Comparing Different Jet Algorithms for DIS at High Q^2*

Boris Lemmer

DESY, MPI-Munich / Justus-Liebig-Universität Gießen

Abstract. To analyze data taken from high Q^2 DIS events with the H1 detector at HERA, it is necessary to obtain very accurate information about the jets in the events. Therefore, the optimal jet algorithm has to be used.

Both clustering (k_T, anti-k_T, Cambridge/Aachen) and cone type (SISCone) algorithms are compared concerning deviations of the most interesting observables between reconstructed and hadron level on the one hand and hadron and parton level on the other hand. Also, jet radius for each of the algorithms was investigated. A matching procedure was used to find the corresponding jet at the other level. The matching efficiency was studied in dependence of the jet radius.

1. MOTIVATION

Within the MIP-Munich group of H1, the cross section of $ep \to jets$ events is studied to obtain information about the proton PDF and the strong coupling constant α_s.

From the energy depositions in the H1 calorimeters and the tracking detector the particle flows (jets) are reconstructed. So called *jet algorithms* perform this task, and they should do so as precisely as possible. Up to now the k_T algorithm is well established and was used for the analyses of HERA data.

But within the last year further infra-red safe algorithms like the SISCone and the anti-k_T algorithm were developed and the older Cambridge/Aachen was rediscovered. With this variety of usable algorithms the question arises whether the k_T algorithm is still state of the art. Currently two PhD students are working on jet analyses for DIS (Roman Kogler[†]) and photoproduction (Aziz Dossanov[‡]). While the the jet study described here deals with DIS events, the study by Clemens Mellein (also a summer student) focuses on photoproduction (1). This work with DIS events is briefly compared to the one with photoproduction events in chapter 6.3.

2. DIS AT HIGH Q^2

The general basics about electron-proton-scattering are introduced first. To describe the kinematics of an event, only a small set of observables is needed, i.e. the four momentum vectors of the proton, incoming electron and outgoing electron. These will be called P, k and k' and . In the limit of high kinetic energies (neglect of rest masses) the total energy in the center of mass system is

$$s = (k + P)^2 \approx 4E_e E_p \tag{2.1}$$

where E_e and E_p, the energies of the electron and proton beams, are fixed. We can define the squared four momentum transfer for scattering events with

$$Q^2 = -q^2 = -(k - k') \approx 2E_e E_{e'} (1 - \cos(\theta)) \tag{2.2}$$

where θ is polar the angle between the beamline and the scattered electron, with the z-axis aligned in the direction of the proton.

Justus-Liebig-Universität Gießen (e-mail: boris.lemmer@physik.uni-giessen.de).

*Supervisor: Günter Grindhammer, MPI-Munich, guenterg@desy.de

[†]kogler@desy.de

[‡]aziza@desy.de

Scattering processes can be more or less inelastic. To describe the degree of inelasticity, we use the variable

$$y = \frac{Pq}{Pk} \approx 1 - \frac{E'_e}{E_e} \sin^2 \frac{\theta}{2} \tag{2.3}$$

The *Bjorken Scaling Variable* is defined as

$$x = \frac{Q^2}{2Pq} = \frac{Q^2}{2yPk} \tag{2.4}$$

For fixed s any two variables of Q^2, y and x suffice to describe inclusive events. In the most basic leading order process in DIS (see figure 1) it can be interpreted as the fraction of the proton's four momentum carried by the interacting parton. In this quark-parton-model view we further assume that the parton has neither a transverse momentum nor rest mass. Furthermore, the different partons in the proton may not interact with each other during the time of scattering interaction which is guaranteed if Q^2 is large enough (or the wavelength of the virtual photon small enough). A suitable reference frame to study effects of the strong interactions, i.e. of Quantum Chromodynamics (QCD) on the basic process of figure 1 is the *Breit frame*. In this frame the scattered parton of the lowest order process carries no transverse momentum. Boosting all events into the Breit frame therefore allows to get rid of events easily by a simple requirement on the P_T of the photons or jets. In the event shown in figure 1 the photon

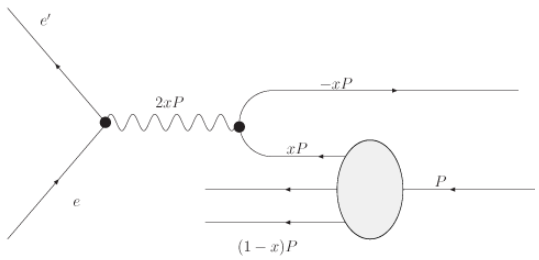

FIG 1. *Breit frame*

does not transmit any energy. What happens is that the four momentum vector xP of the parton changes its direction, but keeps its absolute value.

There are several types of reaction that can occur at a DIS event (see figure 2). The virtual photon can either react directly with a quark from the proton carrying the momentum fraction x. Those events are in the order of α and are named "Born contribution" (see figure figure 2 (a)). The simplest case in the order of $\alpha\alpha_s$ is the "QCD Compton Scattering" shown in figure 2 (b). In this process the proton's quark radiates a gluon before scattering, so the quark carries a momentum fraction smaller than the initial x. Both the gluon and the scattered quark hadronize to jets. Another possible event is "boson-gluon fusion" which is shown in figure 2 (c). A gluon from the proton and the virtual photon "fuse" to a $q\bar{q}$ pair. This process is again in the order of $\alpha\alpha_s$.

As α_s is going to be studied, events with only the Born contribution are not of interest. When a scattered parton hadronizes and forms a jet, it might be hard to decide if the detected jet originates only from a parton (figure 2 (a)) or if it is an overlap of hadronization of the original parton and a radiated gluon (figure 2 (b)). To select events in the order of α_s, a boost to the Breit frame is done.

3. THE JET ALGORITHMS

3.1 Demands on Jet Algorithms

A collision event can be analyzed at different levels. What we see at the H1 event display is what should be called the *reconstructed level*: Data from the trackers and calorimeters is used to reconstruct the track of a particle and identify it. At this level, a jet can be defined as an area with high energy density.

Who asks for the physical reason for the energy detections is guided to the *hadron level*, where one talks directly about the reconstructed particles and their interactions. Here, jets can be defined as sprays of hadrons.

Going one step deeper brings you to the *parton level*, where one deals directly with the hadron's constituents (partons). And here a jet is a formation of one or many close-by partons.

Our algorithms should ideally give the same or very similar results at all levels. To check this, we observe the deviations of the reconstructed jet properties at the different levels (See chapter 6.2).

Further requirements address issues of infrared and collinear (IRC) safety, factorizability , small renormalization scale dependence and small hadronization corrections. IRC safety demands that the algorithm output - namely the jets - are not sensitive to the addition of "soft" or collinear particles.

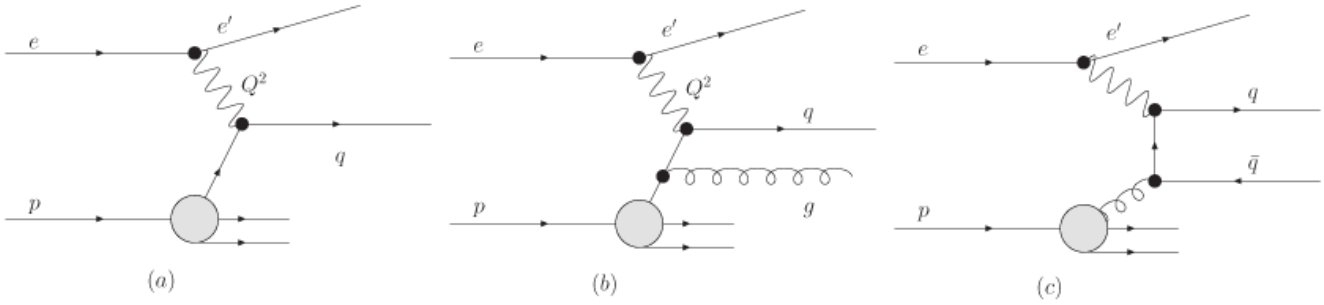

FIG 2. *Deep-inelastic lepton-proton scattering at different orders of α_s: (a) Born contribution $\mathcal{O}(1)$, (b) QCD Compton scattering $\mathcal{O}(\alpha_s)$ and (c) boson-gluon fusion $\mathcal{O}(\alpha_s)$*

There exist two classes of jet algorithms: clustering and cone type algorithms. While clustering algorithms are favored in the analysis of e^+/e^- annihilation events (initial state free of hadrons), cone type algorithms are preferentially used at p/p collisions as the jet areas calculated with cone type algorithms are more regular and hadronic background is easier to subtract. Both types of algorithms are briefly introduced below.

3.2 Clustering Algorithms

Clustering algorithms start calculating the distances d_{ij} of all entities (particles, pseudojets) in the final state and also the distance for each entity to the beam (d_{iB}). Within the list of all d_{ij} and d_{iB} the smallest distance is picked. If it is a d_{ij}, the entities i and j will be merged. If a d_{iB} is smallest, entity i will be defined as a jet.

What is needed for this procedure to run is a clear definition of a *distance* and the merging process. A general definition of the distance for all clustering algorithms is given by

$$d_{ij} = \min\left(k_{Ti}^{2p} k_{Tj}^{2p}\right) \frac{\Delta_{ij}^2}{R_0^2} \tag{3.1}$$

$$d_{iB} = P_{Ti}^{2p} \tag{3.2}$$

with the geometrical distance

$$\Delta_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2 \tag{3.3}$$

and the transverse momenta P_T, the pseudorapidity η and the azimuth ϕ. Each geometrical distance is weighted with the momentum of one of the entities to the power $2p$. The constant p defines a jet class. R_0 is a free parameter for each jet finder algorithm. In this analysis the dependence on the value for R_0 is investigated.

For $p = 1$ we have the k_t *algorithm* (4). Here, the distance is mainly affected by the P_T (all transverse momenta with respect to the beam) of the soft particle/jet. $p = 0$ leads to the *Cambridge/Aachen algorithm* (5), where we only have a dependence on the geometrical distance. Finally, for $p = -1$ we get the *anti-k_t algorithm* (6), where d_{ij} is dominated by the hard particle/jet. The different P_T dominance mainly influences the clustering behavior of soft particles: For the anti-k_t algorithm, the soft particles will first try to attach to a hard jet instead of clustering with each other. Vice versa for the k_t algorithm.

For the merging of two entities the P_T *scheme* is commonly used. As the name suggests, it works the following way:

$$P_{T(ij)} = P_{Ti} + P_{Tj} \tag{3.4}$$

$$\eta_{(ij)} = \frac{P_{Ti}\eta_i + P_{Tj}\eta_j}{P_{Ti} + P_{Tj}} \tag{3.5}$$

$$\phi_{(ij)} = \frac{P_{Ti}\phi_i + P_{Tj}\phi_j}{P_{Ti} + P_{Tj}} \tag{3.6}$$

It ensures that the jets remain massless.

3.3 Cone Type Algorithms

Cone type algorithms work in a different, more geometric way. In a first step, stable cones are found. After that the cones are split and merged into jets.

Stable cones are defined as follows: starting from angular trial cones pointing in the direction of a more or less arbitrary particle ("seed"), all four momenta within the cone are added, which results in a new direction, the direction of the next trial cone.

This iteration is repeated until the cone axis is not changing anymore, and the cone can be considered as stable. All particles within the stable cone are removed from the list of particles in the event, and the next trial cone is treated.

The split-merge procedure works as follows: A P_T cut on the protojets is performed. Out of the remaining protojets, the one with the highest P_T is chosen. If there is no overlap with another protojet, it is defined as the final jet. If there is an overlap, the jets are - depending on the total P_T of the overlap - either merged or split. If they are split, the particles of the overlap are attached to the jet with the closer axis.

A problem with cone type jets starting with seeds is that they are not IRC safe. A way out of this problem is provided by the seedless and IRC safe SISCone algorithm (details in (7)), which are used within this study.

4. SIMULATION AND SETTINGS

To compare data on all levels (reconstructed, hadron and parton level), a set of generated Monte Carlo data is needed. Two generators were used to provide data: DJANGOH (2) and RAPGAP (3). They differ mainly in how they simulate parton showers. Both models use the LUND string model as implemented in PYTHIA for the fragmentation of partons to hadrons and for their decays. Bth models simulate QED radiative effects as implemented in HERACLES. To take only those events from the simulation belonging to a *neutral current DIS events at high* Q^2, the following cuts are applied:

$$150\text{GeV}^2 < Q^2 < 15000\text{GeV}^2 \quad (4.1)$$

$$0.2 < y < 0.7 \quad (4.2)$$

$$45.0\text{GeV} < \sum E - p_z < 65\text{GeV} \quad (4.3)$$

$$P_T^{missing} < 15.0 \quad (4.4)$$

Cut 4.1 selects the high Q^2 region. Together with 4.2 the phase space for further analysis is defined. As a charged current (CC) event would cause missing P_T, 4.4 is applied to select only NC events. Usually, it is also required that the interaction vertex lies within $z = 0 \pm 35$cm (0 means the average interaction point) to reduce contributions from beam-gas interactions and cosmic muons. But as we are here dealing with Monte Carlo data only which contains no such events, this cut is not necessary.

The sum in 4.3 runs over all particles in the event and should be (due to energy and momentum conservation) $2E_e = 55$GeV. For the case of DIS initial state photon radiation the value of the sum drops and the event is rejected.

After the events are selected, our program for jet analysis selects only jets of interest:

$$5\text{GeV} < P_T^{jet} < 50\text{GeV} \text{(in Breit frame)} \quad (4.5)$$

$$-1.0 < \eta_{lab} < 2.5 \quad (4.6)$$

To ensure that the detected jets lie within the acceptance of the LAr calorimeter, cut 4.6 is applied, where η_{lab} is the pseudorapidity in the laboratory frame. The pseudorapidity of a particle is defined as

$$\eta = -\ln \tan \frac{\theta_{jet}}{2} \quad (4.7)$$

No matter how many jets are found within an event, only the first one is of interest in this study. When this first (as all jets in our arrays are ordered by P_T, "first" means "jet with the largest P_T") jet with all its properties like P_T, ϕ and η is compared within the different levels, one has to make sure that really the correct jets are compared. An example: A simulated event contains two jets at hadron level, one with a P_T of 13 GeV, the other one with 14 GeV. As detectors and reconstruction methods are not perfect, the 14 GeV jet might be identified as a 12 GeV jet at detector level ($\equiv$ reconstructed level) and the 13 GeV correctly as a 13 GeV jet. So the "first" jets are no longer the same and might point in different directions.

To make sure that the comparison between the levels is always made correctly, a matching procedure was done as described in the following chapter.

5. THE MATCHING PROCEDURE

As the hadron level is supposed to be the one being "closest to physical reality", the matching always starts with the hardest hadron jet. To find a matching jet on the other level, the following procedure is done (as an example, the reconstructed level is taken as comparison level):

1. Define a maximal matching distance ΔR (as in 3.3) that is fixed during the whole analysis and which will be called ΔR_{max} (which value is optimal will be discussed later in 6.1).

2. **Check**, if the distance between $Jet[0]_{rec}$ (Indexing starts with the hardest jet) and $Jet[0]_{had}$ is smaller than ΔR_{max}.
3. If **yes**, then the two hardest jets are matched.
4. If **no**, **check** if the distance between $Jet[1]_{rec}$ and $Jet[0]_{had}$ is smaller than ΔR_{max}.
5. If **also** $Jet[2]_{rec}$ is close enough and even closer than $Jet[1]_{rec}$, chose $Jet[2]_{rec}$ to compare with.
6. Else, compare with $Jet[1]_{rec}$.
7. If all of the first three jets on the reconstructed level have a distance larger than ΔR_{max}, the event is rejected, and a marker for the jet matching efficiency is set (see 6.1).

This matching procedure differs slightly from the one used in (1), where the jet with the minimal distance to $Jet[0]_{had}$ is chosen. Whereas here, we have defined 2 jets as matching, when they are both the hardest jets and within ΔR_{max}. So only if there is a mismatch between the hardest, the next closest will be used. A significant deviation between both procedures is not expected as the matching jet is in most cases the first one, for which we have to differences in our algorithms.

6. RESULTS

If nothing else is said, all following results were obtained by using the MC generator DJANGOH. Any significant differences between the results from DJANGOH and RAPGAP, they will be explained in chapter 6.2.3.

6.1 Matching efficiency

6.1.1 *An optimal* ΔR_{max}

As described in the matching procedure, an optimal ΔR_{max} has to be found. If it is chosen too large, chances increase to make a matching between two different jets which are just close-by by accident. In contrast, a ΔR_{max} chosen too small will result in low statistics caused by a smaller matching efficiency and may depend on a biased jet sample (for example jets with large electromagnetic fraction).

All further analysis concerning the matching efficiency will be made between hadron and reconstructed level. To get a first impression of what a good ΔR_{max} could be, one can have a look at figure 3. It shows the distance distribution of two matching jets. In this run ΔR_{max} was set to 1.0. A change in the jet radius R_0 from 0.7 to 1.0 (not to be confused with ΔR_{max}) has no significant influence.

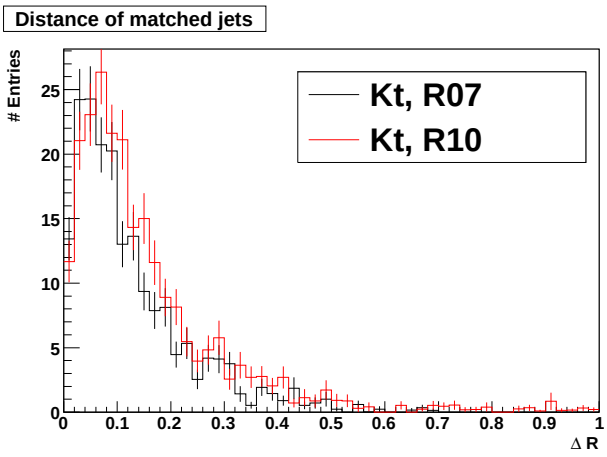

FIG 3. *Distance distribution of two matched jets*

To make a decision for ΔR_{max}, a concrete analysis of the matching efficiency has to be done. The matching efficiency is defined as the fraction of matching trials, where a jet within ΔR_{max} was found on the other level. To calculate it, the marker for rejected events described in chapter 5 was used.

The matching efficiency depending on ΔR_{max} is shown in figure 4 as an example for the anti-k_T algorithm.

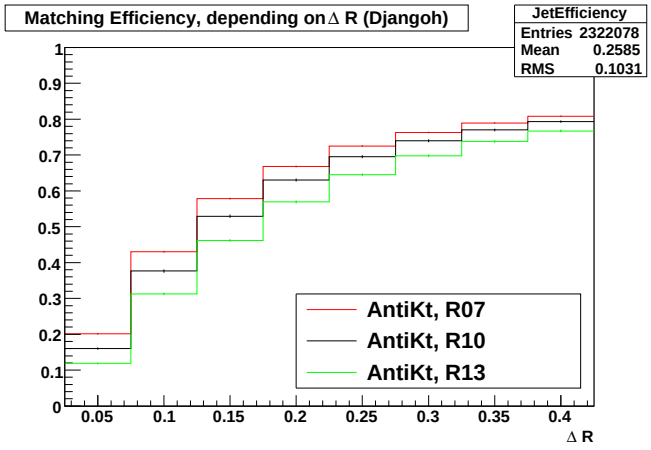

FIG 4. *Matching efficiency depending on* ΔR

An overview for all algorithms is given in figure 11. The results for the other algorithms differ only slightly, except for SISCone, which has in general a lower matching efficiency. If one wants a winner in the competition for efficiency, anti-k_T can be pointed

out.

For smaller jet radii the efficiency is in general better. The idea for this is pretty clear: The more a jet can spread geometrically, the more its boundaries can be smeared out with a given uncertainty.

With the results from this ΔR_{max} dependance of the efficiency, I decided to set it to $\Delta R_{max} = 0.2$ as efficiency does not increase that strongly anymore above 0.2 and in order to still observe sensitivity to R_0.

6.1.2 ***Matching efficiency dependences***

It is checked, how the matching efficiency varies depending on different observables like P_T, η, ϕ, the squared invariant mass of the two hardest jets M_{12}^2 and Q^2. One remark about M_{12}^2: For this observable, no matching procedure was followed. The only demand was that there are two jets on both levels. If so, they were compared according to their index in the jet array, namely in the order of P_T.

- P_T: Efficiency increases for higher P_T which will be analyzed later in chapter 6.3 (see figure 12). It is an intuitive property of QCD that a jet with higher P_T is defined clearer (better collimated) and thereby easier to match in its properties on all levels.
- η: Figure 13 shows the η dependance of the efficiency. It decreases for large $|\eta|$. This effect is caused by inefficiencies in the detector and other effects which are also a reason for the global η cuts.
- ϕ: This is one of the most interesting effects observed during my work (figure 14). Unfortunately it is also the one least understood. It might be an effect of the boost to the Breit frame. Figure 5 shows how a uniform ϕ distribution changes after such a boost.
- M_{12}^2: The effect is similar and comparable to the P_T dependence (figure 15).
- Q^2: Large Q^2 end up with a worse efficiency (figure 16).

One might ask what the reason for the mismatches were. A closer look at the mismatched events showed that in almost all cases the jet which should have been matched (jets with very small ΔR and a P_T close to each other) did not fulfill the necessary cuts for P_T.

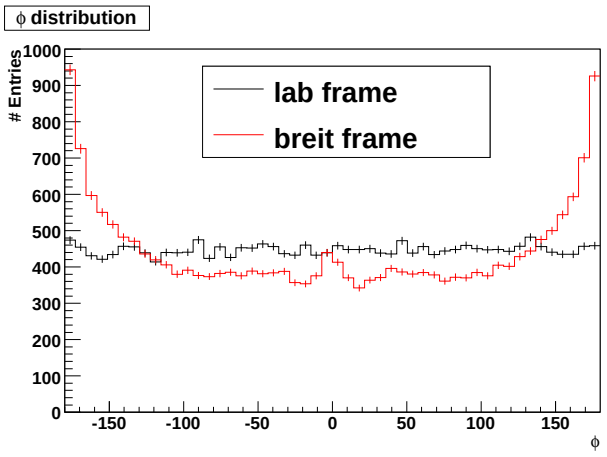

FIG 5. *ϕ distribution in laboratory rest frame and Breit frame, reconstructed level*

6.2 Level Deviations

If one sets the hadronic level as the one being closest to the physical truth, it is important to interpret the data at reconstructed level in that way that the deviations between reconstructed and hadronic level are as small as possible. This is also a claim to our jet algorithm. So we will now first compare the deviations of P_T, η, ϕ and M_{12}^2 on both levels first under a general aspect ("Can we understand them?") and then find a jet algorithm which fits best for each purpose. In the next step, the free parameter of the algorithm - the jet radius R_0 - is checked for its best setting.

What the analyzing program did was the following: The differences of the jet properties were calculated for each event and stored in a two dimensional histogram (an example is given in figure 6 depending on ϕ). Later on the differences will also be analyzed as a function of η, ϕ, M_{12}^2 and Q^2.

The best thing next to a delta function at zero would be a gaussian like distribution around zero. The mean value (ideally zero) and the σ of all those distributions (ideally as small as possible) is then projected on a one dimensional histogram and interpreted.

Later on the same checks will be done for a comparison of hadron with parton level.

6.2.1 ***REC vs. HAD Level***

First, about the deviations in general (all relevant histograms can be found in figures 17 - 28):

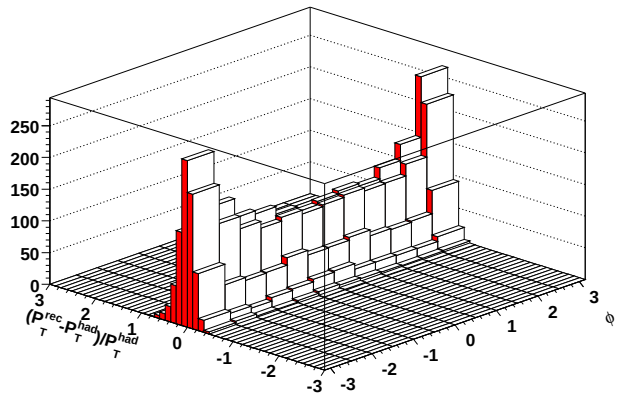

FIG 6. *P_T deviation between reconstructed and hadron level depending on ϕ*

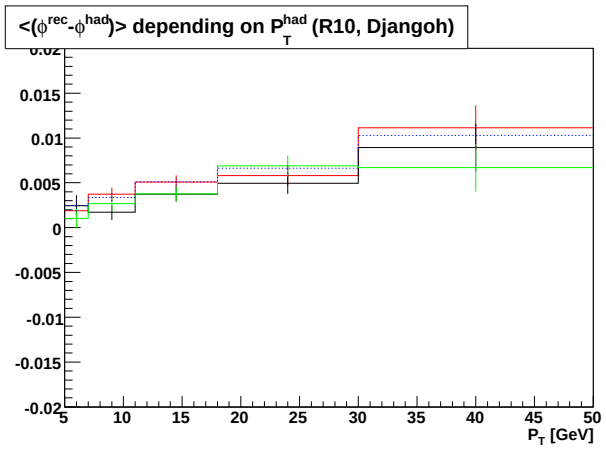

FIG 7. *Shift of ϕ between reconstructed and hadron level*

- ΔP_T: The reconstructed P_T is calculated slightly too high ($R_0 = 1.0$ or $R_0 = 1.3$). The effect increases for small P_T. The bad results for small P_T were already mentioned. For $R_0 = 0.7$, P_T is even too low for $P_T > 10$ GeV. The shape of the distribution is the same but gets shifted to lower values.
- $\Delta\eta$: η is consistently but slightly too small at the reconstructed level.
- $\Delta\phi$: A small deviation of $\Delta\phi$ from zero is observed. An example is given in figure 7. The larger P_T gets, the larger the deviation becomes (the reconstructed data is too large). The origin of this deviation is not clear.
- ΔM_{12}^2: The shape of the M_{12}^2 deviation stays the same but is shifted for different jet radii. This means that for small P_T it is always reconstructed too high and for larger P_T ($> 10GeV$) it is either reconstructed too low ($R_0 = 0.7$, $R_0 = 1.0$) or too high ($R_0 = 1.3$). In all cases, the width of the distributions is much higher than it is for the other observables as there are now two jets involved and no matching is done.

