in section 5.

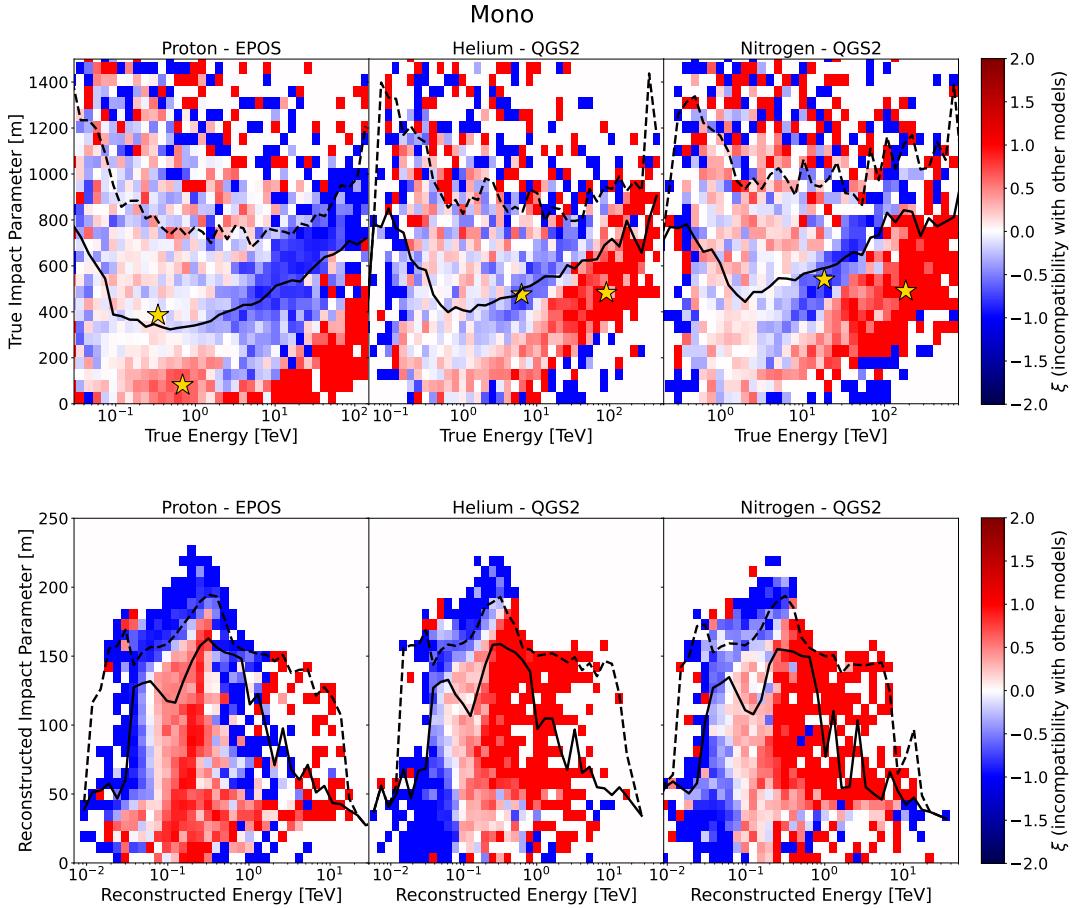


Figure 9. Energy and impact parameter distribution of the (in)compatibility with other hadronic models for a mono configuration. The color scale represents the value of $\xi = (n_{\text{lowBDT}} - n_{\text{highBDT}})/(n_{\text{lowBDT}} + n_{\text{highBDT}})$, for which large positive (red) values mean incompatibility with the other two hadronic models, while large negative (blue) values mean compatibility with the other two hadronic models. The three panels show the results for proton, helium, and nitrogen, respectively. For each primary, the model whose BDT result outliers from the other two is shown. The full and dashed black lines show the 68% and 95% containment radius for the Monte Carlo true impact parameter. The yellow stars point out regions of interest further explored in section 5. The top and bottom panels show the space for true and reconstructed parameters, respectively.

4.2 Hybrid

The same procedure was repeated for the hybrid configurations, but now using the standard deviation of the variables as input.

Figure 10 shows the results for proton. Once more EPOS-LHC is found to be the outlier model. No significant difference between the models is found for the other primaries. This is expected due to the strict preselection cuts we have chosen for the hybrid configuration.

Finally, figure 11 shows the distribution of incompatibility in the parameter phase space. A clear behavior is seen, with most incompatibility for EPOS-LHC being found in events with small impact parameters. This is consistent with the results found for the mono configuration.

