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

**Boris Lemmer**

DESY, H1, MPIM / 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 reconstruction and hadron level on the one hand and hadron and parton level on the other hand. Also, the optimal parameter (jet radius) for the algorithm was searched. For this, a matching procedure was used, trying to find a corresponding jet at another level. The efficiency was studied in dependence of the jet radius.

## 1. MOTIVATION

Within the MIPM 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 $\alpha_s$.

From the energy depositions in the H1 calorimeters the particle flows (jets) have to be reconstructed in a right way. So called *jet algorithms* do this work, and they should do as precisely as possible. Up to now the $k_T$ algorithm is well established and was used for the analysis of HERA and LEP data.

But within the last year the SISCone and the anti-$k_T$ algorithm were developed and the older Cambridge/Aachen was rediscovered. So this variety of usable algorithms let the question rise if the $k_T$ algorithm is still state of the art. Currently two PhD students are working on jet analysis for DIS (Roman Kogler†) and photoproduction (Aziz Dossanov‡). While the DIS analysis is done in this work, Clemens Mellein (also a Summer Student) focuses on photoproduction (1). This work about DIS should be briefly compared to his work about photoproduction in chapter 6.3.

## 2. DIS AT HIGH $Q^2$

The general basics about electron-proton-scattering should be introduced first. To describe the kinematics of an event, only a small set of observables is needed. The four momentum vectors of the proton, incoming electron and outgoing electron will be called $k$, $k'$ and $P$. 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 $E_p$ as the energies of electron and proton beam are fixed. We can define the negative 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 $\theta$ is the angle between the beamline and the scattered electron. The z-axis is aligned in the direction of the proton.

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

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

*Supervisor: Günter Grindhammer, guenterg@desy.de

†kogler@desy.de

‡aziza@desy.de

variable

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

The *Bjorken Scale Variable* x with

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

can be interpreted as the fraction of the proton's four momentum which the parton that scatters with electron carries. To say so, we have to assume that the parton has neither a transverse momentum nor a rest mass. Furthermore, the different partons may not interact with each other during the time of scattering interaction. A suitable reference frame which fulfills all those conditions is the *Breit frame* (Fig. 1). Here the photon does not transmit energy. What

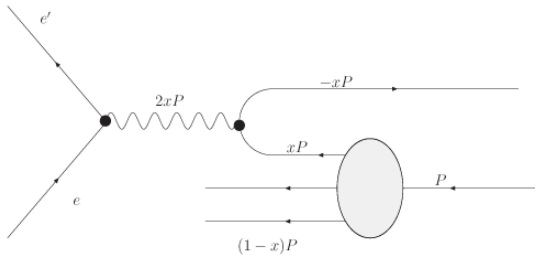


FIG 1. *Breit frame*

happens is that the four momentum vector $xP$ of the partons changes its direction but keeps its absolute value.

Next to the nice interpretation of $x$ the Breit frame has another interesting property which is the reason why we boost every event into this frame before we analyze it. Let us first have a look at what can happen when the electron scatters on a proton (Fig. 2). As $\alpha_s$ is going to be studied, events with only a 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 (Fig. 2 (a)) or if it is an overlap of hadronization of the original parton and a radiated gluon (Fig. 2 (b)). To select events in the order of $\alpha_s$, a boost to the Breit frame is done. Having a look at Fig. 1, one can see that the scattered parton carries no transverse momentum. So a suitable cut on $P_t$ will provide only jets originating from a $\mathcal{O}(\alpha_s)$ process.

## 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 *reconstruction 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 many close-by partons.

Our algorithms should now give the same results at all levels. To check this, we observe the deviations of jet properties at the different levels (See Chapter 6.2).

Requirements are furthermore infrared and collinear (IRC) safety, factorizability and a small renormalization scale dependence. IRC safety claims that the algorithm output - namely the jets - does not contain objects which guide to divergent propagators terms ("soft" particles with small $k$ and jets with a very small angle) when making QCD calculations.

There exist two classes of jet algorithms: clustering and cone type algorithms. While clustering algorithms are favored in the analysis of dilepton collisions (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 typed should briefly be introduced.

### 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.

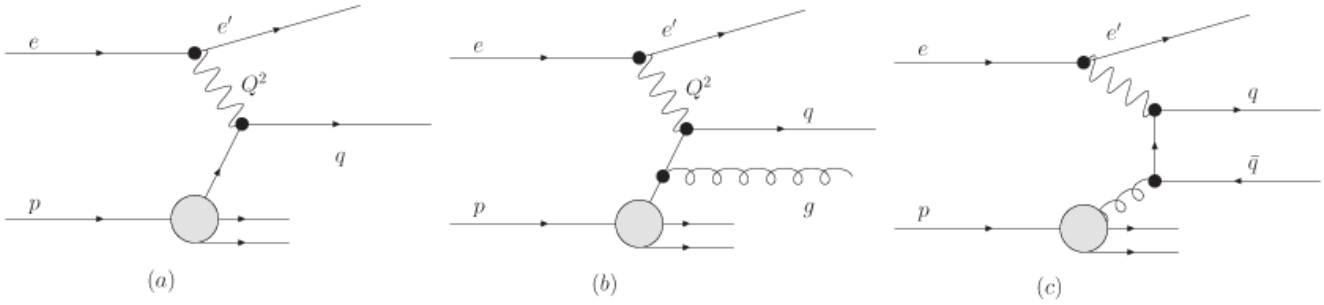


FIG 2. *Deep-inelastic lepton-proton scattering at different orders of $\alpha_s$: (a) Born contribution $\mathcal{O}(1)$, (b) QCD Compton scattering $\mathcal{O}(\alpha_s)$ and (c) boson-gluon fusion $\mathcal{O}(\alpha_s)$*

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

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

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

with the geometrical distance

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

and the transverse momenta $k_t$, the pseudorapidity $\eta$ and the azimuth $\phi$. Each geometrical distance is weighted with a powered momentum. The constant $p$ defines a jet class. $R_0$ is a free parameter for each jet finder algorithm which defines the radius of the jet. In the analysis done within this project, an optimal value for $R_0$ is going to be found.

For $p = 1$ we have a $k_t$ *algorithm* (4). Here, the distance is mainly affected by the $k_t$ of the soft particle/jet. $p = 0$ leads to the *Cambridge/Aachen algorithm* (5), where we only get 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 $k_t$ 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 a merging of two entities the $P_t$ *weighted scheme* is commonly used. As the name suggests, it works the following way:

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

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

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

### 3.3 Cone Type Algorithms

Cone type algorithms work on 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 the following: Starting from angular trial cones pointing in a direction of a more or less arbitrary particle ("seed"), all four momenta within the cone are added and then point in the direction of the next trial cone. This iteration is repeated until the cone axis is not changing anymore and the cone can be defined as stable. All particles within the stable cone are removed from the list of particles in the event and iteration goes on.

The split-merge procedure works as the following: A $P_t$ cut for the protojets is performed. Out of the remaining protojets, the one with highest $P_t$ is chosen. If there is no overlap with another protojet, it's defined as a 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 given by the seedless and IRC safe SISCone algorithm (details in (7)), which will also be used within this study.

## 4. SIMULATION AND SETTINGS

To compare data on all levels (reconstruction, 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 calculating the hadronization procedure. To take only those events from the simulation belonging to an *neutral current DIS at high* $Q^2$ event, 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 $Q^2$. 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 claimed 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, it 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:

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

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

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

$$\eta = \frac{1}{2} \ln \frac{E + P_z}{E - P_z} \quad (4.7)$$

The detector efficiency for an $\eta$ out of this range gets too bad which makes this cut necessary.

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 $P_T$, $\phi$ and $\eta$ 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 (≡ reconstruction 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 reconstruction level is taken as comparison level):

1. Define a maximal matching distance $\Delta R$ (as in 3.3) that is fixed during the whole analysis and will be called $\Delta R_{max}$ (which one 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 $\Delta R_{max}$.
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 $\Delta R_{max}$.
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 other level have a distance larger than $\Delta R_{max}$, the event is rejected and a marker for the jet 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 immediately defined 2 jets as matching, when they are both hardest jets and within the $\Delta R$. So only if there is a mismatch between the hardest, the next closest will be used. A deviation between both pro-

cedures is not expected.

## 6. RESULTS

If nothing else is said, all the following results where obtained by using the MC generator DJANGOH. If there have been 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*** $\Delta R_{max}$ As described in the matching procedure, an optimal $\Delta R_{max}$ 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 $\Delta R_{max}$ chosen too small will end in low statics caused by a smaller matching efficiency.

All further analysis concerning the matching efficiency will be made between hadron and reconstructions level. To get a first impression of what a good $\Delta R_{max}$ could be, one can have a look at figure 3. It shows the distance distribution of two matching

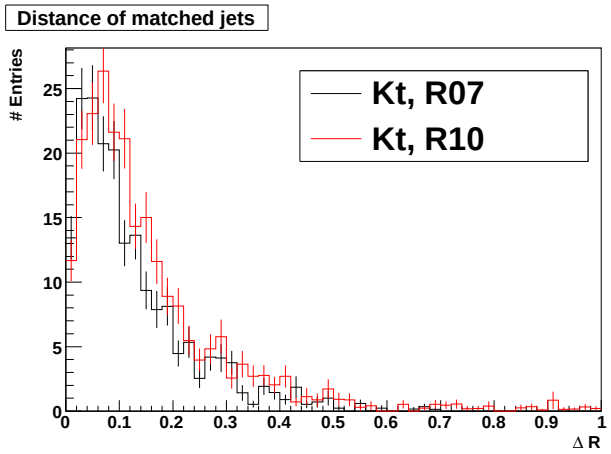


FIG 3. *Distance distribution of two matched jets*

jets. In this run $\Delta R_{max}$ was set to 1.0. A change in the jet radius $R_0$ from 0.7 to 1.0 (please do not get confused with $\Delta R_{max}$) has no significant influence.

To make a decision for $\Delta R_{max}$, 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 $\Delta R_{max}$ 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 $\Delta R_{max}$ is shown in figure 4 as an example for the anti-$k_T$ algorithm.

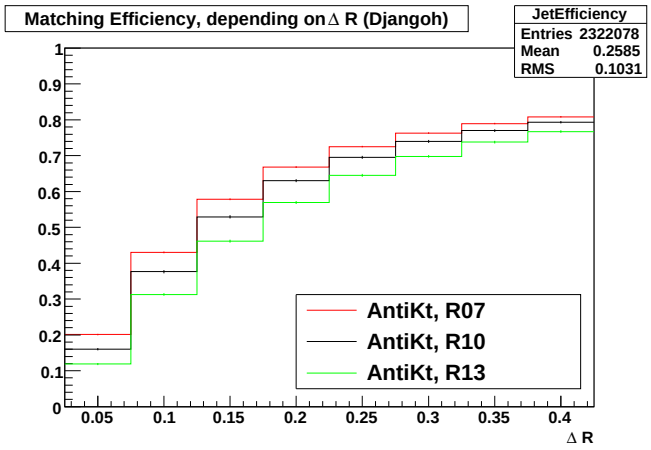


FIG 4. *Matching efficiency depending on* $\Delta R$

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 of 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 $\Delta R_{max}$ dependance of the efficiency, I decided to set it to $\Delta R_{max} = 0.2$ as efficiency does not increase that strong anymore above 0.2.

*6.1.2* ***Matching efficiency dependances*** It should be checked how the matching efficiency varies depending on different observables like $P_T$, $\eta$, $\phi$, 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 has been made. The only question was if there were two jets on both level. If so, they were compared according their indexing in the jet array, namely in the order of $P_T$.

- $P_T$: Efficiency increases for higher $P_T$ what will be analyzed later in chapter 6.3 (see figure 12). It is intuitive that a jet with higher $P_T$ is defined clearer and thereby easier to match in its properties on all levels.

- $\eta$: Figure 13 shows the $\eta$ dependance of the efficiency. It decreases for large $|\eta|$. This effect is caused by inefficiencies in the detector which are also a reason for the global $\eta$ cuts.
- $\phi$: 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 $\phi$ distribution changes after such a boost.
- $M_{12}^2$: The effect is similar and comparable to the $P_T$ dependance (Figure 15).
- $Q^2$: Large $Q^2$ end up in a worse efficiency (Figure 16).

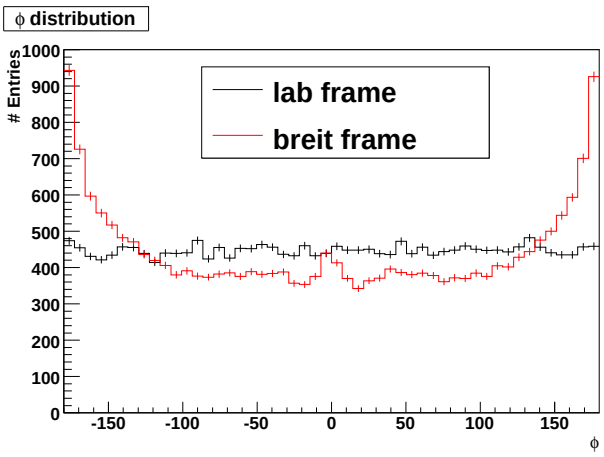


FIG 5. *$\phi$ distribution in laboratory rest frame and Breit frame*

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 be matched did not fulfill the necessary cuts.

## 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 reconstruction level in that way that the deviations between reconstruction 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 $P_T$, $\eta$, $\phi$ 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 6 depending on $\phi$). Later on the differences will also be analyzed under the aspect of $\eta$, $\phi$, $M_{12}^2$ and $Q^2$ dependance.

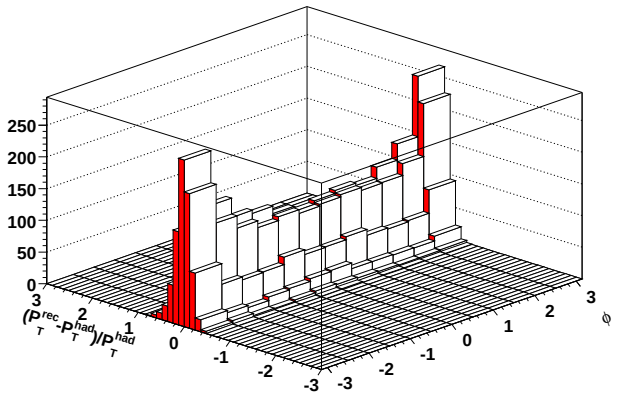

FIG 6. *$P_T$ deviation between reconstruction and hadron level depending on $\phi$*

The best thing next to a delta function at zero would be a gaussian like distribution around zero. The mean value (wanted to be zero) and the $\sigma$ of all those distributions (wanted to be 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 histograms according to the following analysis can be found in figures 17 - 28):

- $\Delta P_T$: 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.
- $\Delta\eta$: $\eta$ is both constantly and only slightly too small at the reconstruction level.
- $\Delta\phi$: The deviation of $\phi$ is something which is not at all obvious. An example is given is figure 7. The larger $P_T$ gets, to larger the deviation becomes (where the reconstructed data is too

large). The origin for that might be a problem of the track reconstruction in the event generator, but a further investigation has not been made yet.

- $\Delta M_{12}^2$: 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 was for the other observables as there were now even two jets involved and no matching was done.

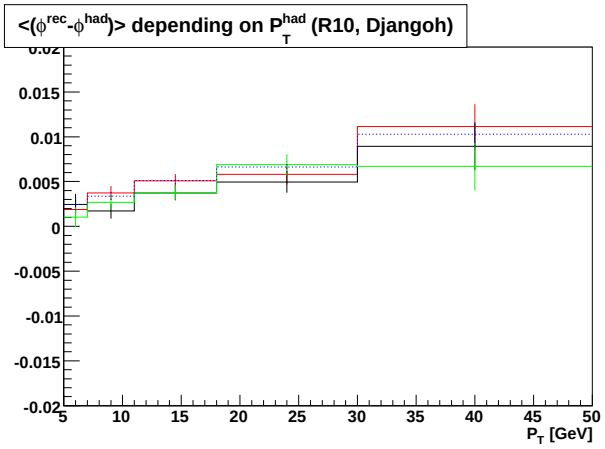


FIG 7. *Shift of $\phi$ between reconstruction and hadron level*

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 $A$ 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 $\phi$ deviations, all algorithms behave similar.
- anti-$k_T$ is closest to zero concerning the $\eta$ 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 $\sigma$.
- As $\sigma$ for the $M_{12}^2$ is very radius sensitive, SISCone is worst in this part.

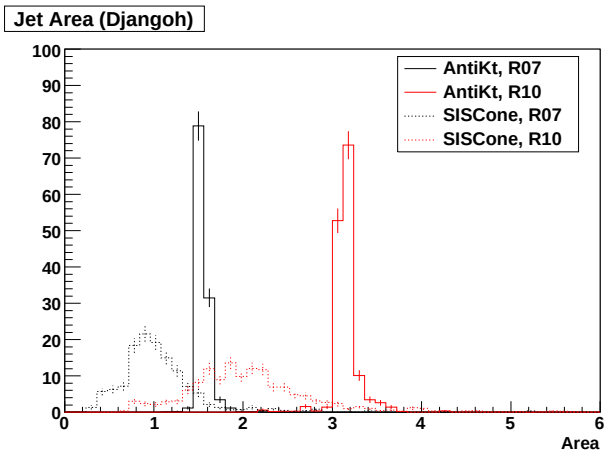


FIG 8. *Comparison of the jet areas*

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

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

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

- $\Delta P_T$: $\eta > 1.0$ gives main contribution for the $P_T$ deviations. The effect gets eased by a smaller $R_0$.
- $\Delta\eta$: The main contribution for the too low $\eta$ on reconstruction level can definitely be assigned to low $Q^2$ events.
- $\Delta\phi$: Again (as for $\Delta P_T$), $\eta > 1.0$ gives main contribution. The higher $\eta$, the worse.
- $\Delta M_{12}^2$: Low $Q^2$ ($Q^2 < 600\text{GeV}^2$) events cause largest deviations. There are also strong effects for both angles, but unfortunately I have no idea how to interpret them.

*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 reconstruction and hadron level was done again. But we will focus only on the first part with the $P_T$ dependance of the deviations. First about the deviations in general (corresponding plots can be found in figures 45 - 56):

- $\Delta P_T$: Small $P_T$ ($< 10$ GeV) are calculated too low at the hadron level. The rest is very close to zero.
- $\Delta\eta$: Independent of $P_T$, $\eta$ is constantly about 4-5% too high on hadron level.
- $\Delta\phi$: Constantly no deviation.
- $\Delta M_{12}^2$: Very interesting observations were made. The smaller $P_T$ gets, the more drastic $M_{12}^2$ gets too small on the 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 the hadron level. And the best radius setting for SISCone is the worst for the clustering ones.

Now about the advantages of each algorithm:

- All clustering algorithms behave similar. $k_T$ is slightly better in the $P_T$ matching sector. SISCone still 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 $\sigma$: The SISCone distribution is much broader than those of the clustering algorithms.

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

- $\Delta P_T$: $R_0 = 1.0$ leads to best results.
- $\Delta\eta$: A $R_0$ as low as possible is preferred.
- $\Delta\phi$: No preference.
- $\Delta M_{12}^2$: 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". So I will focus only on systematic differences between both.

- The problem of $\eta$ which was constantly too low on the reconstruction level (compared to hadron level) is "mirrored" for RAPGAP. Now $\eta$ is constantly too high, the absolute value of the deviation stays almost the same ($0.005 - 0.01$).
- Remember that $\phi$ deviated in that way that it was too large on reconstruction level. The effect got more intensive for higher $P_T$. Now for RAPGAP the effect is lowered 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 $\sigma$ is no longer so horrible for the SISCone with RAPGAP.

So especially the big $\phi$ deviation mystery could be solved when having a closer look at the differences between RAPGAP and DJANGOH. Unfortunately, this goes beyond the scope of this study.

### 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) (low $Q^2$, $P_T$ cut for 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 $\phi$ shift mystery. As it was weakened with the RAPGAP dataset, one could

compare PYTHIA which was used in (1) with DJANGOH (same behavior) and RAPGAP.

- We both got worse efficiencies and larger deviations at a high $\eta$, which then has to be allocated to the bad detector efficiency in this region.
- Where in (1) the matching efficiency had no $\phi$ dependance, I observed a strong variation with $\phi$. I already mentioned that there might be a connection to the boost. This theory 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. Where in this study we get values between 60 and 70%, in (1) values of more than 90% were 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 detailled in figures 57 and 58. As I also studied high $Q^2$ events and efficiency

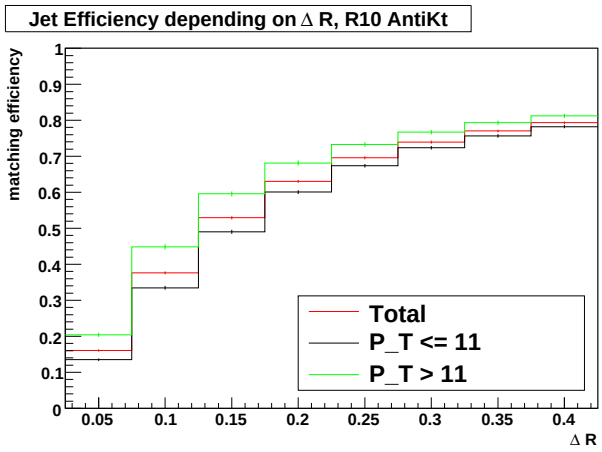


FIG 9. $P_T$ *dependance of the matching efficiency*

decreased for higher $Q^2$, we get the other big contribution for the lower 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 $\sigma$. 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) $\sigma$ and the best matching efficiency, which is also somehow related. The $k_T$ behaves similar but its strengths are not as distinct. Also, runtime was not considered.

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

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

The $\phi$ shifting mystery needs clarification. An analysis of the MC generators the their differences might help.