Now the question is: Which algorithm is best for which purpose? First of all, I should mention something about the SISCone. All used jet algorithms except SISCone are of the clustering type. The split-merge procedure used in the SISCone leads to a smaller effective jet radius than used as input for the algorithm. This means: If we use SISCone with $R_0 = 1.0$, it should also be compared to a clustering algorithm at a smaller radius; in our case $R = 0.7$. So the discussion about the optimal R_0 will be skipped for SISCone. Figure 8 is used to visualize this effect. Expected is a peak for the jet area $A = \pi R_0^2$ at the

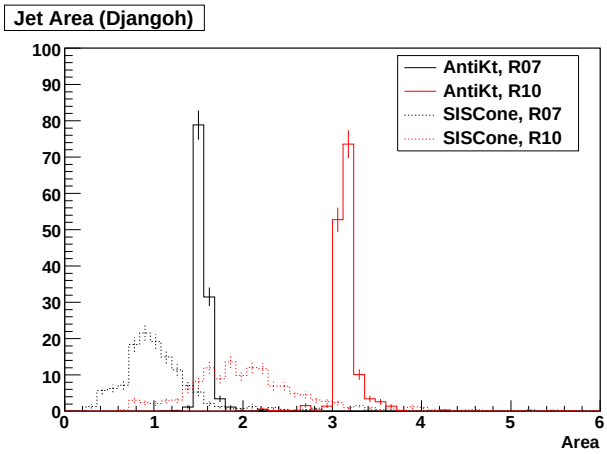

FIG 8. *Comparison of the jet areas*

A belonging to each R_0. One can see that for SISCone, the jet radius given as parameter is not the one obtained from the jet area.

Back to the question which algorithm to use:

- SISCone gives smaller (so in some case "better") values for P_T. But the effect can be explained with the different radius behavior.
- For the ϕ deviations, all algorithms behave similar.
- anti-k_T is closest to zero concerning the η deviations.
- For M_{12}^2 SISCone is again lower (so sometimes

better) due to the radius effect. Anti-k_T slightly worse than the rest concerning the mean value of deviations, but best concerning σ.
- As σ for the M_{12}^2 is very radius sensitive, SISCone is worst in this part.

Now we look for a "good" R_0 for the jet algorithms depending on the demands of the analysis:

- ΔP_T: $R_0 = 1.0$ leads to results closest to zero. Only for small P_T, where it is always given too high on reconstructed level, one might wish a smaller R_0 to decrease the deviations.
- $\Delta\eta$: The higher, the better.
- $\Delta\phi$: Same behavior for all radii.
- ΔM_{12}^2: Exactly the same argument as for P_T.

As mentioned before, these deviations of P_T, η, ϕ and M_{12}^2 were stored in 2D diagrams depending on different observables. Up to now they were only shown as a function of P_T. I will now also show how they depend on the other observables. But as the algorithms themselfes do not differ that much, the results will only be discussed for the anti-k_T algorithm. The corresponding plots are found in figures 37 - 44. For all deviations of observables the most significant dependences are summarized:

- ΔP_T: $\eta > 1.0$ gives the main contribution for the P_T deviations. The effect decreases for a smaller R_0.
- $\Delta\eta$: The main part of the η deviations can be assigned to $\eta > 1.5$ (see figure 37) events and events with a low Q^2 (see figure 43).
- $\Delta\phi$: Again (as for ΔP_T), $\eta > 1.0$ gives main contribution. The higher η, the worse.
- ΔM_{12}^2: Low Q^2 ($Q^2 < 600\text{GeV}^2$) events cause largest deviations. The larger deviations can be explained by the lower P_T of jets in low Q^2 events.

6.2.2 ***HAD vs. PAR Level*** As there are now no more detector effects, one expects that the deviations between hadron and parton level are small. And so they were. In general, the same analysis as for reconstructed and hadron level was done again. Here, a matching between the hardest jet at hadron level and one of the three hardest jets on parton level was done. But we will focus only on the first part, the P_T dependence of the deviations. First about the deviations in general (corresponding plots can be found in figures 45 - 56):

- ΔP_T: Small P_T (< 10 GeV) jets have too little P_T at hadron level. The rest is very close to zero.
- $\Delta\eta$: Independent of P_T and η, but constantly about 4-5% too high on hadron level.
- $\Delta\phi$: No deviation observed.
- ΔM_{12}^2: Very interesting observations are made. The smaller P_T gets, the more drastically M_{12}^2 gets too small on hadron level. This effect depends very much on the jet finder radius. But all this holds only for the clustering algorithms. For SISCone, M_{12}^2 is always too high on hadron level. And the best radius setting for SISCone is the worst for the clustering ones.

The advantages of each jet algorithm:

- All clustering algorithms behave similar. k_T is slightly better in the P_T matching sector. SISCone - as already explained - behaves in the way that its actual radius in smaller than the one from the setting.
- An exception is the M_{12}^2 comparison. SISCone is quite good for $R_0 = 0.7$. About the same quality one can reach with the clustering algorithms for $R_0 = 1.3$. To make a decision, one has to take a look at σ: The SISCone distribution is much broader than those of the clustering algorithms.

The question concerning the optimal jet finder radius R_0 can be answered the following:

- ΔP_T: $R_0 = 1.0$ leads to best results.
- $\Delta\eta$: A R_0 as low as possible is preferred.
- $\Delta\phi$: No preference.
- ΔM_{12}^2: If one neglects low P_T jets, $R_0 = 1.0$ fits best, but only for the clustering algorithms (see discussion above).

6.2.3 ***Differences Between the MC Sets*** We should now have a look at the differences of the two used MC generators, DJANGOH and RAPGAP. Of course, there are a lot of differences of the kind "here a little more, here a litte less". I will focus only on systematic differences between both.

- The problem with η being constantly too low on reconstructed level (compared to hadron level) is "mirrored" for RAPGAP. Now η is consistently too high, the absolute value of the devi-

ation stays almost the same $(0.005 - 0.01)$.

- Remember that ϕ deviated in that way that it was too large on reconstructed level. The effect became more pronounced for higher P_T. Now for RAPGAP, the effect is smaller and even disappears for $R_0 = 1.3$.
- For the M_{12}^2 comparison between hadron and parton level, SISCone and the clustering algorithms behave more similar. Especially σ is no longer so horrible for the SISCone with RAPGAP.

6.3 Comparing DIS with Photoproduction

To compare the results of this study with the photoproduction study done by C. Mellein (1), I have to mention first that in my case many significant effects were obtained for either low P_T or high Q^2. Both cases cannot be compared to (1) ($Q^2 < 4\text{GeV}^2$ and jet $P_T > 25$ GeV). Also, one has to keep in mind that in (1) no boost was made. So we have to focus on something independent. This is a comparison of our results:

- We both observed the ϕ shift mystery. As it was weaker with the RAPGAP dataset, one could compare PYTHIA which was used in (1) with DJANGOH (same behavior) and RAPGAP.
- We both observed worse efficiencies and larger deviations at high η, which then has to be blamed on worse detector efficiency and other effects in this region.
- Whereas in (1) the matching efficiency had no ϕ dependence, I observed a strong variation with ϕ. I already mentioned that there might be a connection to the boost. This idea now seems to be supported.
- Both of us came to the conclusion that the k_T and anti-k_T are always slightly better than the rest.

One last remark about the matching efficiency. Whereas in this study we get values between 60 and 70%, in (1) values of more than 90% are quite common. In chapter 6.1.2 the observation was made that the efficiency increased for higher P_T. This is nicely shown in figure 9 or more detailed in figures 57 and 58. As I studied high Q^2 events and the efficiency decreased for increasing Q^2, this presumably provides the other big contribution for the lower efficiency.

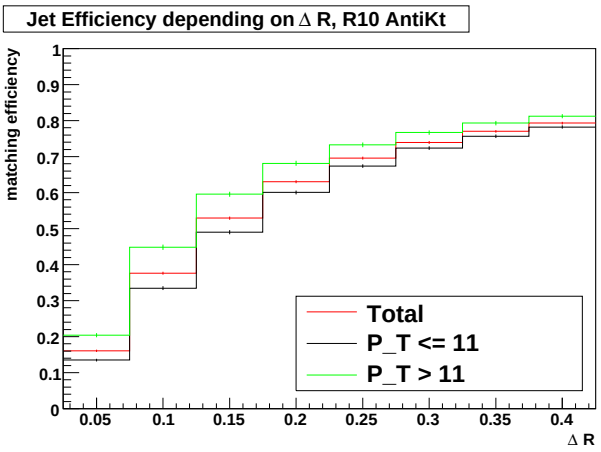

FIG 9. P_T *dependance of the matching efficiency*

6.4 Final Conclusion and Outlook

Unfortunately, there is no conclusion saying "You get perfect results if you just use the algorithm x with the radius y. What can be said in general: SISCone is very hard to compare with the others and has in general a worse behavior, especially concerning σ. SISCone's advantages to other algorithms are usually no real ones and originate only from the different radius handling. Cambridge/Aachen has no special strengths, but anti-k_T and k_T do: anti-k_T has in general the best (smallest) σ and the best matching efficiency, which is also somehow related. The k_T behaves similar but its strengths are not as distinct. Concerning runtime, SISCone was the slowest of the algorithms.

The jet radius should be treated a little more specifically. A small jet radius ($R_0 = 0.7$) is most effective in providing a high matching efficiency. When the distribution is not consistently shifted away from zero (as for P_T), $R_0 = 1.0$ is always the best choice for the average of all jets. If one wants to analyze a specific range for P_T, one of the "extreme" radii (0.7 and 1.3) would be the preferred choice. For $\Delta\eta$, which is constant, $R_0 = 1.0$ is again the best choice as each extreme which leads to better results at one level makes it worse on the other level (see level comparison).

To sum up: If anti-k_T is not much slower than k_T, I would prefer it. For an overall analysis, $R_0 = 1.0$ is the best compromise.

An extension of the deviation dependency analysis to other algorithms than anti-k_T seems not to be

necessary.

The ϕ shift mystery needs clarification. The fact that the deviation does not appear using a RAPGAP Monte Carlo with $R_0 = 1.3$ (see figure 55) even though it is still observable for the same setting with DJANGOH (figure 49) might give a first hint.

ACKNOWLEDGEMENTS

During my stay at DESY I was able to live a purely scientific life: Working up to 24 hours a day, seven days a week, getting filled with coffee and tea all the time, being served cake and chocolate, having the nicest people and best scientists around me. Without those circumstances this work would not have been possible to that extend. For this I would like to thank all members of my working group: Adil, Aziz, Clemens, Günter, Roman and Zuzana. You are the best! Many thanks also to Andrea Schrader and Joachim Meyer for organizing the whole summer student program, the lecturers for feeding us with knowledge and all the summer students for the entertainment. What may not be forgotten are my friends at home giving me mental support, when I got clobbered over the head with work. And last but not least: The countless amount of cats giving me inspiration, and the snack machine providing fresh sugar and carbohydrates.

FIG 10. *My Team*

REFERENCES

[1] C. MELLEIN (2008).
Comparing different Jet-Algorithms for photoproduction Monte Carlo data
DESY Summer School 2008 Report
[2] K. CHARCHULA, G. A. SCHULER AND H. SPIESBERGER.
DJANGOH V1.4
Comput. Phys. Commun. 81 (1994) 381
[3] H. JUNG.
RAPGAP V2.08
Comput. Phys. Commun. 86 (1995) 147
[4] S. CATANI ET AL..
Longitudinally-invariant k_t-clustering algorithms for hadronhadron collisions
Nucl. Phys. B. 406 (1993) 187
[5] Y. L. DOKSHITZER, G. D. LEDER, S. MORETTI AND B. R. WEBBER.
Better Jet Clustering Algorithms
JHEP 9708, 001 (1997) [hep-ph/9707323]
[6] MATTEO CACCIARI AND GAVIN P. SALAM.
The anti-kt jet clustering algorithm
arXiv:0802.1189v1 [hep-ph]
[7] GAVIN P. SALAM AND GREGORY SOYEZ.
A practical seedless infrared-safe cone jet algorithm
JHEP05(2007)086

APPENDIX A: MATCHING EFFICIENCY PLOTS

- **Figure 11:** Jet efficiency depending on the maximal jet distance ΔR_{max}
- **Figure 12:** Jet efficiency depending on P_T
- **Figure 13:** Jet efficiency depending on η
- **Figure 14:** Jet efficiency depending on ϕ
- **Figure 15:** Jet efficiency depending on M_{12}^2
- **Figure 16:** Jet efficiency depending on Q^2

APPENDIX B: DEVIATIONS BETWEEN RECONSTRUCTION AND HADRON LEVEL, COMPARING ALGORITHMS

- **Figure 17:** Deviations of P_T, η and ϕ between reconstructed and hadron level, DJANGOH data, $R_0 = 0.7$
- **Figure 18:** Deviation of M_{12}^2 between reconstructed and hadron level, DJANGOH data, $R_0 = 0.7$
- **Figure 19:** Deviations of P_T, η and ϕ between reconstructed and hadron level, DJANGOH data, $R_0 = 1.0$
- **Figure 20:** Deviation of M_{12}^2 between reconstructed and hadron level, DJANGOH data, $R_0 = 1.0$

- **Figure 21:** Deviations of P_T, η and ϕ between reconstructed and hadron level, DJANGOH data, $R_0 = 1.3$
- **Figure 22:** Deviation of M_{12}^2 between reconstructed and hadron level, DJANGOH data, $R_0 = 1.3$
- **Figure 23:** Deviations of P_T, η and ϕ between reconstructed and hadron level, RAPGAP data, $R_0 = 0.7$
- **Figure 24:** Deviation of M_{12}^2 between reconstructed and hadron level, RAPGAP data, $R_0 = 0.7$
- **Figure 25:** Deviations of P_T, η and ϕ between reconstructed and hadron level, RAPGAP data, $R_0 = 1.0$
- **Figure 26:** Deviation of M_{12}^2 between reconstructed and hadron level, RAPGAP data, $R_0 = 1.0$
- **Figure 27:** Deviations of P_T, η and ϕ between reconstructed and hadron level, RAPGAP data, $R_0 = 1.3$
- **Figure 28:** Deviation of M_{12}^2 between reconstructed and hadron level, RAPGAP data, $R_0 = 1.3$

APPENDIX C: DEVIATIONS BETWEEN RECONSTRUCTION AND HADRON LEVEL DEPENDING ON P_T, COMPARING RADII

- **Figure 29:** Deviations of P_T, η and ϕ between reconstructed and hadron level, DJANGOH data, Anti-k_T
- **Figure 30:** Deviation of M_{12}^2 between reconstructed and hadron level, DJANGOH data, Anti-k_T
- **Figure 31:** Deviations of P_T, η and ϕ between reconstructed and hadron level, DJANGOH data, Cambridge/Aachen
- **Figure 32:** Deviation of M_{12}^2 between reconstructed and hadron level, DJANGOH data, Cambridge/Aachen
- **Figure 33:** Deviations of P_T, η and ϕ between reconstructed and hadron level, DJANGOH data, k_T
- **Figure 34:** Deviation of M_{12}^2 between reconstructed and hadron level, DJANGOH data, k_T
- **Figure 35:** Deviations of P_T, η and ϕ between reconstructed and hadron level, DJANGOH data, SISCone
- **Figure 36:** Deviation of M_{12}^2 between reconstructed and hadron level, DJANGOH data, SISCone

APPENDIX D: DEVIATIONS BETWEEN RECONSTRUCTION AND HADRON LEVEL DEPENDING ON ϕ, η, M_{12}^2 AND Q^2, COMPARING RADII

- **Figure 37:** Deviations of P_T, η and ϕ between reconstructed and hadron level, DJANGOH data, η dependance
- **Figure 38:** Deviation of M_{12}^2 between reconstructed and hadron level, DJANGOH data, η dependance
- **Figure 39:** Deviations of P_T, η and ϕ between reconstructed and hadron level, DJANGOH data, ϕ dependance
- **Figure 40:** Deviation of M_{12}^2 between reconstructed and hadron level, DJANGOH data, ϕ dependance
- **Figure 41:** Deviations of P_T, η and ϕ between reconstructed and hadron level, DJANGOH data, M_{12}^2 dependance
- **Figure 42:** Deviation of M_{12}^2 between reconstructed and hadron level, DJANGOH data, M_{12}^2 dependance
- **Figure 43:** Deviations of P_T, η and ϕ between reconstructed and hadron level, DJANGOH data, Q^2 dependance
- **Figure 44:** Deviation of M_{12}^2 between reconstructed and hadron level, DJANGOH data, Q^2 dependance

APPENDIX E: DEVIATIONS BETWEEN HADRON AND PARTON LEVEL, COMPARING ALGORITHMS

- **Figure 45:** Deviations of P_T, η and ϕ between parton and hadron level, DJANGOH data, $R_0 = 0.7$
- **Figure 46:** Deviation of M_{12}^2 between parton and hadron level, DJANGOH data, $R_0 = 0.7$
- **Figure 47:** Deviations of P_T, η and ϕ between parton and hadron level, DJANGOH data, $R_0 = 1.0$
- **Figure 48:** Deviation of M_{12}^2 between parton and hadron level, DJANGOH data, $R_0 = 1.0$
- **Figure 49:** Deviations of P_T, η and ϕ between parton and hadron level, DJANGOH

data, $R_0 = 1.3$

- **Figure 50:** Deviation of M_{12}^2 between parton and hadron level, DJANGOH data, $R_0 = 1.3$
- **Figure 51:** Deviations of P_T, η and ϕ between parton and hadron level, RAPGAP data, $R_0 = 0.7$
- **Figure 52:** Deviation of M_{12}^2 between parton and hadron level, RAPGAP data, $R_0 = 0.7$
- **Figure 53:** Deviations of P_T, η and ϕ between parton and hadron level, RAPGAP data, $R_0 = 1.0$
- **Figure 54:** Deviation of M_{12}^2 between parton and hadron level, RAPGAP data, $R_0 = 1.0$
- **Figure 55:** Deviations of P_T, η and ϕ between parton and hadron level, RAPGAP data, $R_0 = 1.3$
- **Figure 56:** Deviation of M_{12}^2 between parton and hadron level, RAPGAP data, $R_0 = 1.3$