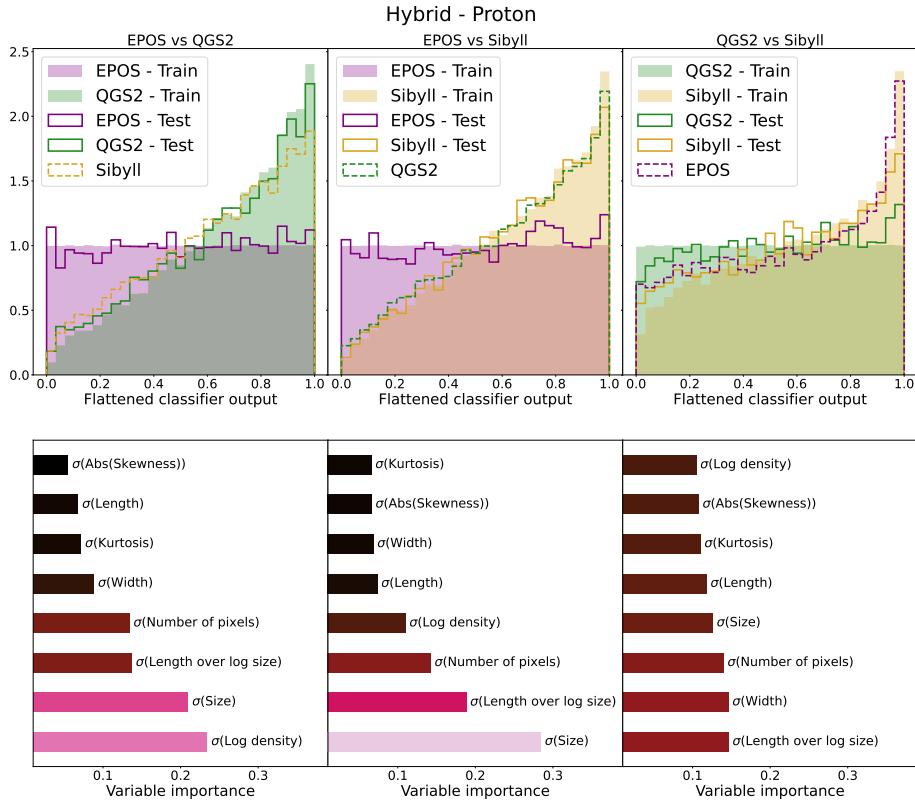


Figure 10. Results of the hybrid BDT trained for separating each pair of hadronic models. Only primary protons are considered. The top panels show the flattened classifier output distribution for the *train* (shaded areas) and *test* (lines) data as well as for the third model not involved in the training. The bottom panels show the variable importance for each case.

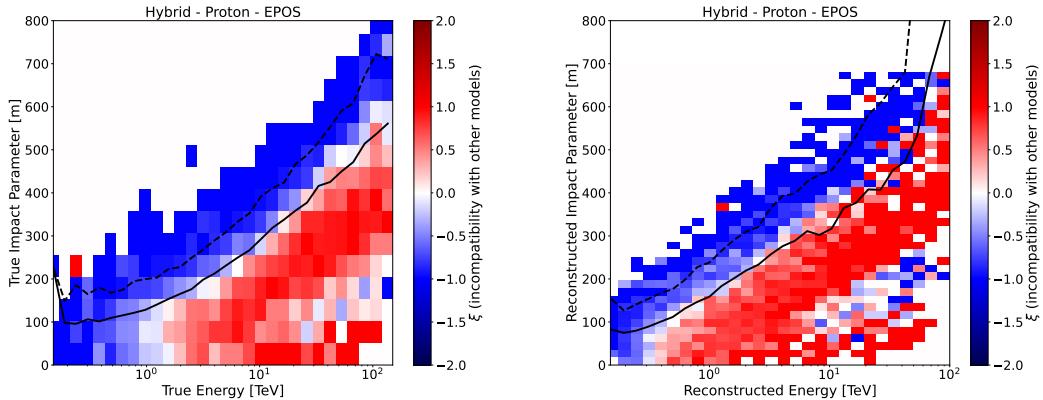


Figure 11. Energy and impact parameter distribution of the (in)compatibility with other hadronic models for a hybrid configuration. The color scale represents the value of $\xi = (n_{\text{lowBDT}} - n_{\text{highBDT}}) / (n_{\text{lowBDT}} + n_{\text{highBDT}})$, for which large positive (red) values mean incompatibility with the other two hadronic models, while large negative (blue) values mean compatibility with the other two hadronic models. The three panels show the results for proton, helium, and nitrogen, respectively. For each primary, the model whose BDT result outliers from the other two is shown. The full and dashed black lines show the 68% and 95% containment radius for the Monte Carlo true impact parameter. The left and right panels show the space for true and reconstructed parameters, respectively.