## 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 with cake and chocolate, having the nicest people and best scientists around me. Without those circumstances this work would not have been possible in 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 $\Delta R_{max}$
- **Figure 12:** Jet efficiency depending on $P_T$
- **Figure 13:** Jet efficiency depending on $\eta$
- **Figure 14:** Jet efficiency depending on $\phi$
- **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 $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, DJANGOH data, $R_0 = 0.7$
- **Figure 18:** Deviation of $M_{12}^2$ between reconstruction and hadron level, DJANGOH data, $R_0 = 0.7$
- **Figure 19:** Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, DJANGOH data, $R_0 = 1.0$
- **Figure 20:** Deviation of $M_{12}^2$ between reconstruction and hadron level, DJANGOH data, $R_0 = 1.0$
- **Figure 21:** Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, DJANGOH data, $R_0 = 1.3$
- **Figure 22:** Deviation of $M_{12}^2$ between reconstruction and hadron level, DJANGOH data, $R_0 = 1.3$
- **Figure 23:** Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, RAPGAP data, $R_0 = 0.7$
- **Figure 24:** Deviation of $M_{12}^2$ between reconstruction and hadron level, RAPGAP data, $R_0 = 0.7$
- **Figure 25:** Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, RAPGAP data, $R_0 = 1.0$
- **Figure 26:** Deviation of $M_{12}^2$ between reconstruction and hadron level, RAPGAP data, $R_0 = 1.0$
- **Figure 27:** Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, RAPGAP data, $R_0 = 1.3$
- **Figure 28:** Deviation of $M_{12}^2$ between reconstruction 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 $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, DJANGOH data, Anti-$k_T$
- **Figure 30:** Deviation of $M_{12}^2$ between reconstruction and hadron level, DJANGOH data,

Anti-$k_T$

- **Figure 31:** Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, DJANGOH data, Cambridge/Aachen
- **Figure 32:** Deviation of $M_{12}^2$ between reconstruction and hadron level, DJANGOH data, Cambridge/Aachen
- **Figure 33:** Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, DJANGOH data, $k_T$
- **Figure 34:** Deviation of $M_{12}^2$ between reconstruction and hadron level, DJANGOH data, $k_T$
- **Figure 35:** Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, DJANGOH data, SISCone
- **Figure 36:** Deviation of $M_{12}^2$ between reconstruction and hadron level, DJANGOH data, SISCone

## APPENDIX D: DEVIATIONS BETWEEN RECONSTRUCTION AND HADRON LEVEL DEPENDING ON $\phi$, $\eta$, $M_{12}^2$ AND $Q^2$, COMPARING RADII

- **Figure 37:** Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, DJANGOH data, $\eta$ dependance
- **Figure 38:** Deviation of $M_{12}^2$ between reconstruction and hadron level, DJANGOH data, $\eta$ dependance
- **Figure 39:** Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, DJANGOH data, $\phi$ dependance
- **Figure 40:** Deviation of $M_{12}^2$ between reconstruction and hadron level, DJANGOH data, $\phi$ dependance
- **Figure 41:** Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, DJANGOH data, $M_{12}^2$ dependance
- **Figure 42:** Deviation of $M_{12}^2$ between reconstruction and hadron level, DJANGOH data, $M_{12}^2$ dependance
- **Figure 43:** Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, DJANGOH data, $Q^2$ dependance
- **Figure 44:** Deviation of $M_{12}^2$ between reconstruction and hadron level, DJANGOH data, $Q^2$ dependance

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

- **Figure 45:** Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, DJANGOH data, $R_0 = 0.7$
- **Figure 46:** Deviation of $M_{12}^2$ between reconstruction and hadron level, DJANGOH data, $R_0 = 0.7$
- **Figure 47:** Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, DJANGOH data, $R_0 = 1.0$
- **Figure 48:** Deviation of $M_{12}^2$ between reconstruction and hadron level, DJANGOH data, $R_0 = 1.0$
- **Figure 49:** Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, DJANGOH data, $R_0 = 1.3$
- **Figure 50:** Deviation of $M_{12}^2$ between reconstruction and hadron level, DJANGOH data, $R_0 = 1.3$
- **Figure 51:** Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, RAPGAP data, $R_0 = 0.7$
- **Figure 52:** Deviation of $M_{12}^2$ between reconstruction and hadron level, RAPGAP data, $R_0 = 0.7$
- **Figure 53:** Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, RAPGAP data, $R_0 = 1.0$
- **Figure 54:** Deviation of $M_{12}^2$ between reconstruction and hadron level, RAPGAP data, $R_0 = 1.0$
- **Figure 55:** Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, RAPGAP data, $R_0 = 1.3$
- **Figure 56:** Deviation of $M_{12}^2$ between reconstruction and hadron level, RAPGAP data, $R_0 = 1.3$