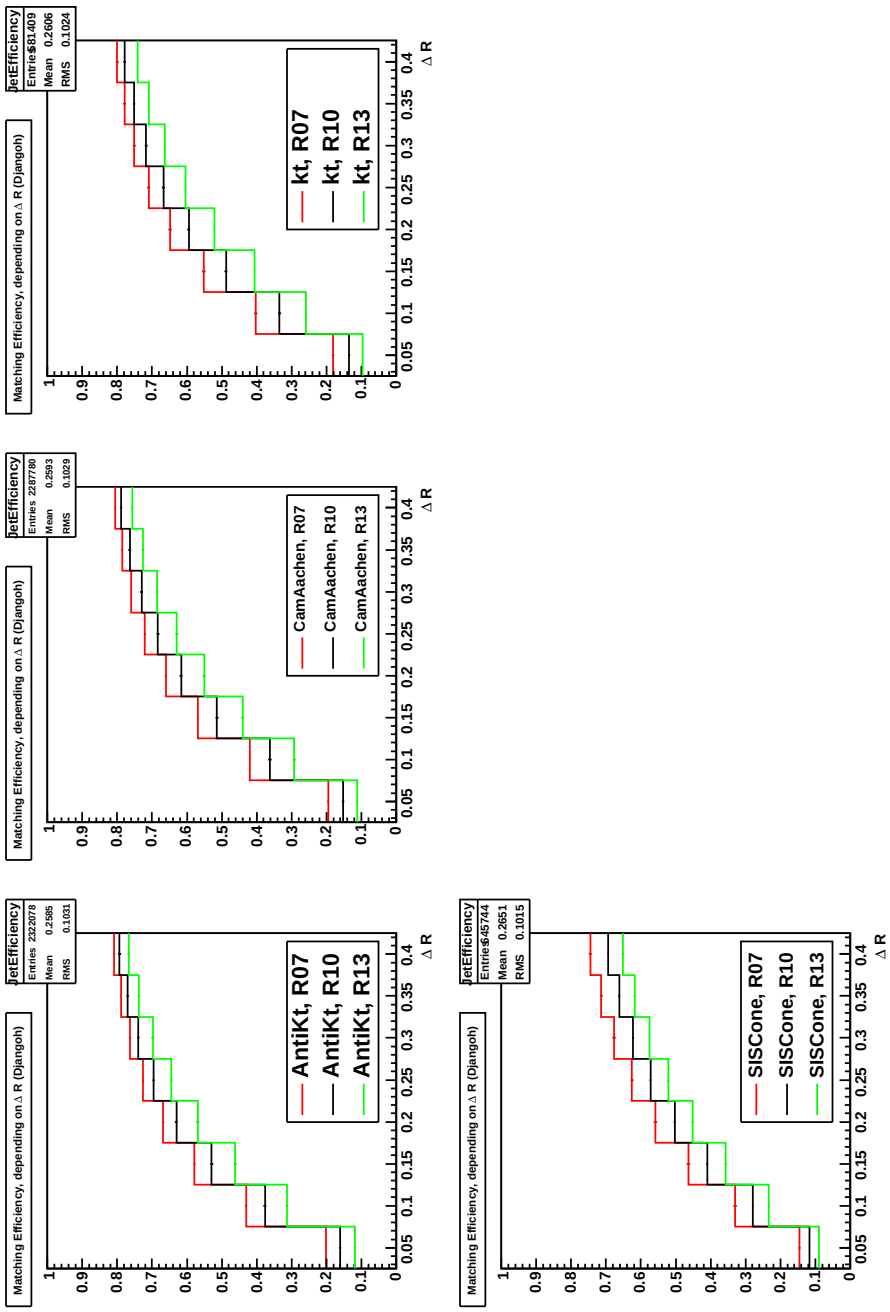

FIG 11. *Jet efficiency depending on the maximal jet distance* ΔR_{max}

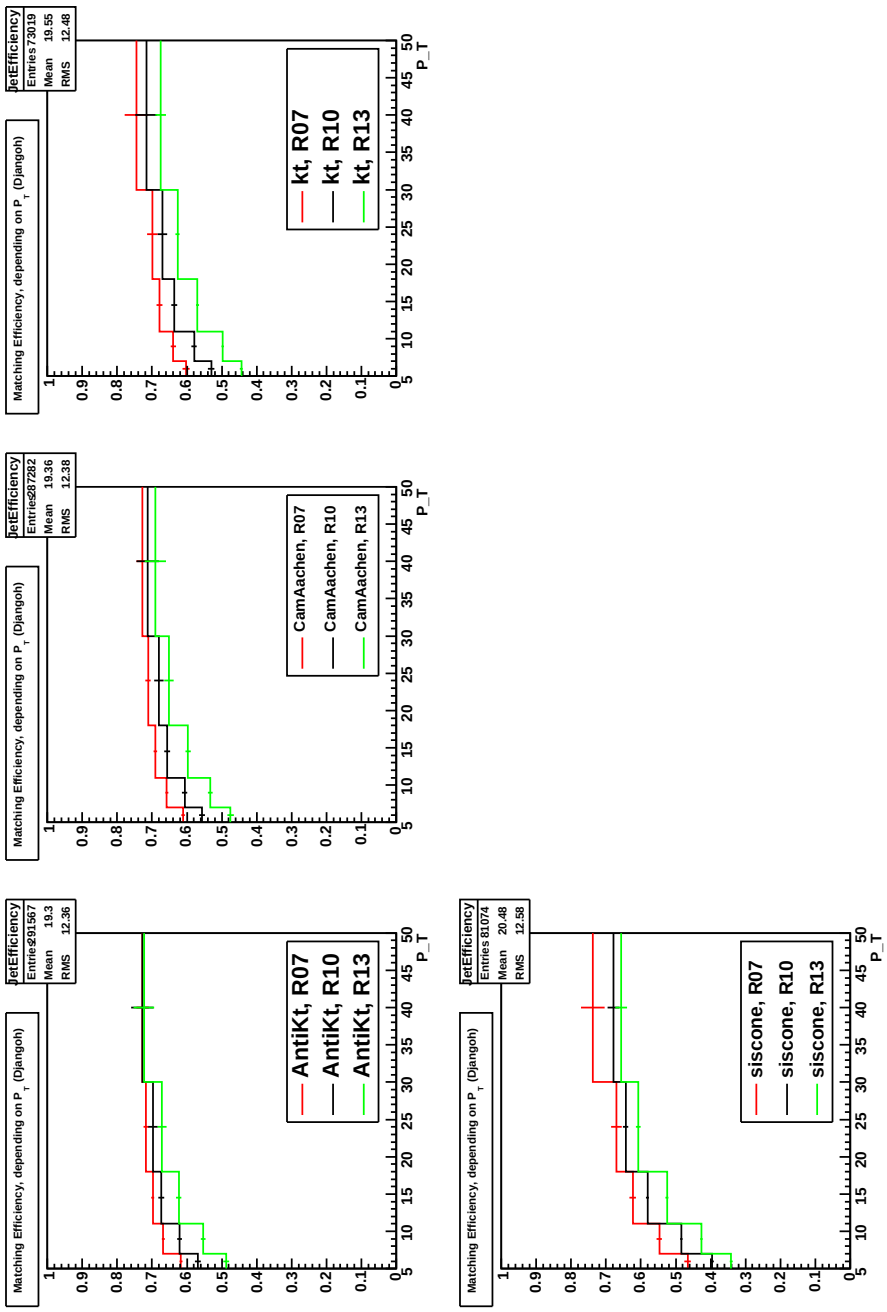

FIG 12. *Jet efficiency depending on P_T*

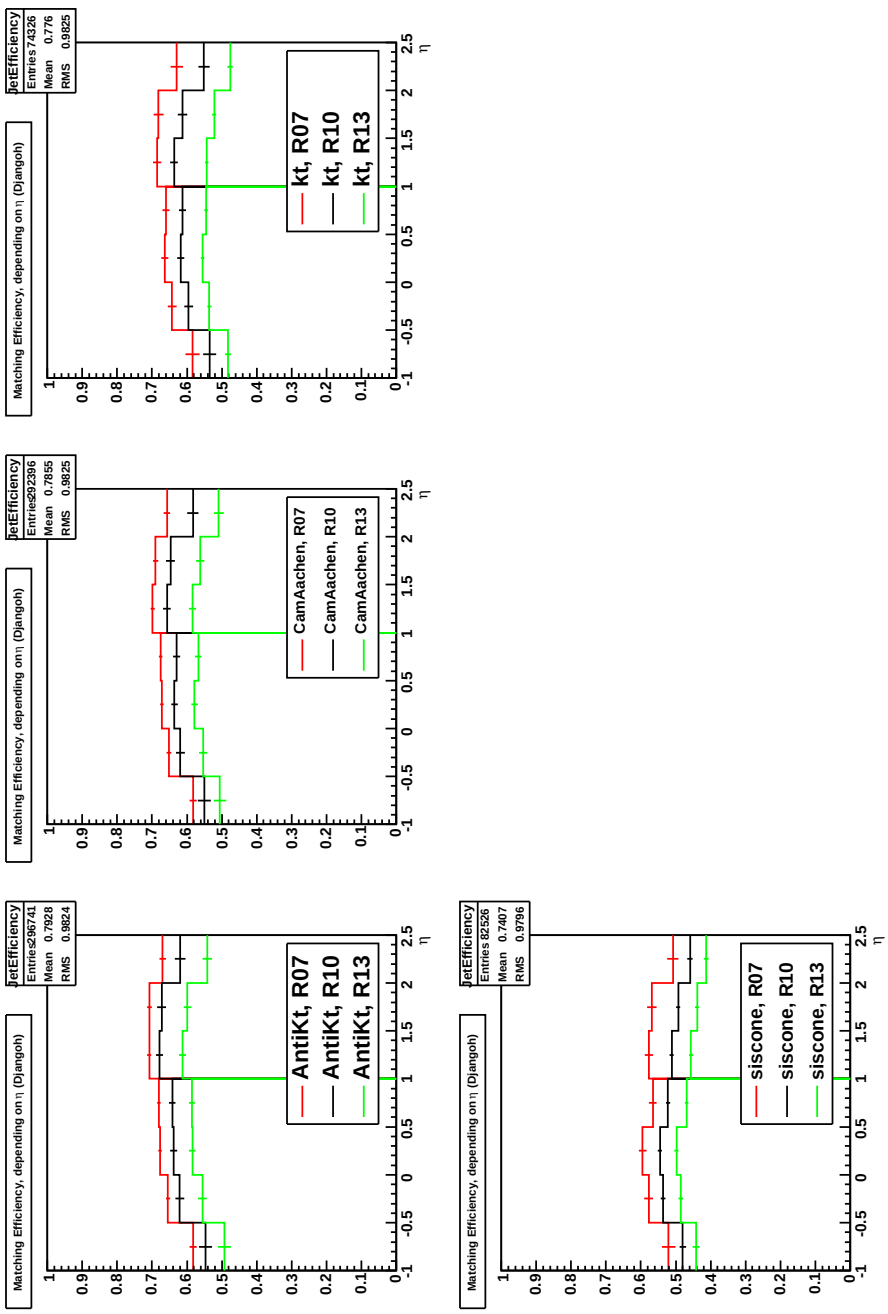

FIG 13. *Jet efficiency depending on* η

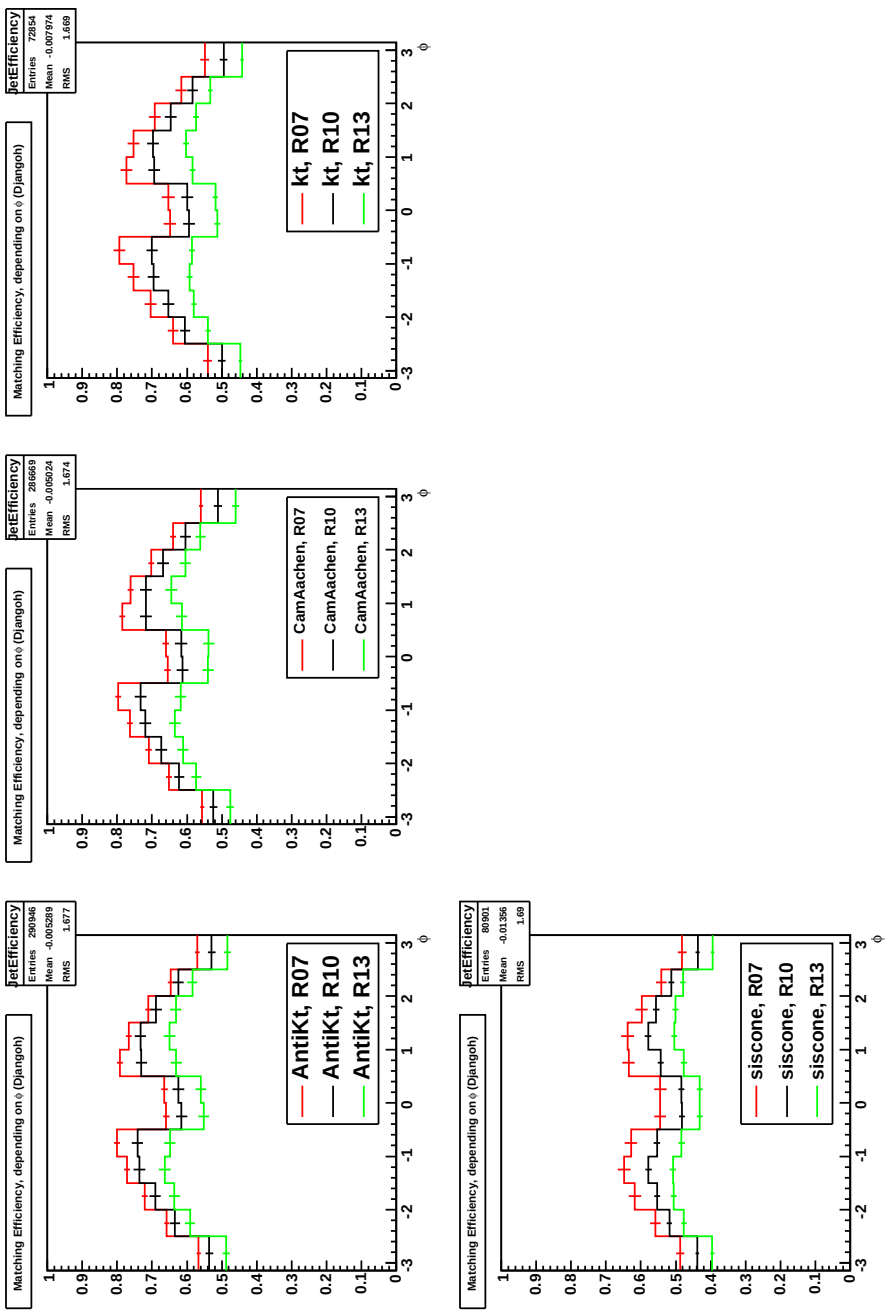

FIG 14. *Jet efficiency depending on* ϕ

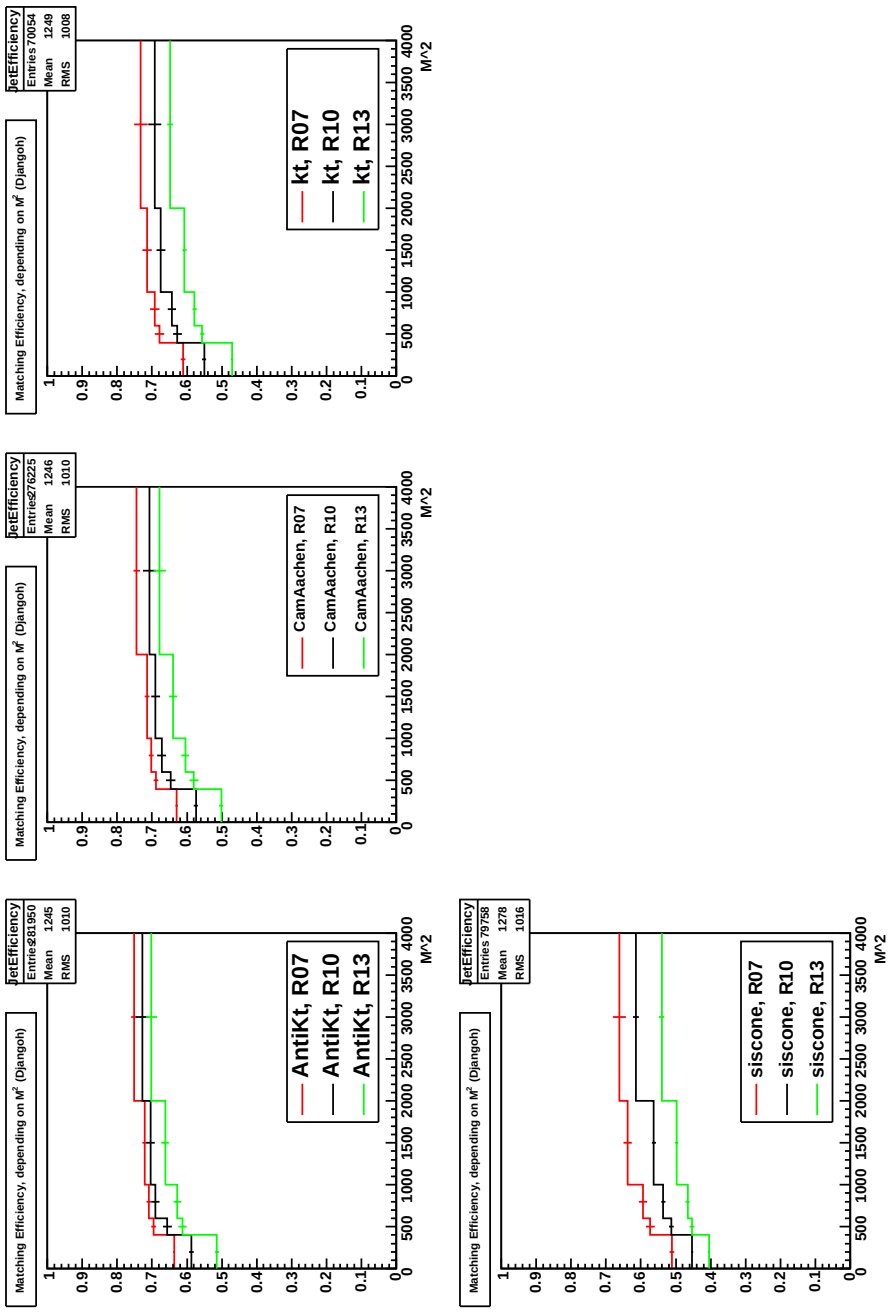

FIG 15. *Jet efficiency depending on* M_{12}^2

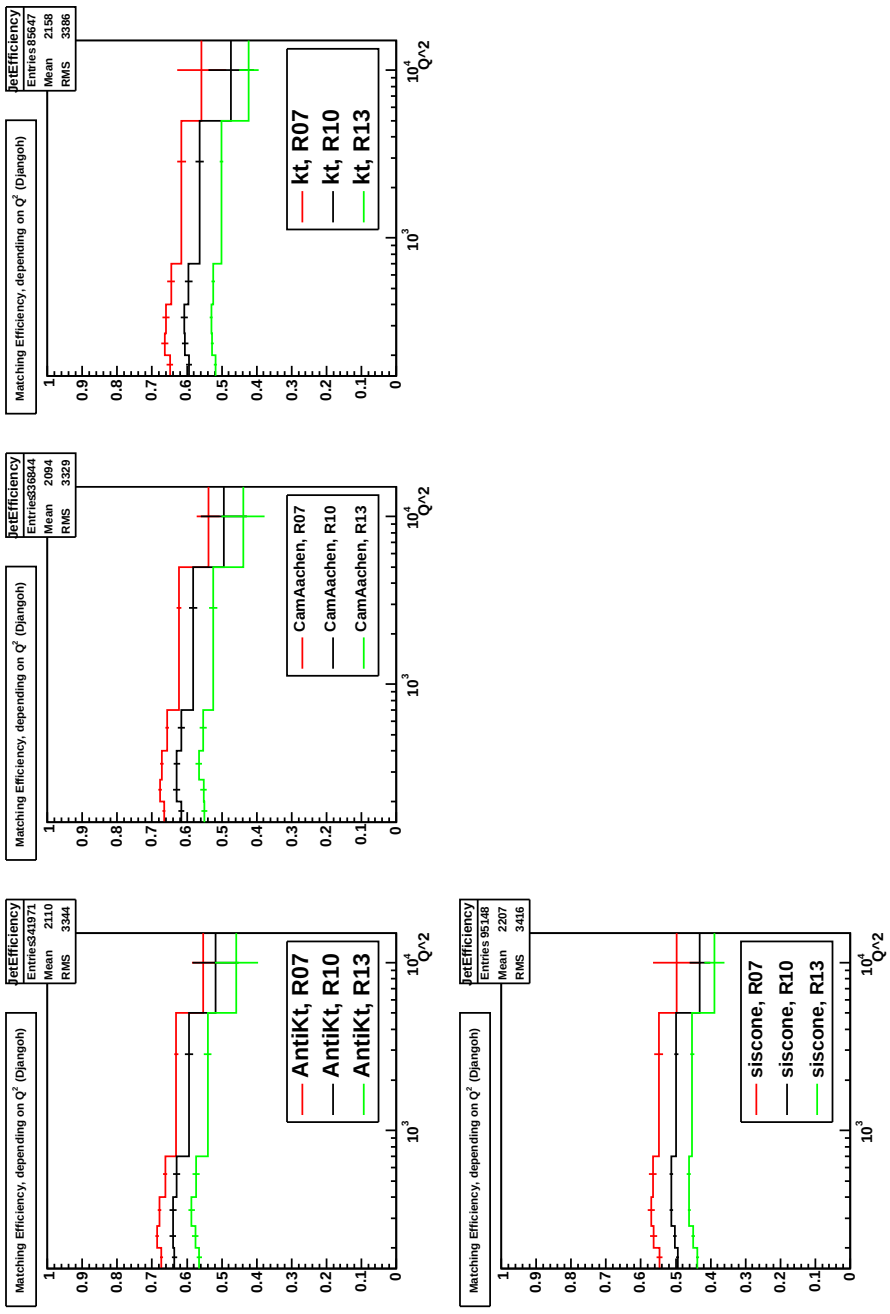

FIG 16. *Jet efficiency depending on* Q^2

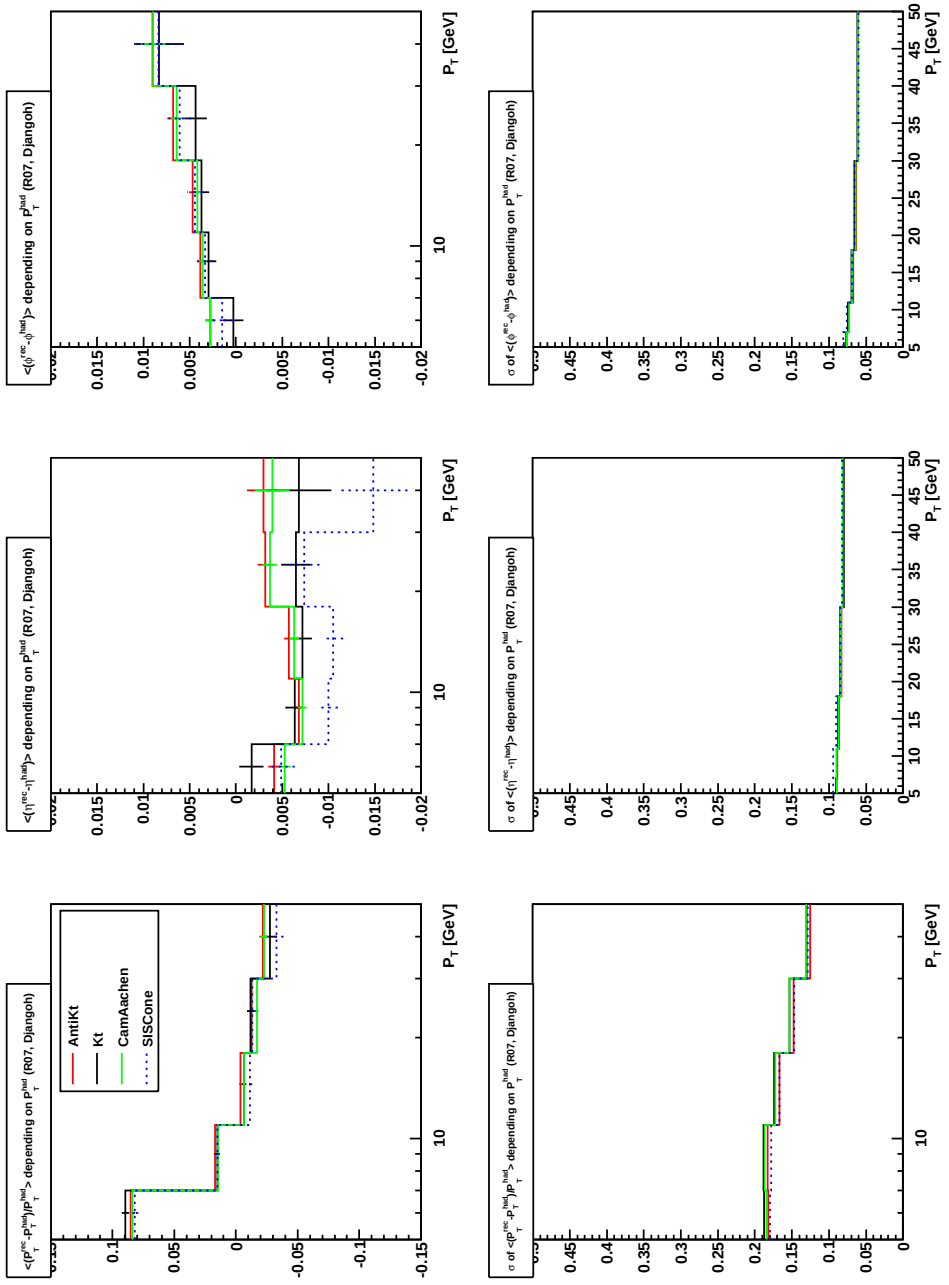

FIG 17. *Deviations of P_T, η and ϕ between reconstructed and hadron level, DJANGOH data, $R_0 = 0.7$*

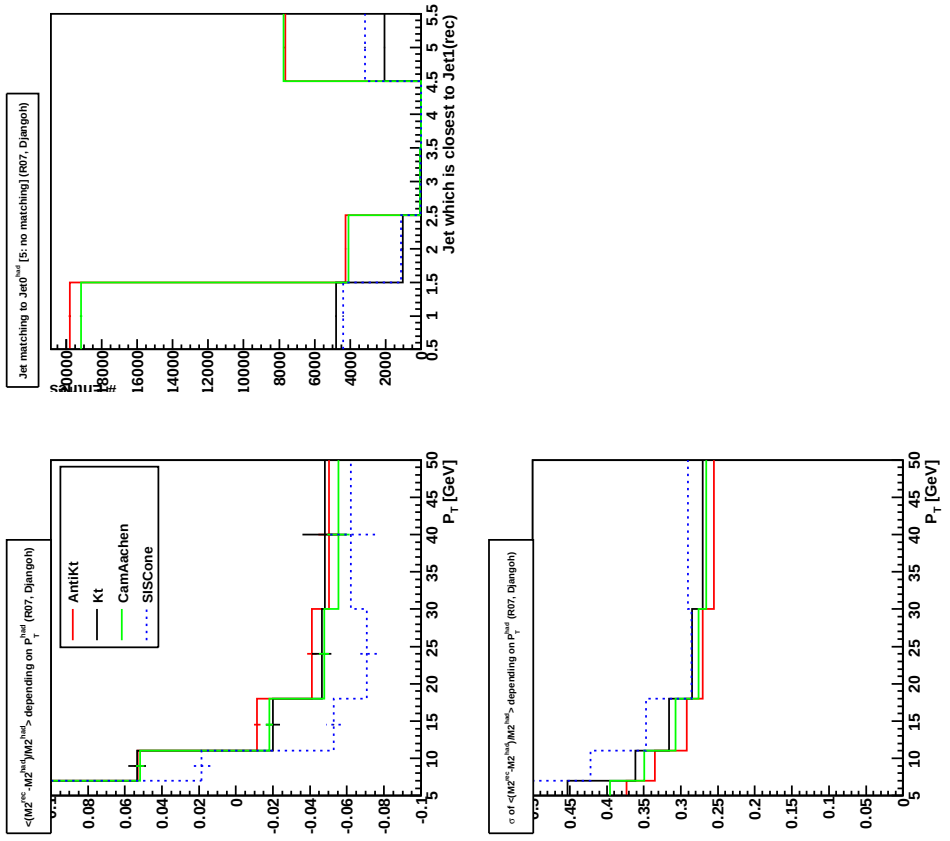

FIG 18. *Deviation of M_{12}^2 between reconstructed and hadron level, DJANGOH data, $R_0 = 0.7$*

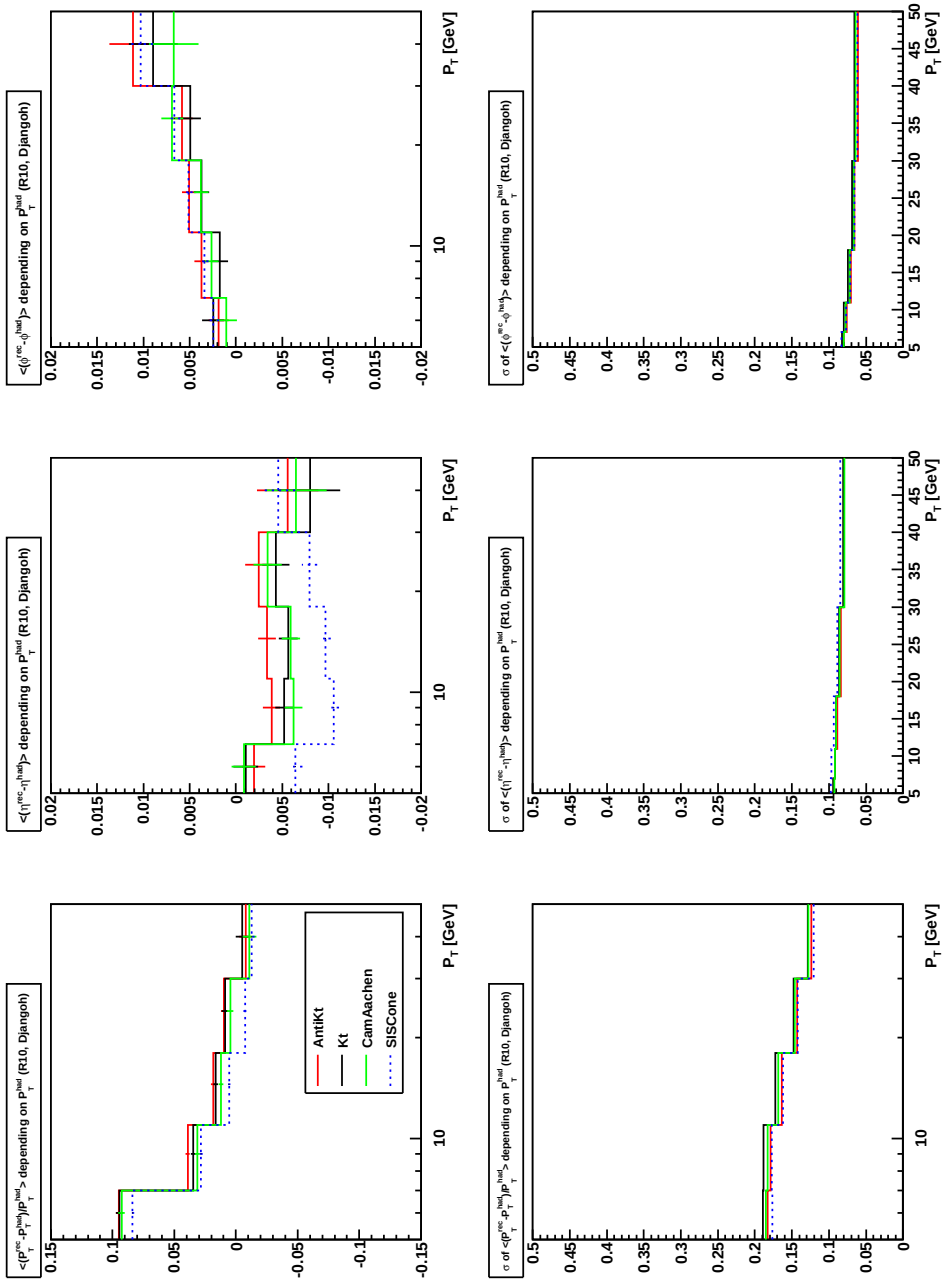

FIG 19. *Deviations of P_T, η and ϕ between reconstructed and hadron level, DJANGOH data, $R_0 = 1.0$*

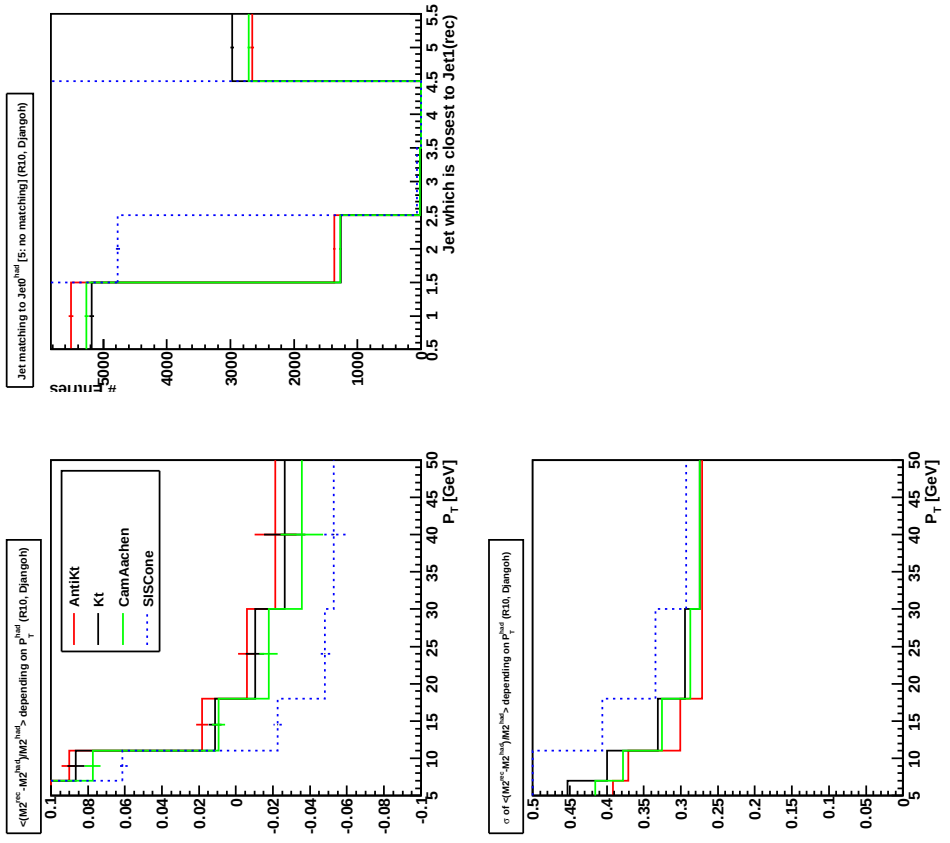

FIG 20. *Deviation of M_{12}^2 between reconstructed and hadron level, DJANGOH data, $R_0 = 1.0$*