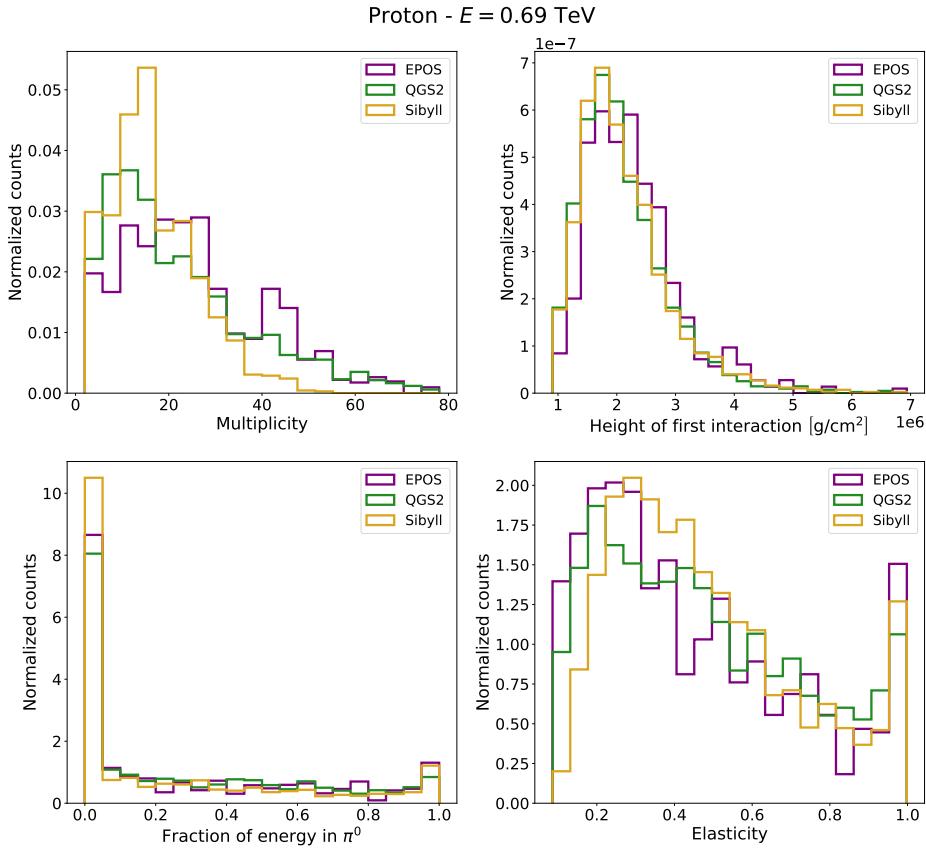


Figure 12. Distribution of parameters of the first shower interaction for the region of interest with large incompatibility for protons. Each color represents a different hadronic model and the panels show distributions for multiplicity, height of first interaction, fraction of energy on π^0 and elasticity, respectively.

5 Correlations between shower and Hillas parameter distributions

With the understanding of the most affected parameters of the IACT images and the corresponding regions of phase space of the shower where larger inconsistencies are found, the next natural step is to build some insight into the difference in shower physics leading to these discrepancies. To do so, we have selected two points of the phase space for each primary in the mono analysis, one in which a large discrepancy is found and one in which a small discrepancy is found. These are represented by a yellow star in figure 9. For each point, i.e., fixed energy and impact parameter, 10^4 showers were simulated for each hadronic interaction model. The species and four momenta of the particles generated in the first interaction, as well as the height of the first interaction, were saved for each shower. From that, it was calculated: (a) multiplicity, the number of particles generated, (b) elasticity, the fraction of the energy carried by the most energetic particle, and (c) fraction of energy carried by neutral pions.

Figure 12 shows the distributions for these quantities for the point of interest with larger incompatibility for protons, i.e., $E = 0.69$ TeV. EPOS-LHC was shown in the previous section to be the most discrepant model for this primary. A larger multiplicity and a deeper first interaction are obtained on average with this model. The average elasticity, on the other

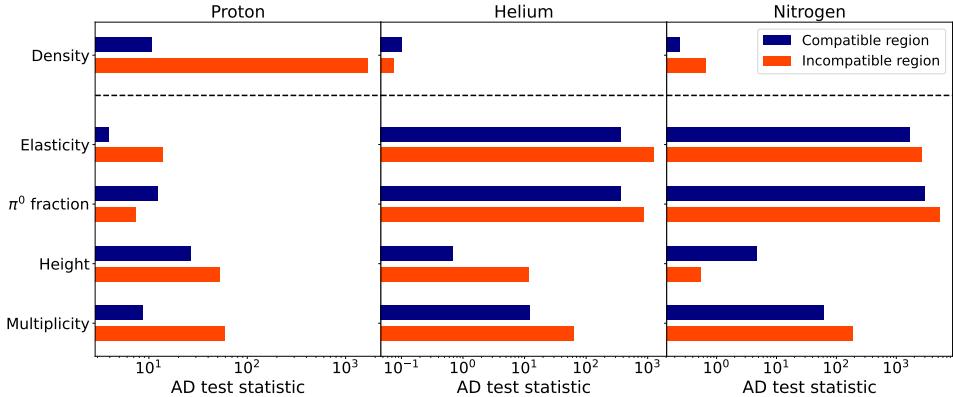


Figure 13. Anderson-Darling statistical test between QGSJET-II04 and EPOS-LHC showers in the same region of the phase space of true parameters. Larger values of the statistical test represent larger intrinsic differences in the distributions. The selected regions are those highlighted by yellow stars in figure 9, representing a region of high compatibility and a region of high incompatibility for each primary. The dashed line separates Hillas parameters (density) from first shower interaction parameters (elasticity, multiplicity, fraction of energy carried by neutral pions, and elasticity). Each panel represents results for a different primary.

hand, is smaller, which is expected in an interaction with larger multiplicity, as more particles are present to share the energy.

A more quantitative estimation of the discrepancies is then investigated by performing an Anderson-Darling test for the distributions of EPOS-LHC and QGS-JET-II04 for each of the six points of interest and the four first interaction distributions as well as Hillas density. The results are shown in figure 13. For protons, the largest differences are found in the height and multiplicity of the first interaction. For helium and nitrogen, on the other hand, the most significant differences come from elasticity and fraction of the energy carried by neutral pions. As expected, for every significant distribution, a larger discrepancy is seen in the incompatible region with the respect to the one seen in the compatible region.

6 Conclusions

Understanding hadronic interactions is a crucial problem for both high-energy physics and astroparticle physics and IACT observatories can be useful in this regard. In this work, we have investigated the possibility of testing hadronic interaction models with IACTs. We performed a range of simulations for extensive air showers with the three most commonly used hadronic models, EPOS-LHC, QGSJET-II04 and Sibyll 2.3d, and IACT response using H.E.S.S. as our show-case scenario.