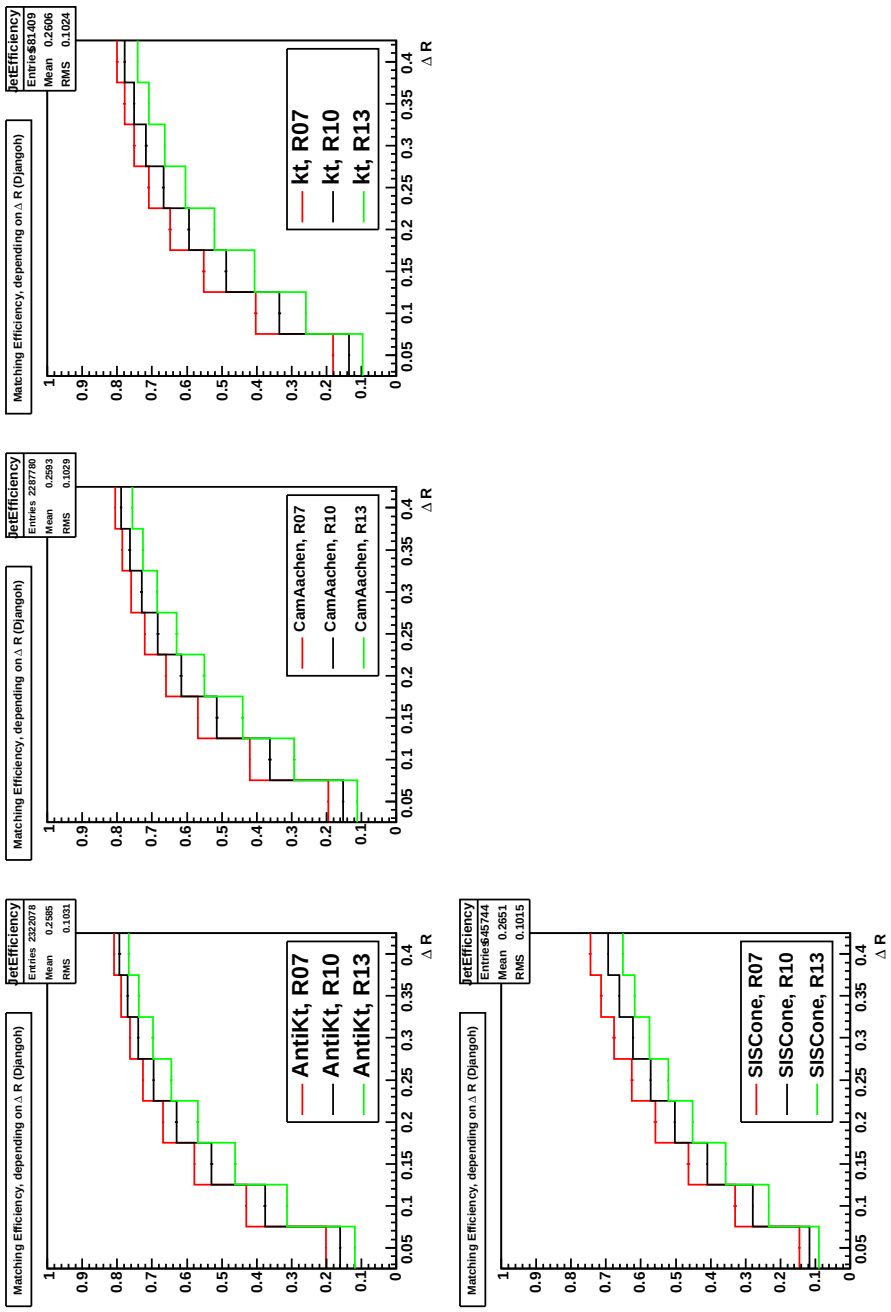


FIG 11. *Jet efficiency depending on the maximal jet distance* $\Delta R_{max}$

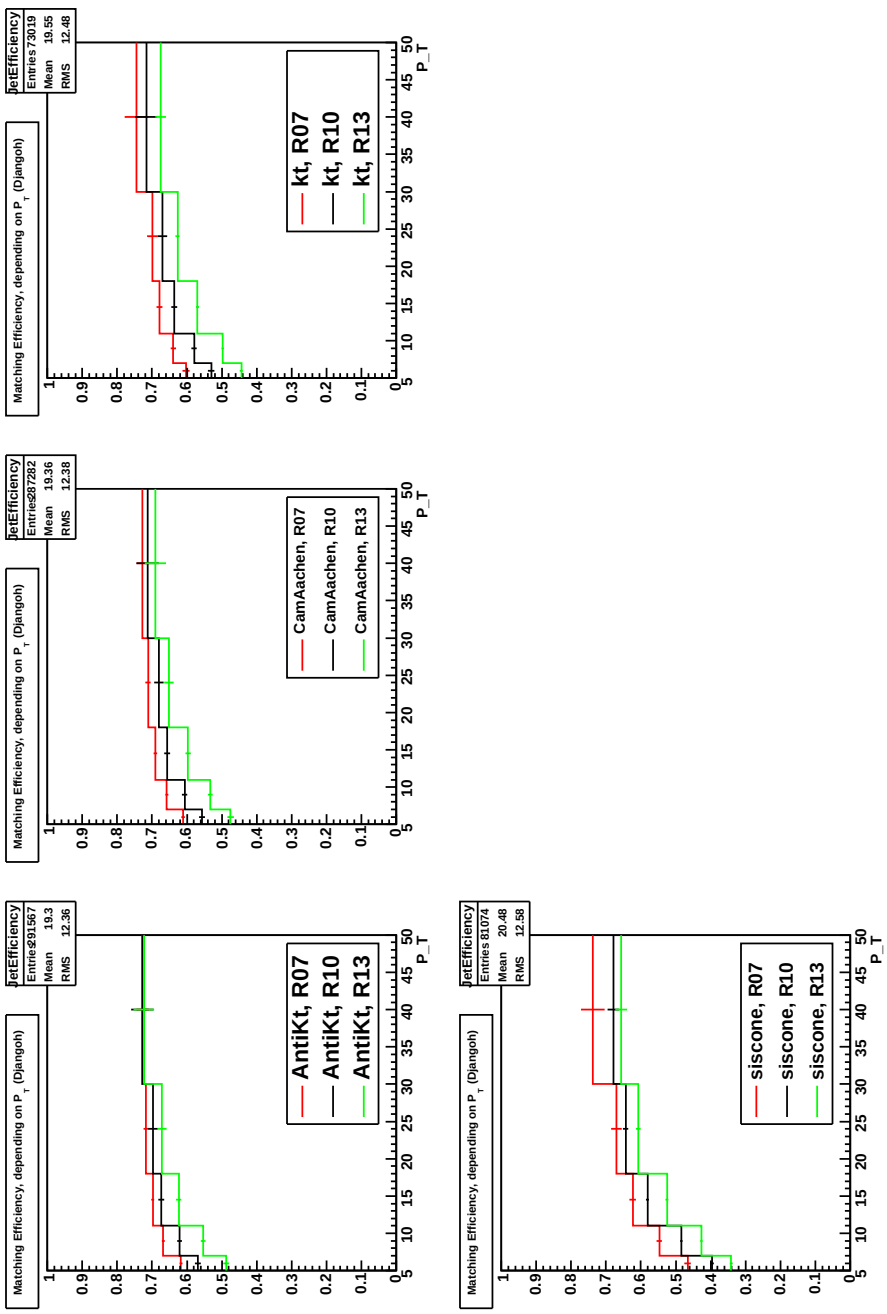


FIG 12. *Jet efficiency depending on $P_T$*

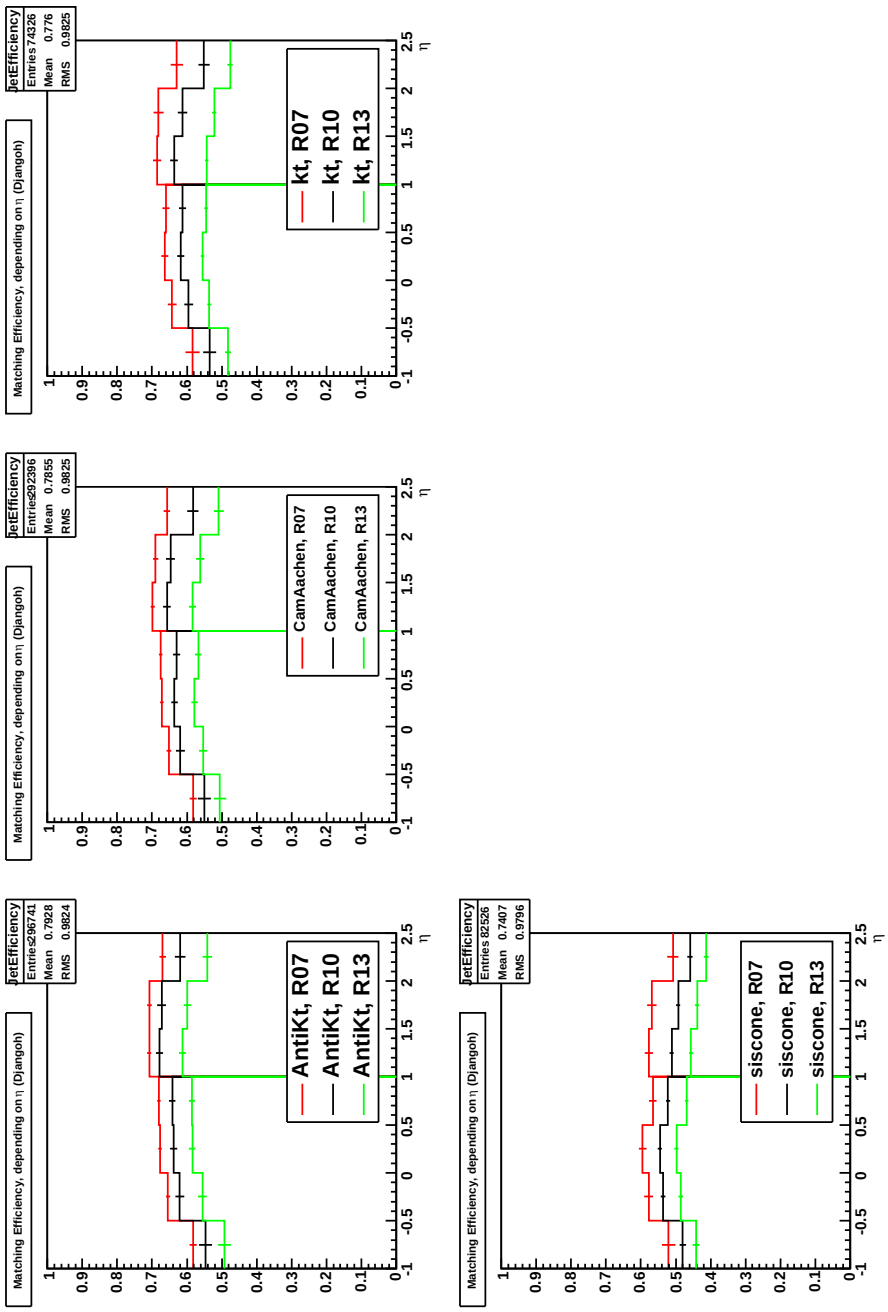


FIG 13. *Jet efficiency depending on η*

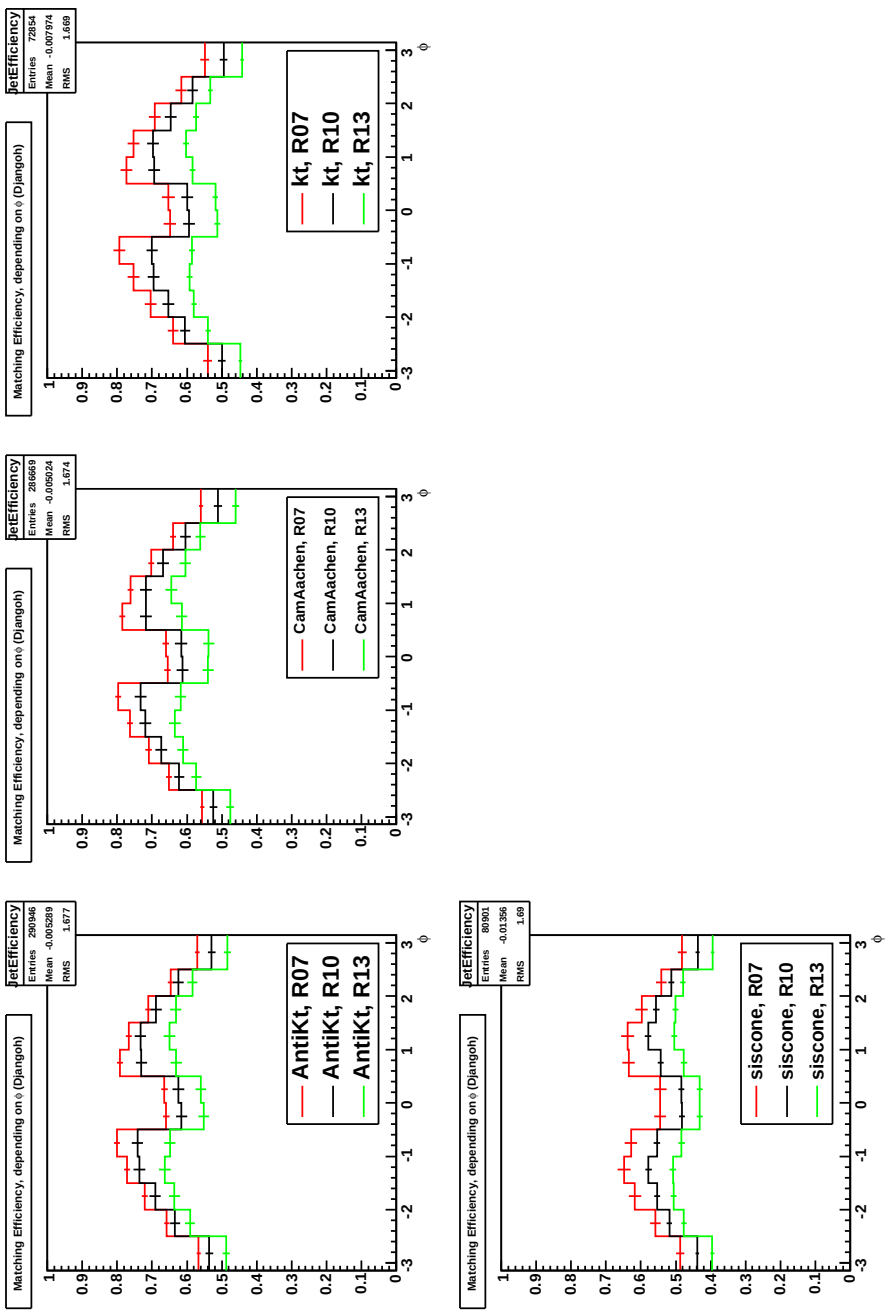


FIG 14. *Jet efficiency depending on $\phi$*

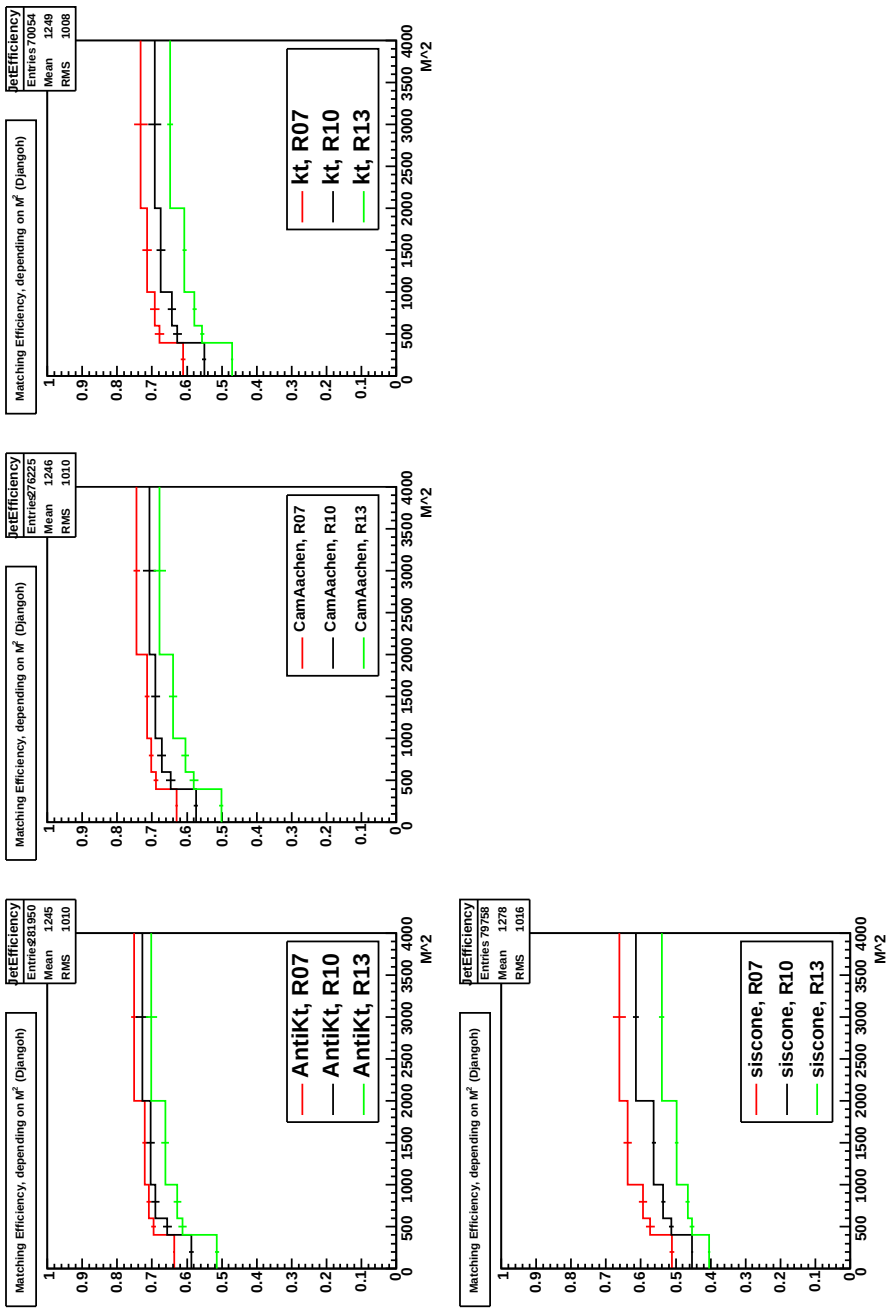


FIG 15. *Jet efficiency depending on* $M_{12}^2$

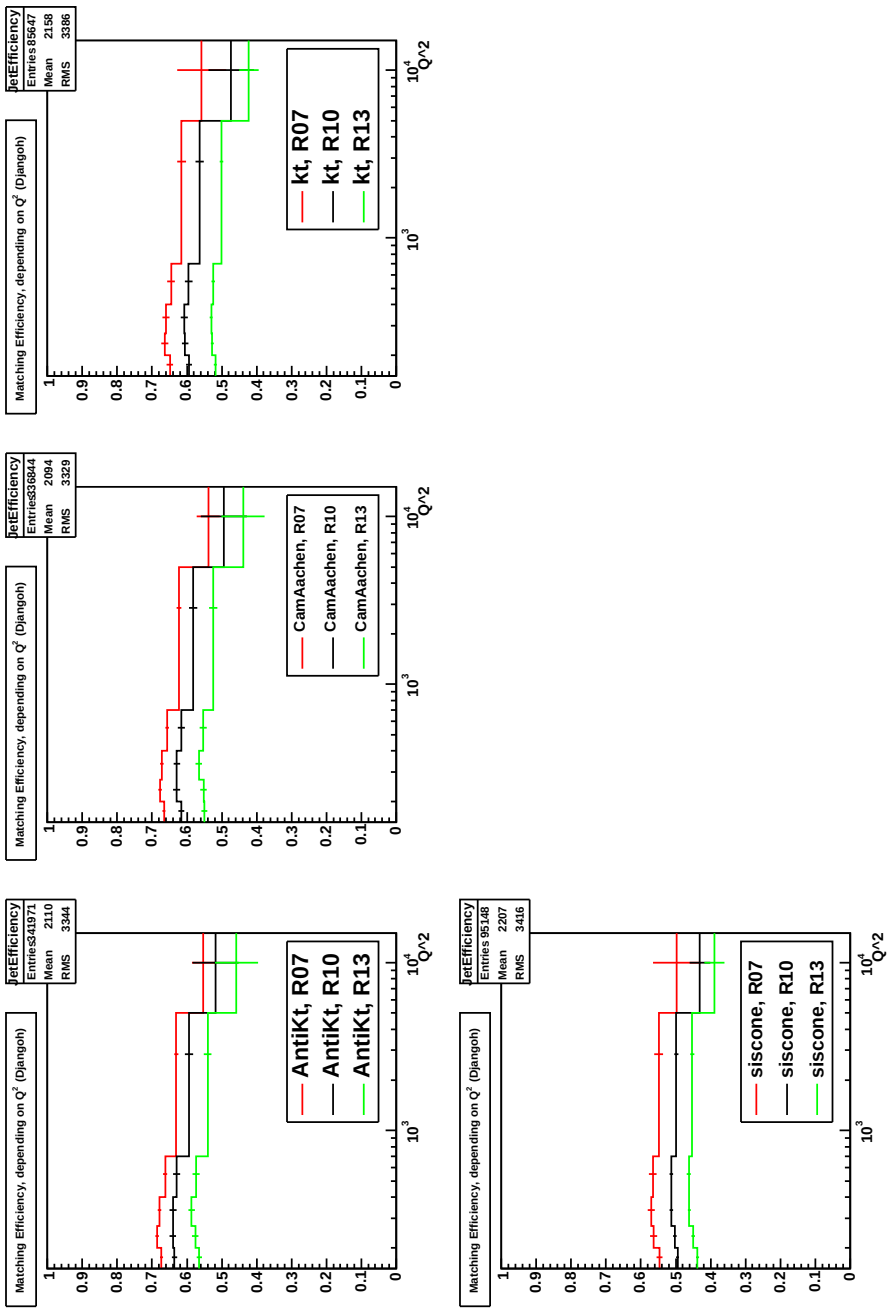


FIG 16. *Jet efficiency depending on* $Q^2$

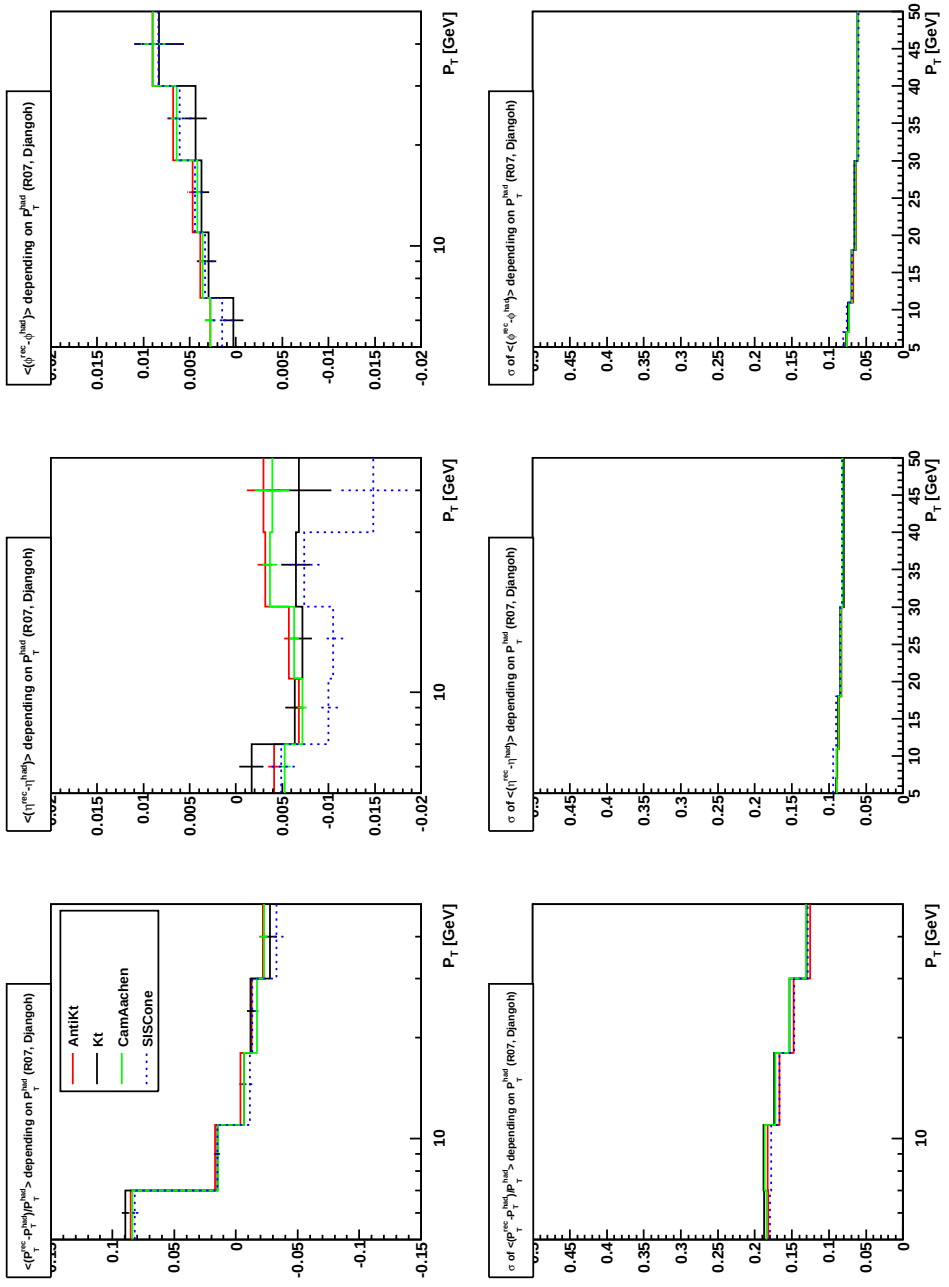


FIG 17. *Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, DJANGOH data, $R_0 = 0.7$*

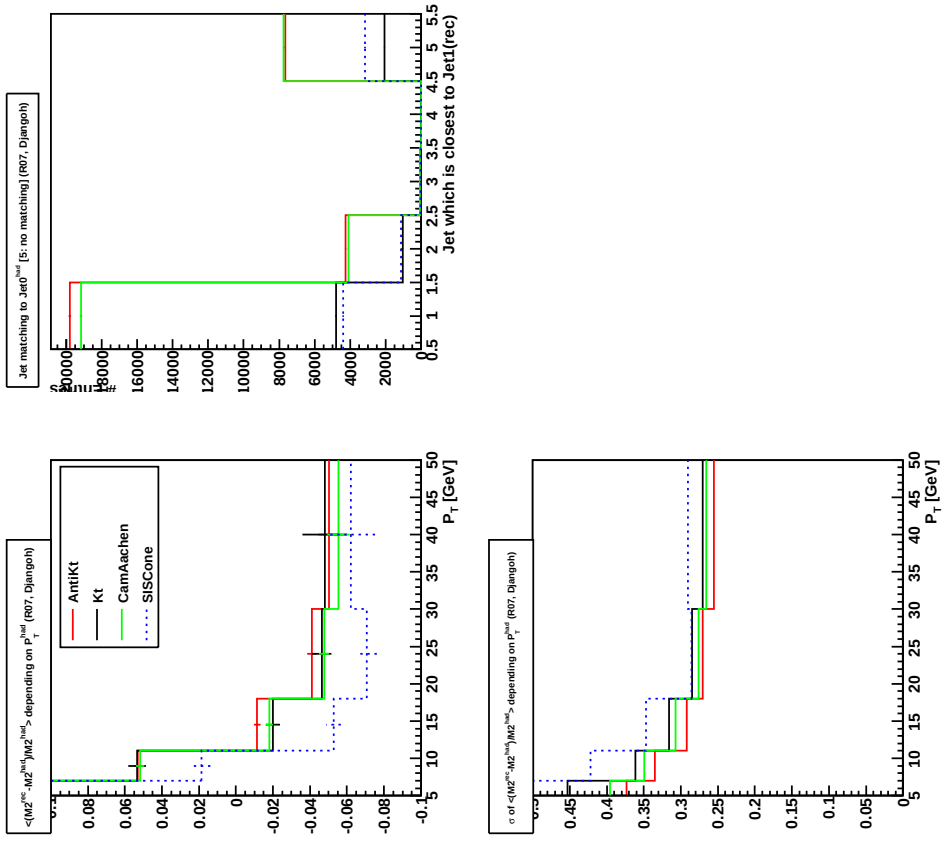


FIG 18. *Deviation of $M_{12}^2$ between reconstruction and hadron level, DJANGOH data, $R_0 = 0.7$*

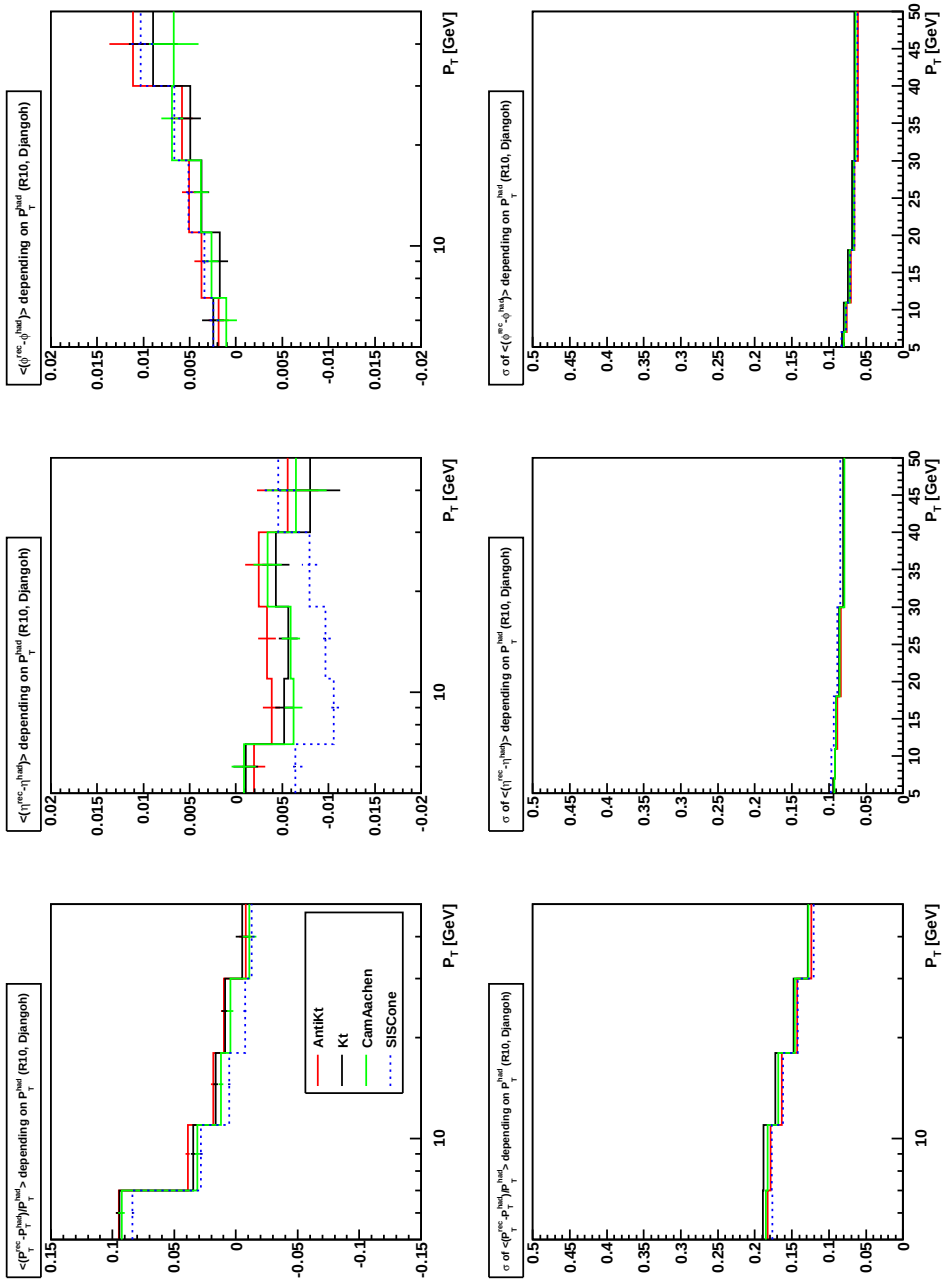


FIG 19. *Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, DJANGOH data, $R_0 = 1.0$*

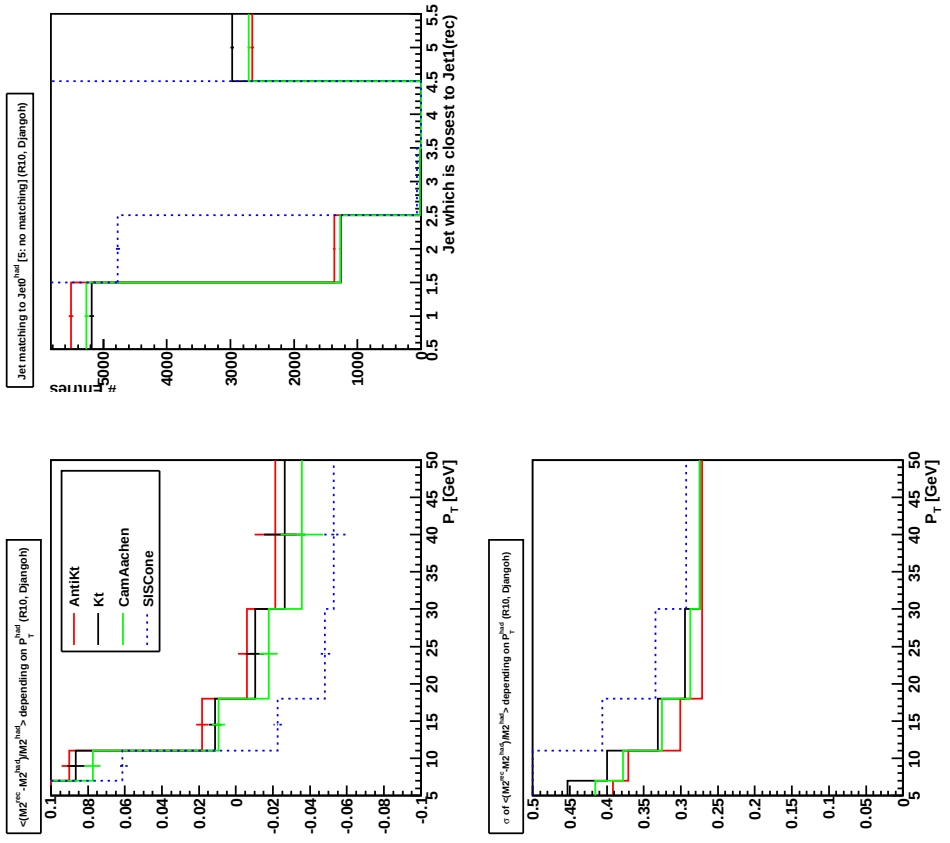


FIG 20. *Deviation of $M_{12}^2$ between reconstruction and hadron level, DJANGOH data, $R_0 = 1.0$*

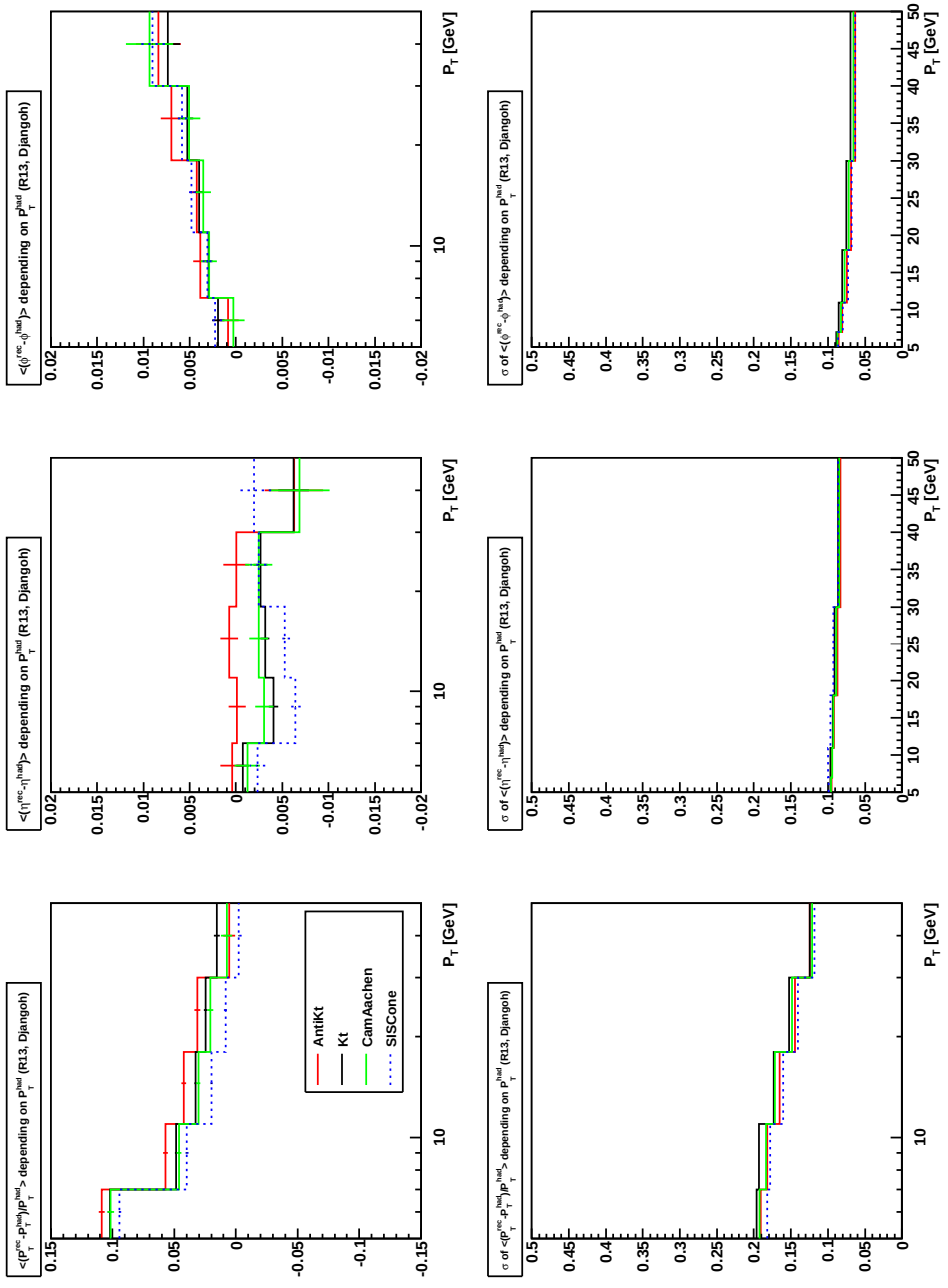


FIG 21. *Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, DJANGOH data, $R_0 = 1.3$*

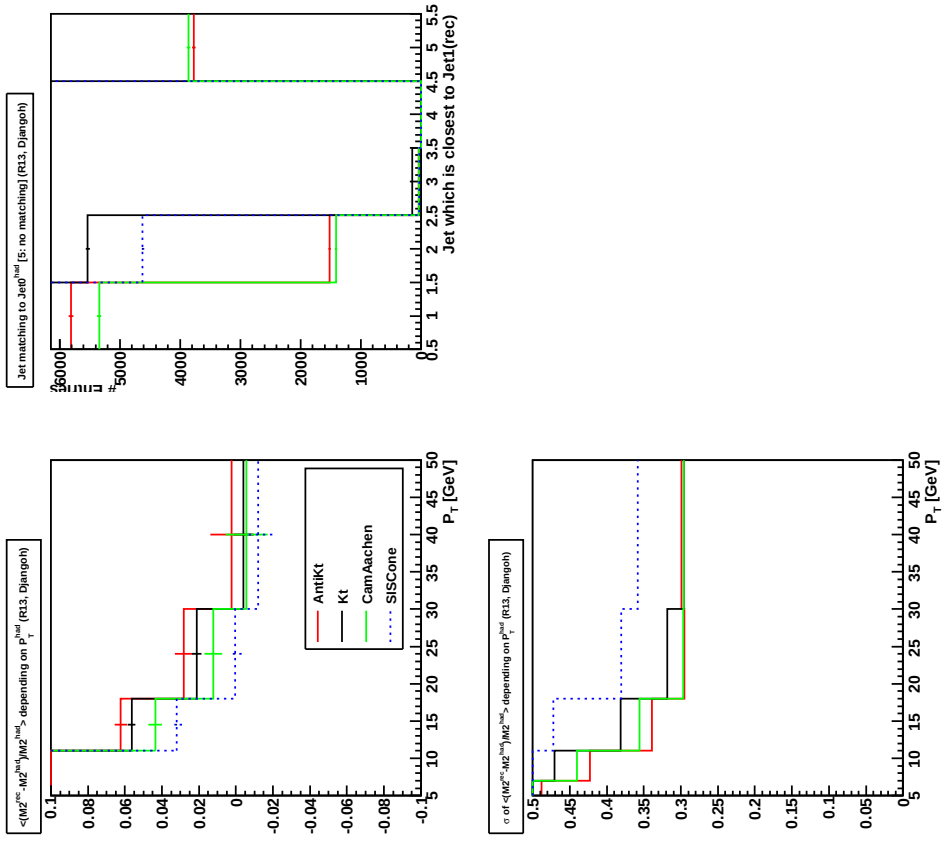


FIG 22. *Deviation of $M_{12}^2$ between reconstruction and hadron level, DJANGOH data, $R_0 = 1.3$*

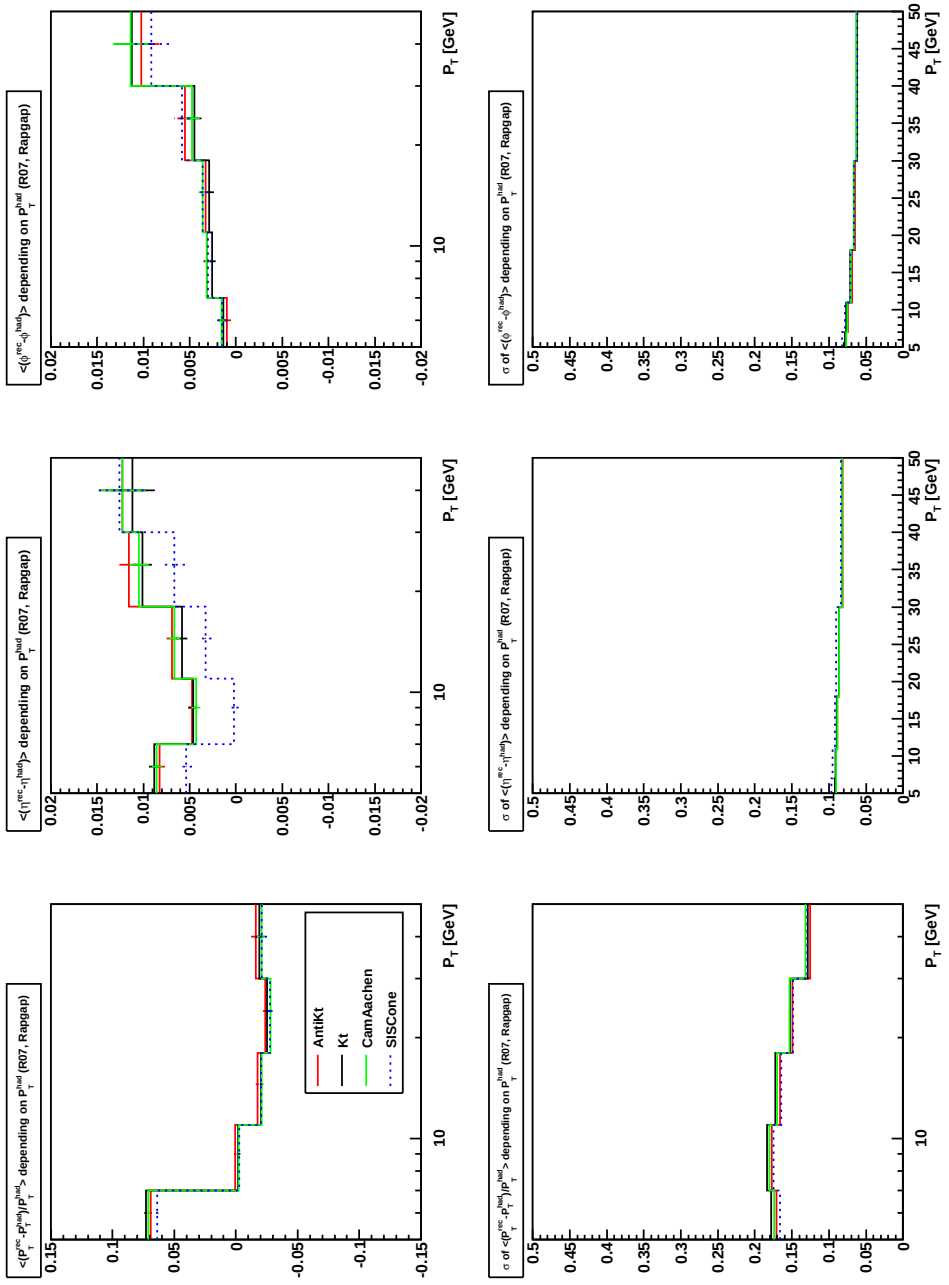


FIG 23. *Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, RAPGAP data, $R_0 = 0.7$*