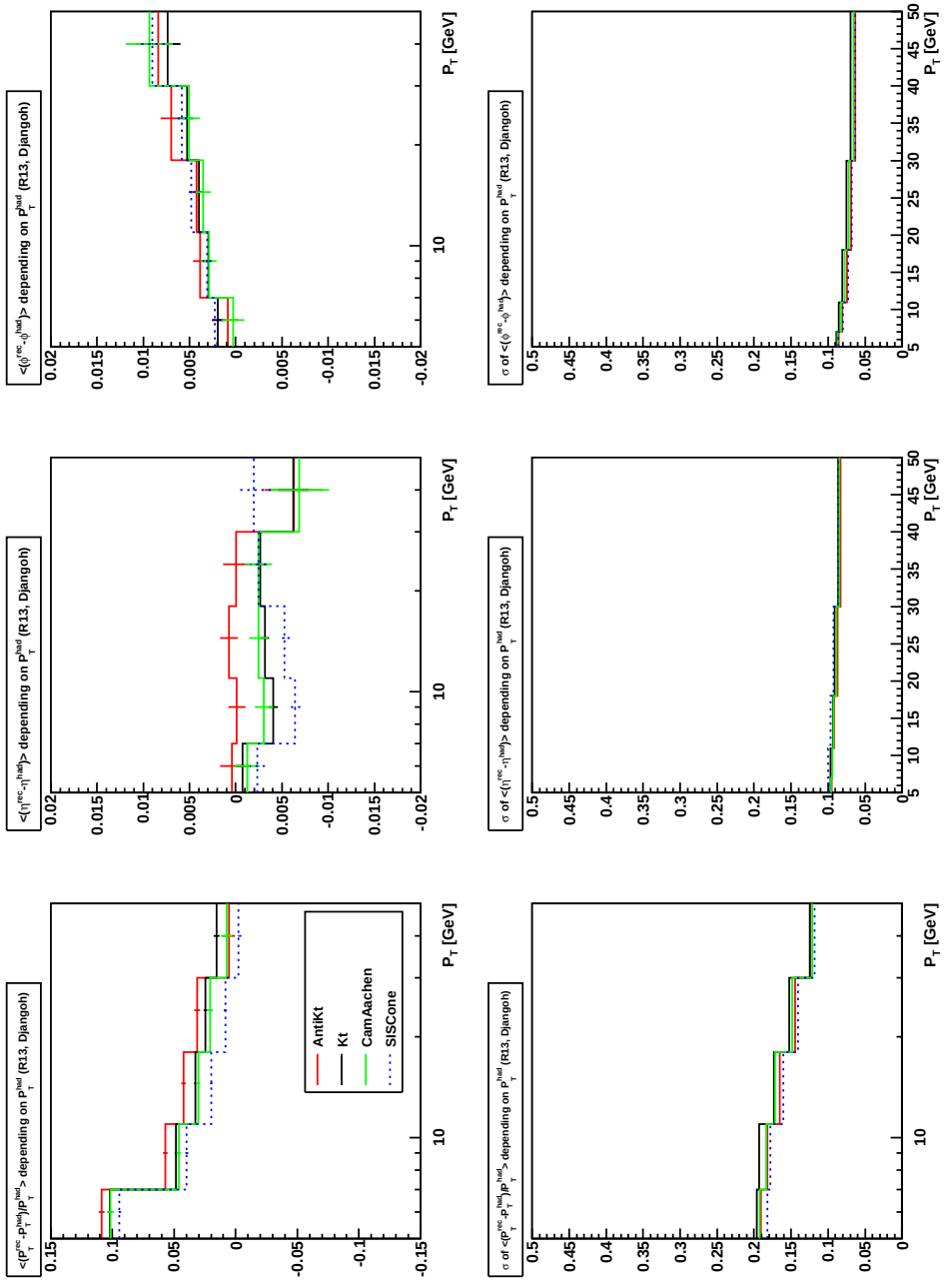

FIG 21. *Deviations of P_T, η and ϕ between reconstructed and hadron level, DJANGOH data, $R_0 = 1.3$*

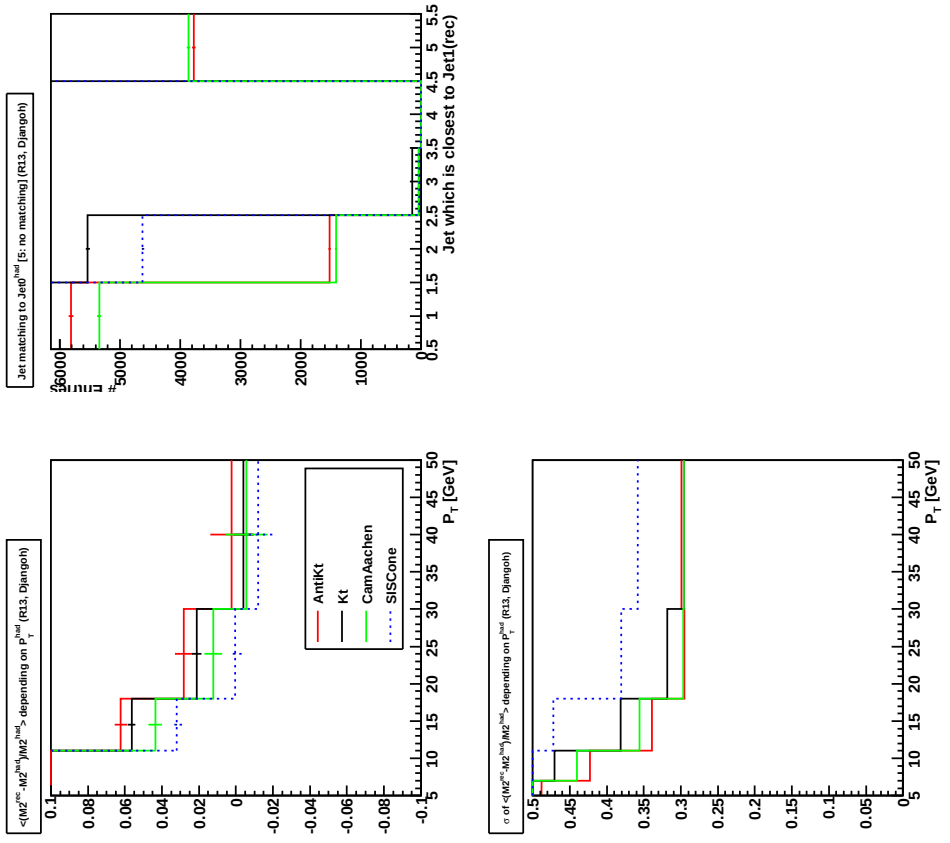

FIG 22. *Deviation of M_{12}^2 between reconstructed and hadron level, DJANGOH data, $R_0 = 1.3$*

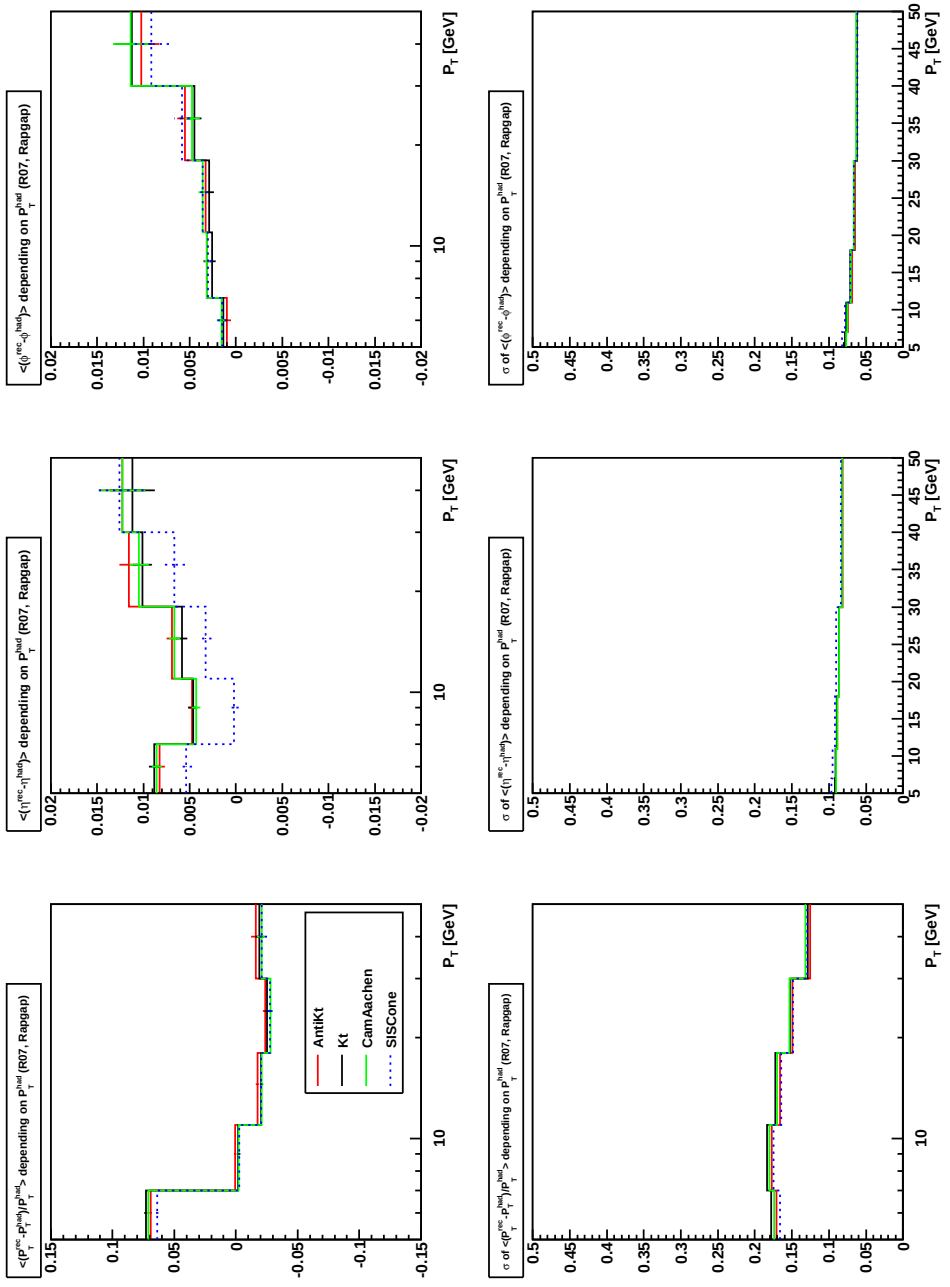

FIG 23. *Deviations of P_T, η and ϕ between reconstructed and hadron level, RAPGAP data, $R_0 = 0.7$*

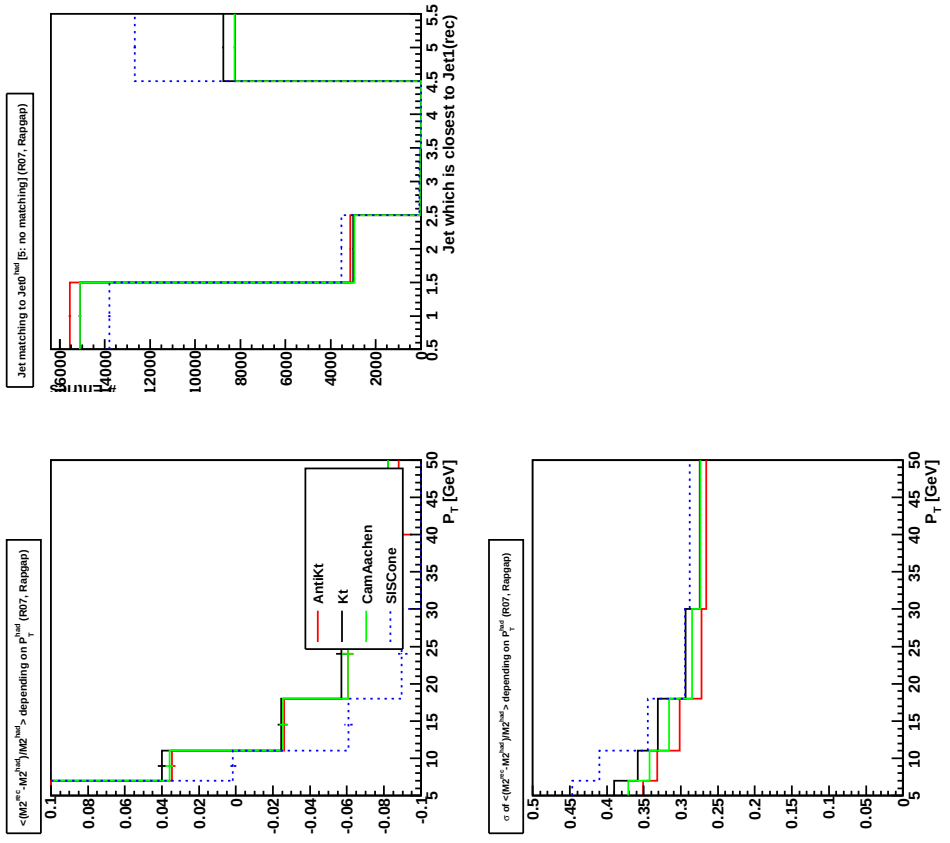

FIG 24. *Deviation of M_{12}^2 between reconstructed and hadron level, RAPGAP data, $R_0 = 0.7$*

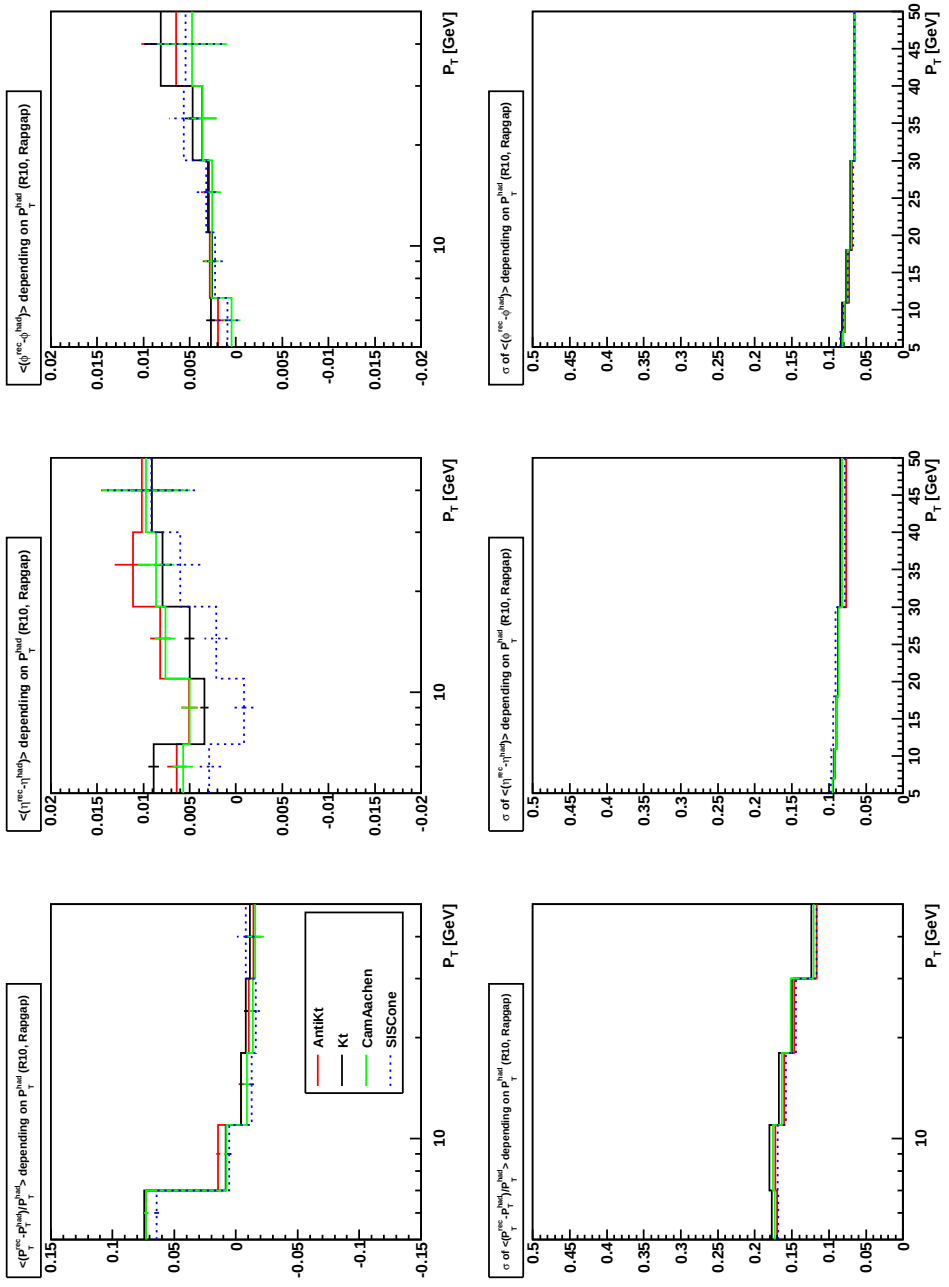

FIG 25. *Deviations of P_T, η and ϕ between reconstructed and hadron level, RAPGAP data, $R_0 = 1.0$*

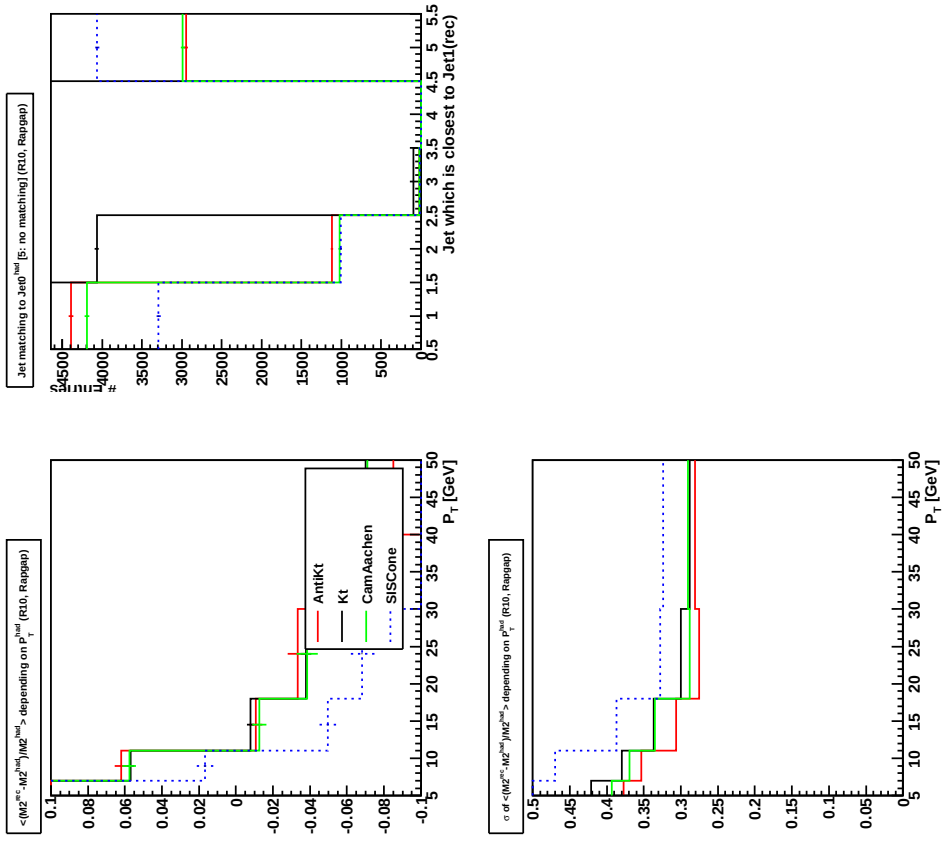

FIG 26. *Deviation of M_{12}^2 between reconstructed and hadron level, RAPGAP data, $R_0 = 1.0$*

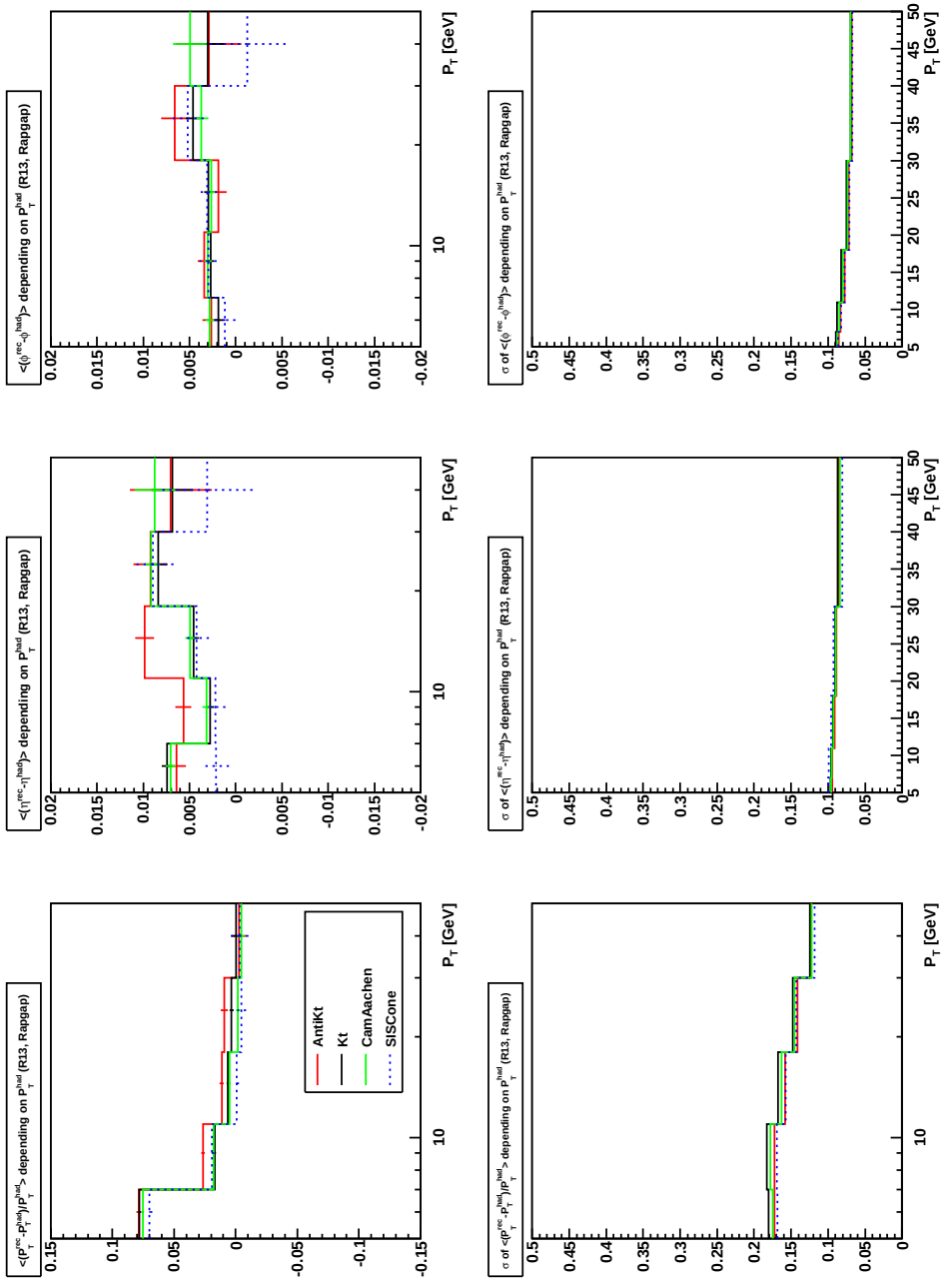

FIG 27. *Deviations of P_T, η and ϕ between reconstructed and hadron level, RAPGAP data, $R_0 = 1.3$*

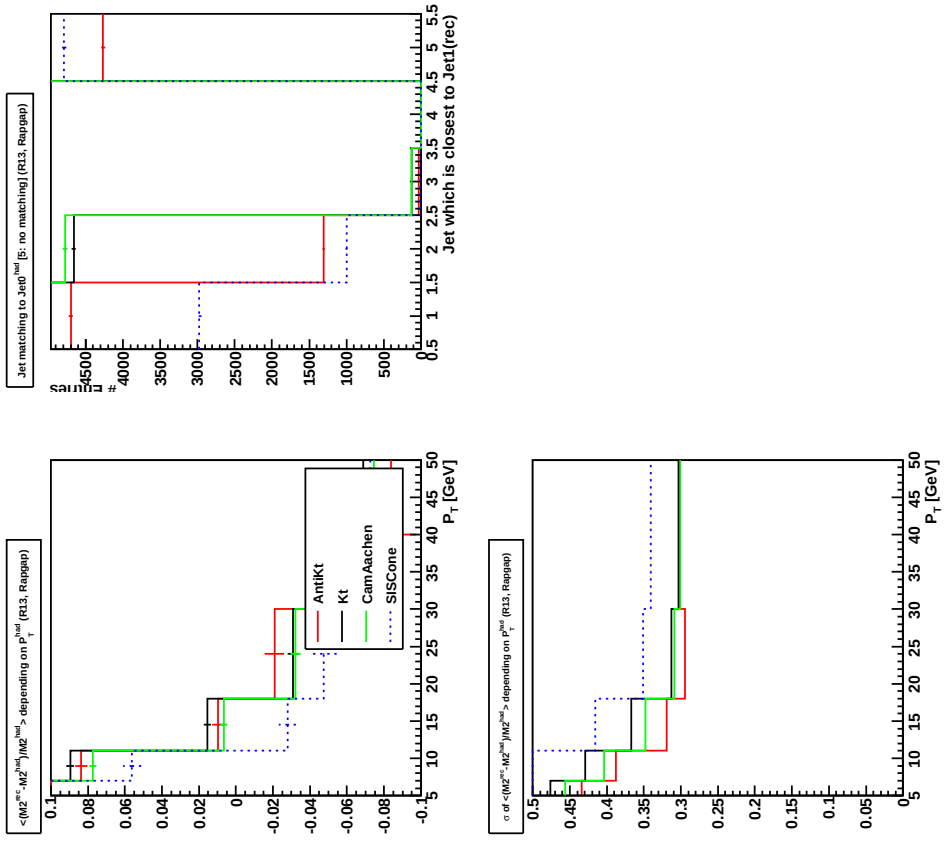

FIG 28. *Deviation of M_{12}^2 between reconstructed and hadron level, RAPGAP data, $R_0 = 1.3$*