The main focus of the work was on evaluating the effects of the choice of the hadronic model in the IACT camera images, in particular, in the distribution of the Hillas parameters. The work was divided into two main approaches. First, the distributions of different Hillas parameters were investigated independently. Two different configurations, five different primaries and the three possible combinations of each pair of models were evaluated. Number of pixels, size and density were shown to be the variables with the most significant differences.

EPOS-LHC showed the largest differences to the other two models for protons and **QGSJET-II04** for helium, nitrogen, and silicon. No clear difference was seen for iron. These results can shine a light on understanding possible differences coming from the different assumptions made by each of the models considered.

A further investigation was performed by combining all the Hillas parameters into a BDT. Even though an event-by-event classification is not possible (and was not envisioned), a significant and stable difference was found in the BDT distributions for some cases. It was once more verified that **EPOS-LHC** stands out for protons, while **QGSJET-II04** stands out for helium and nitrogen. The incompatibility of the models was explored for different regions of the phase space of reconstructed and true parameters. From the reconstructed parameters, it was possible to determine target regions to be explored by further studies using measured data.

The true phase space was then used to determine interesting regions in which the models were further investigated. Distributions of first interaction parameters were investigated showing that the main discrepancies between **EPOS-LHC** and **QGSJET-II04** are in the multiplicity and height of first interaction for protons and elasticity and fraction of energy carried by neutral pions for helium and nitrogen.

Therefore, we propose this method as a novel tool for using IACT data to explore the compatibility of the hadronic models with experimental data, allowing further constraining of the models. We also show that a more targeted search could be fruitful by separately analyzing the data in certain regions of the phase space of energy and impact parameter. While H.E.S.S. is used as a show-case scenario, the methods here presented can easily be applied to other IACT experiments as well as to the forthcoming CTAO.

Acknowledgments

We thank the H.E.S.S. Collaboration for allowing us to use H.E.S.S. simulations for this publication. The idea for this work was spawned during a series of two joint FAPESP/BAYLAT workshops titled “Astroparticle physics in the era of CTA and SWGO” at the Friedrich-Alexander-Universität Erlangen-Nürnberg and the Instituto de Física de São Carlos, Universidade de São Paulo in 2023 and 2024. We acknowledge the generous support of these workshops by FAPESP (through grant number 2022/01271-7) and BAYLAT. VdS also thanks FAPESP grant number 2021/01089-1. The authors would like to thank Sam Spencer and Alison Mitchell for valuable feedback on the draft of this manuscript. The authors also thank the referee for valuable comments that lead to a deeper investigation of the subject in this work.