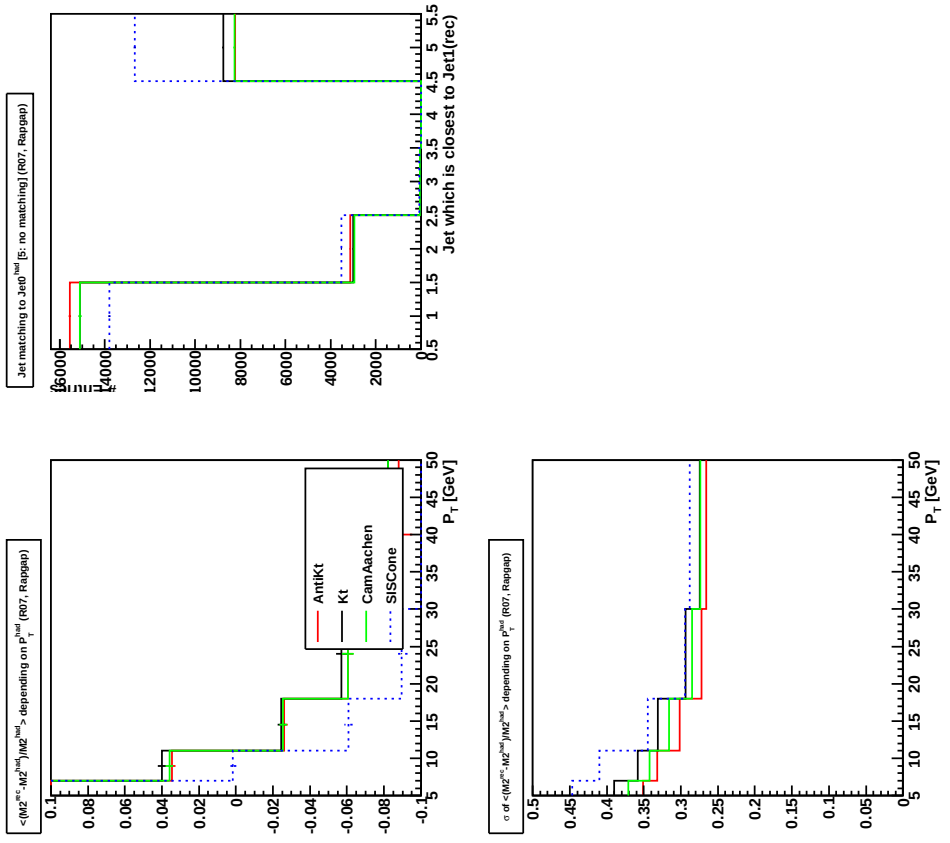


FIG 24. *Deviation of $M_{12}^2$ between reconstruction and hadron level, RAPGAP data, $R_0 = 0.7$*

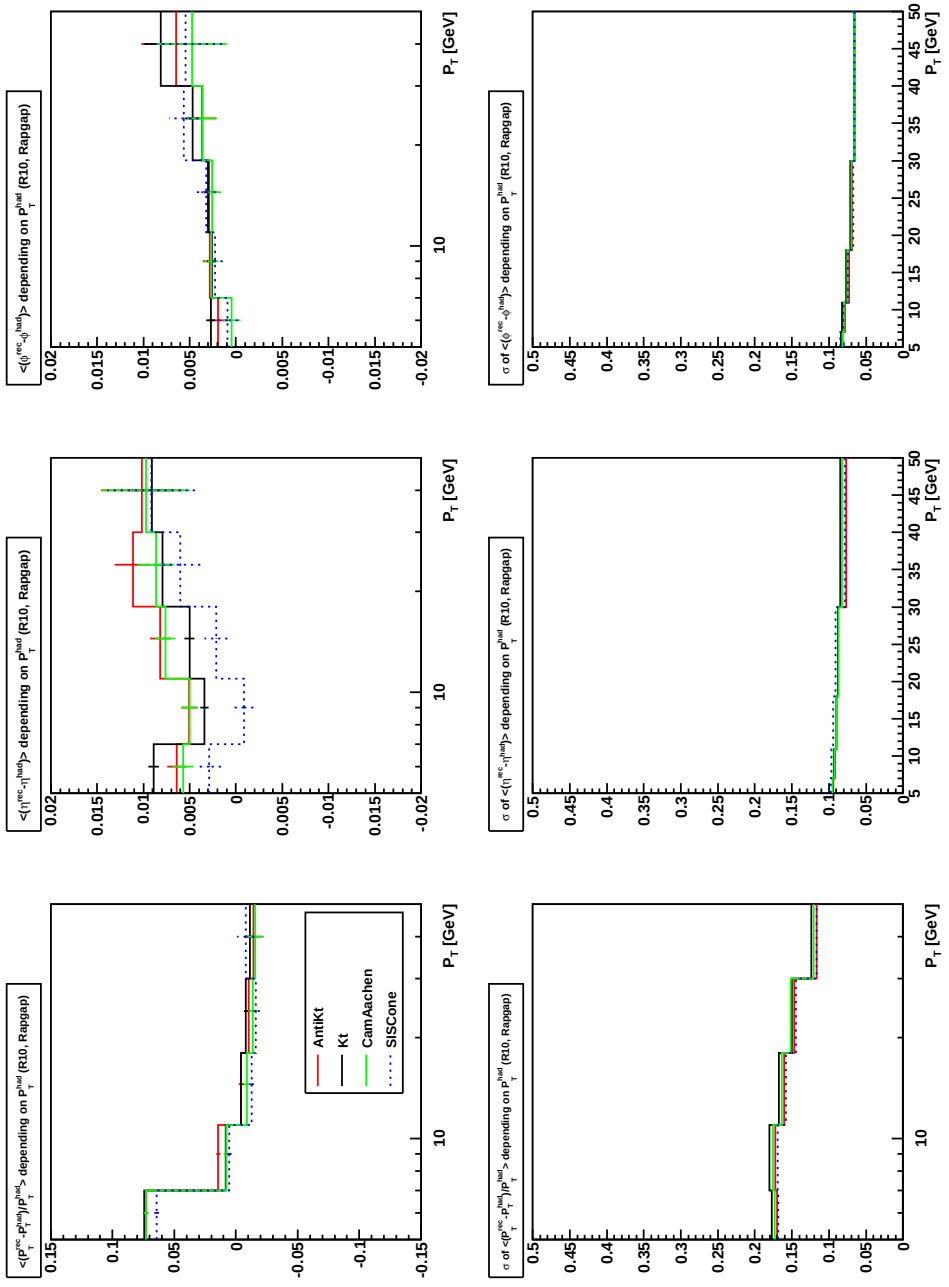


FIG 25. *Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, RAPGAP data, $R_0 = 1.0$*

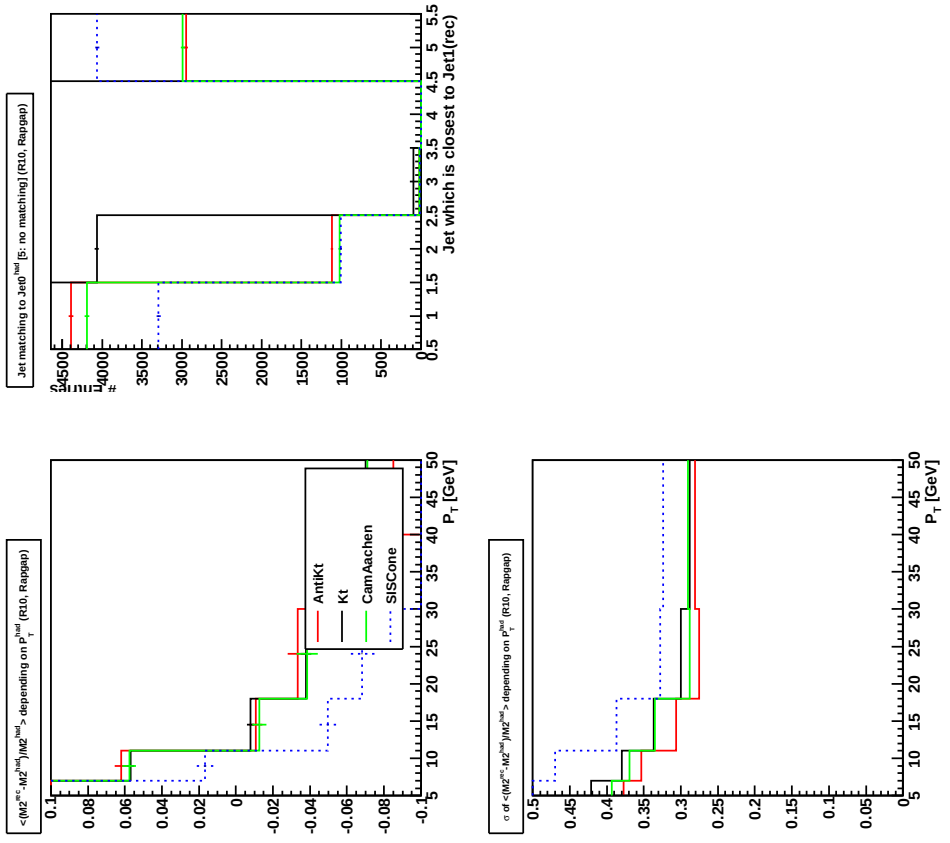


FIG 26. *Deviation of $M_{12}^2$ between reconstruction and hadron level, RAPGAP data, $R_0 = 1.0$*

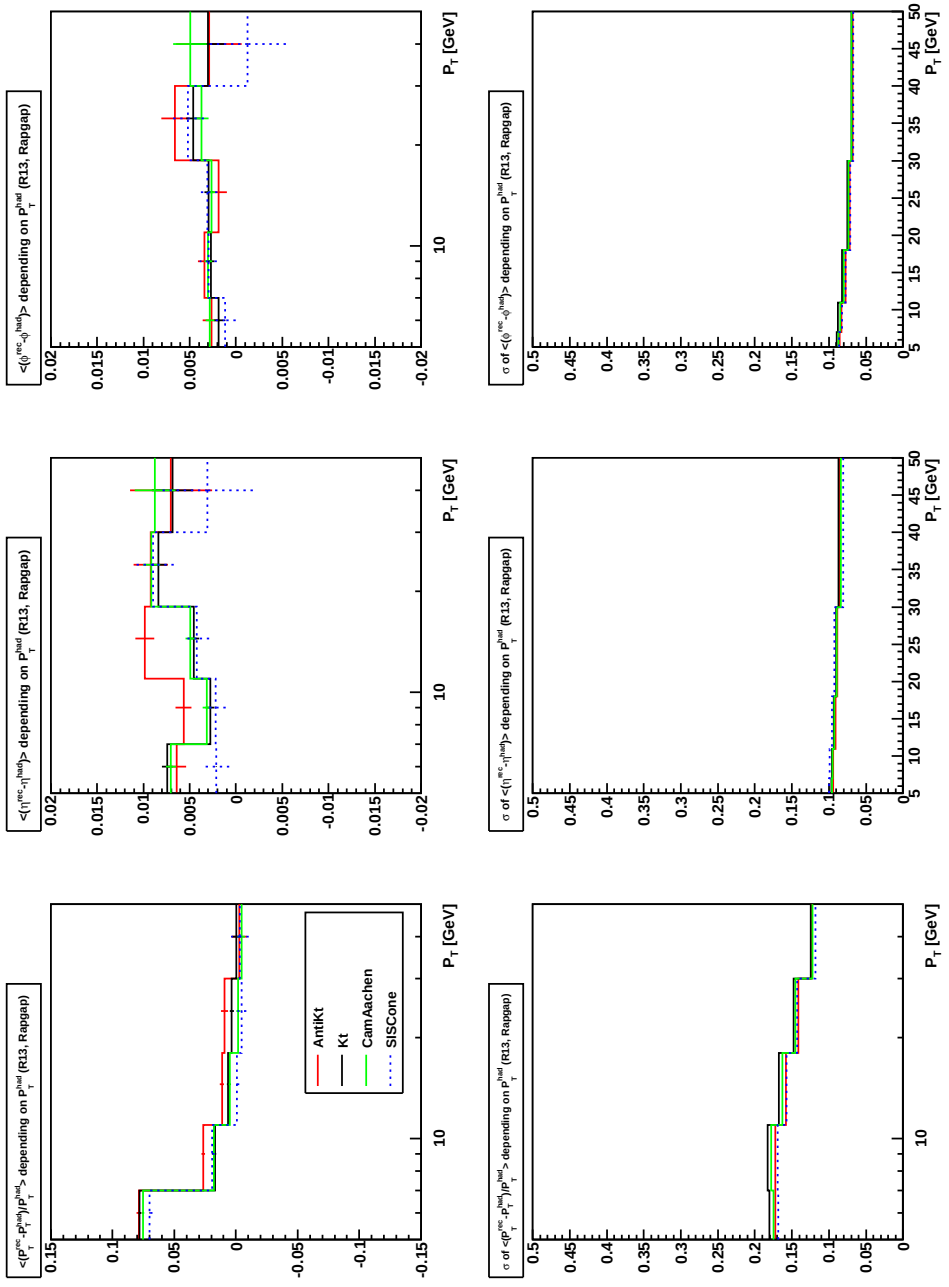


FIG 27. *Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, RAPGAP data, $R_0 = 1.3$*

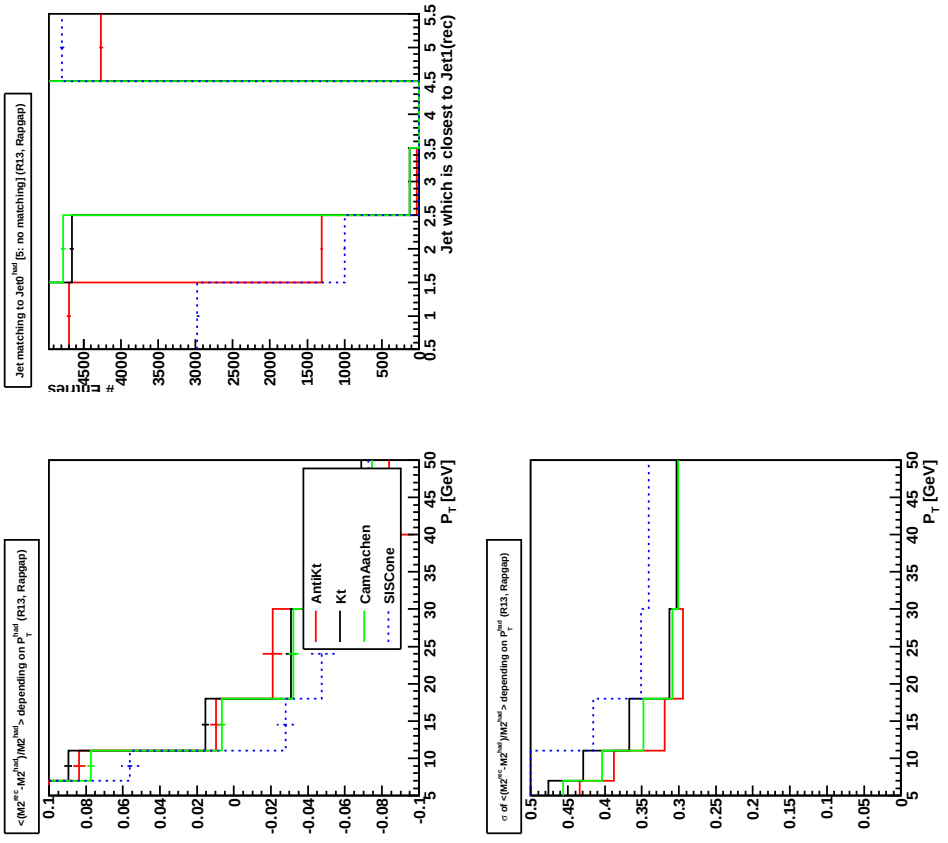


FIG 28. *Deviation of $M_{12}^2$ between reconstruction and hadron level, RAPGAP data, $R_0 = 1.3$*

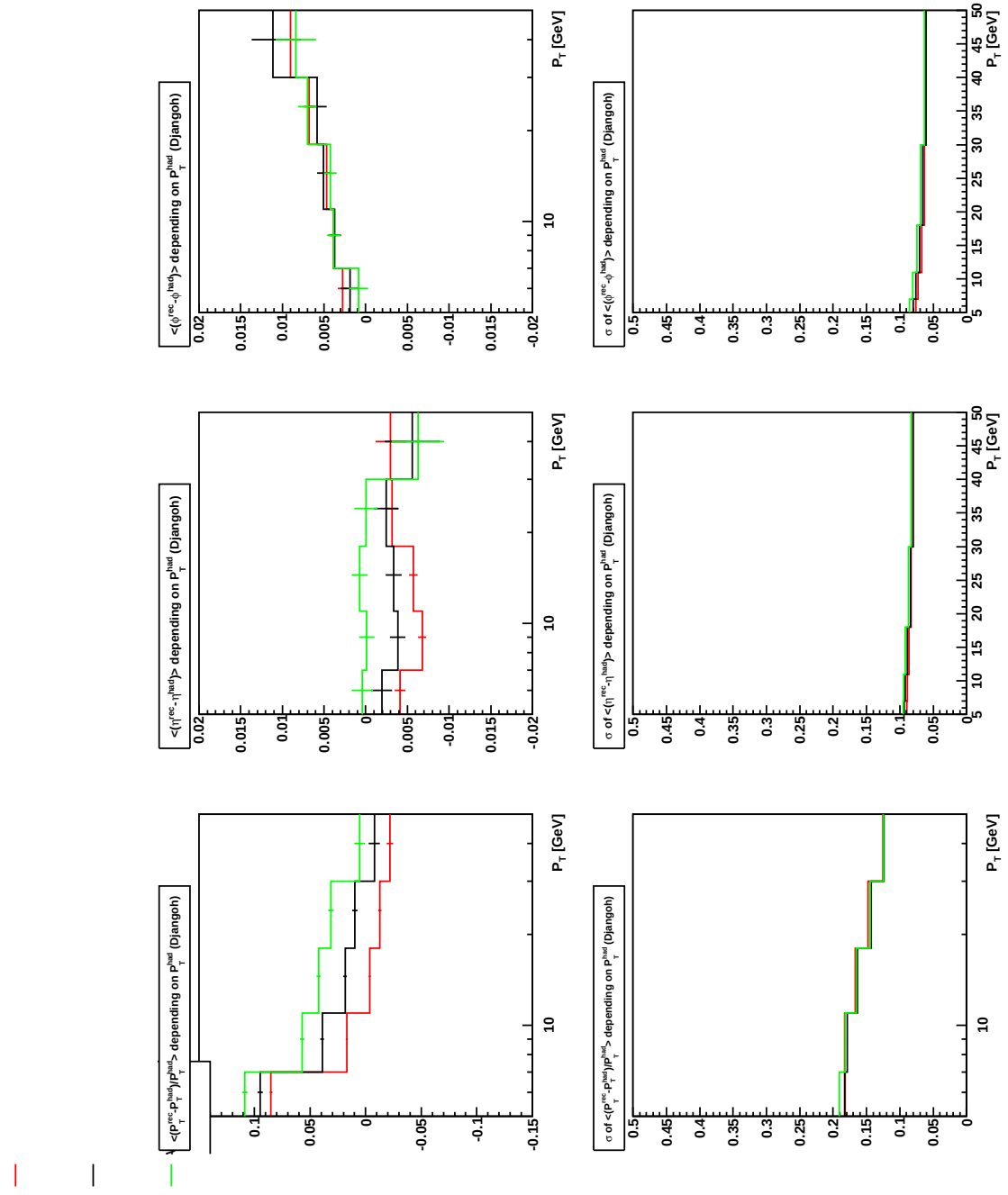


FIG 29. *Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, DJANGOH data, Anti-$k_T$*

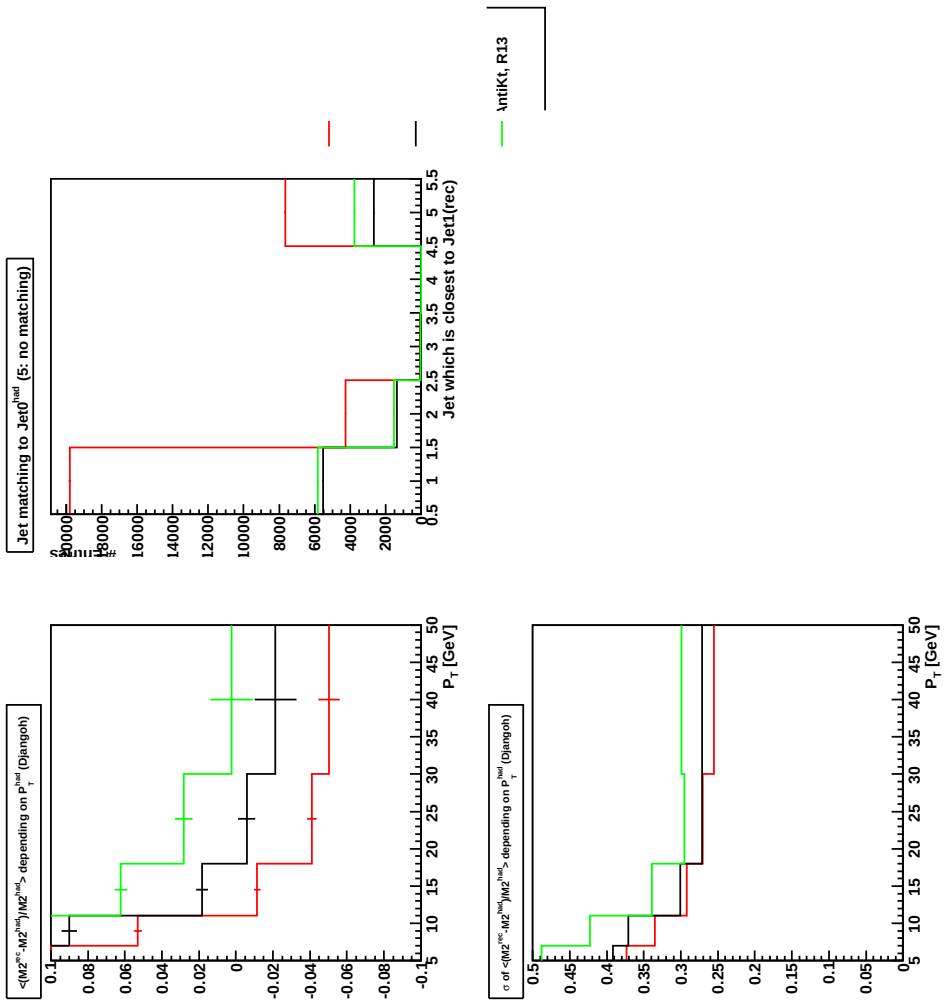


FIG 30. *Deviation of $M_{12}^2$ between reconstruction and hadron level, DJANGOH data, Anti-$k_T$*