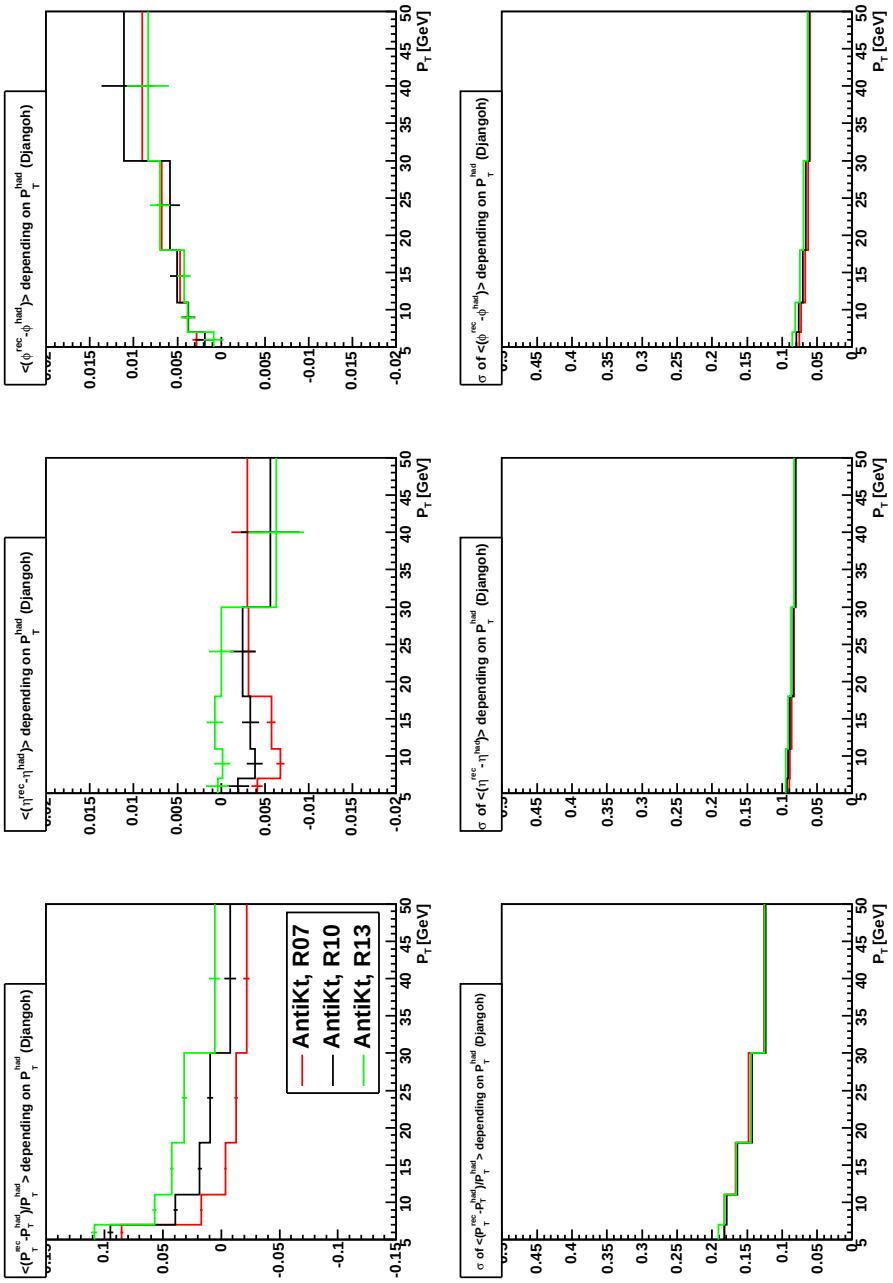

FIG 29. *Deviations of P_T, η and ϕ between reconstructed and hadron level, DJANGOH data, Anti-k_T*

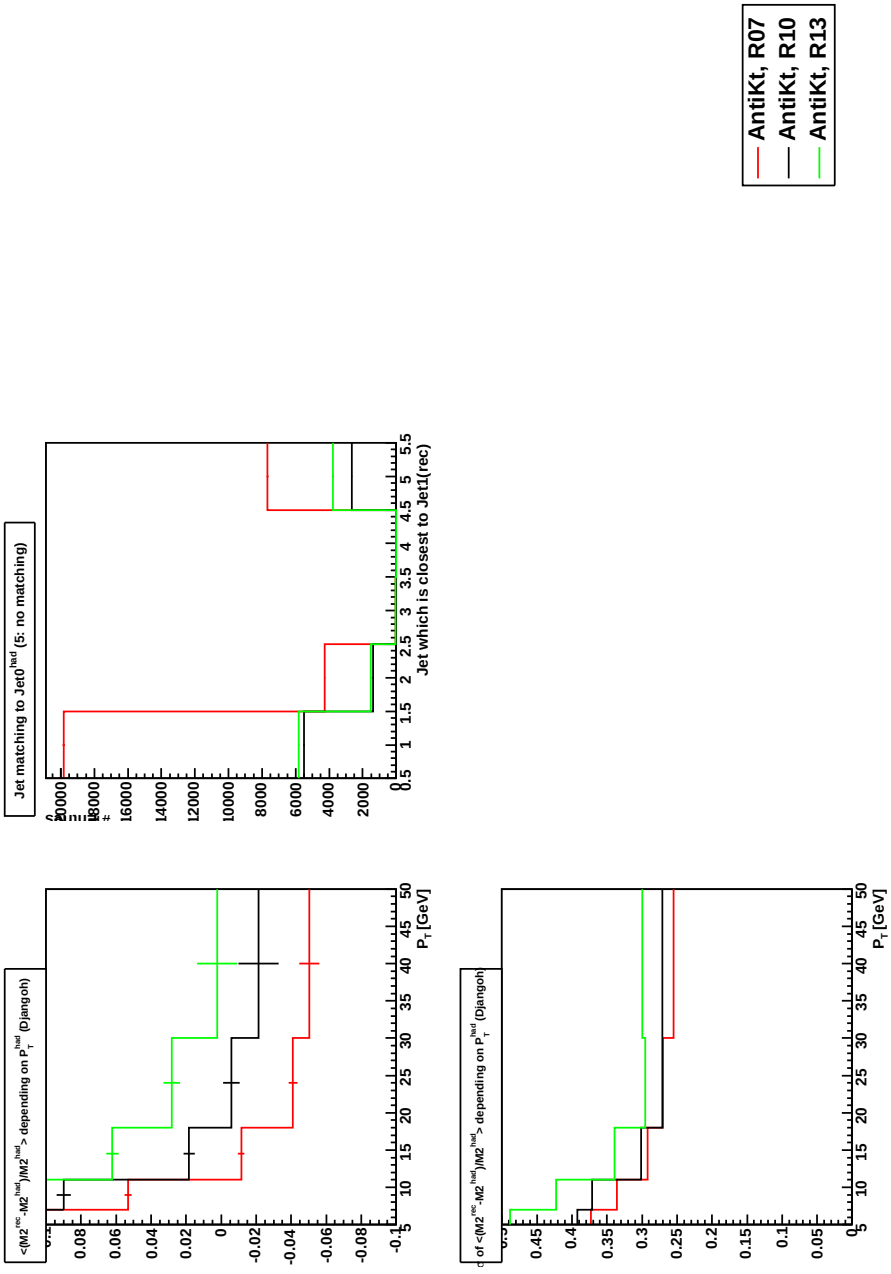

FIG 30. *Deviation of M_{12}^2 between reconstructed and hadron level, DJANGOH data, Anti-k_T*

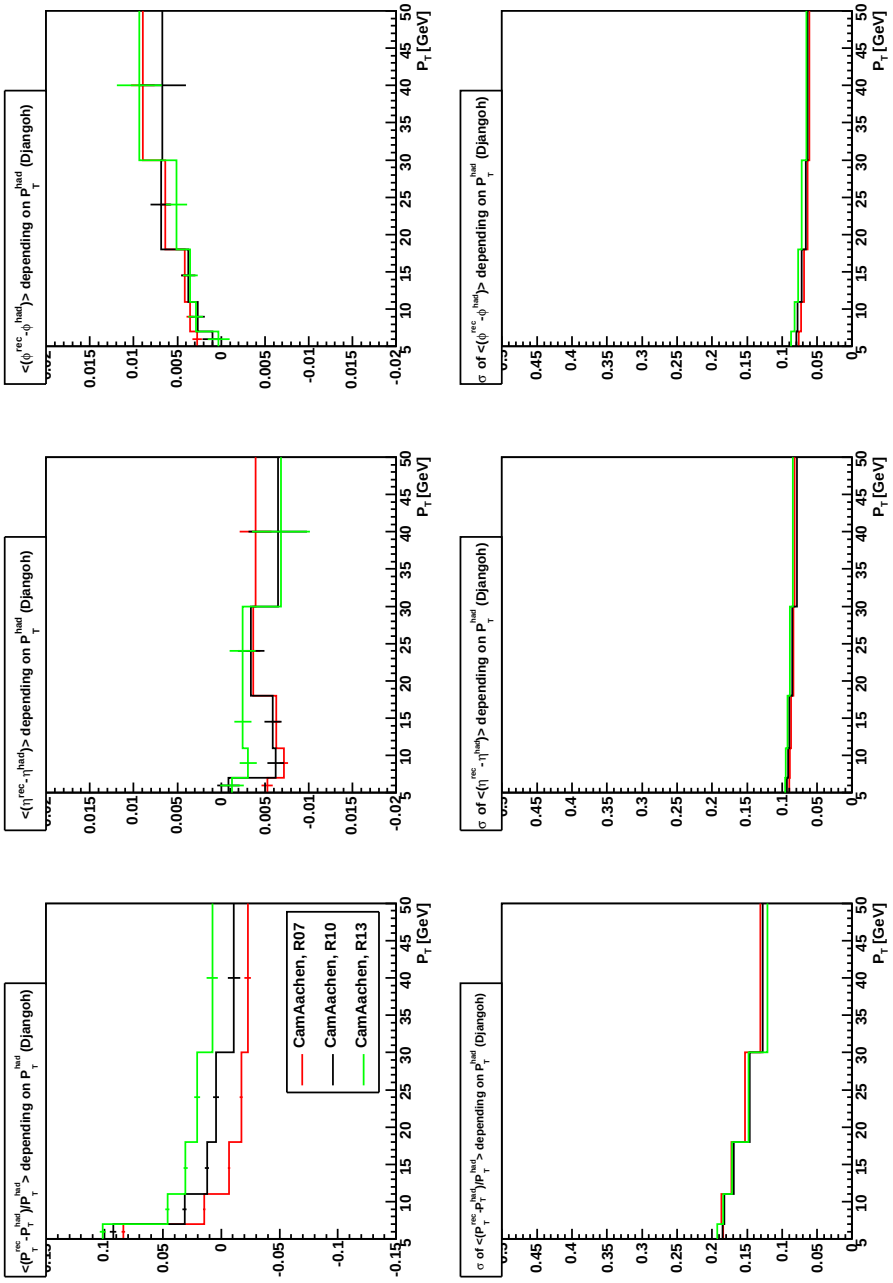

FIG 31. *Deviations of P_T, η and ϕ between reconstructed and hadron level, DJANGOH data, Cambridge/Aachen*

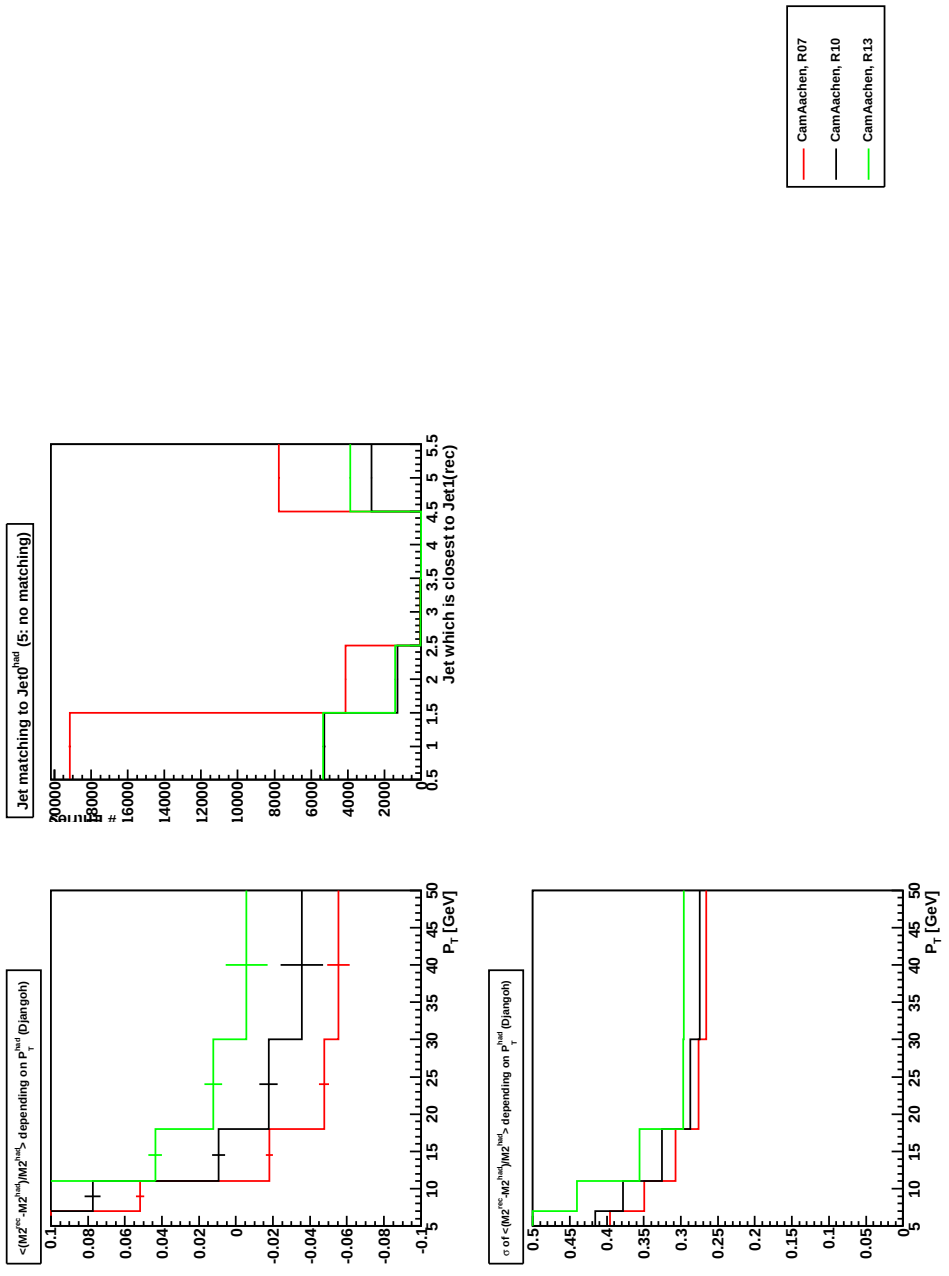

FIG 32. *Deviation of M_{12}^2 between reconstructed and hadron level, DJANGOH data, Cambridge/Aachen*

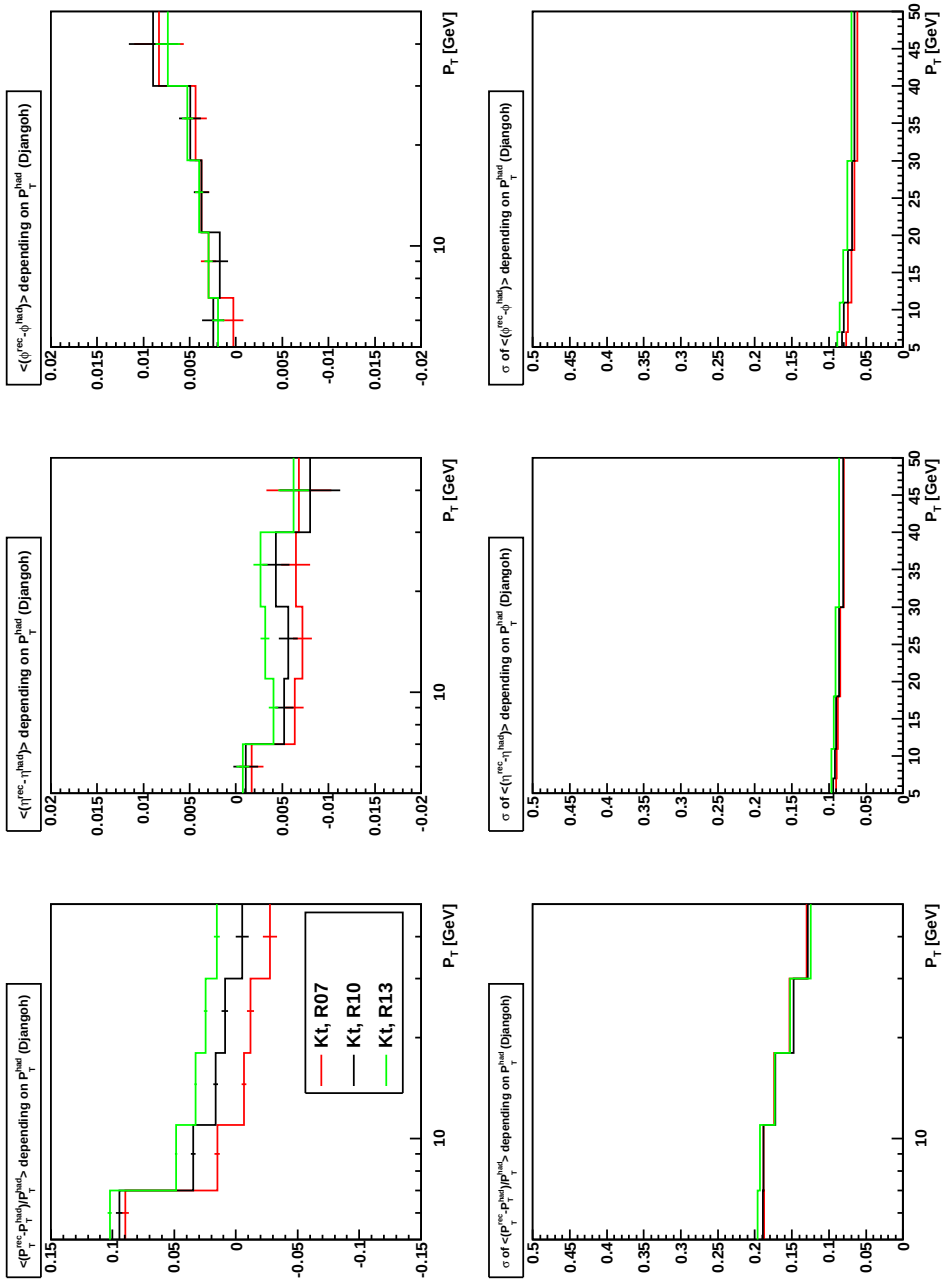

FIG 33. *Deviations of P_T, η and ϕ between reconstructed and hadron level, DJANGOH data, k_T*

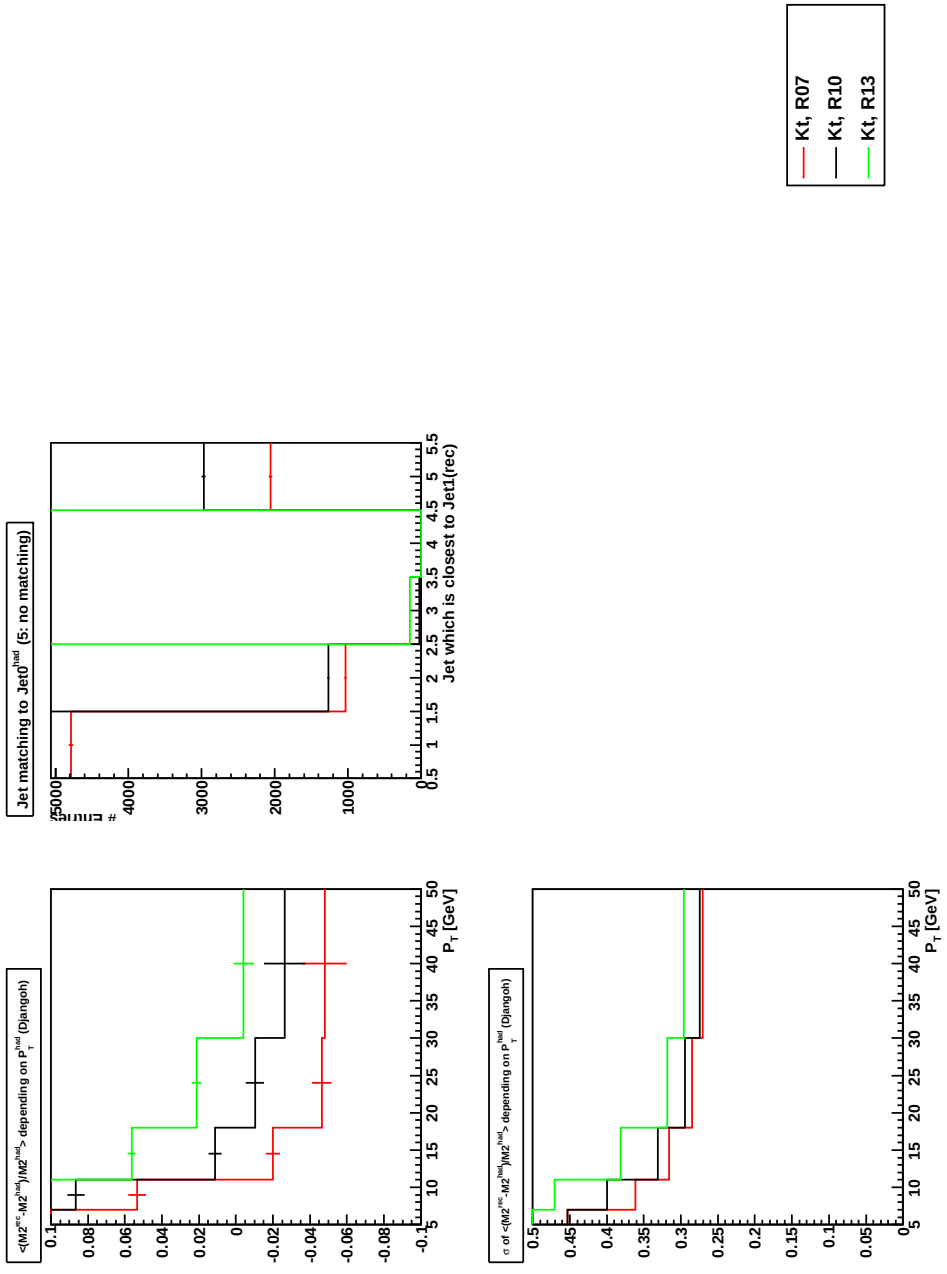

FIG 34. *Deviation of M_{12}^2 between reconstructed and hadron level, DJANGOH data, k_T*

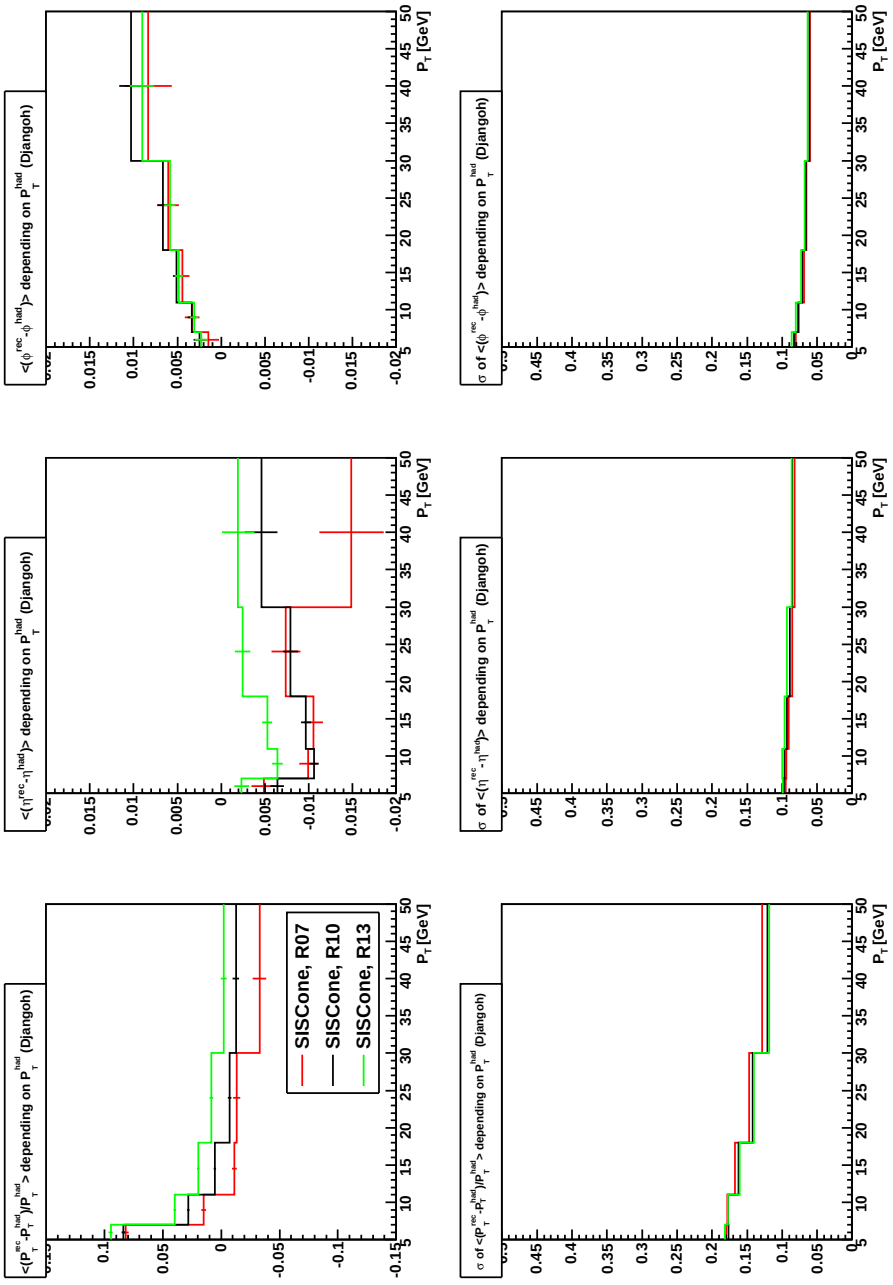

FIG 35. *Deviations of P_T, η and ϕ between reconstructed and hadron level, DJANGOH data, SISCone*