A Distributions of low-level variables for helium

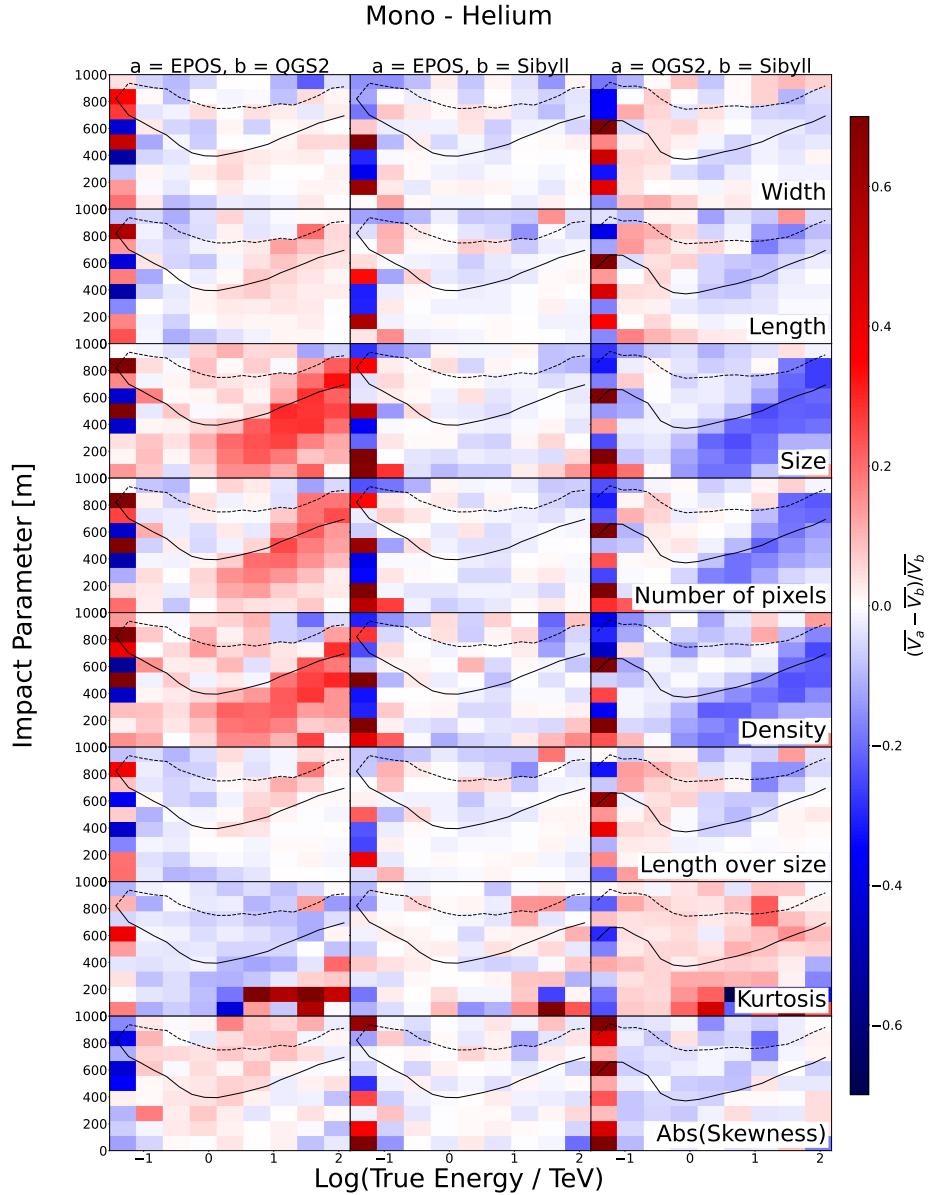


Figure 14. Given a pair of models, a and b, and indicating with V the mean of a given variable, this grid-plot shows the 2D histogram of the $\frac{V_a - \bar{V}_b}{\bar{V}_b}$ in impact parameter and true energy bins for the mono configuration. Each row corresponds to a low-level variable, while each column refers to a pair of compared models, specified at the top of the corresponding column. Solid and dashed lines indicate the 68% and 95% containment radius for the Monte Carlo true impact parameter. Only helium primaries are shown.

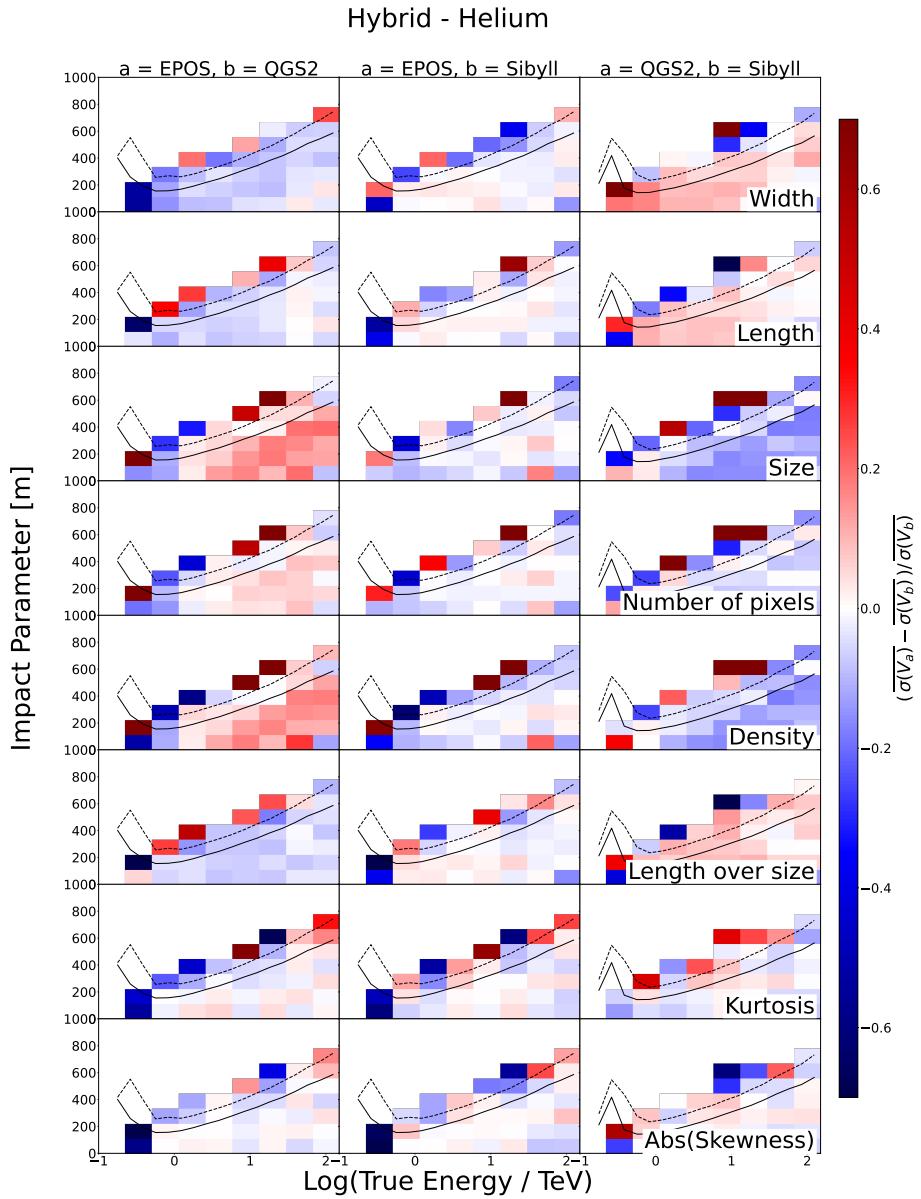


Figure 15. Given a pair of models, a and b, and indicating with V the mean of the standard deviation among CT1-4 of a given variable, this grid-plot shows the 2D histogram of the $\frac{\sigma(V_a) - \sigma(V_b)}{\sigma(V_b)}$ in impact parameter and true energy bins for the hybrid configuration. Each row corresponds to a low-level variable, while each column refers to a pair of compared models, specified at the top of the corresponding column. Solid and dashed lines indicate the 68% and 95% containment radius for the Monte Carlo true impact parameter. Only helium primaries are shown.