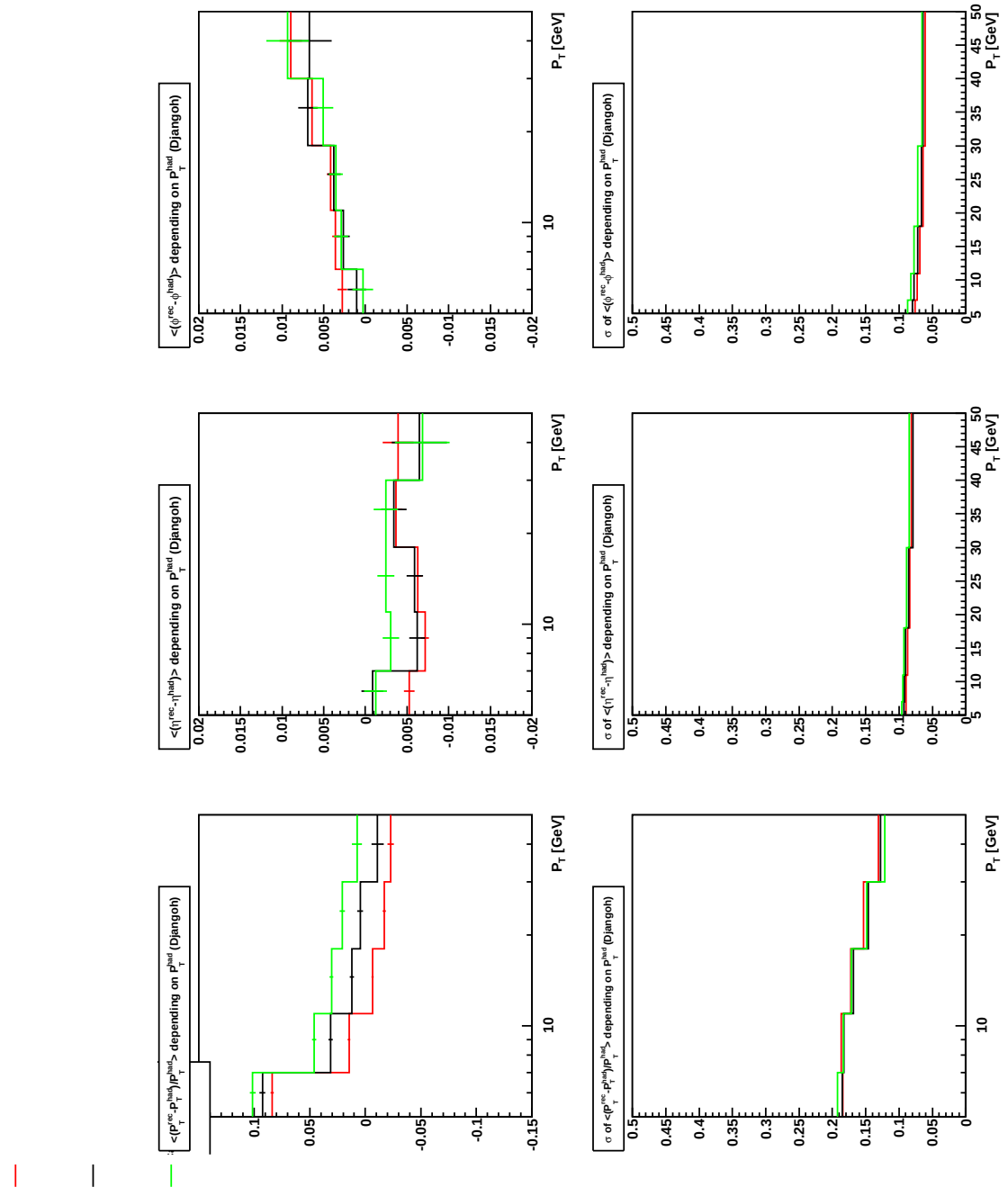


FIG 31. *Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, DJANGOH data, Cambridge/Aachen*

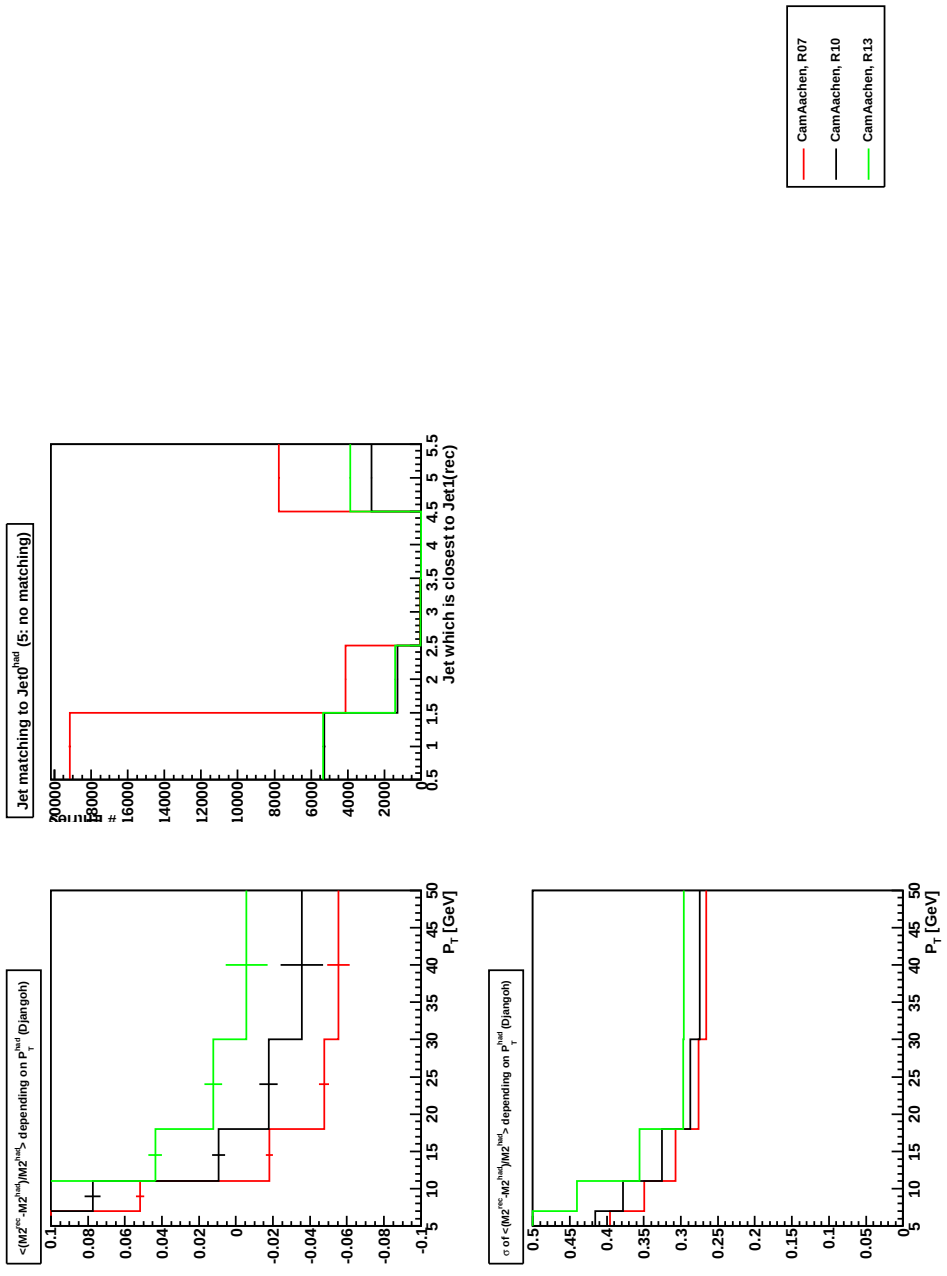


FIG 32. *Deviation of $M_{12}^2$ between reconstruction and hadron level, DJANGOH data, Cambridge/Aachen*

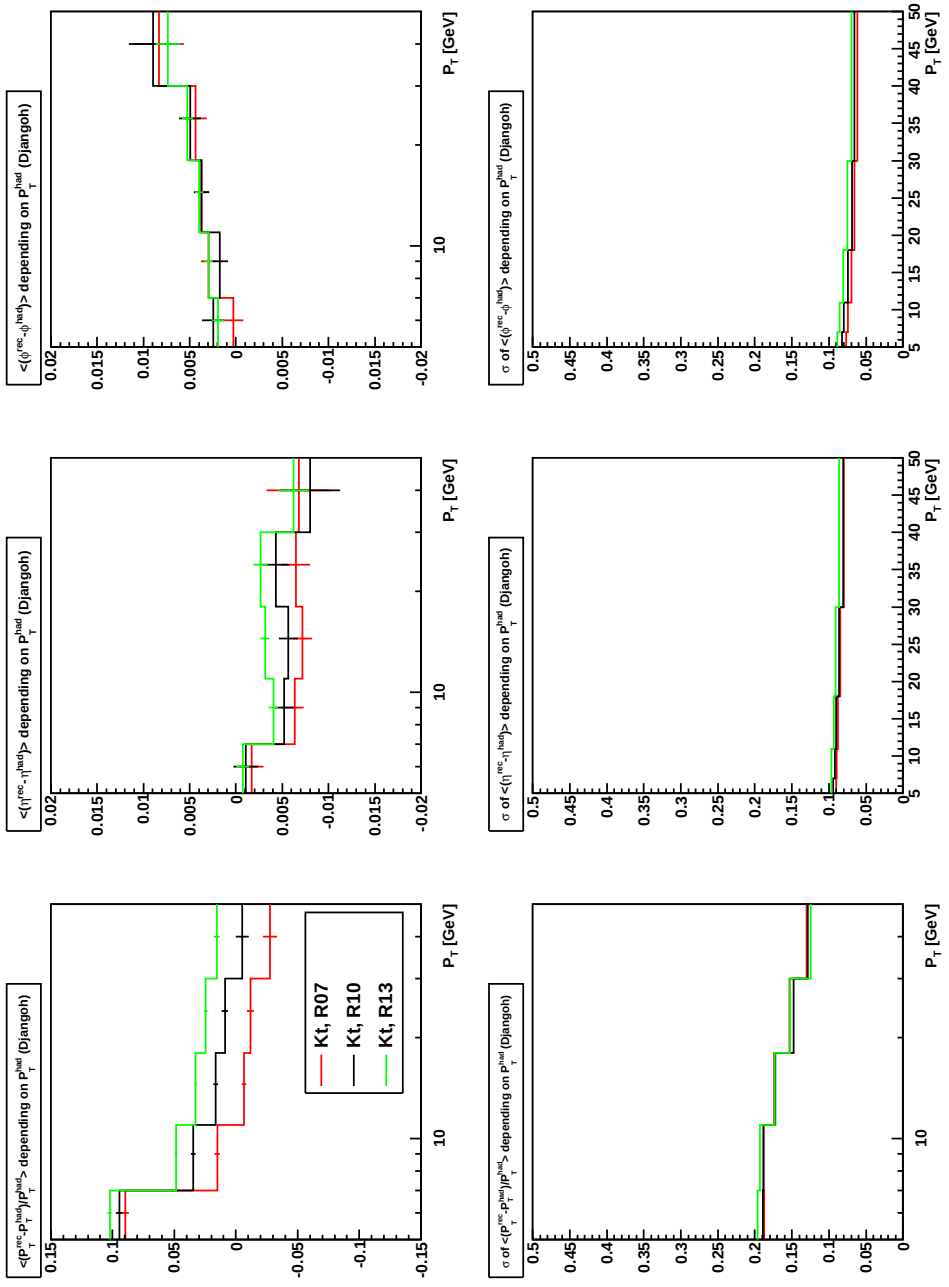


FIG 33. *Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, DJANGOH data, $k_T$*

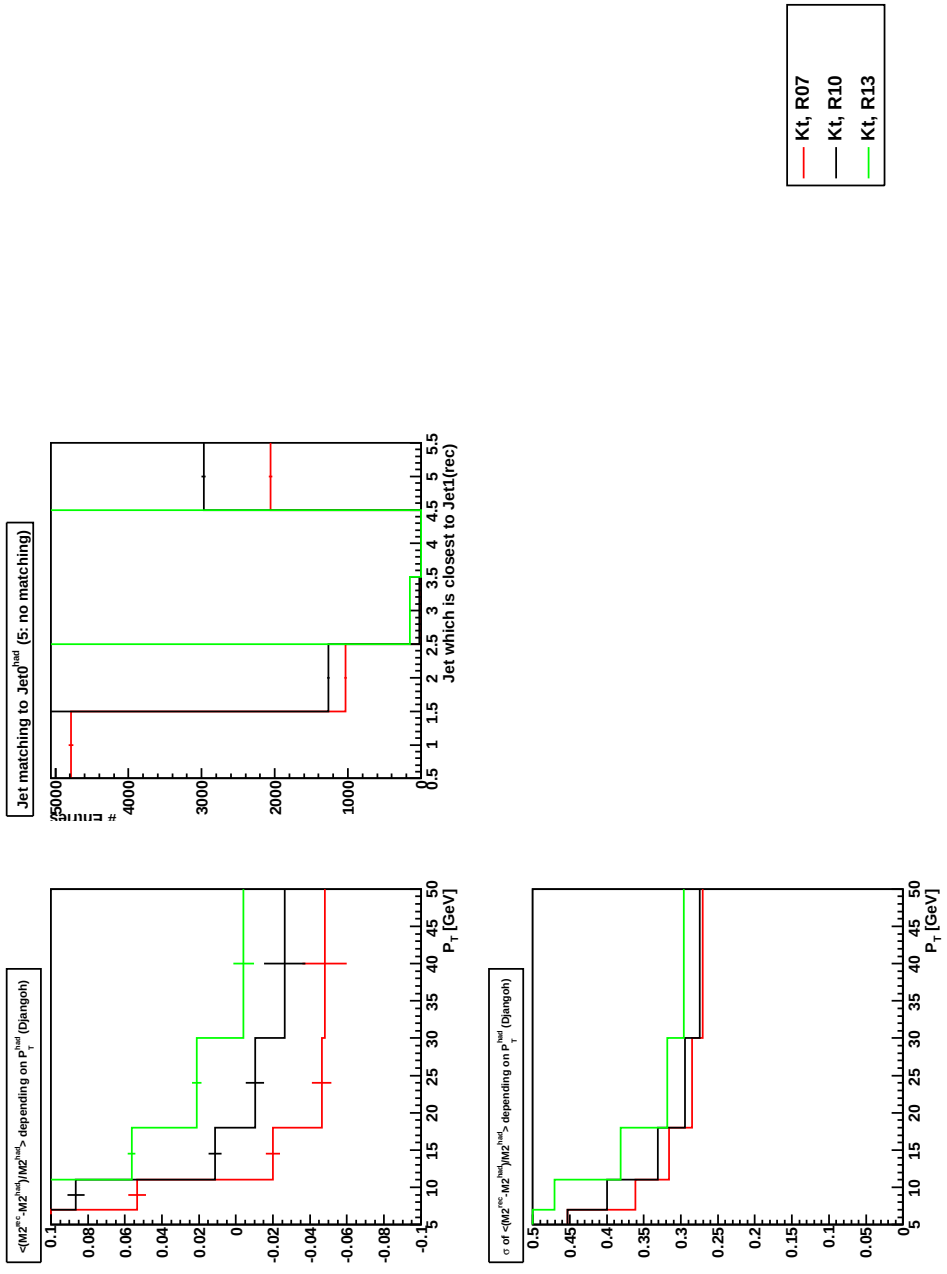


FIG 34. *Deviation of $M^2_{12}$ between reconstruction and hadron level, DJANGOH data, $k_T$*

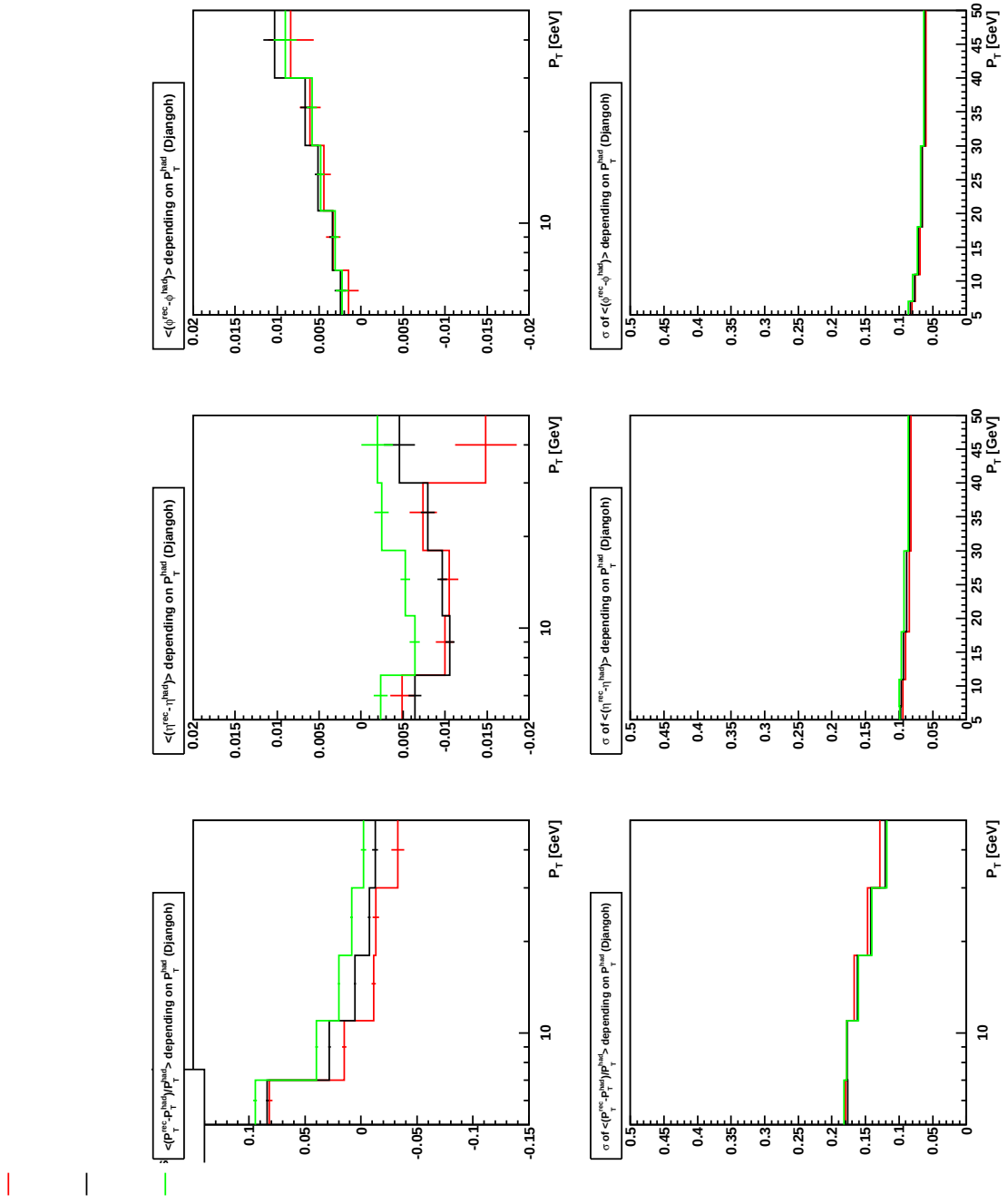


FIG 35. *Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, DJANGOH data, SISCone*

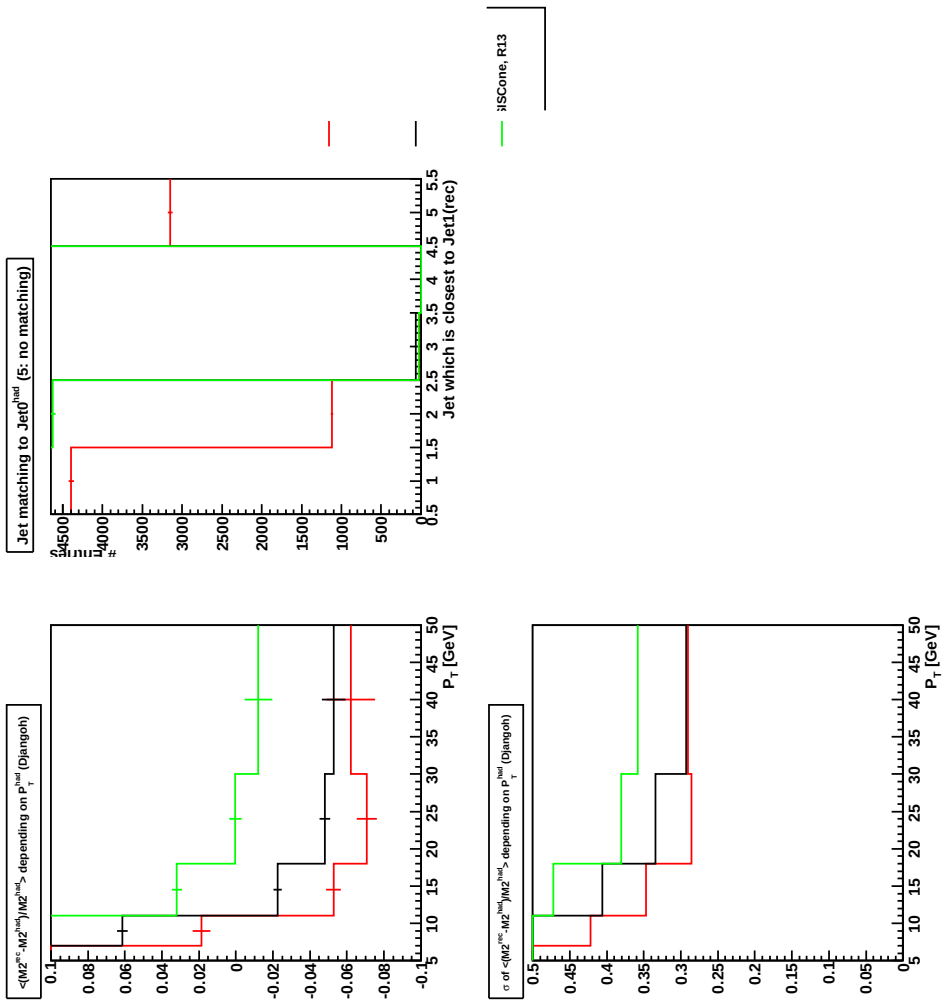


FIG 36. *Deviation of $M_{12}^2$ between reconstruction and hadron level, DJANGOH data, SISCone*