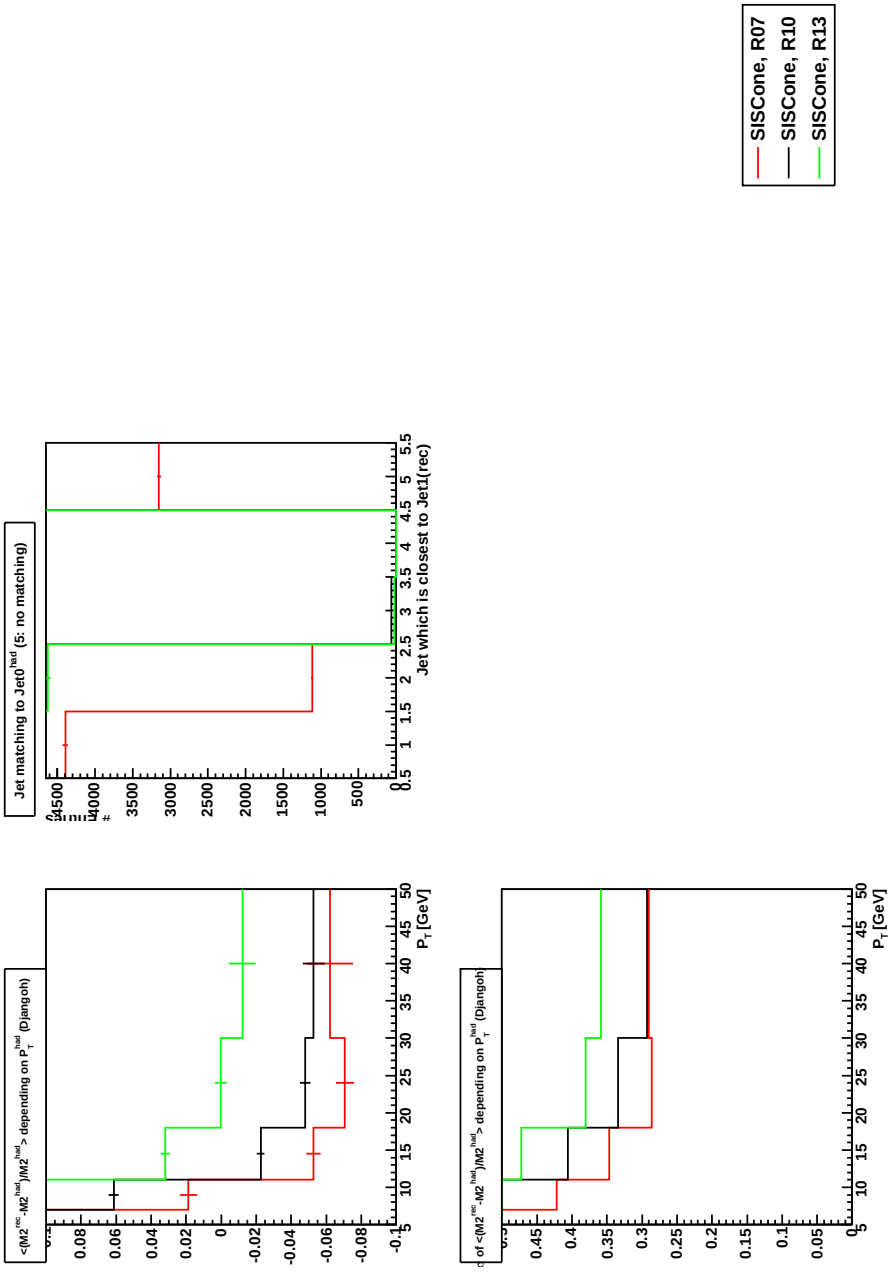

FIG 36. *Deviation of M_{12}^2 between reconstructed and hadron level, DJANGOH data, SISCone*

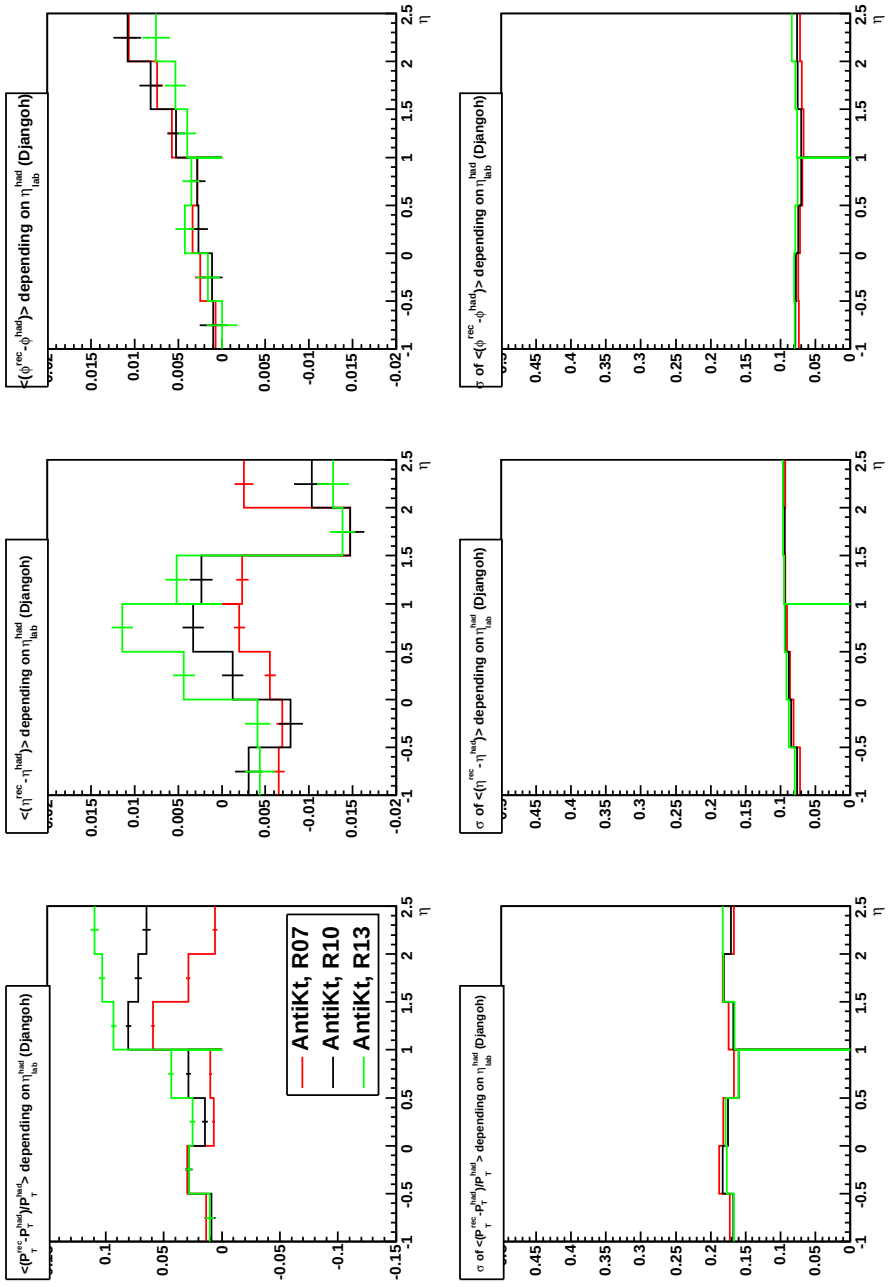

FIG 37. *Deviations of P_T, η and ϕ between reconstructed and hadron level, DJANGOH data, η dependance*

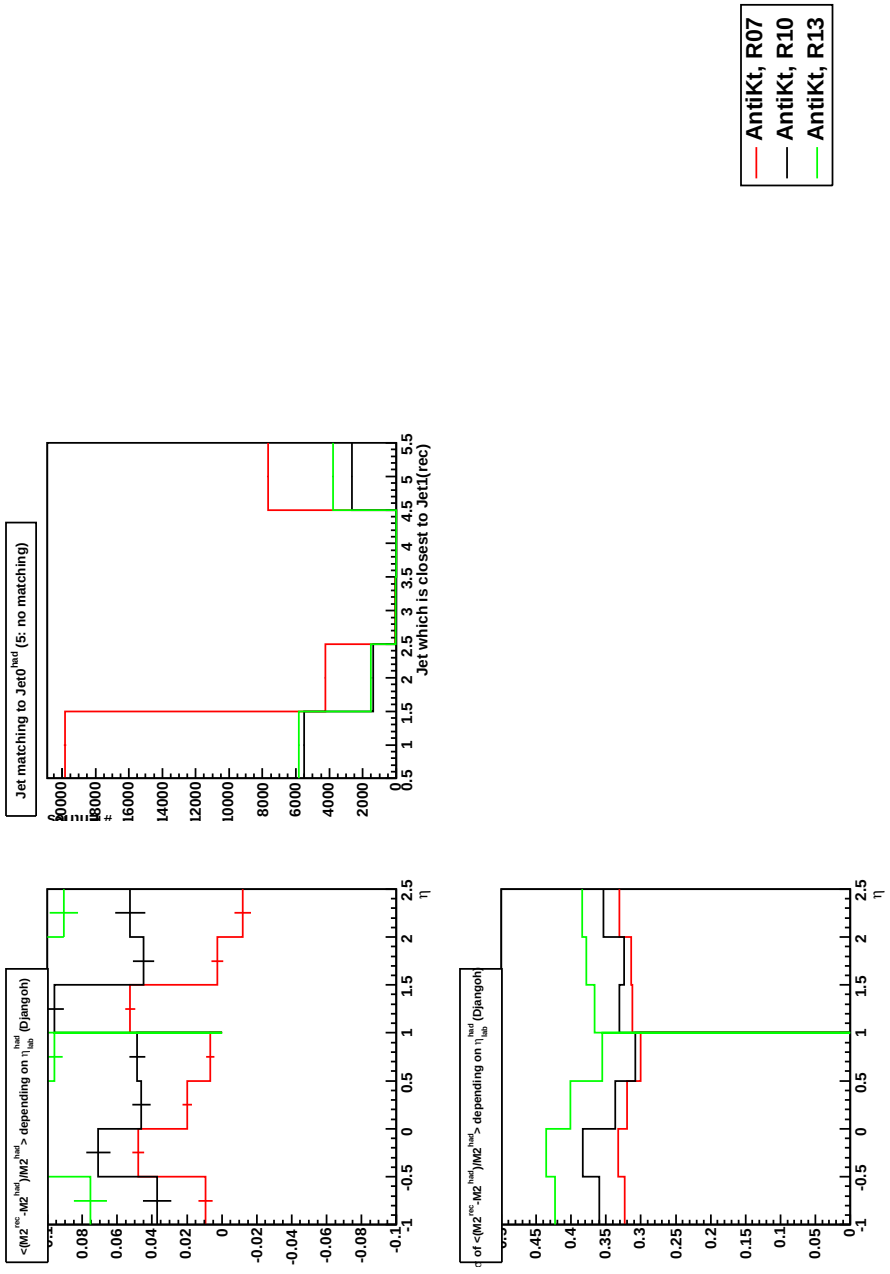

FIG 38. *Deviation of M_{12}^2 between reconstructed and hadron level, DJANGOH data, η dependance*

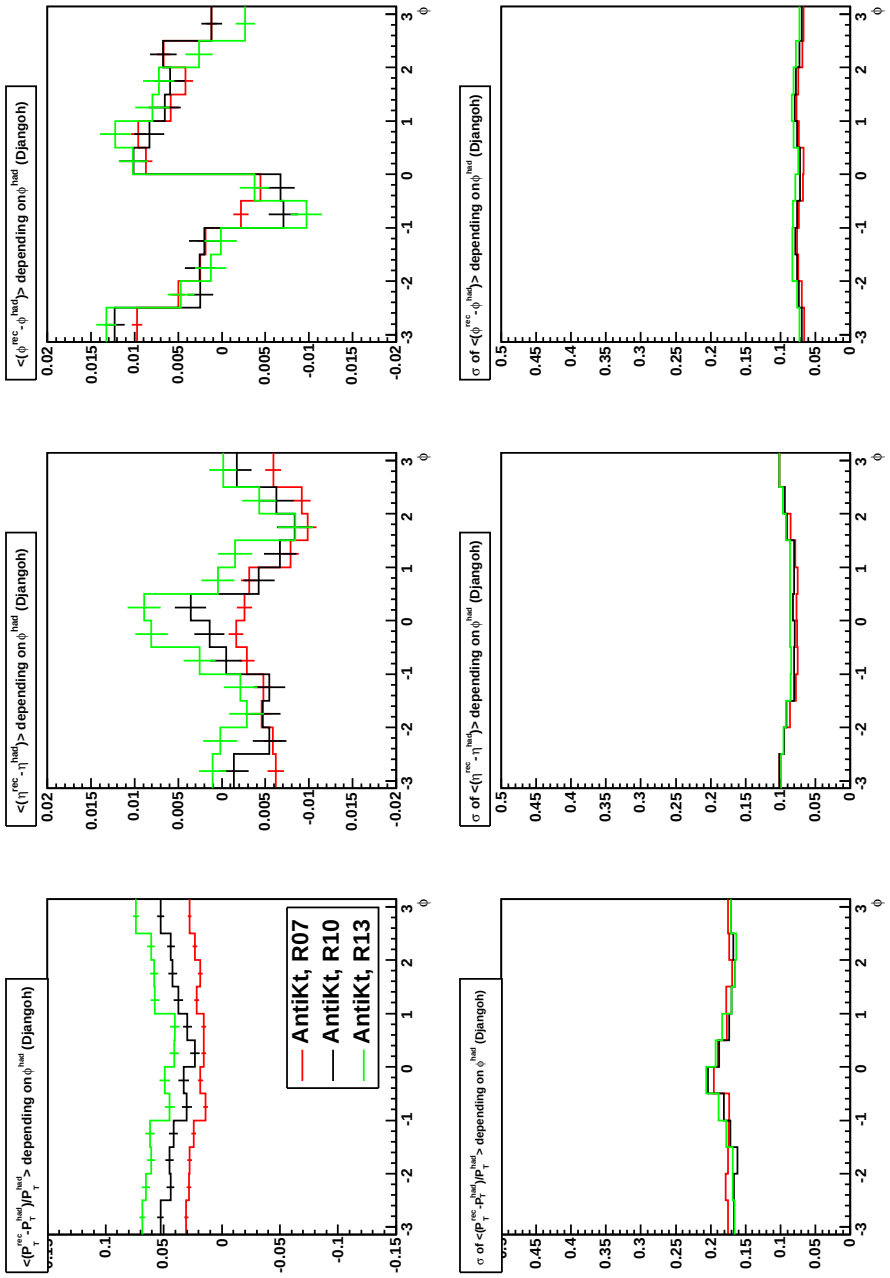

FIG 39. *Deviations of P_T, η and ϕ between reconstructed and hadron level, DJANGOH data, ϕ dependance*

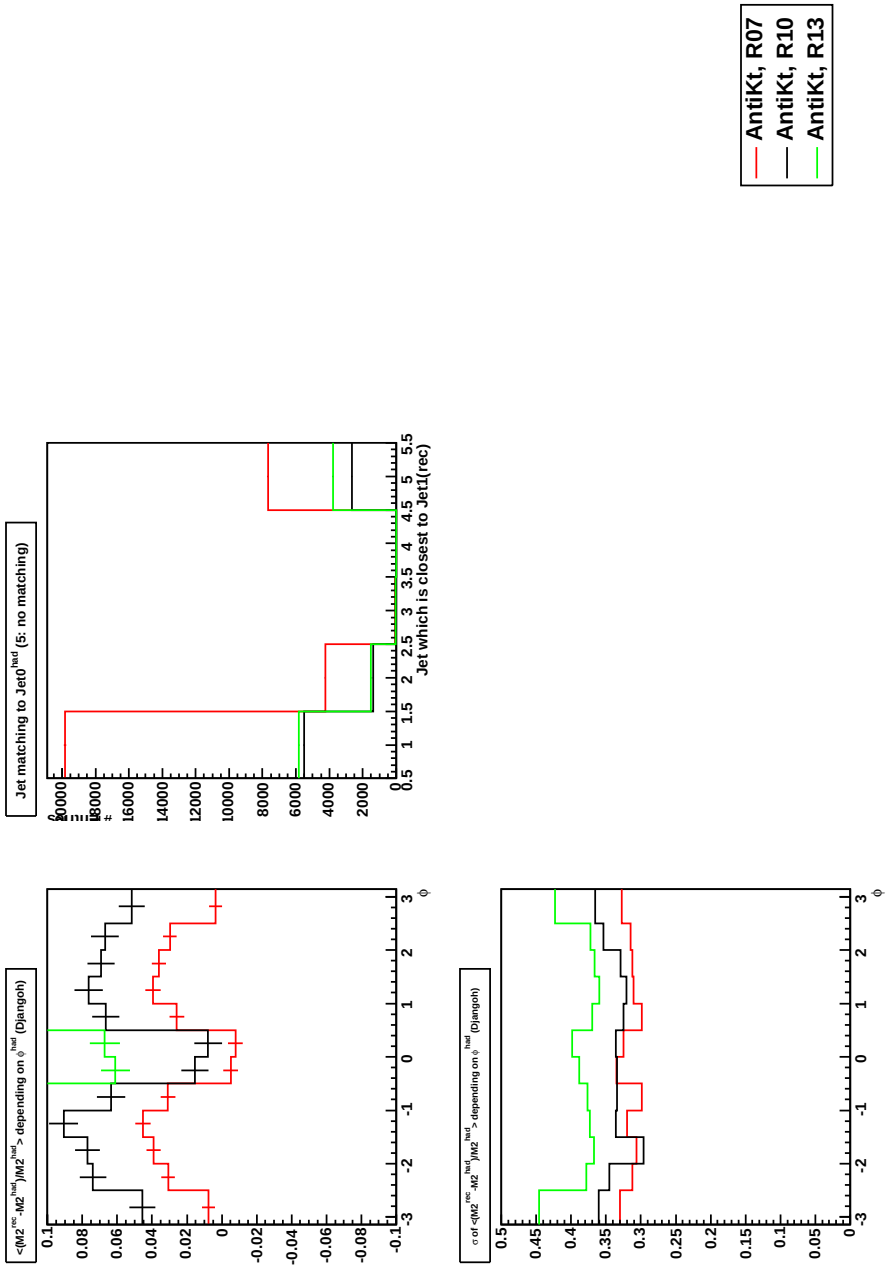

FIG 40. *Deviation of M_{12}^2 between reconstructed and hadron level, DJANGOH data, ϕ dependance*

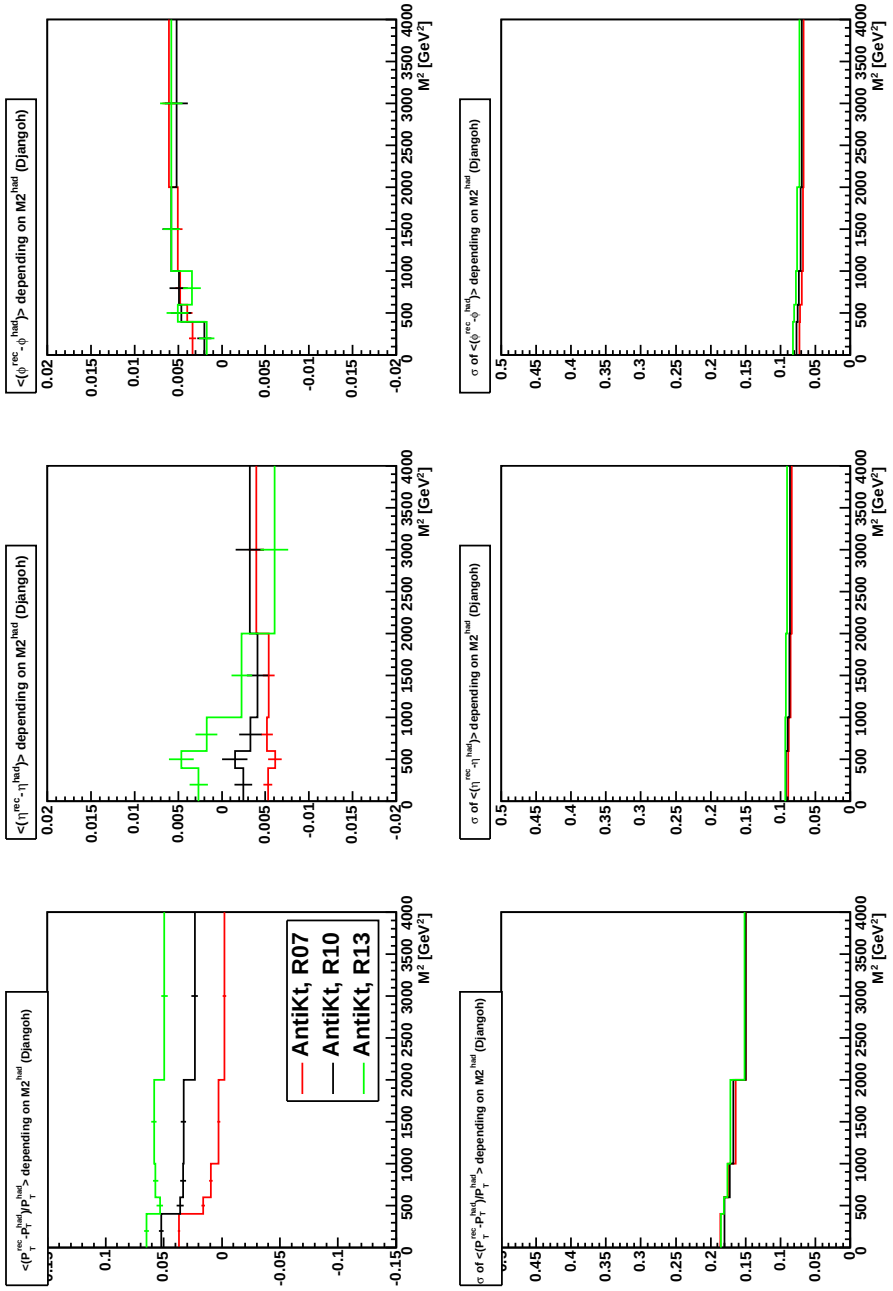

FIG 41. *Deviations of P_T, η and ϕ between reconstructed and hadron level, DJANGOH data, M^2_{12} dependance*

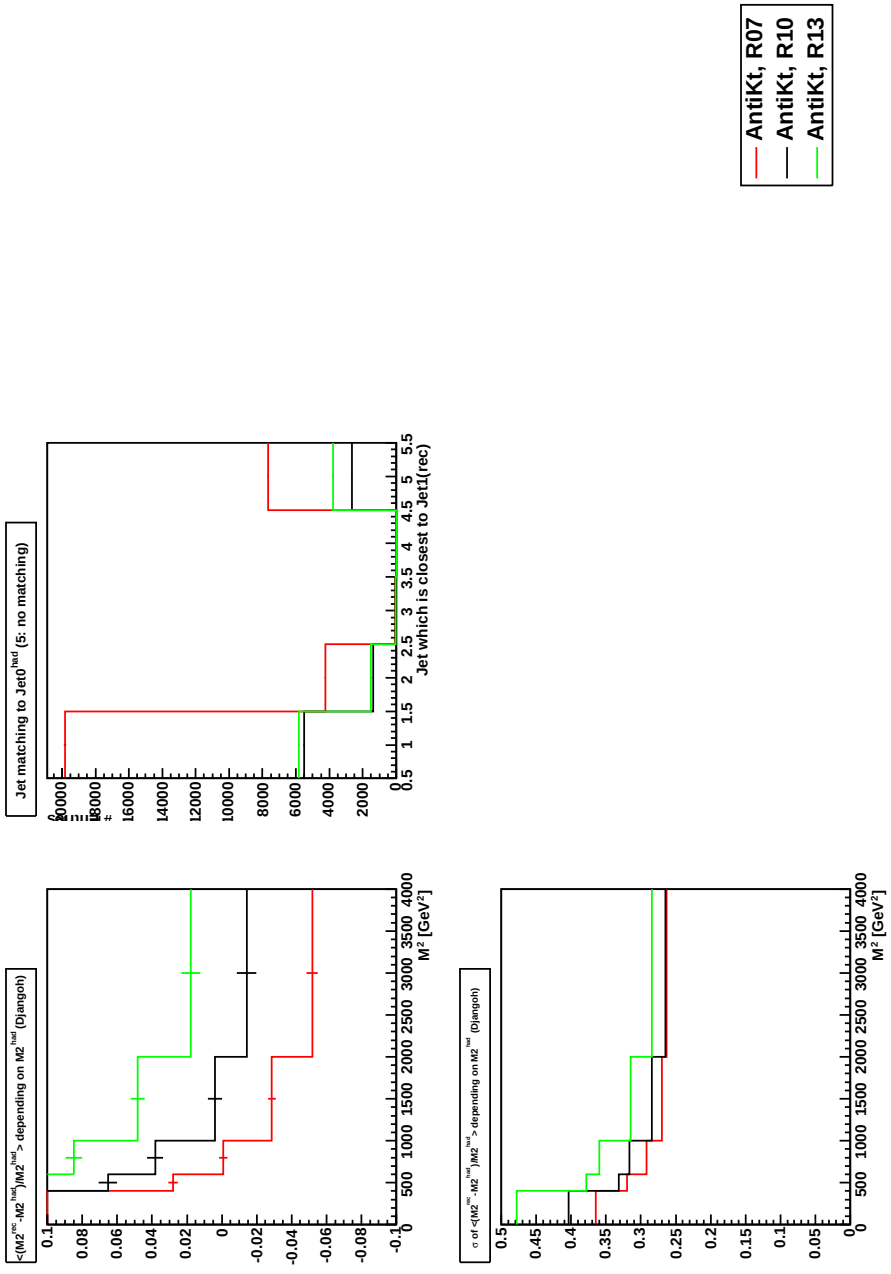

FIG 42. *Deviation of M_{12}^2 between reconstructed and hadron level, DJANGOH data, M_{12}^2 dependance*