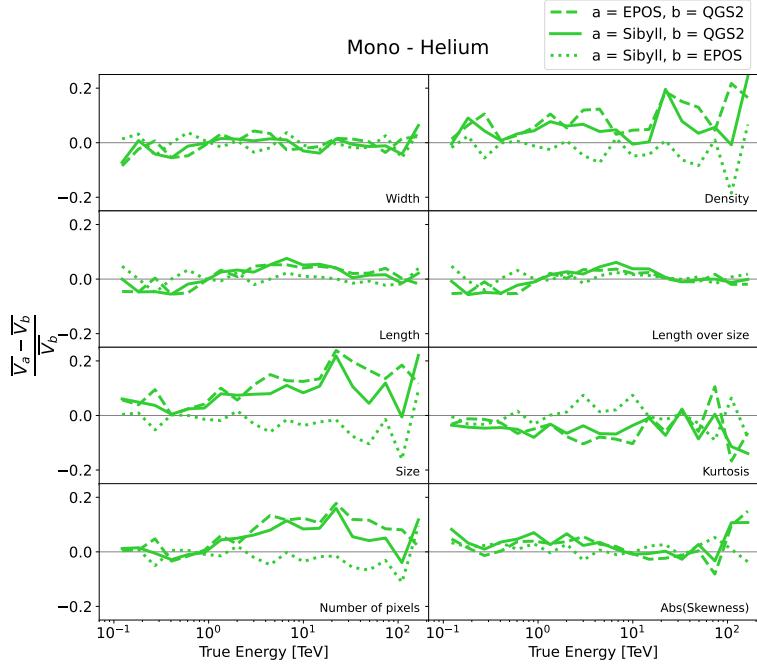


Figure 16. The same quantity for the same set of variables as in figure 14 is shown here only in energy bins. In each bin, the values are calculated by selecting events around 0.5σ of the mean of the impact parameter. Solid, dashed, and dotted lines indicate the pair of different models being compared, according to the legend on the top right of the plot. Only helium primaries are shown.

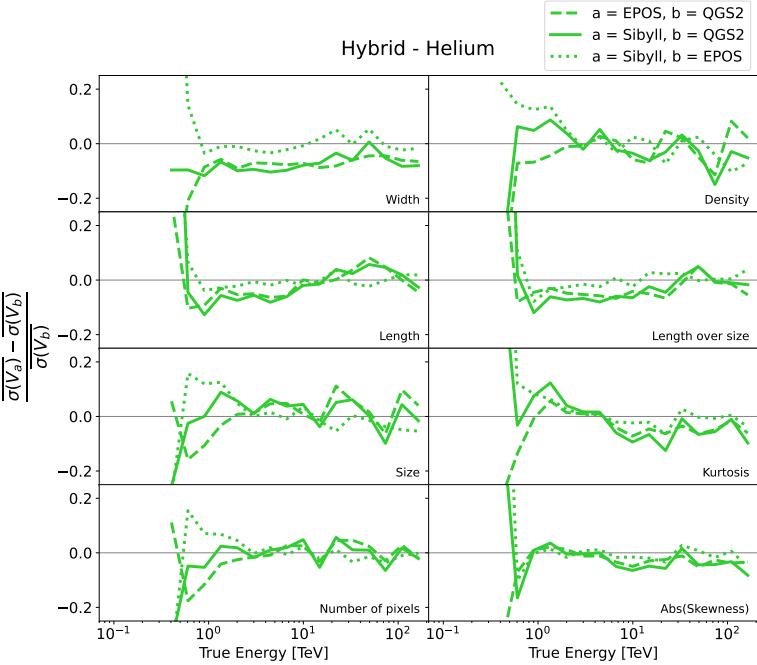


Figure 17. The same ratio for the same set of variables as in figure 15 is shown here only in energy bins. In each bin, the values are calculated by selecting events around 1σ of the mean of the impact parameter. Solid, dashed, and dotted lines indicate the pair of different models being compared, according to the legend on the top right of the plot. Only helium primaries are shown.