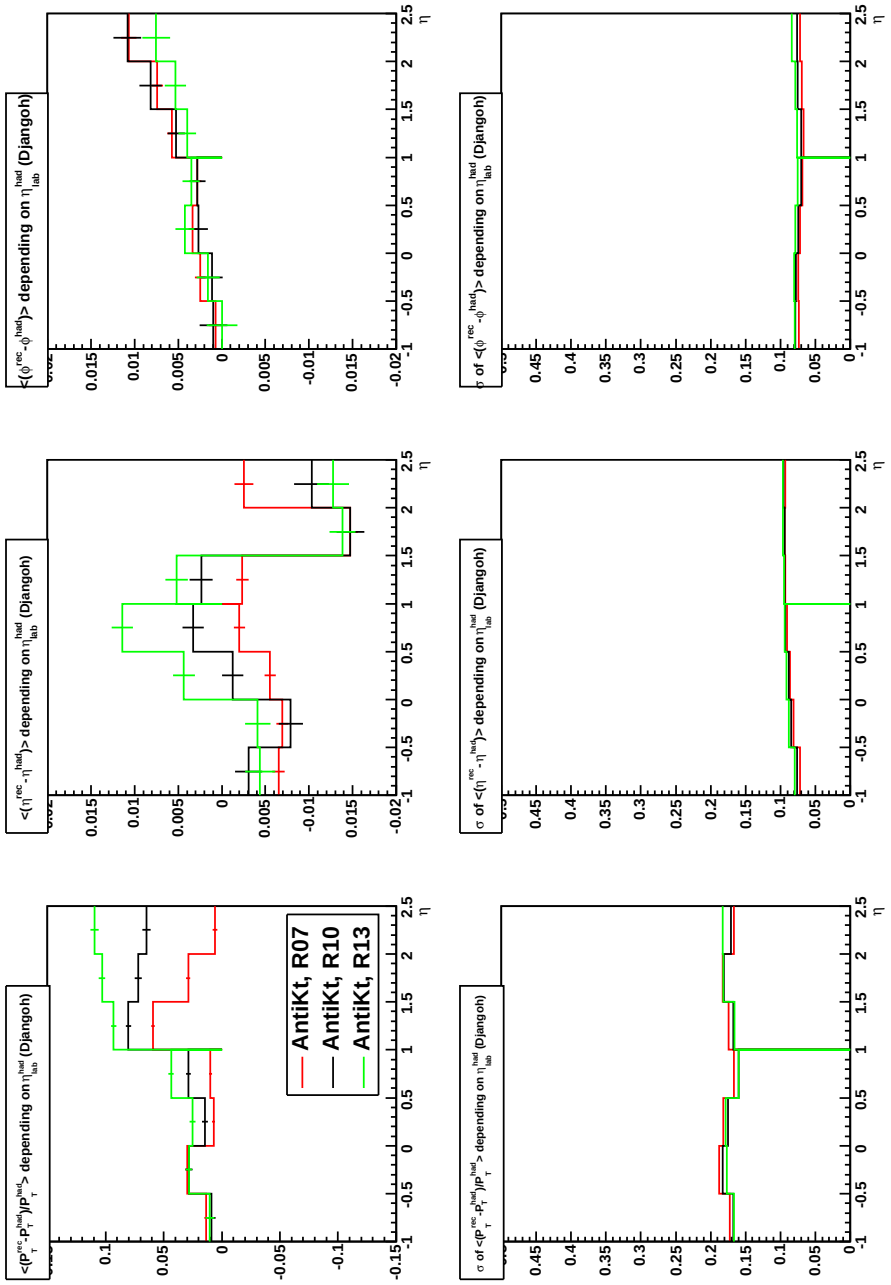


FIG 37. *Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, DJANGOH data, $\eta$ dependance*

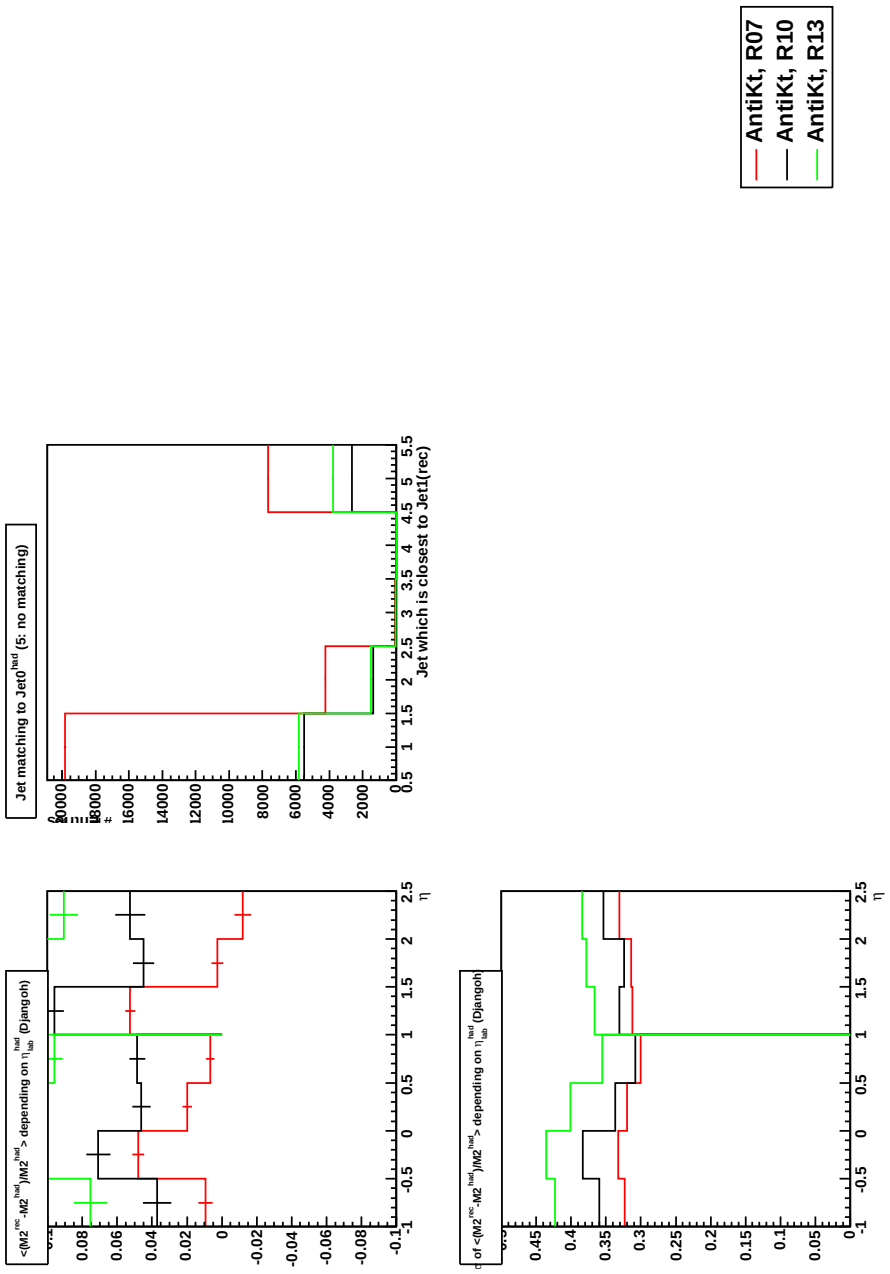


FIG 38. *Deviation of $M^2_{12}$ between reconstruction and hadron level, DJANGOH data, $\eta$ dependance*

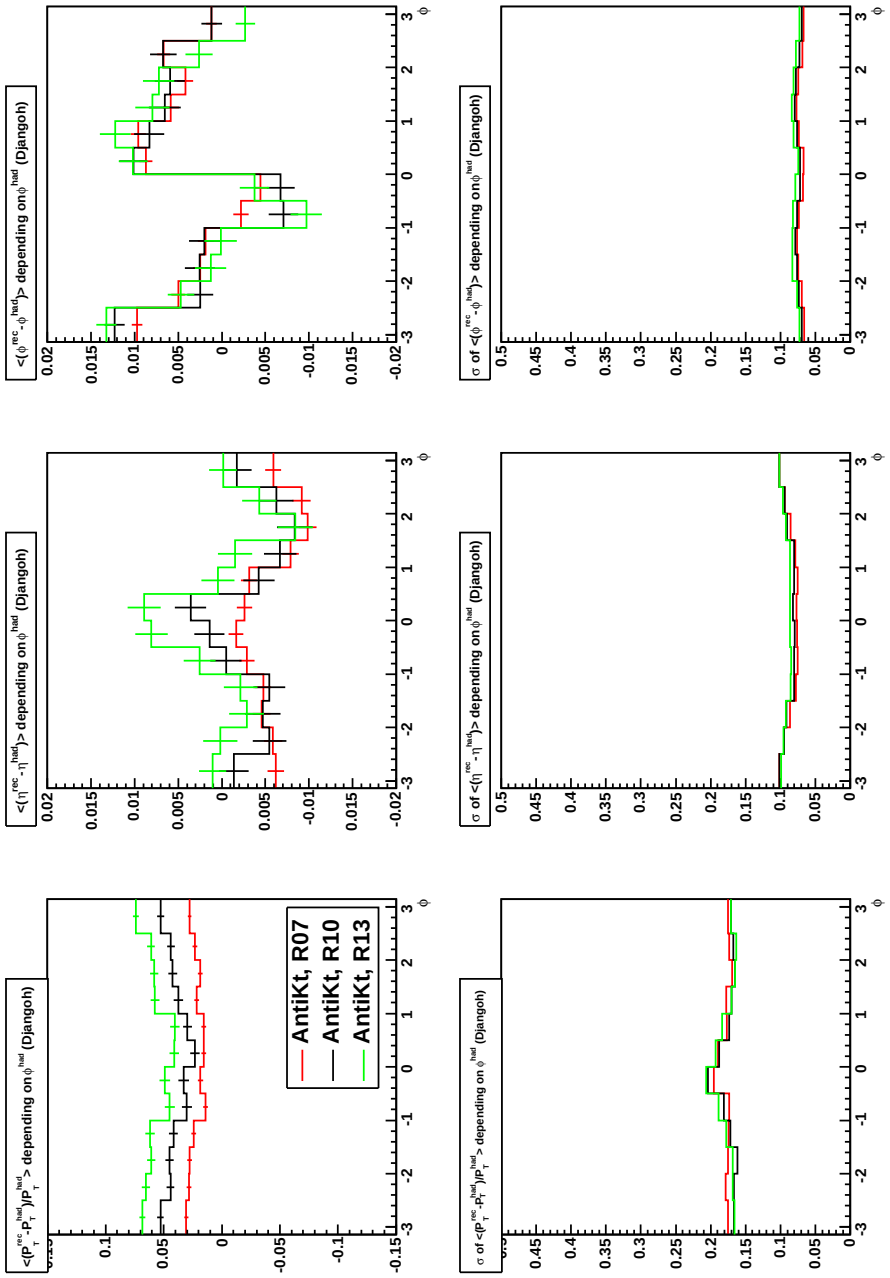


FIG 39. *Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, DJANGOH data, $\phi$ dependance*

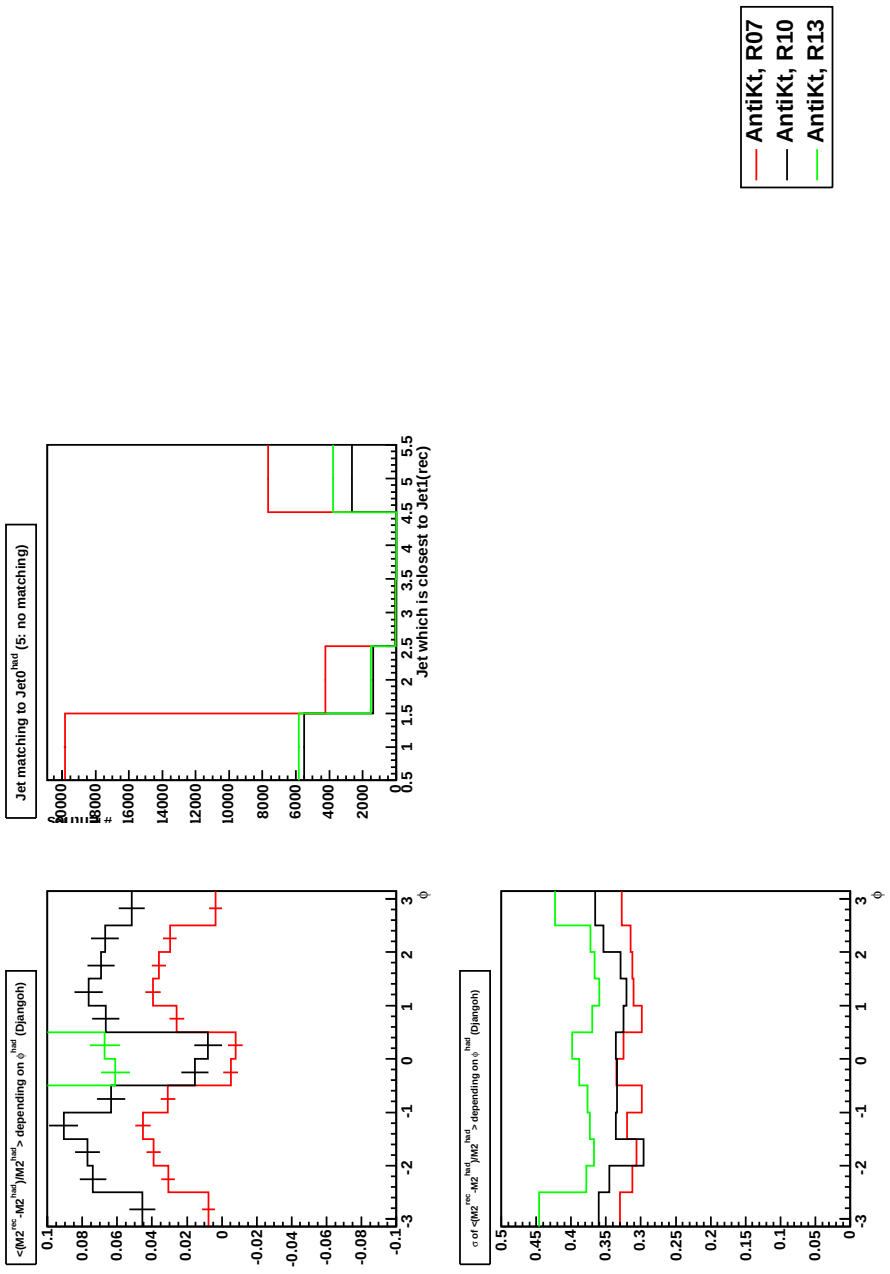


FIG 40. *Deviation of $M^2_{12}$ between reconstruction and hadron level, DJANGOH data, $\phi$ dependance*

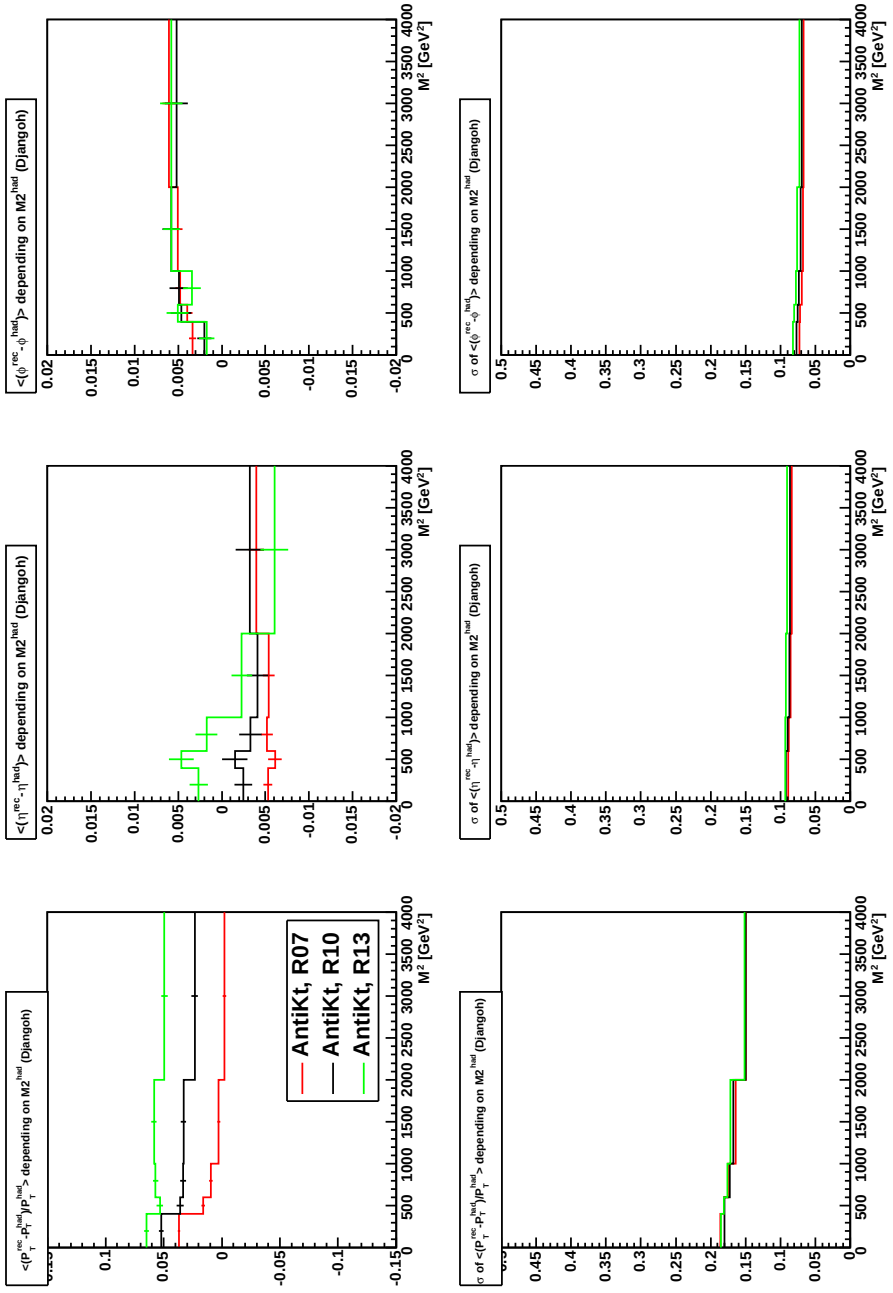


FIG 41. *Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, DJANGOH data, $M_{12}^2$ dependance*

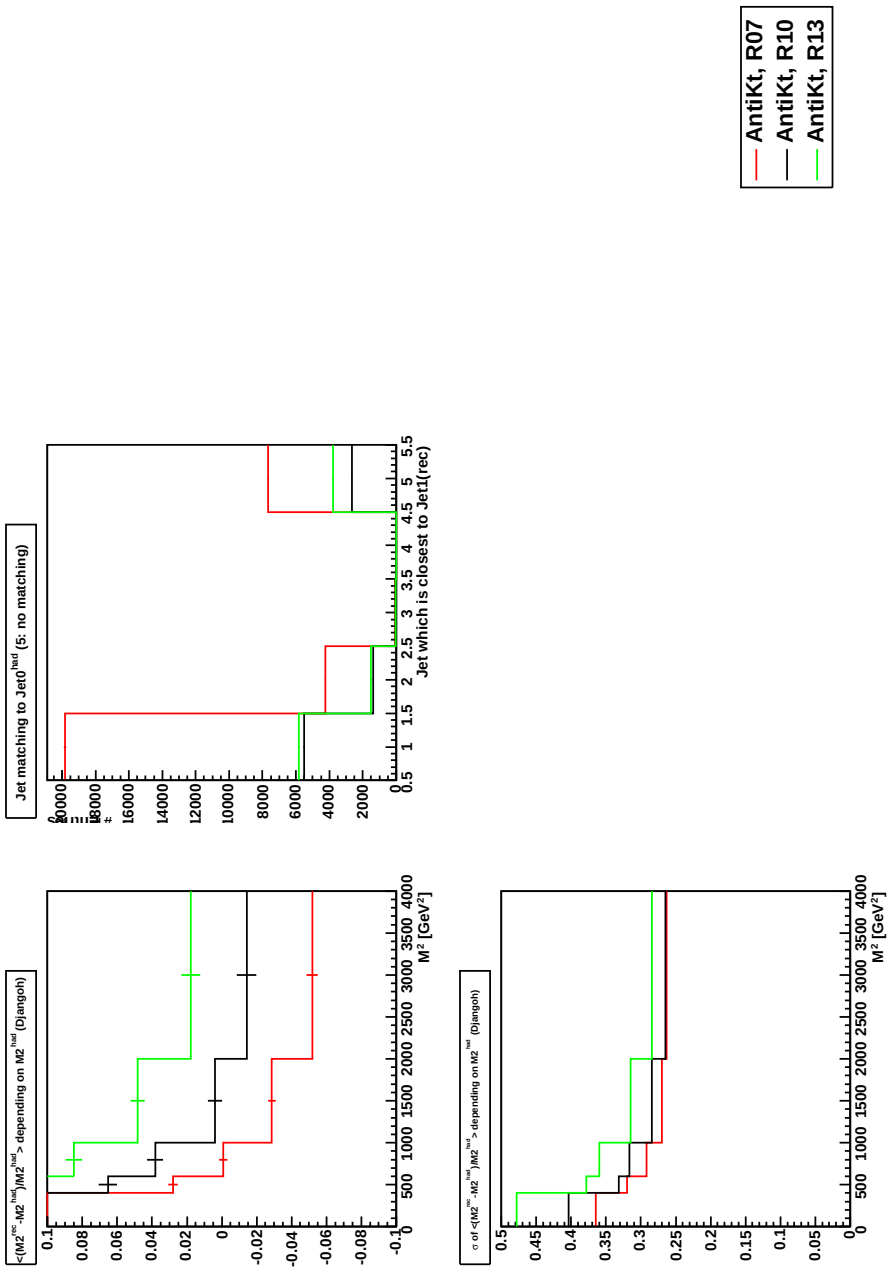


FIG 42. *Deviation of $M^2_{12}$ between reconstruction and hadron level, DJANGOH data, $M^2_{12}$ dependance*

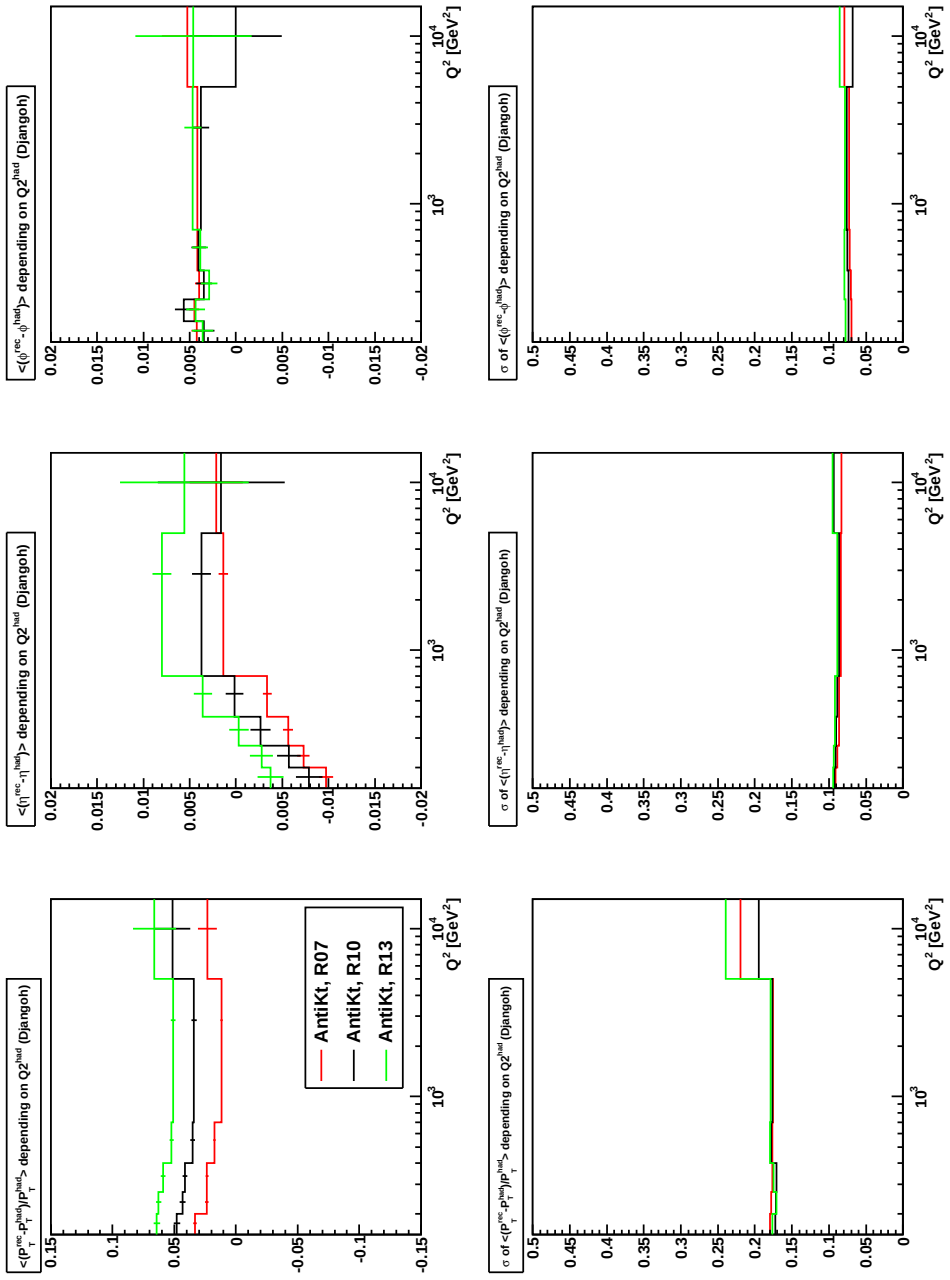


FIG 43. *Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, DJANGOH data, $Q^2$ dependance*