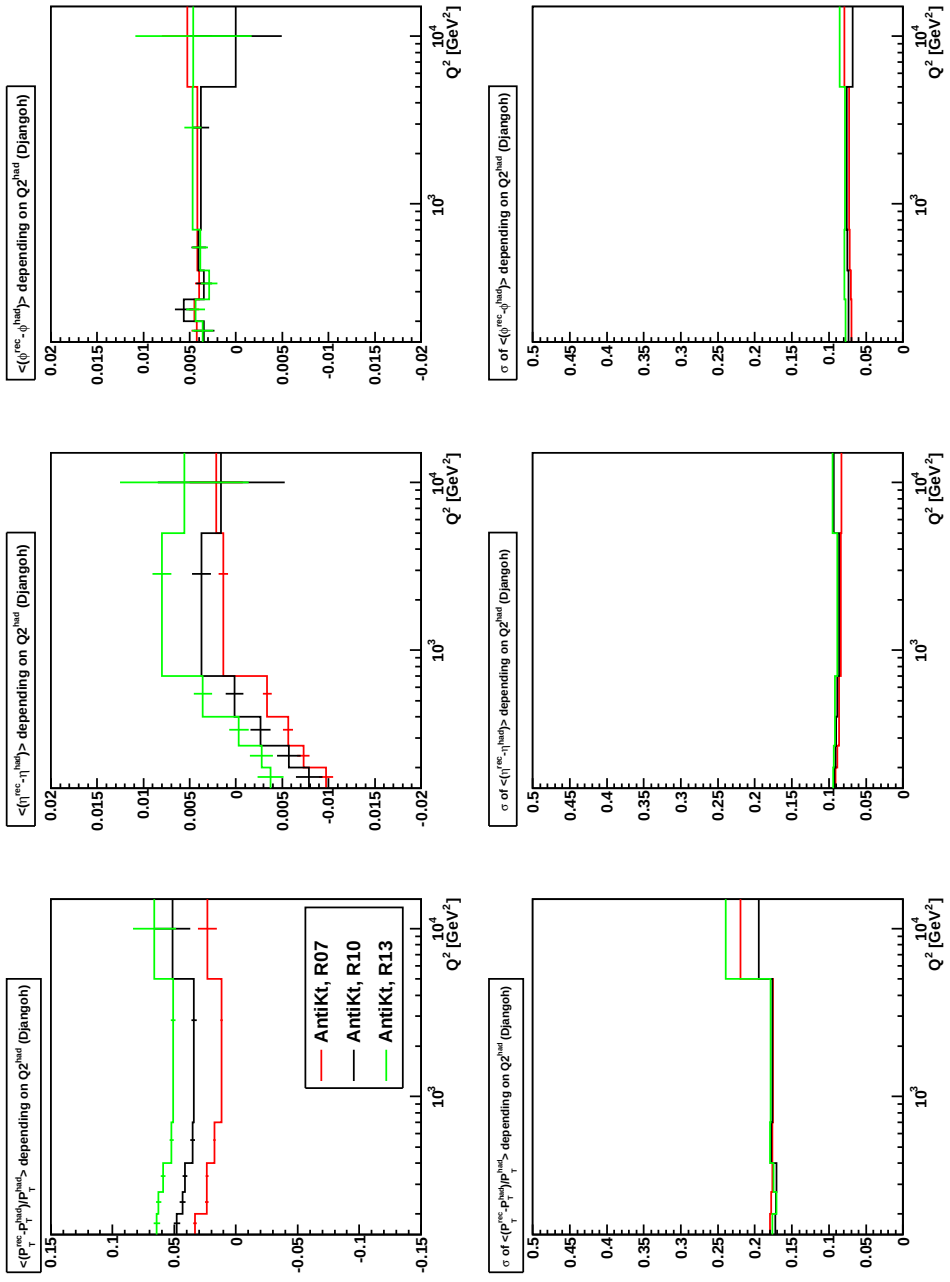

FIG 43. *Deviations of P_T, η and ϕ between reconstructed and hadron level, DJANGOH data, Q^2 dependance*

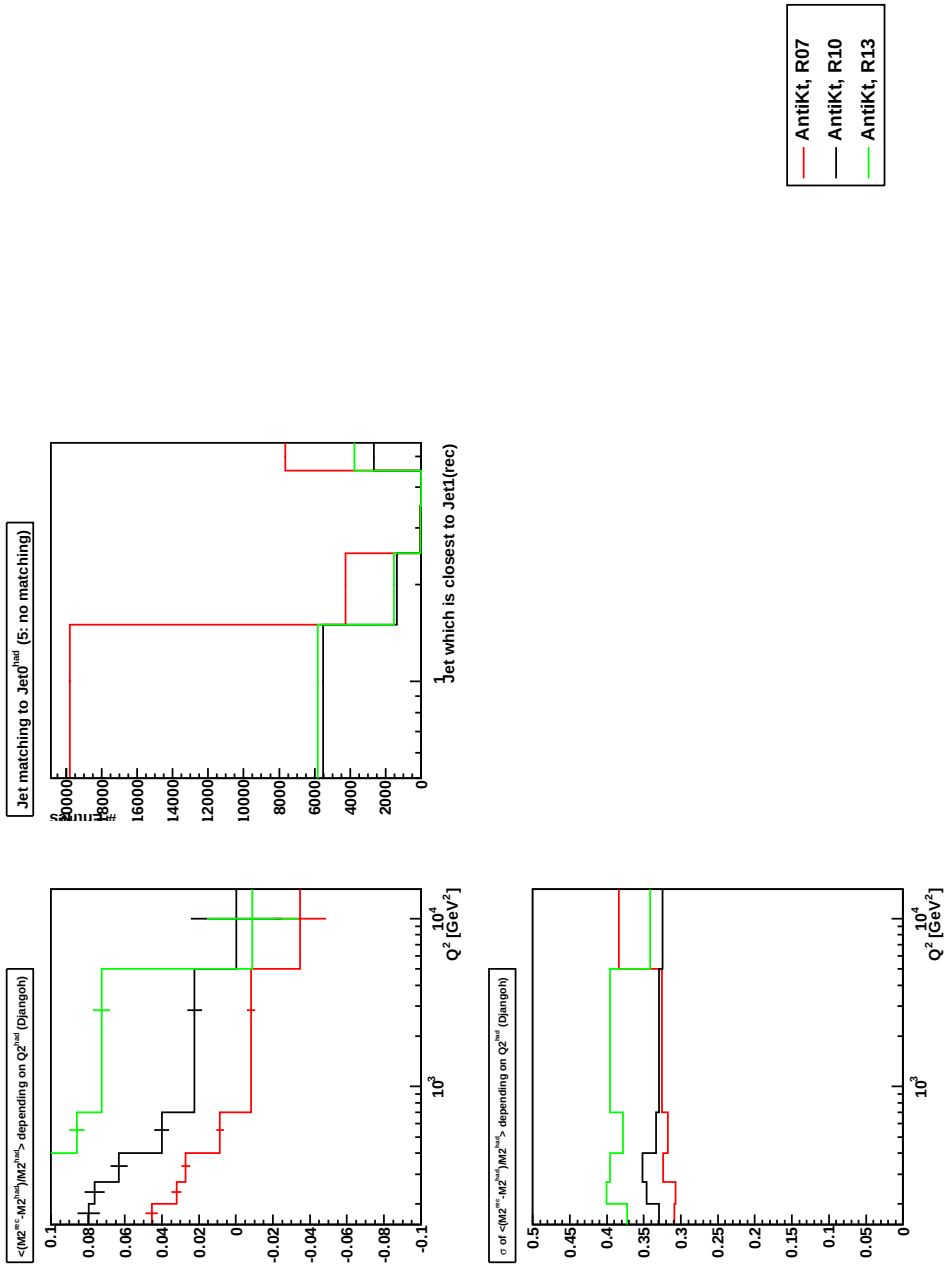

FIG 44. *Deviation of M_{12}^2 between reconstructed and hadron level, DJANGOH data, Q^2 dependance*

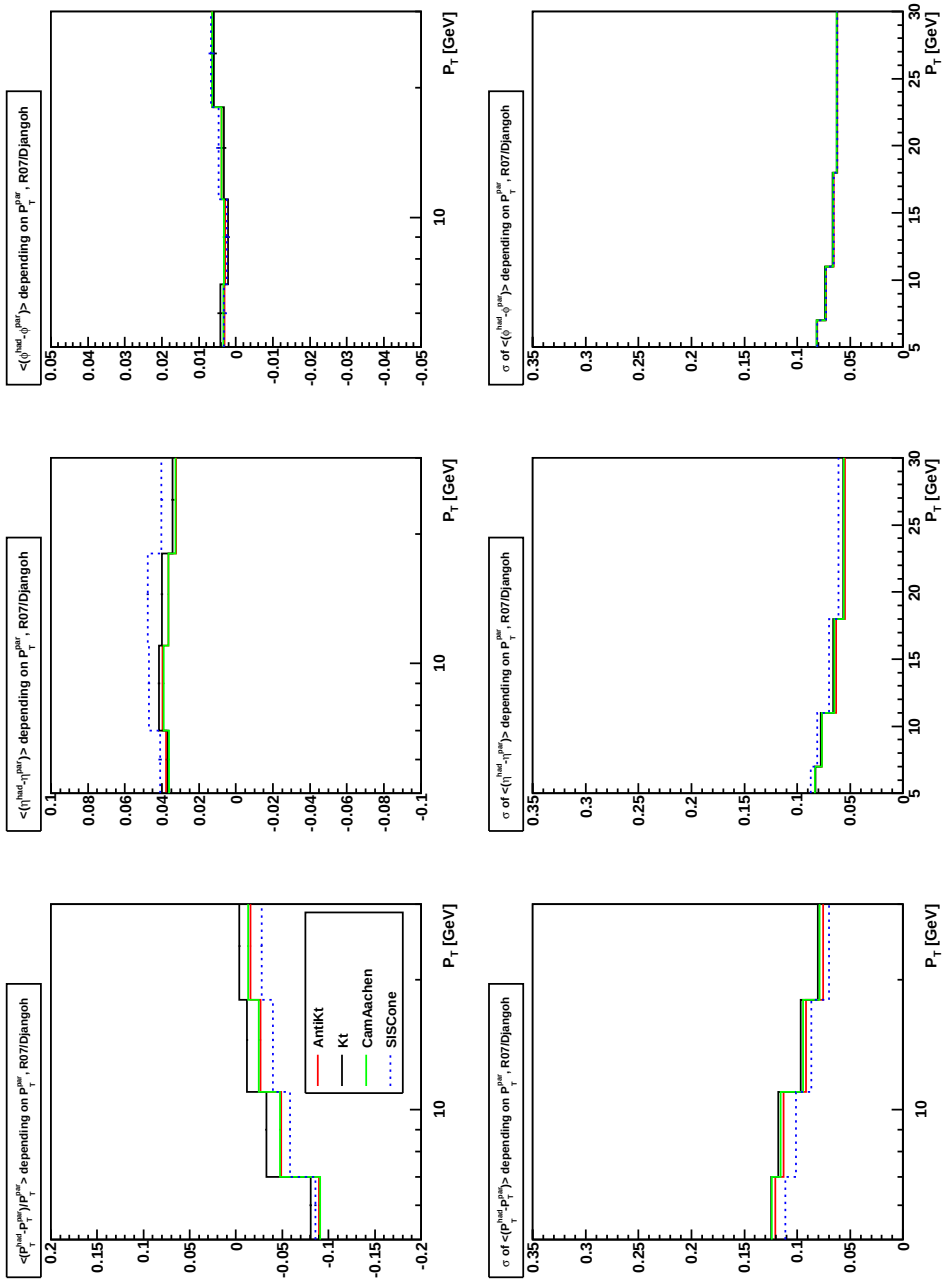

FIG 45. *Deviations of P_T, η and ϕ between parton and hadron level, DJANGOH data, $R_0 = 0.7$*

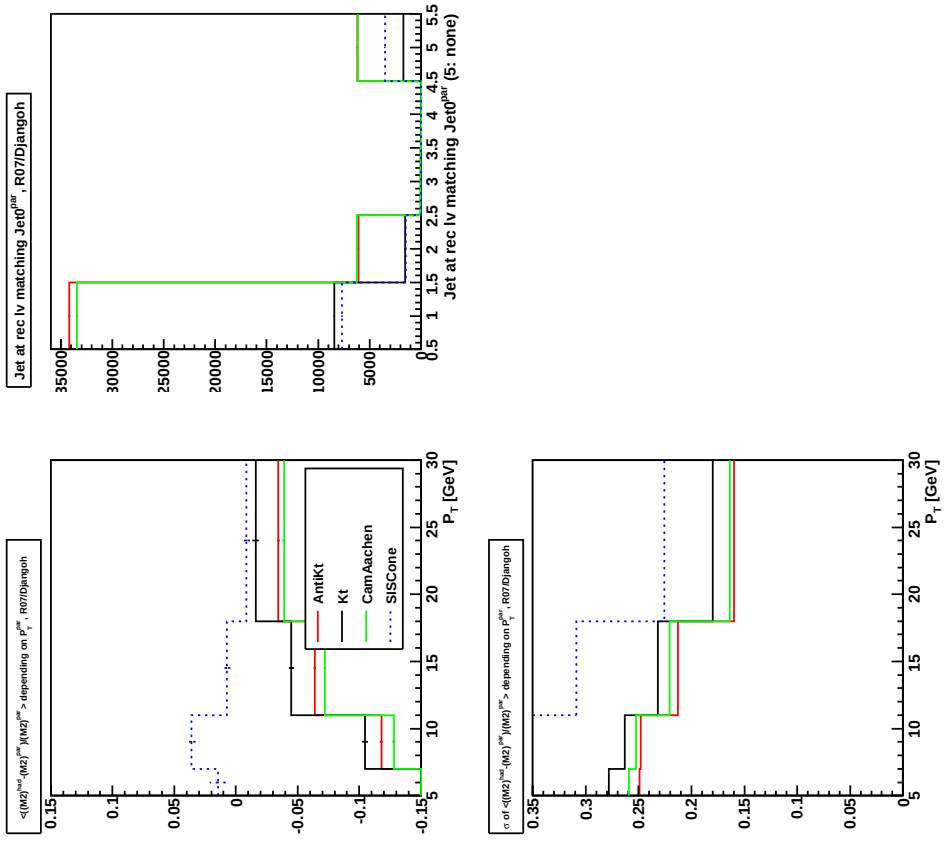

FIG 46. *Deviation of M_{12}^2 between parton and hadron level, DJANGOH data, $R_0 = 0.7$*

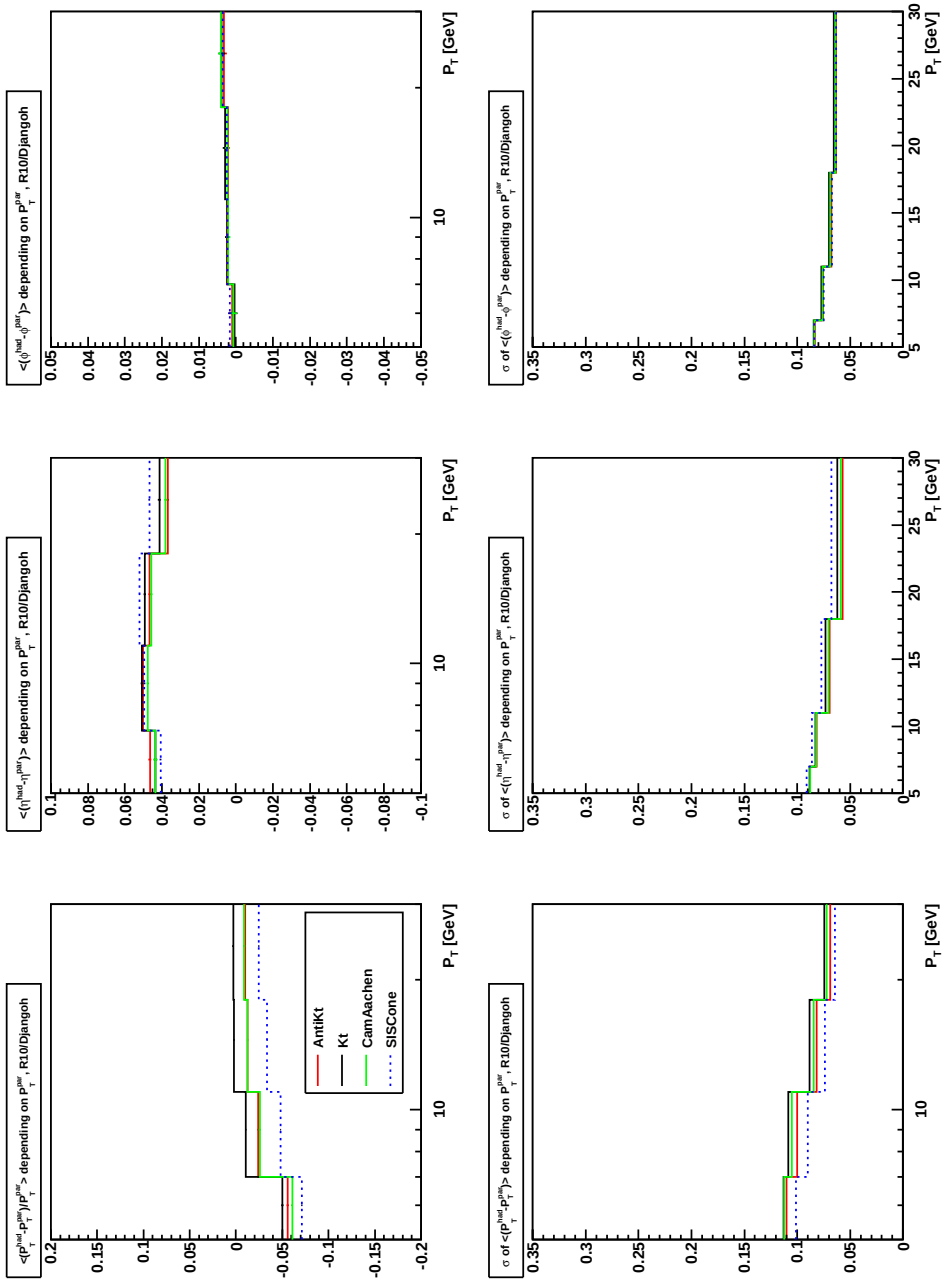

FIG 47. *Deviations of P_T, η and ϕ between parton and hadron level, DJANGOH data, $R_0 = 1.0$*

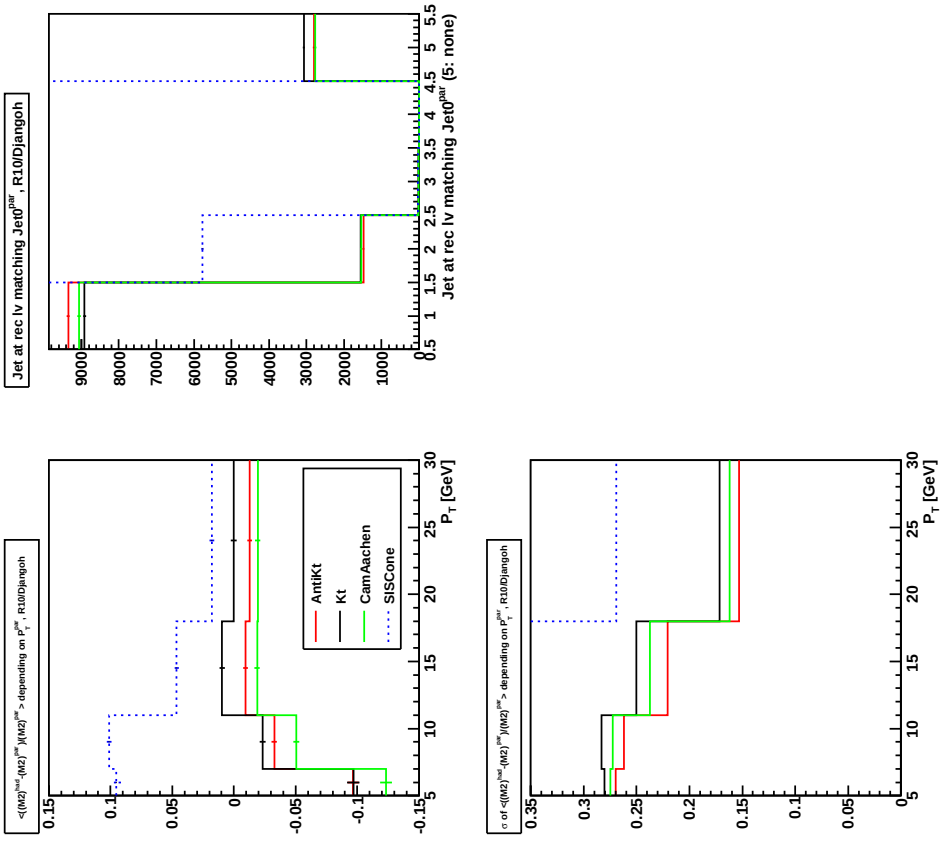

FIG 48. *Deviation of M_{12}^2 between parton and hadron level, DJANGOH data, $R_0 = 1.0$*

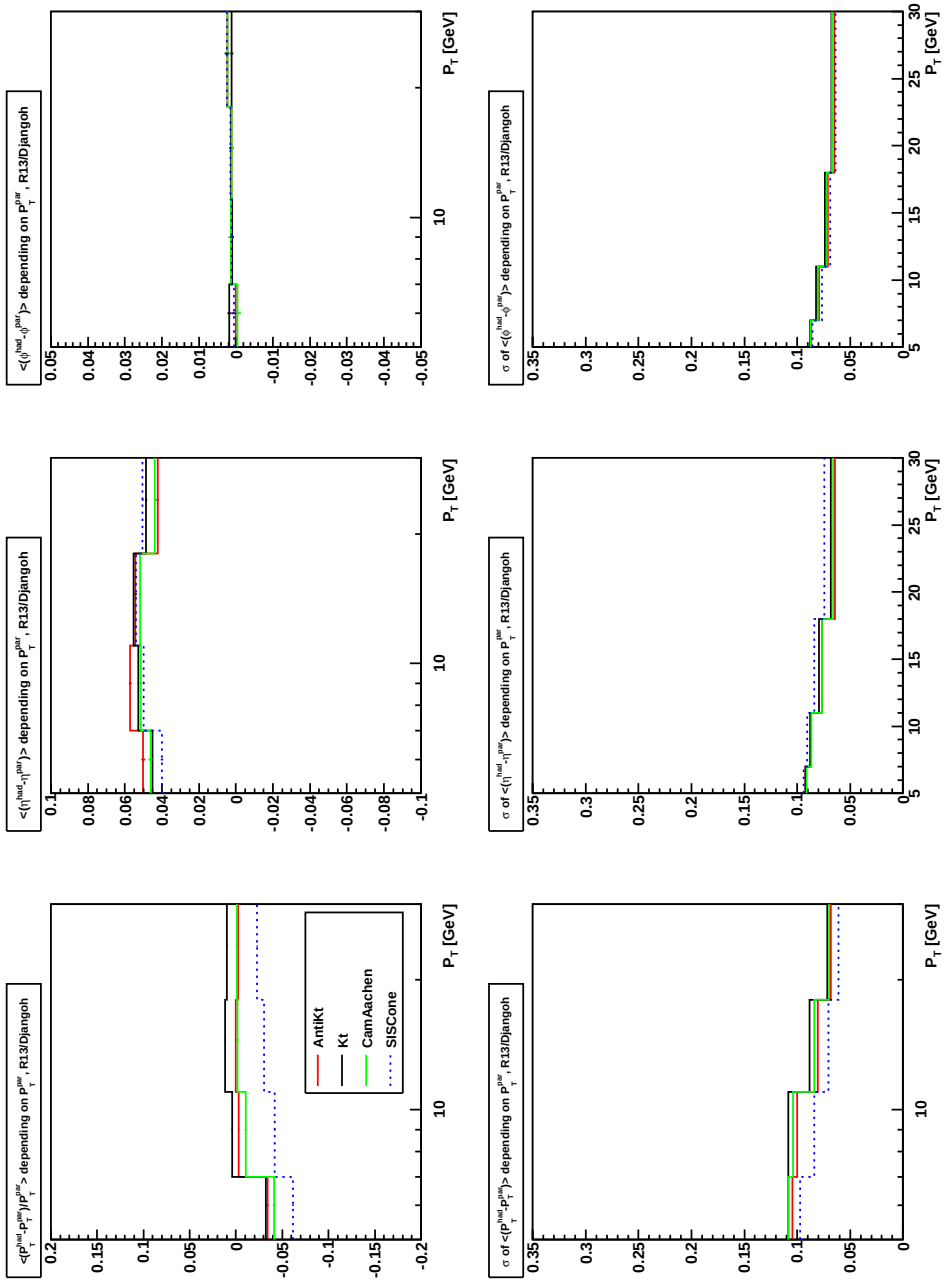

FIG 49. *Deviations of P_T, η and ϕ between parton and hadron level, DJANGOH data, $R_0 = 1.3$*

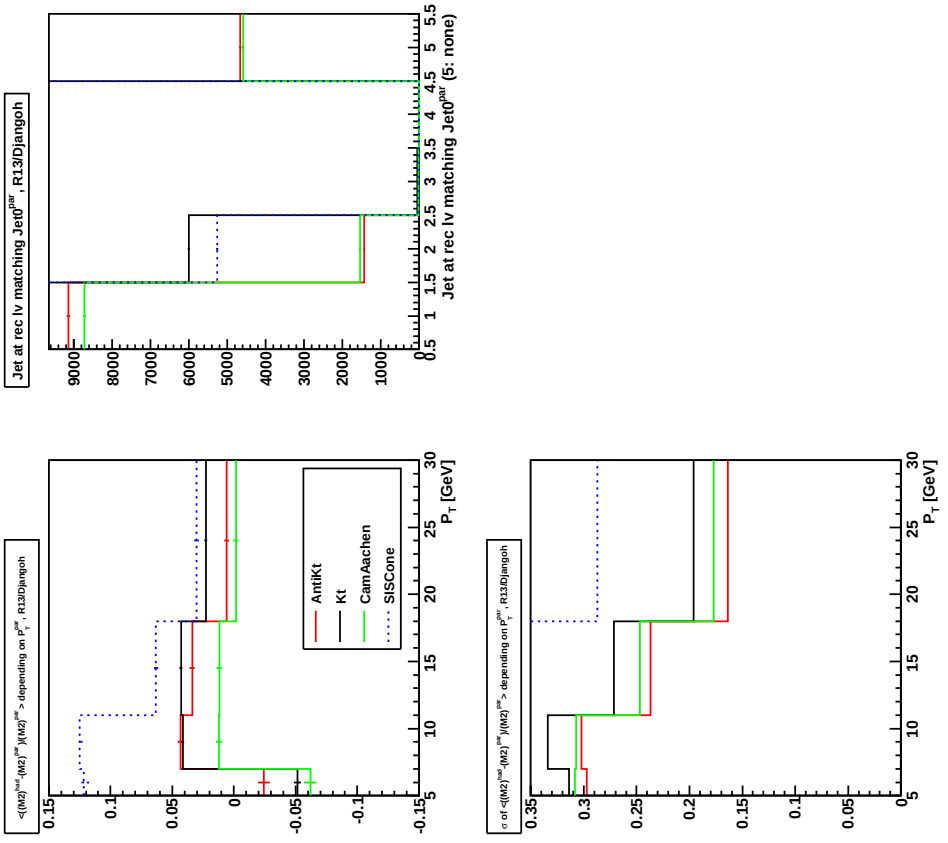

FIG 50. *Deviation of M_{12}^2 between parton and hadron level, DJANGOH data, $R_0 = 1.3$*