References

- [1] HAWC collaboration, *Extended gamma-ray sources around pulsars constrain the origin of the positron flux at Earth*, *Science* **358** (2017) 911 [[arXiv:1711.06223](https://arxiv.org/abs/1711.06223)] [[INSPIRE](#)].
- [2] CTA CONSORTIUM collaboration, *Science with the Cherenkov Telescope Array*, World Scientific (2018) [[DOI:10.1142/10986](https://doi.org/10.1142/10986)] [[INSPIRE](#)].
- [3] A. Acharyya et al., *Monte Carlo studies for the optimisation of the Cherenkov Telescope Array layout*, *Astropart. Phys.* **111** (2019) 35 [[arXiv:1904.01426](https://arxiv.org/abs/1904.01426)] [[INSPIRE](#)].
- [4] FERMI-LAT collaboration, *The Spectrum and Morphology of the Fermi Bubbles*, *Astrophys. J.* **793** (2014) 64 [[arXiv:1407.7905](https://arxiv.org/abs/1407.7905)] [[INSPIRE](#)].
- [5] H.E.S.S. collaboration, *The energy spectrum of cosmic-ray electrons at TeV energies*, *Phys. Rev. Lett.* **101** (2008) 261104 [[arXiv:0811.3894](https://arxiv.org/abs/0811.3894)] [[INSPIRE](#)].
- [6] H.E.S.S. collaboration, *Observations of the Crab Nebula with H.E.S.S.*, *Astron. Astrophys.* **457** (2006) 899 [[astro-ph/0607333](https://arxiv.org/abs/astro-ph/0607333)] [[INSPIRE](#)].
- [7] MAGIC collaboration, *The major upgrade of the MAGIC telescopes. Part II: A performance study using observations of the Crab Nebula*, *Astropart. Phys.* **72** (2016) 76 [[arXiv:1409.5594](https://arxiv.org/abs/1409.5594)] [[INSPIRE](#)].
- [8] KASCADE collaboration, *The Cosmic ray experiment KASCADE*, *Nucl. Instrum. Meth. A* **513** (2003) 490 [[INSPIRE](#)].
- [9] L.B. Arbeletche, V.P. Goncalves and M.A. Muller, *Investigating the influence of diffractive interactions on ultrahigh-energy extensive air showers*, *Int. J. Mod. Phys. A* **33** (2018) 1850153 [[arXiv:1701.05238](https://arxiv.org/abs/1701.05238)] [[INSPIRE](#)].
- [10] VERITAS collaboration, *Measurement of Cosmic-ray Electrons at TeV Energies by VERITAS*, *Phys. Rev. D* **98** (2018) 062004 [[arXiv:1808.10028](https://arxiv.org/abs/1808.10028)] [[INSPIRE](#)].
- [11] D. Berge, S. Funk and J. Hinton, *Background Modelling in Very-High-Energy γ -ray Astronomy*, *Astron. Astrophys.* **466** (2007) 1219 [[astro-ph/0610959](https://arxiv.org/abs/astro-ph/0610959)] [[INSPIRE](#)].
- [12] K. Bernlohr, *Simulation of Imaging Atmospheric Cherenkov Telescopes with CORSIKA and sim_telarray*, *Astropart. Phys.* **30** (2008) 149 [[arXiv:0808.2253](https://arxiv.org/abs/0808.2253)] [[INSPIRE](#)].
- [13] L. Calcagni, C.A. García Canal, S.J. Sciutto and T. Tarutina, *LHC updated hadronic interaction packages analyzed up to cosmic-ray energies*, *Phys. Rev. D* **98** (2018) 083003 [[arXiv:1711.04723](https://arxiv.org/abs/1711.04723)] [[INSPIRE](#)].
- [14] Y. Coadou, *Boosted Decision Trees and Applications*, *EPJ Web Conf.* **55** (2013) 02004 [[INSPIRE](#)].
- [15] D. d'Enterria et al., *Constraints from the first LHC data on hadronic event generators for ultra-high energy cosmic-ray physics*, *Astropart. Phys.* **35** (2011) 98 [[arXiv:1101.5596](https://arxiv.org/abs/1101.5596)] [[INSPIRE](#)].
- [16] D.J. Fegan, *gamma/hadron separation at TeV energies*, *J. Phys. G* **23** (1997) 1013 [[INSPIRE](#)].
- [17] J. Glombitza, V. Joshi, B. Bruno and S. Funk, *Application of graph networks to background rejection in Imaging Air Cherenkov Telescopes*, *JCAP* **11** (2023) 008 [[arXiv:2305.08674](https://arxiv.org/abs/2305.08674)] [[INSPIRE](#)].
- [18] H.E.S.S. collaboration, *Resolving the Crab pulsar wind nebula at teraelectronvolt energies*, *Nature Astron.* **4** (2020) 167 [[arXiv:1909.09494](https://arxiv.org/abs/1909.09494)] [[INSPIRE](#)].
- [19] D. Heck et al., *CORSIKA: A Monte Carlo Code to Simulate Extensive Air Showers*, Technical Report (1998), <https://hal.in2p3.fr/in2p3-00005094>.

[20] A.M. Hillas, *Cerenkov light images of EAS produced by primary gamma*, <https://api.semanticscholar.org/CorpusID:118390321> (1985).

[21] A.M. Hillas, *Cerenkov Light Images of EAS Produced by Primary Gamma Rays and by Nuclei*, in the proceedings of the 19th International Cosmic Ray Conference (ICRC19), San Diego, U.S.A. (1985), pg. 445.

[22] H.E.S.S. collaboration, *Observations of the Crab Nebula with H.E.S.S. Phase II*, *PoS ICRC2015* (2016) 847 [[arXiv:1509.02902](https://arxiv.org/abs/1509.02902)] [[INSPIRE](#)].

[23] M. Jacquemont et al., *GammaLearn: a Deep Learning framework for IACT data*, *PoS ICRC2019* (2020) 705 [[INSPIRE](#)].

[24] M. Krause, E. Pueschel and G. Maier, *Improved γ /hadron separation for the detection of faint γ -ray sources using boosted decision trees*, *Astropart. Phys.* **89** (2017) 1 [[arXiv:1701.06928](https://arxiv.org/abs/1701.06928)] [[INSPIRE](#)].

[25] S. Le Bohec et al., *A new analysis method for very high definition imaging atmospheric Cherenkov telescopes as applied to the CAT telescope*, *Nucl. Instrum. Meth. A* **416** (1998) 425 [[astro-ph/9804133](https://arxiv.org/abs/astro-ph/9804133)] [[INSPIRE](#)].