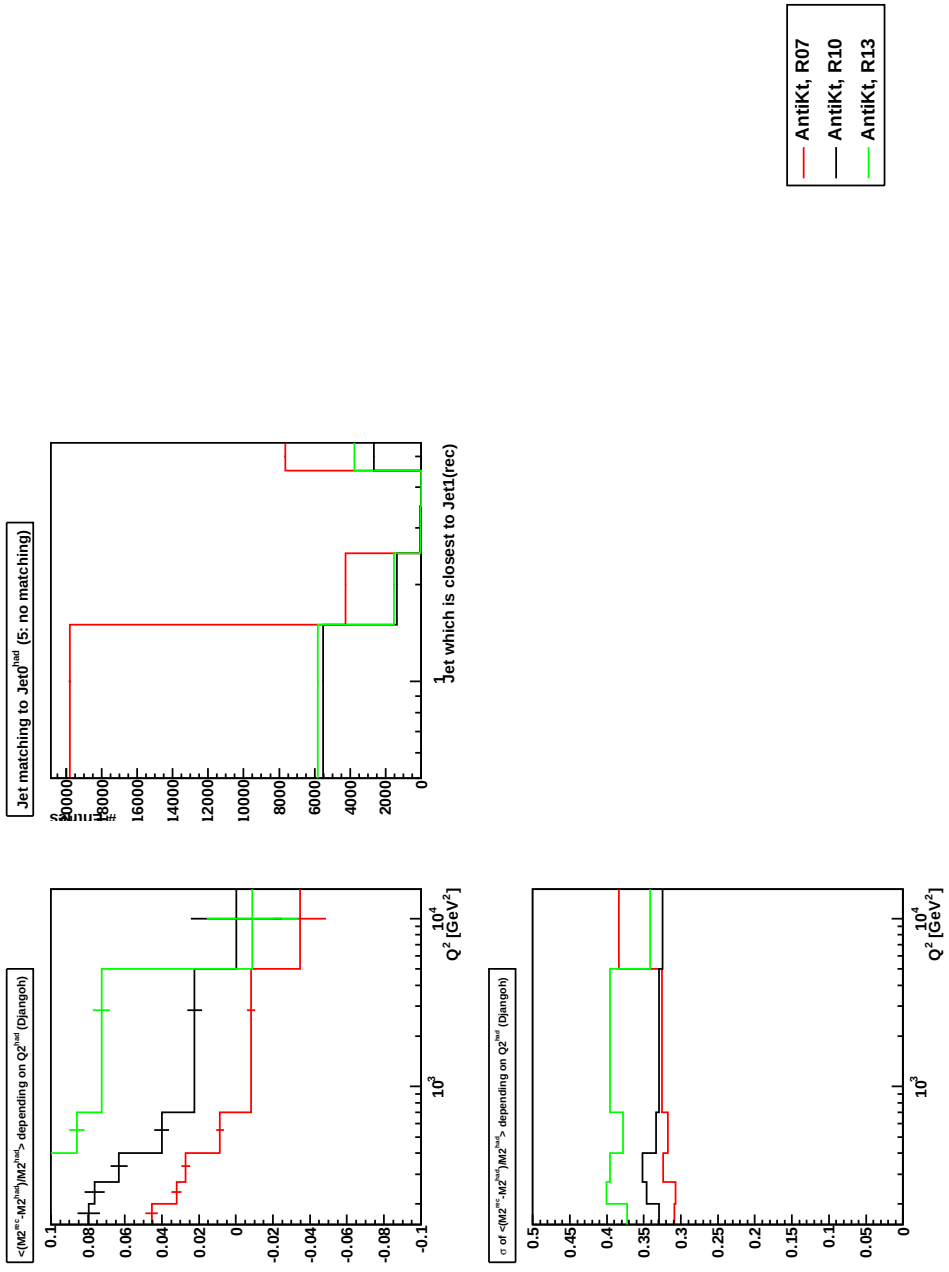


FIG 44. *Deviation of $M_{12}^2$ between reconstruction and hadron level, DJANGOH data, $Q^2$ dependance*

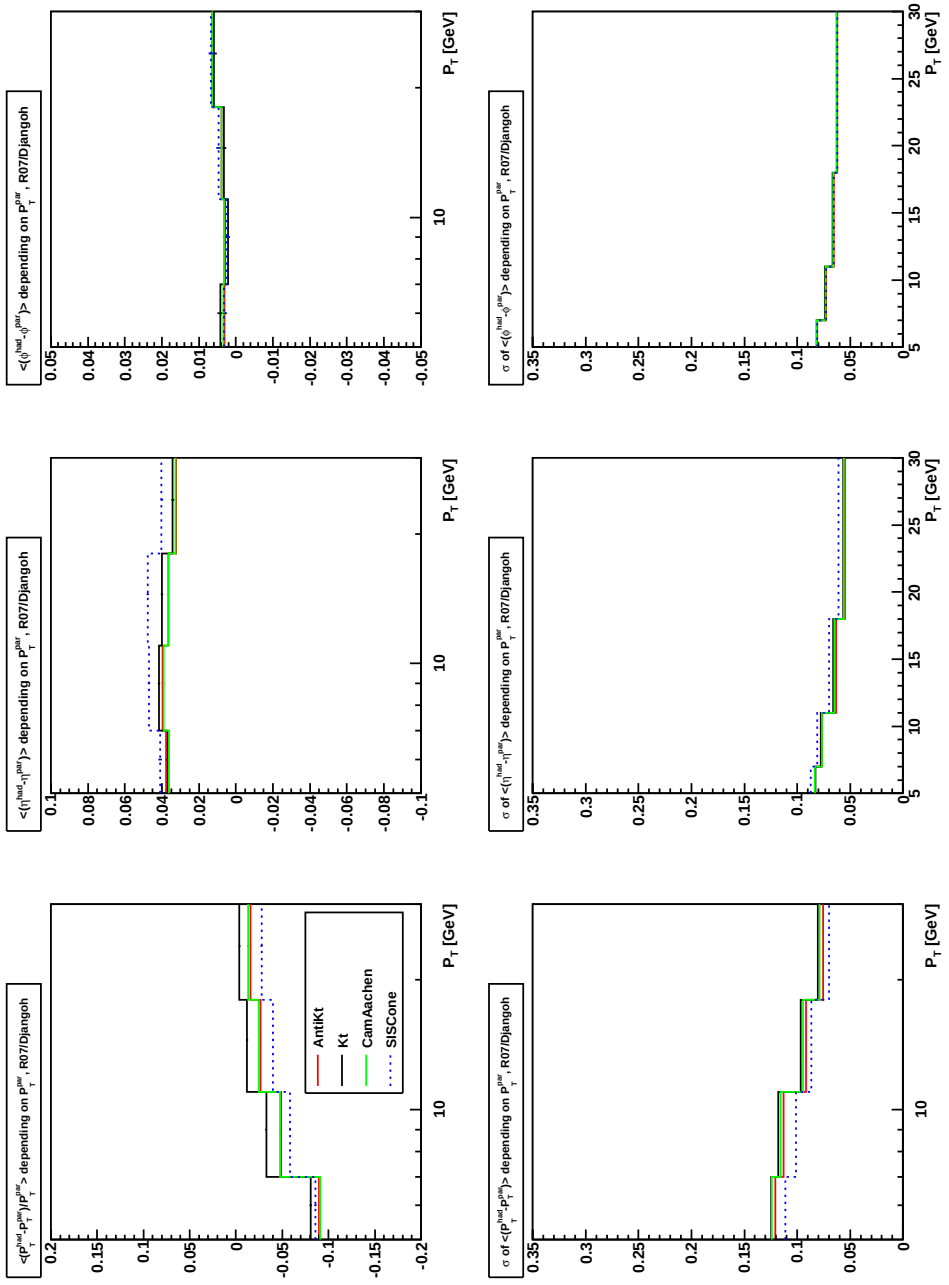


FIG 45. *Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, DJANGOH data, $R_0 = 0.7$*

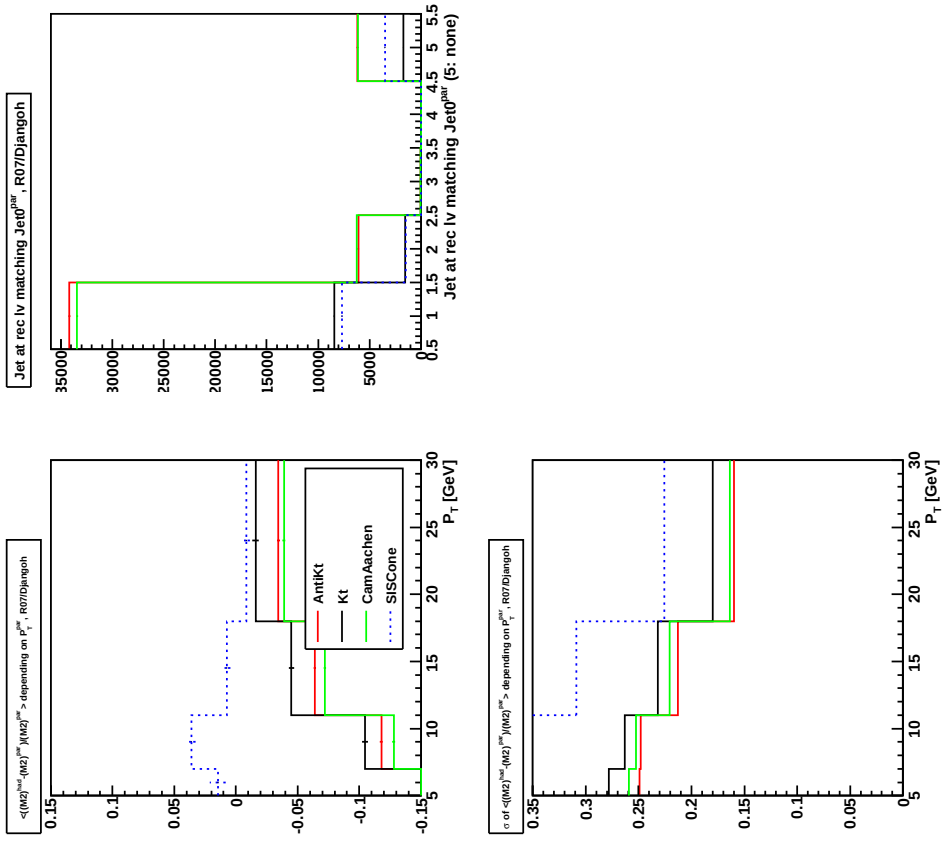


FIG 46. *Deviation of $M_{12}^2$ between reconstruction and hadron level, DJANGOH data, $R_0 = 0.7$*

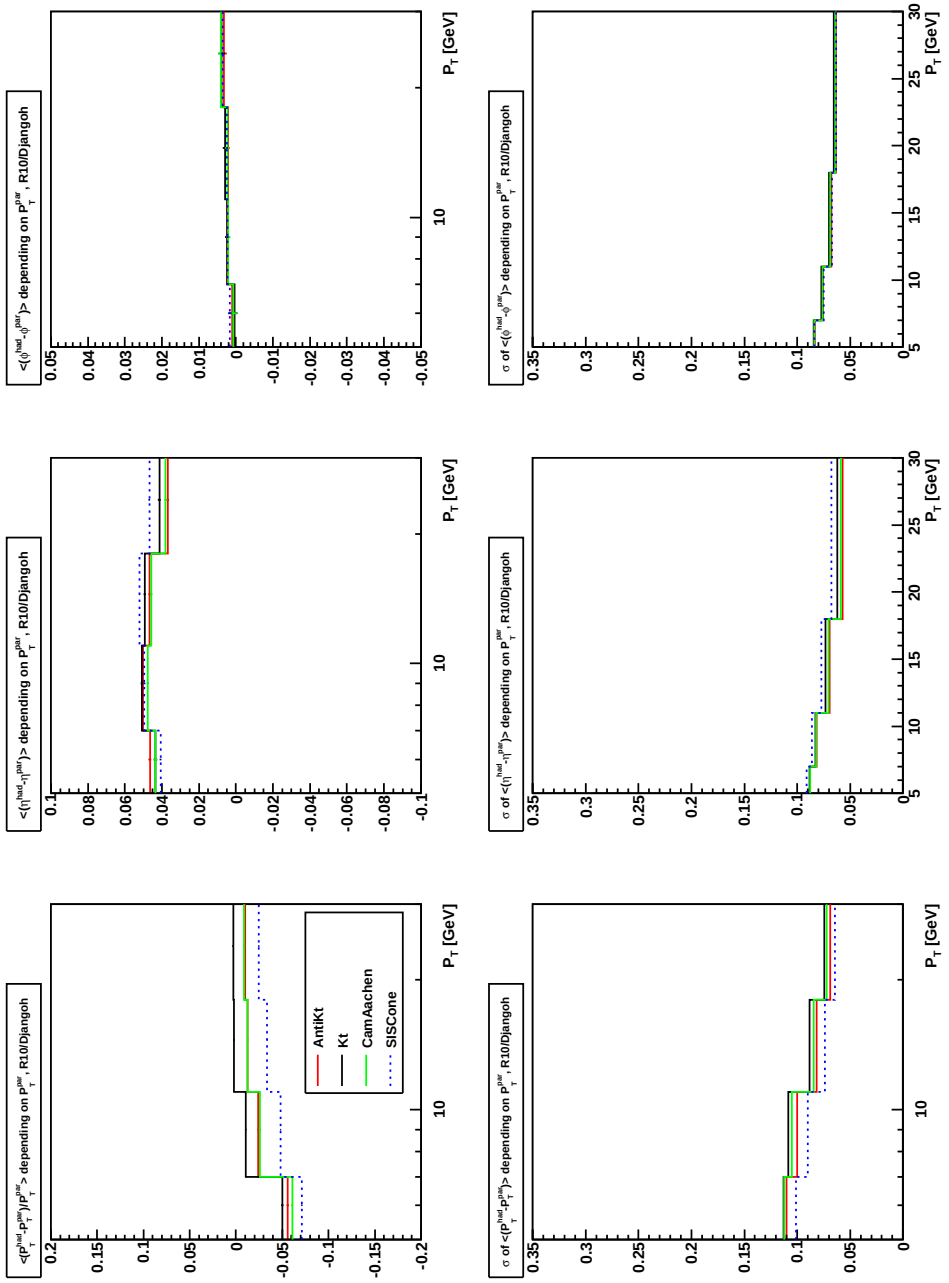


FIG 47. *Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, DJANGOH data, $R_0 = 1.0$*

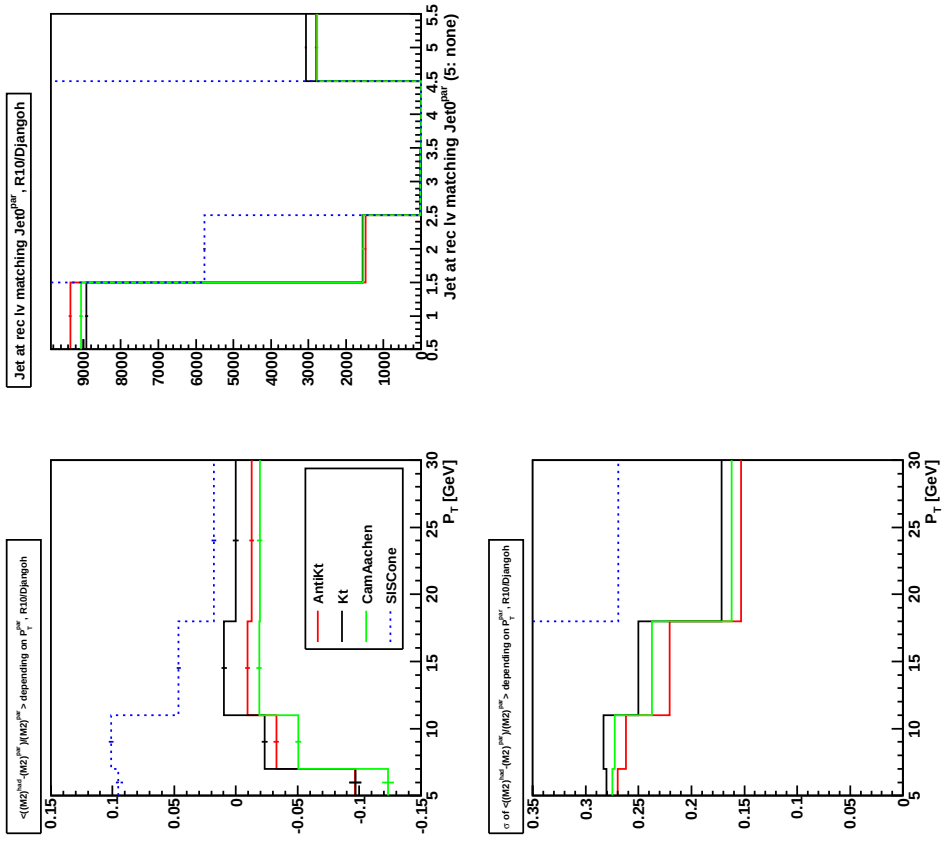


FIG 48. *Deviation of $M_{12}^2$ between reconstruction and hadron level, DJANGOH data, $R_0 = 1.0$*

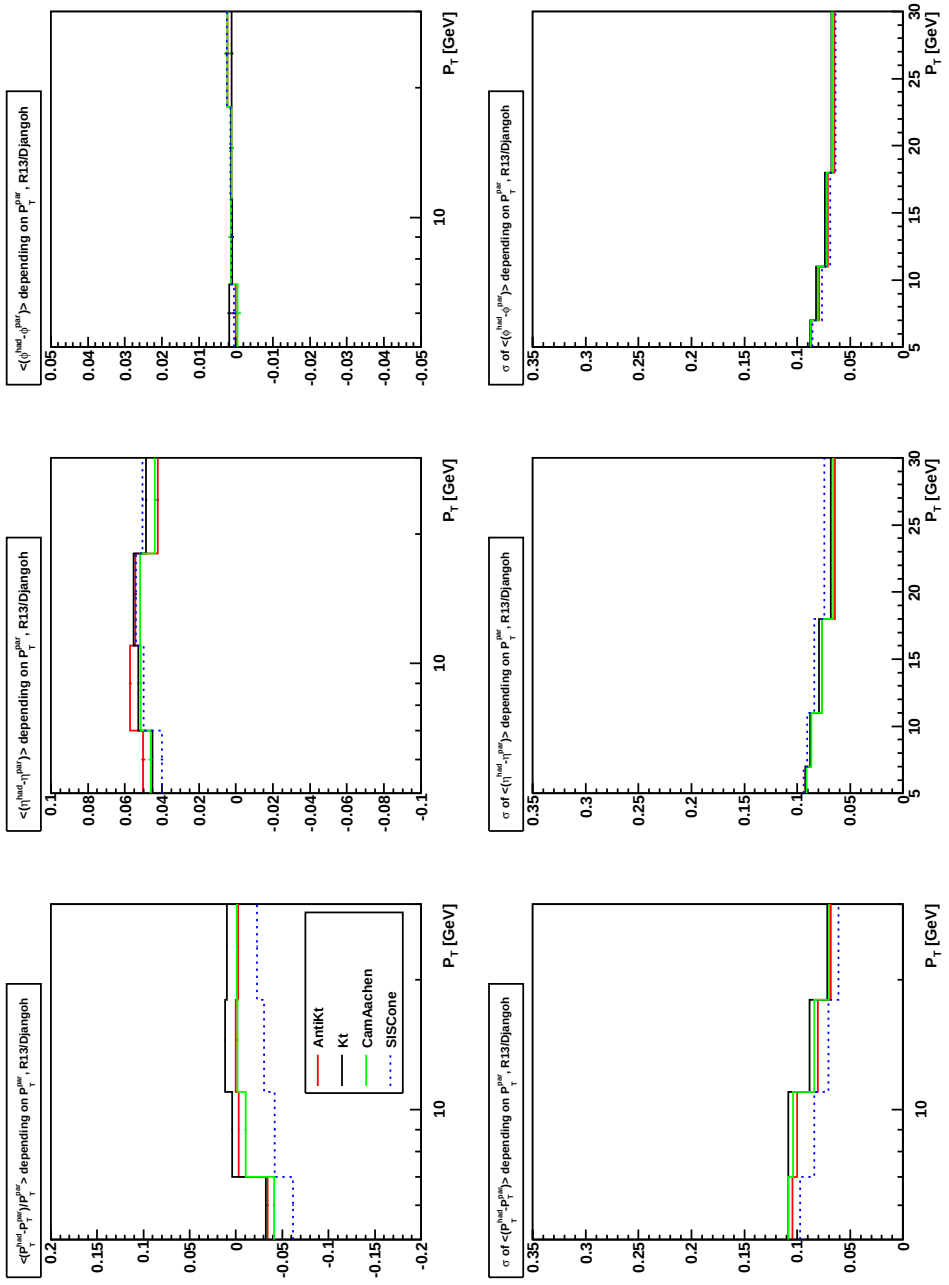


FIG 49. *Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, DJANGOH data, $R_0 = 1.3$*

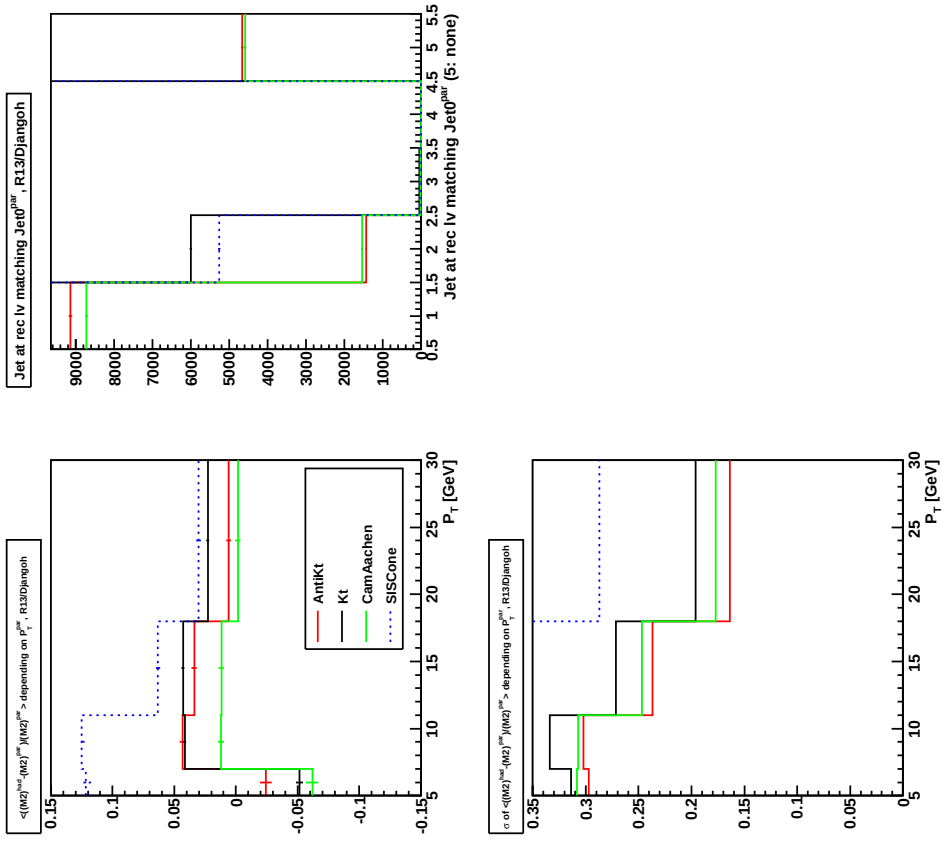


FIG 50. *Deviation of $M_{12}^2$ between reconstruction and hadron level, DJANGOH data, $R_0 = 1.3$*