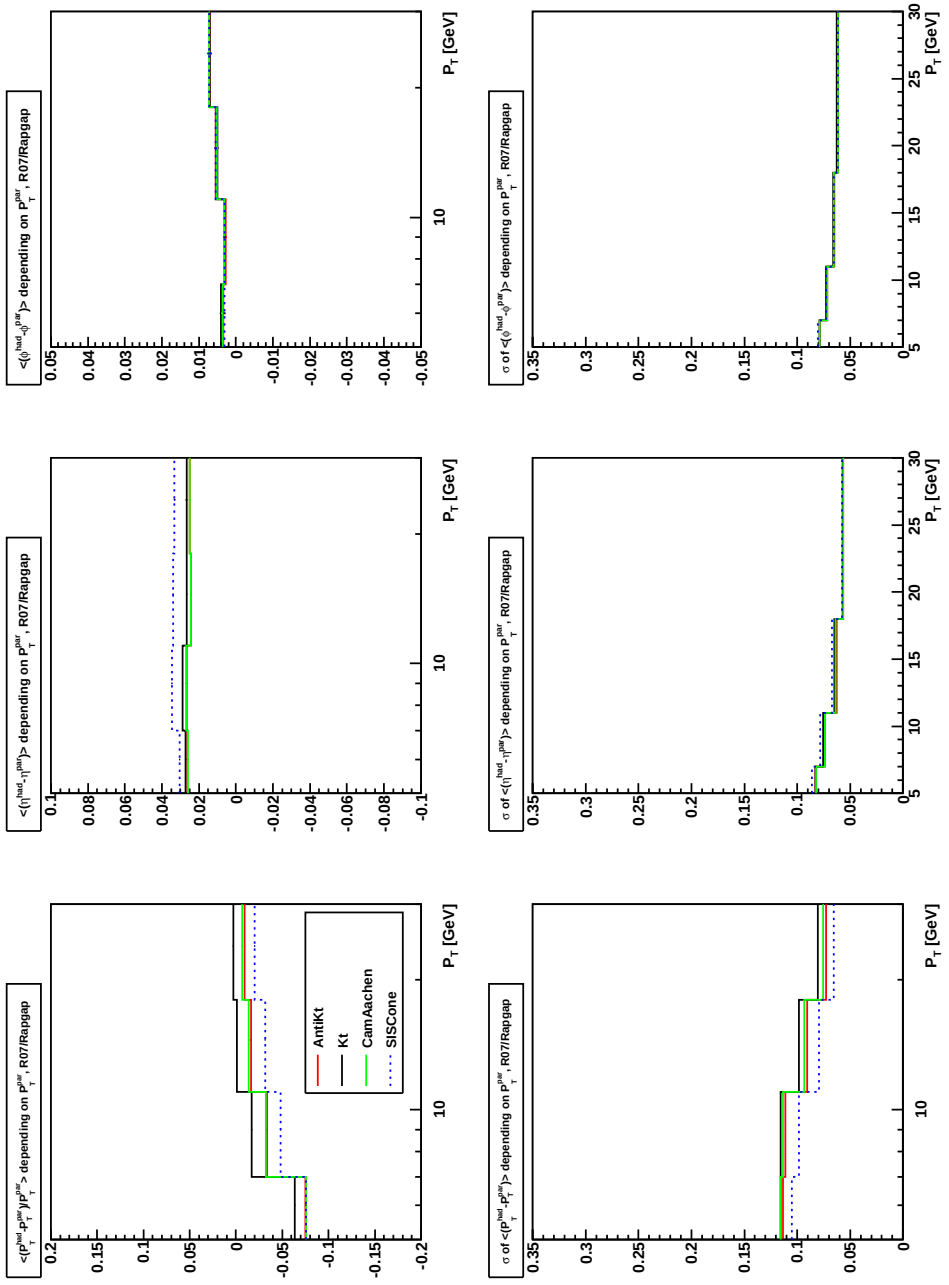

FIG 51. *Deviations of P_T, η and ϕ between parton and hadron level, RAPGAP data, $R_0 = 0.7$*

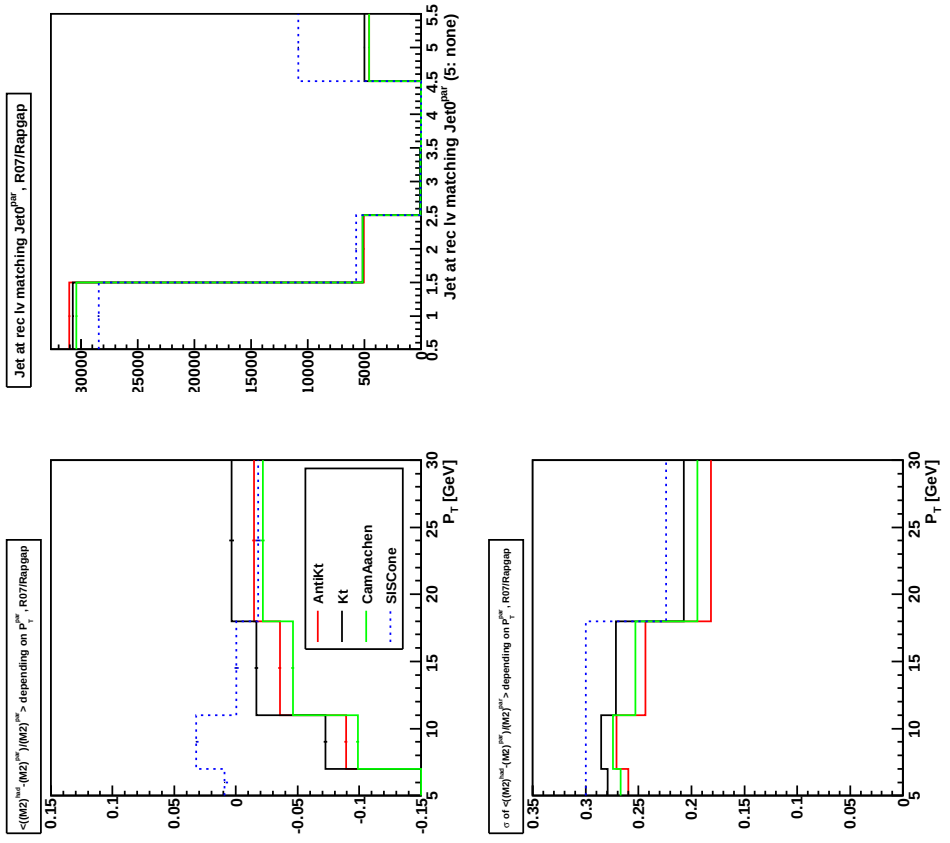

FIG 52. *Deviation of M_{12}^2 between parton and hadron level, RAPGAP data, $R_0 = 0.7$*

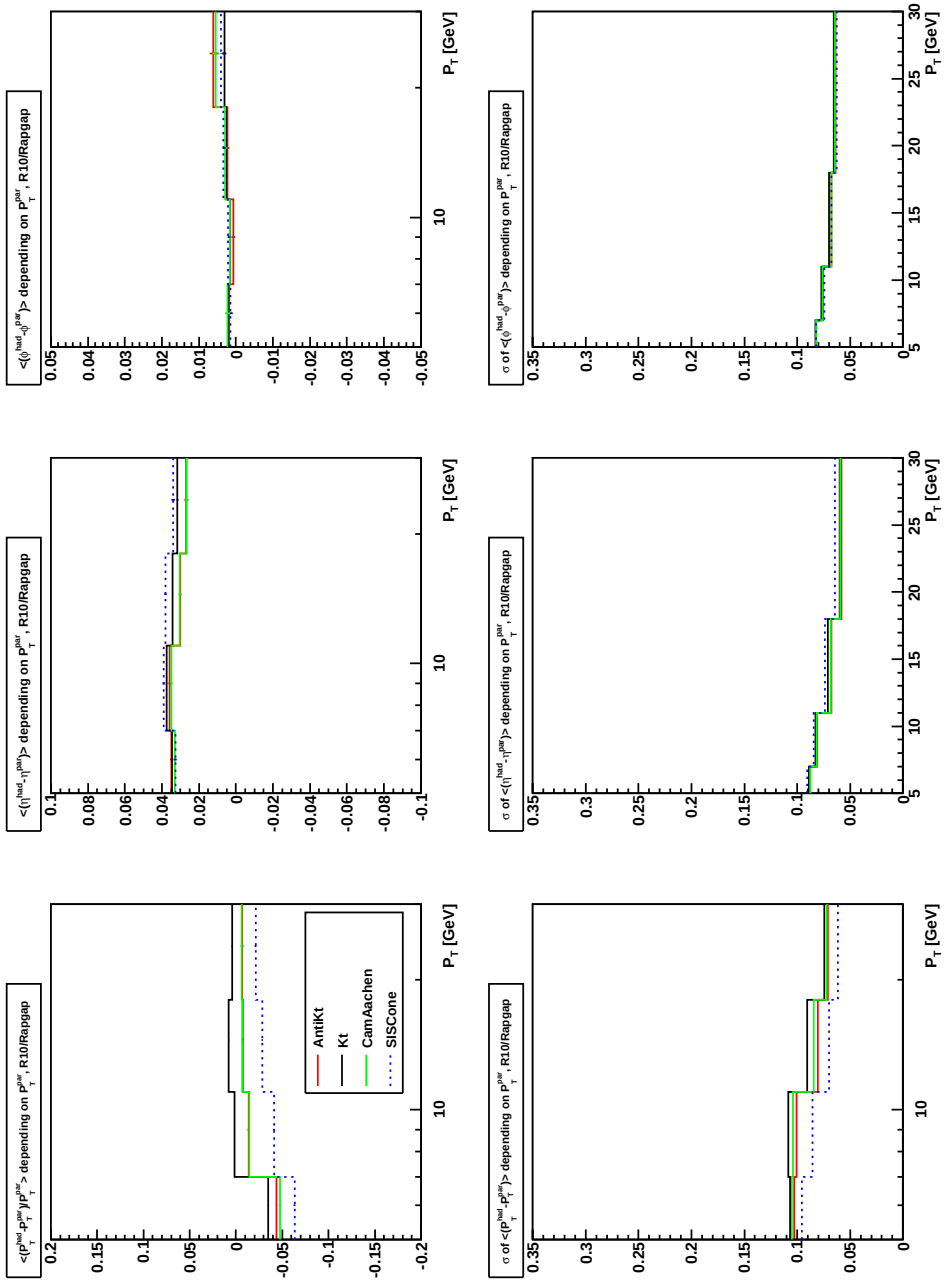

FIG 53. *Deviations of P_T, η and ϕ between parton and hadron level, RAPGAP data, $R_0 = 1.0$*

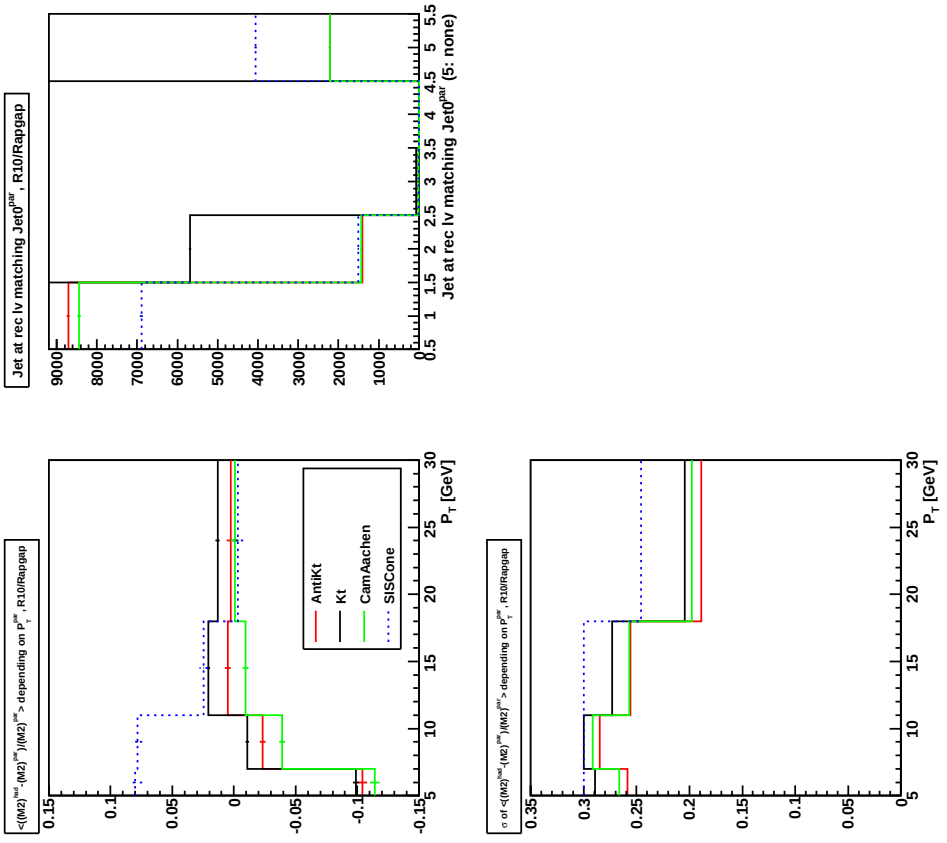

FIG 54. *Deviation of M_{12}^2 between parton and hadron level, RAPGAP data, $R_0 = 1.0$*

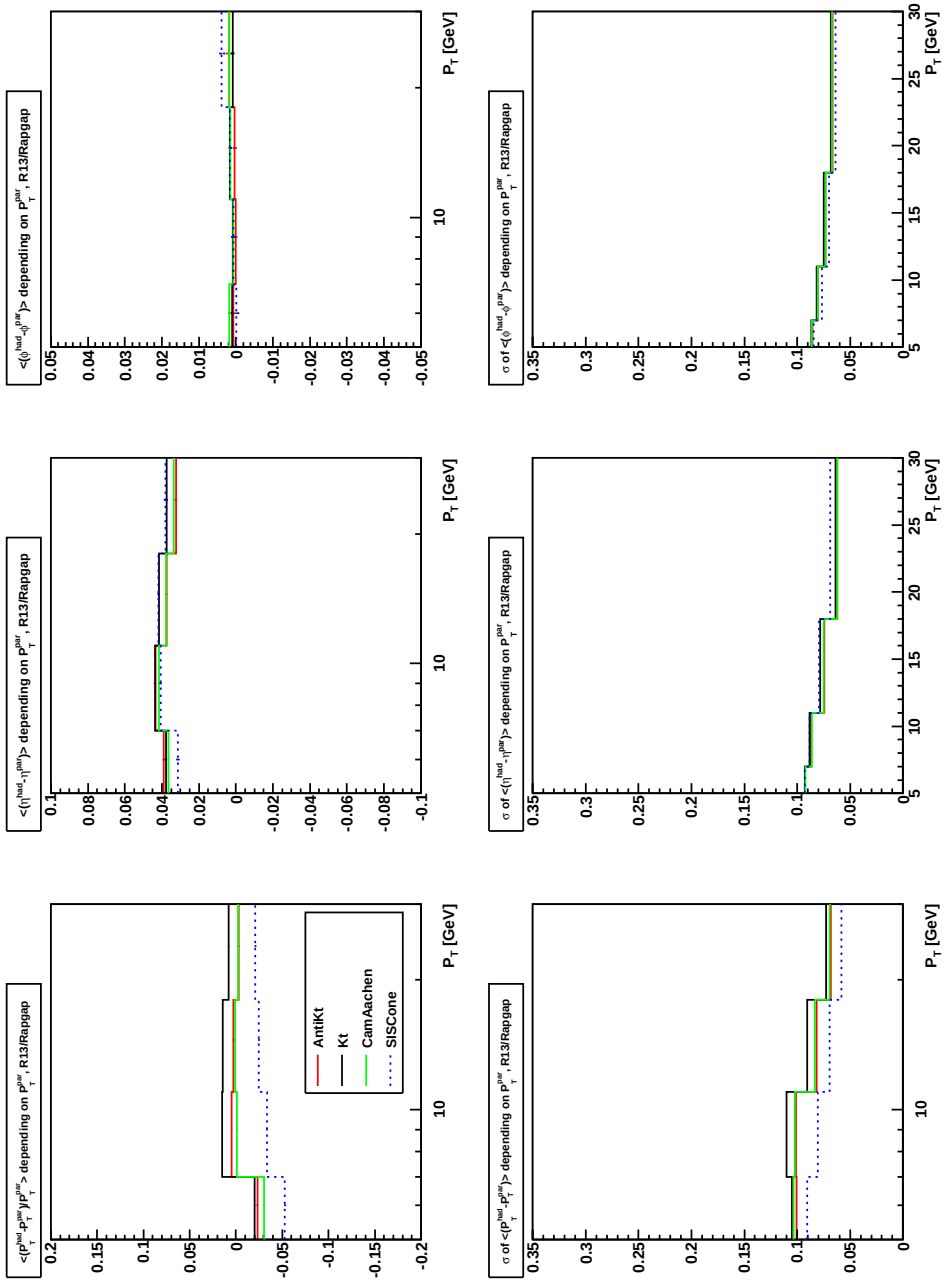

FIG 55. *Deviations of P_T, η and ϕ between parton and hadron level, RAPGAP data, $R_0 = 1.3$*

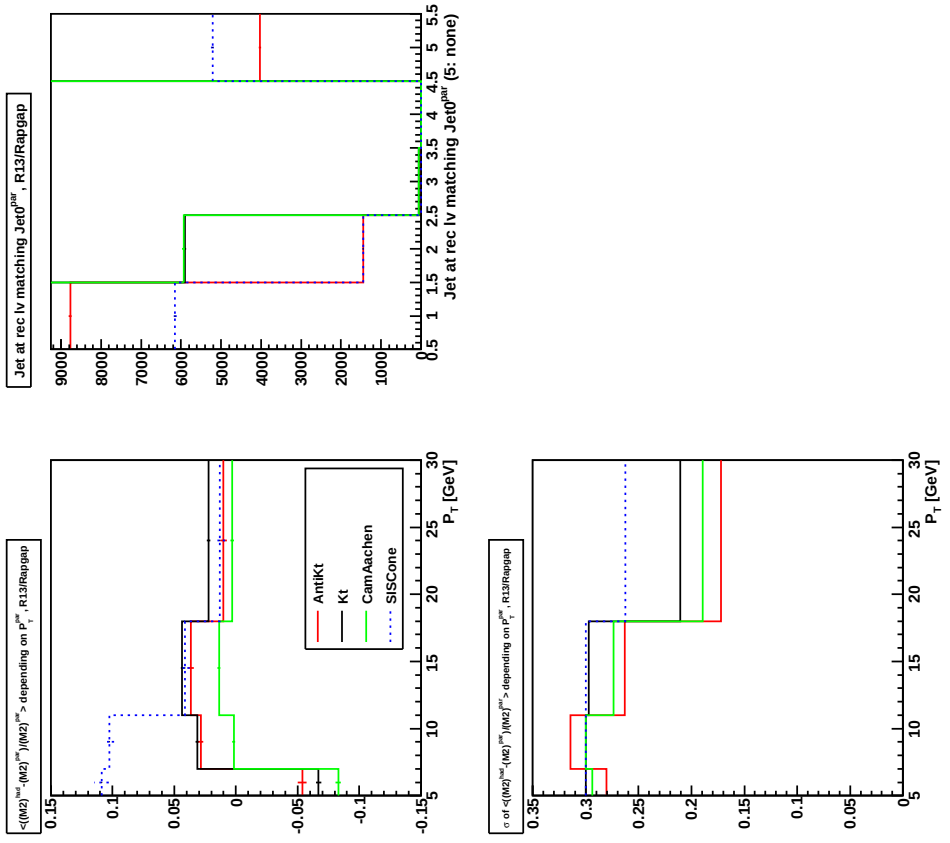

FIG 56. *Deviation of M_{12}^2 between parton and hadron level, RAPGAP data, $R_0 = 1.3$*

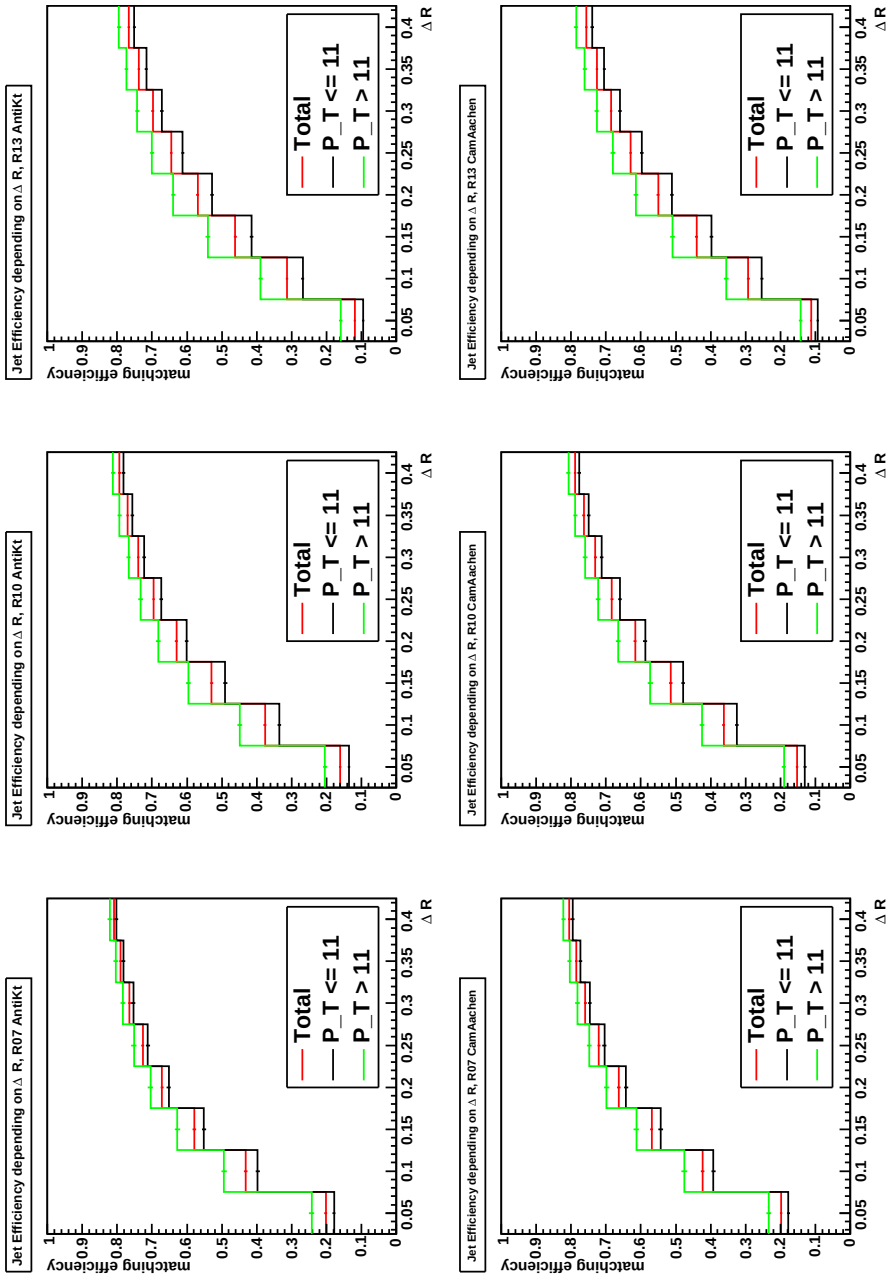

FIG 57. *P_T dependance of the matching efficiency for the anti-k_T and Cambridge/Aachen algorithm*

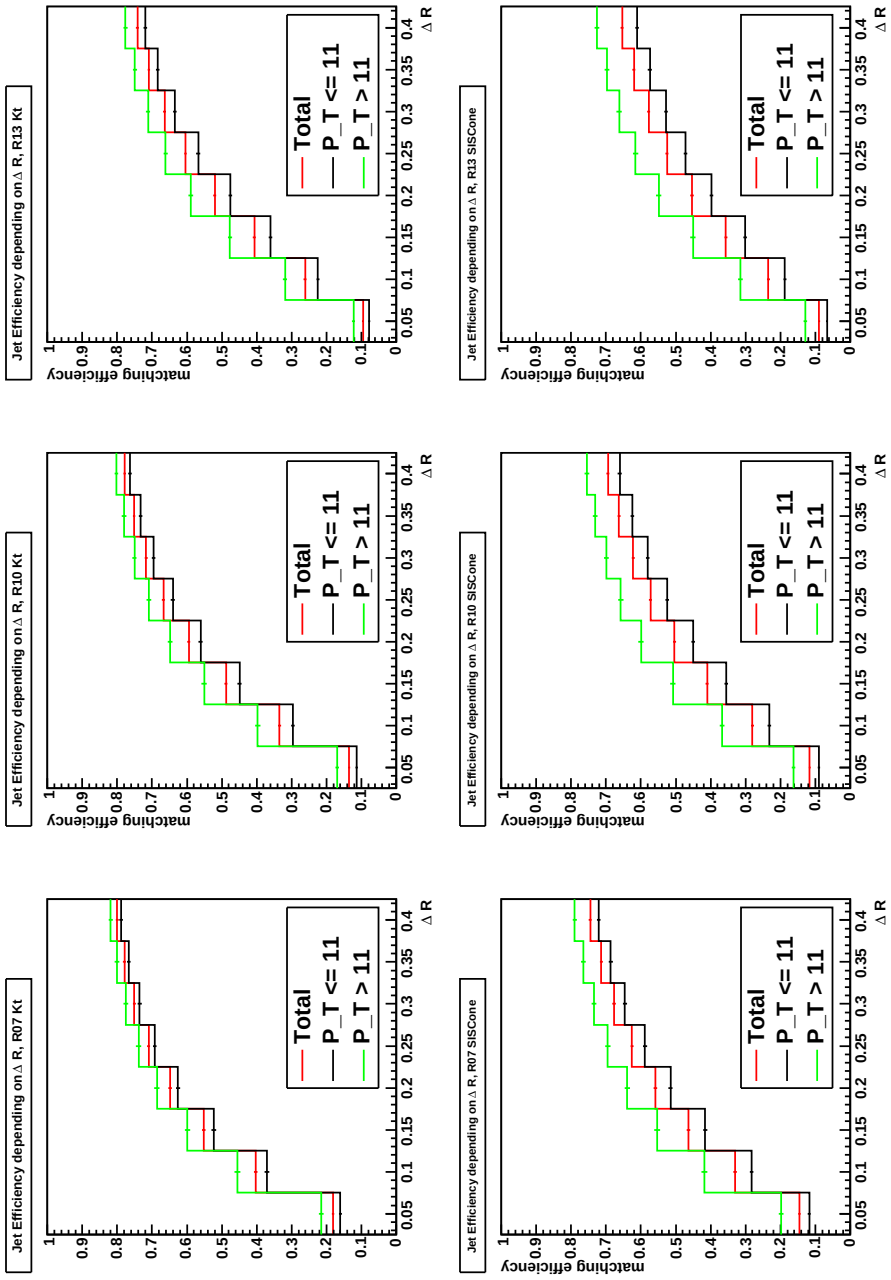

FIG 58. *P_T dependance of the matching efficiency for the k_T and SISCone*