[26] G. Maier and J. Knapp, *Cosmic-Ray Events as Background in Imaging Atmospheric Cherenkov Telescopes*, *Astropart. Phys.* **28** (2007) 72 [[arXiv:0704.3567](https://arxiv.org/abs/0704.3567)] [[INSPIRE](#)].

[27] G. Maier and J. Holder, *Eventdisplay: An Analysis and Reconstruction Package for Ground-based Gamma-ray Astronomy*, *PoS ICRC2017* (2018) 747 [[arXiv:1708.04048](https://arxiv.org/abs/1708.04048)] [[INSPIRE](#)].

[28] T. Miener et al., *Reconstruction of stereoscopic CTA events using deep learning with CTLearn*, *PoS ICRC2021* (2021) 730 [[arXiv:2109.05809](https://arxiv.org/abs/2109.05809)] [[INSPIRE](#)].

[29] A.M.W. Mitchell, H.P. Dembinski and R.D. Parsons, *Potential for measuring the longitudinal and lateral profile of muons in TeV air showers with IACTs*, *Astropart. Phys.* **111** (2019) 23 [[arXiv:1903.12040](https://arxiv.org/abs/1903.12040)] [[INSPIRE](#)].

[30] L. Mohrmann et al., *Validation of open-source science tools and background model construction in γ -ray astronomy*, *Astron. Astrophys.* **632** (2019) A72 [[arXiv:1910.08088](https://arxiv.org/abs/1910.08088)] [[INSPIRE](#)].

[31] M. Ohishi et al., *Effect of the uncertainty in the hadronic interaction models on the estimation of the sensitivity of the Cherenkov telescope array*, *J. Phys. G* **48** (2021) 075201 [[arXiv:2105.05822](https://arxiv.org/abs/2105.05822)] [[INSPIRE](#)].

[32] S. Ohm, C. van Eldik and K. Egberts, *Gamma-Hadron Separation in Very-High-Energy gamma-ray astronomy using a multivariate analysis method*, *Astropart. Phys.* **31** (2009) 383 [[arXiv:0904.1136](https://arxiv.org/abs/0904.1136)] [[INSPIRE](#)].

[33] S. Ostapchenko, *Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: I. QGSJET-II model*, *Phys. Rev. D* **83** (2011) 014018 [[arXiv:1010.1869](https://arxiv.org/abs/1010.1869)] [[INSPIRE](#)].

[34] R.D. Parsons, C. Bleve, S.S. Ostapchenko and J. Knapp, *Systematic uncertainties in air shower measurements from high-energy hadronic interaction models*, *Astropart. Phys.* **34** (2011) 832 [[arXiv:1102.4603](https://arxiv.org/abs/1102.4603)] [[INSPIRE](#)].

[35] R.D. Parsons and J.A. Hinton, *A Monte Carlo Template based analysis for Air-Cherenkov Arrays*, *Astropart. Phys.* **56** (2014) 26 [[arXiv:1403.2993](https://arxiv.org/abs/1403.2993)] [[INSPIRE](#)].

[36] R.D. Parsons and H. Schoorlemmer, *Systematic Differences due to High Energy Hadronic Interaction Models in Air Shower Simulations in the 100 GeV-100 TeV Range*, *Phys. Rev. D* **100** (2019) 023010 [[arXiv:1904.05135](https://arxiv.org/abs/1904.05135)] [[INSPIRE](#)].

- [37] T. Pierog et al., *EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider*, *Phys. Rev. C* **92** (2015) 034906 [[arXiv:1306.0121](https://arxiv.org/abs/1306.0121)] [[INSPIRE](#)].
- [38] F. Riehn et al., *A new version of the event generator Sibyll*, *PoS ICRC2015* (2016) 558 [[arXiv:1510.00568](https://arxiv.org/abs/1510.00568)] [[INSPIRE](#)].
- [39] J. Schäfer, *Simulation of IceACT With TARGET, Validation of the H.E.S.S. Simulation Chain, and a New H.E.S.S. 3D Analysis of the Crab Using the ABRIR Method for Improved Background Rejection*, Ph.D. thesis, Friedrich-Alexander University of Erlangen-Nuremberg, Nuremberg, Germany (2023) [[INSPIRE](#)].
- [40] J. Sitarek et al., *Nature of the low-energy, γ -like background for the Cherenkov Telescope Array*, *Astropart. Phys.* **97** (2018) 1 [[arXiv:1710.06980](https://arxiv.org/abs/1710.06980)] [[INSPIRE](#)].
- [41] D. Sobczynska, *Natural limit to the γ /hadron separation for a stand alone air Cherenkov telescope*, *J. Phys. G* **34** (2007) 2279 [[astro-ph/0702562](https://arxiv.org/abs/astro-ph/0702562)] [[INSPIRE](#)].
- [42] S.F. Steinmaßl, *Probing particle acceleration in stellar binary systems using gamma-ray observations.*, Ph.D. thesis, Heidelberg University, Heidelberg, Germany (2023) [[INSPIRE](#)].
- [43] H.J. Völk and K. Bernlöhr, *Imaging Very High Energy Gamma-Ray Telescopes*, *Exper. Astron.* **25** (2009) 173 [[arXiv:0812.4198](https://arxiv.org/abs/0812.4198)] [[INSPIRE](#)].
- [44] T.C. Weekes et al., *Observation of TeV gamma rays from the Crab nebula using the atmospheric Cerenkov imaging technique*, *Astrophys. J.* **342** (1989) 379 [[INSPIRE](#)].