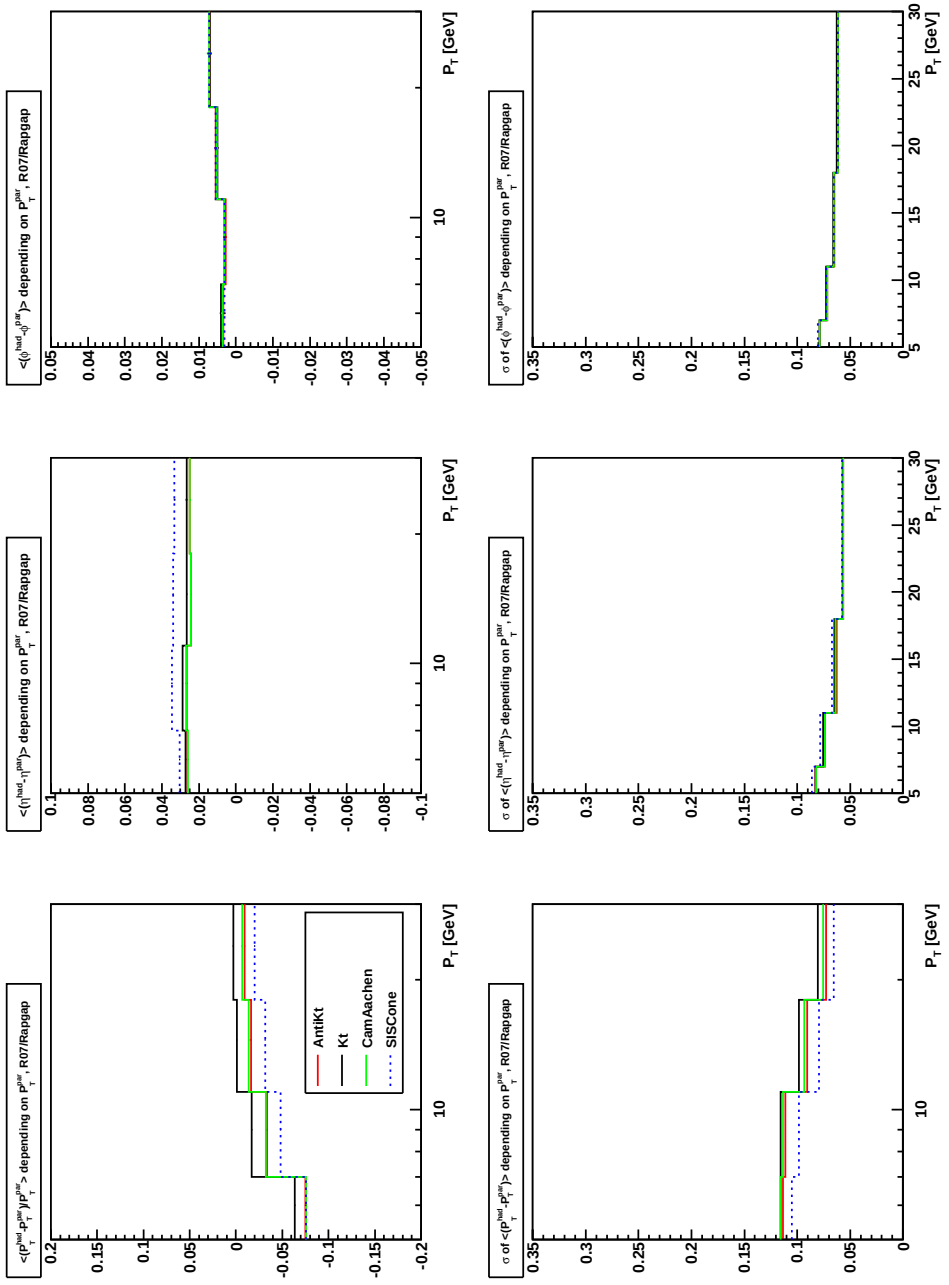


FIG 51. *Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, RAPGAP data, $R_0 = 0.7$*

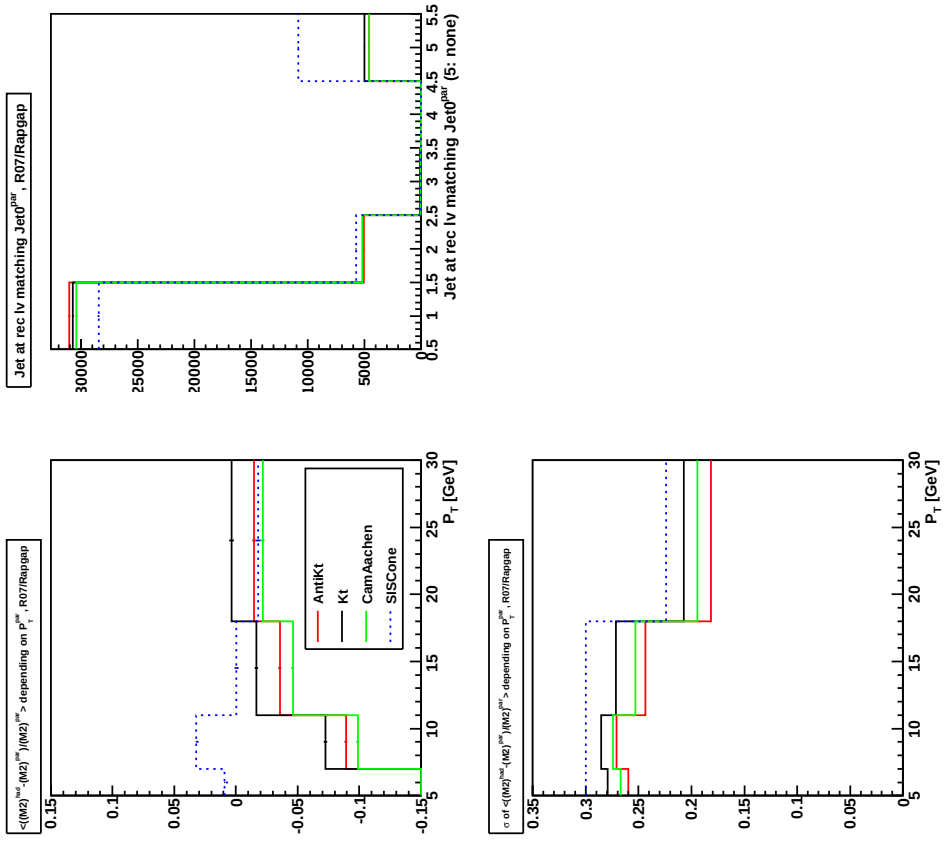


FIG 52. *Deviation of $M_{12}^2$ between reconstruction and hadron level, RAPGAP data, $R_0 = 0.7$*

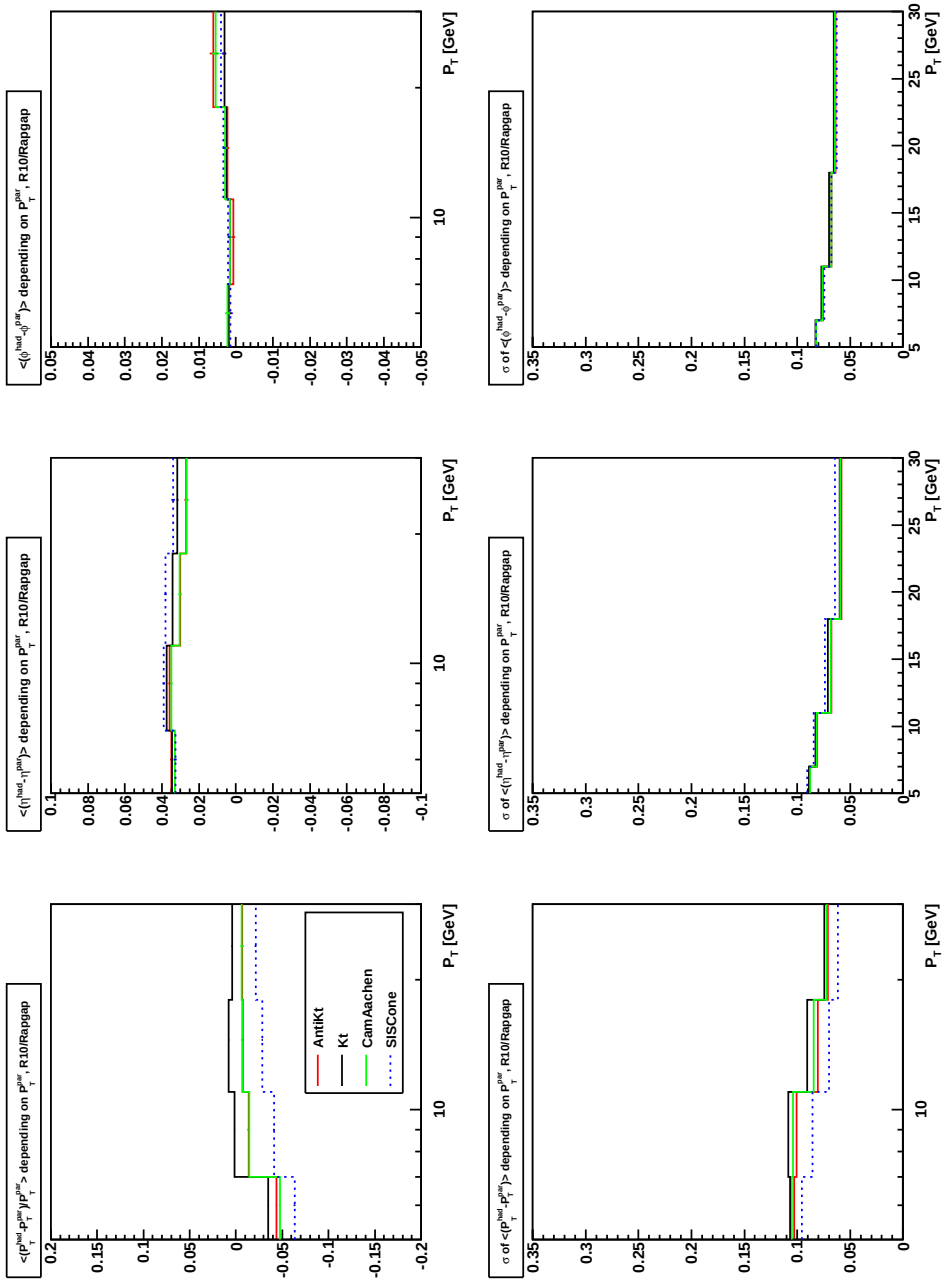


FIG 53. *Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, RAPGAP data, $R_0 = 1.0$*

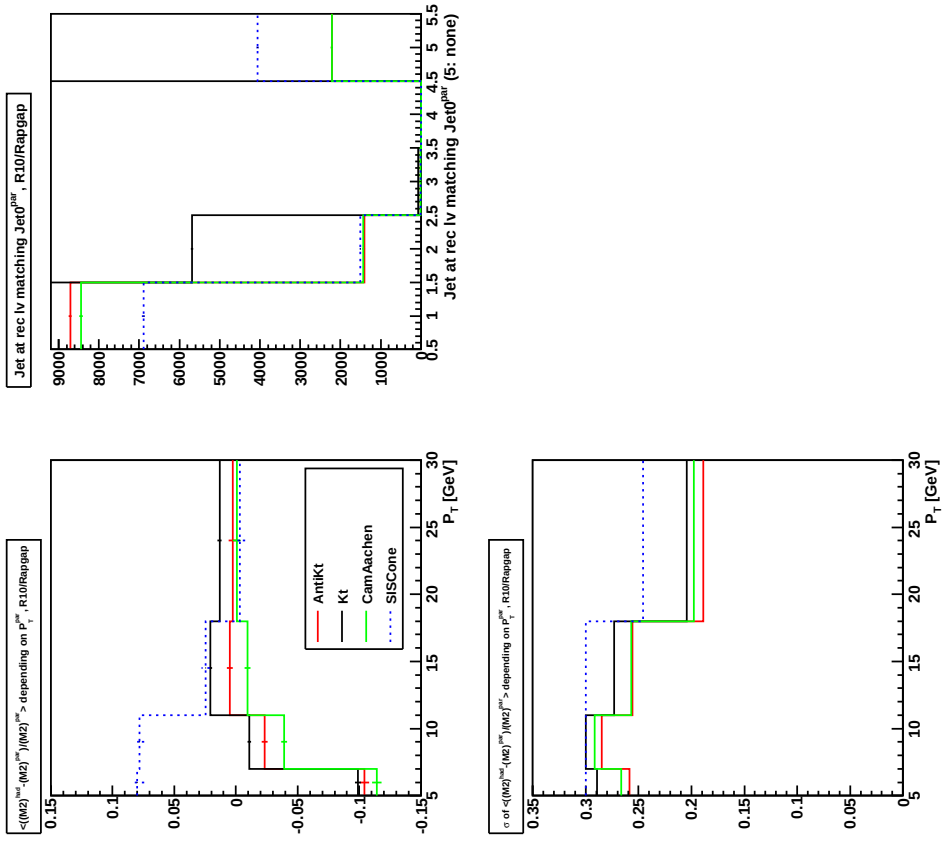


FIG 54. *Deviation of $M_{12}^2$ between reconstruction and hadron level, RAPGAP data, $R_0 = 1.0$*

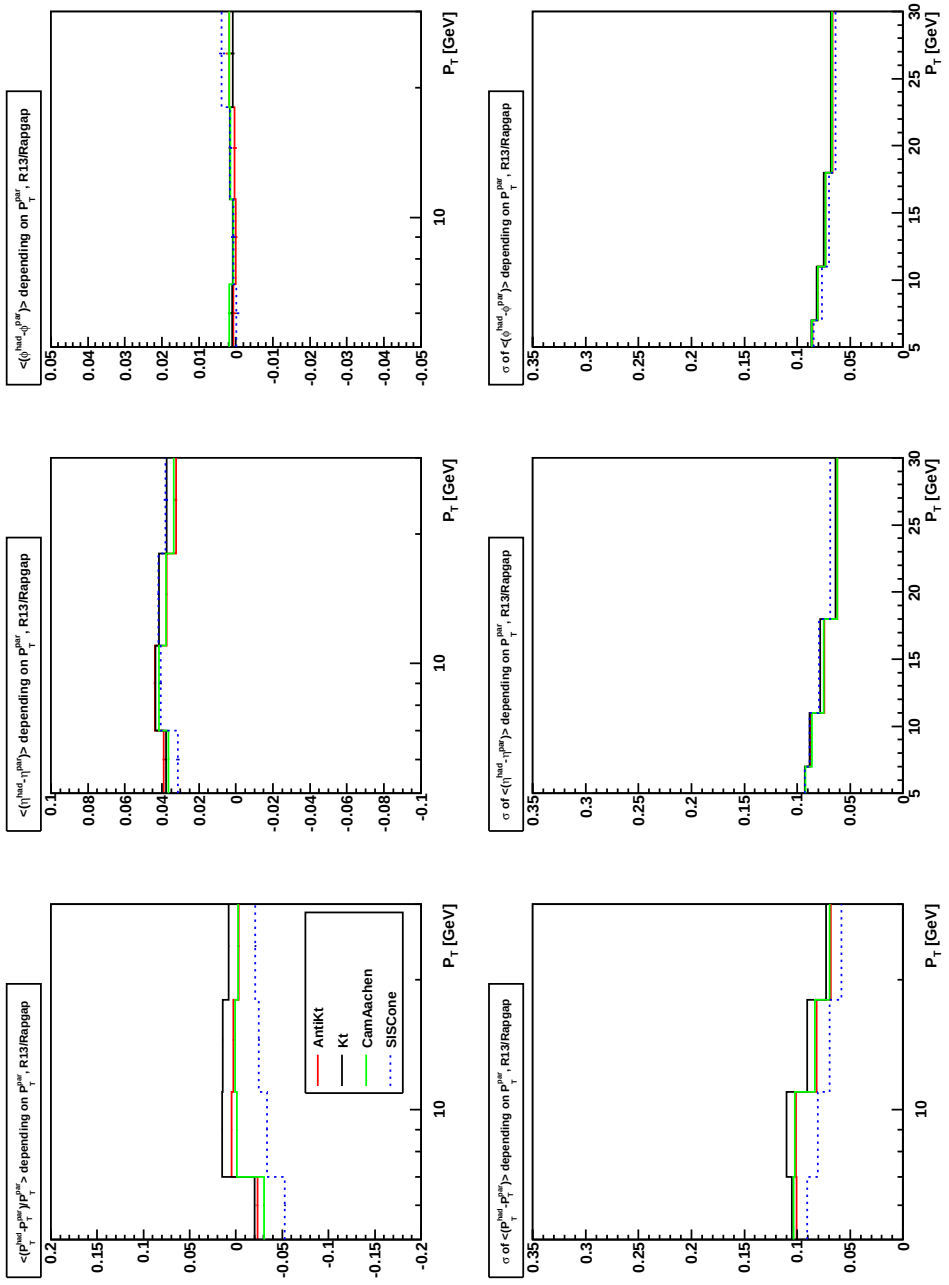


FIG 55. *Deviations of $P_T$, $\eta$ and $\phi$ between reconstruction and hadron level, RAPGAP data, $R_0 = 1.3$*

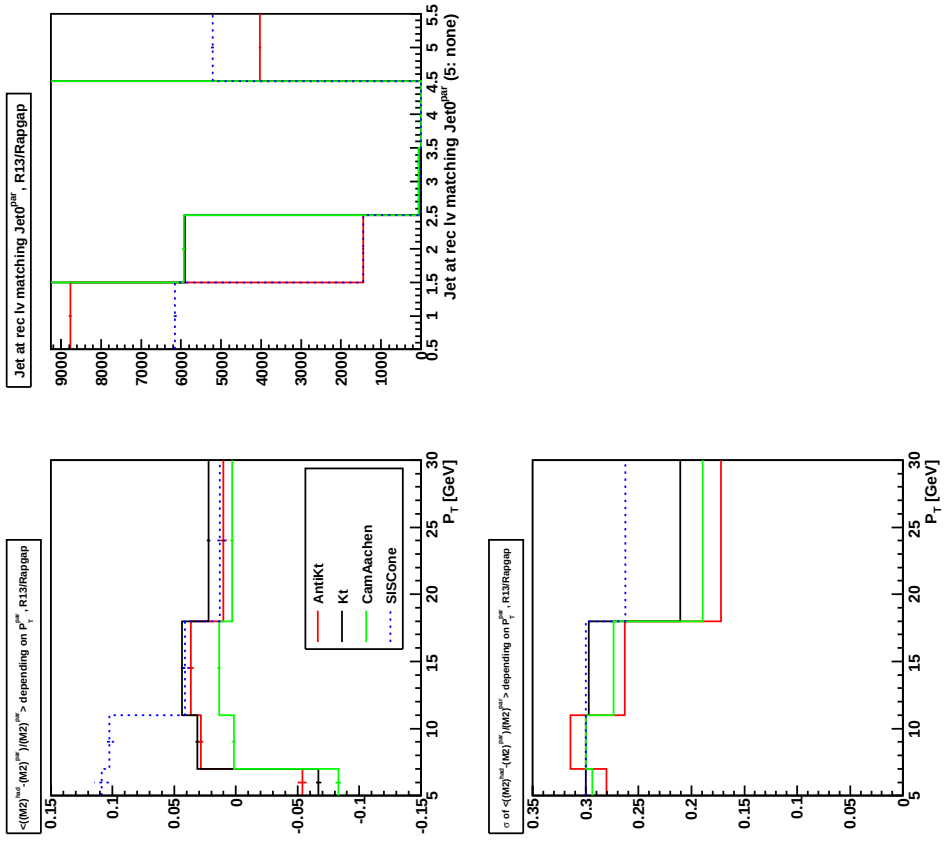


FIG 56. *Deviation of $M_{12}^2$ between reconstruction and hadron level, RAPGAP data, $R_0 = 1.3$*

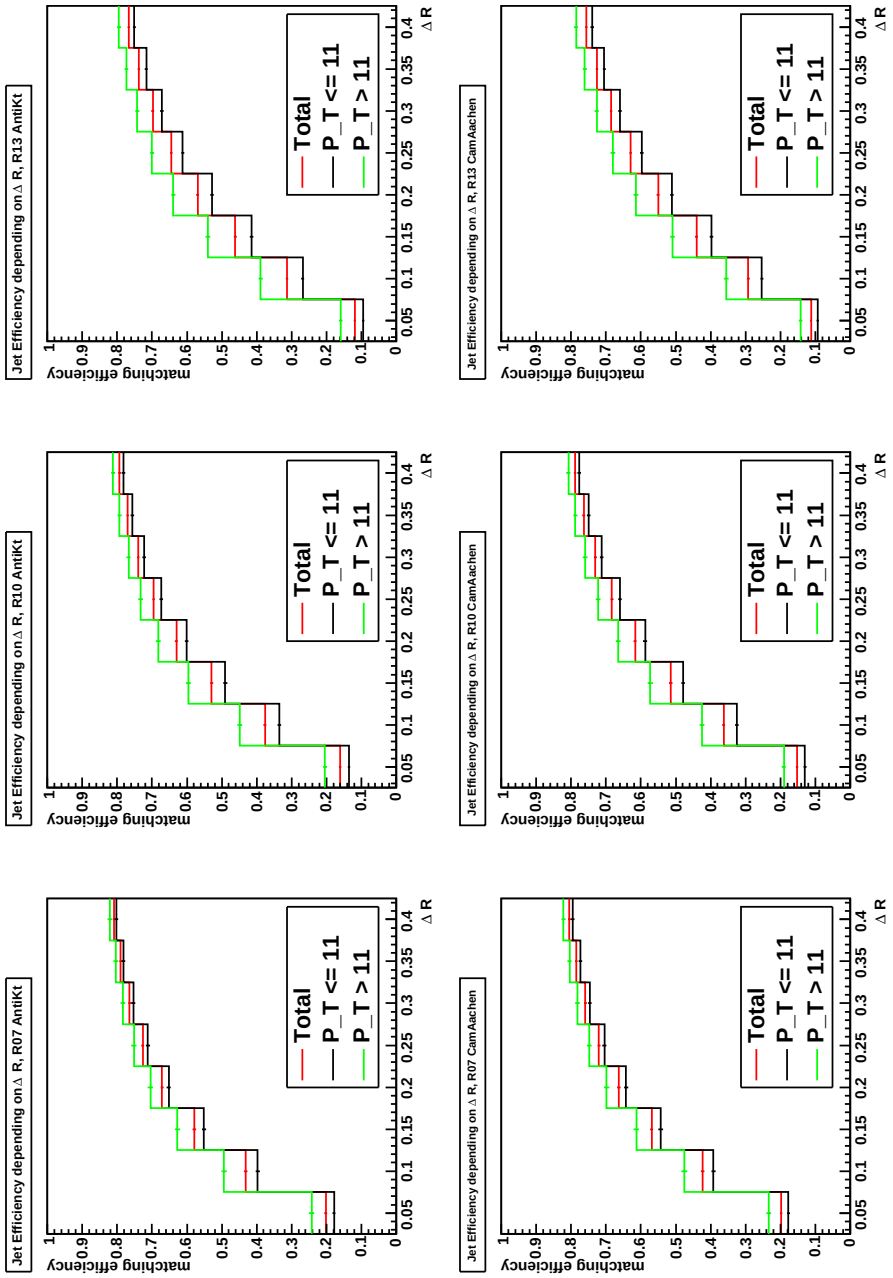


FIG 57. *$P_T$ dependance of the matching efficiency for the anti-$k_T$ and Cambridge/Aachen algorithm*

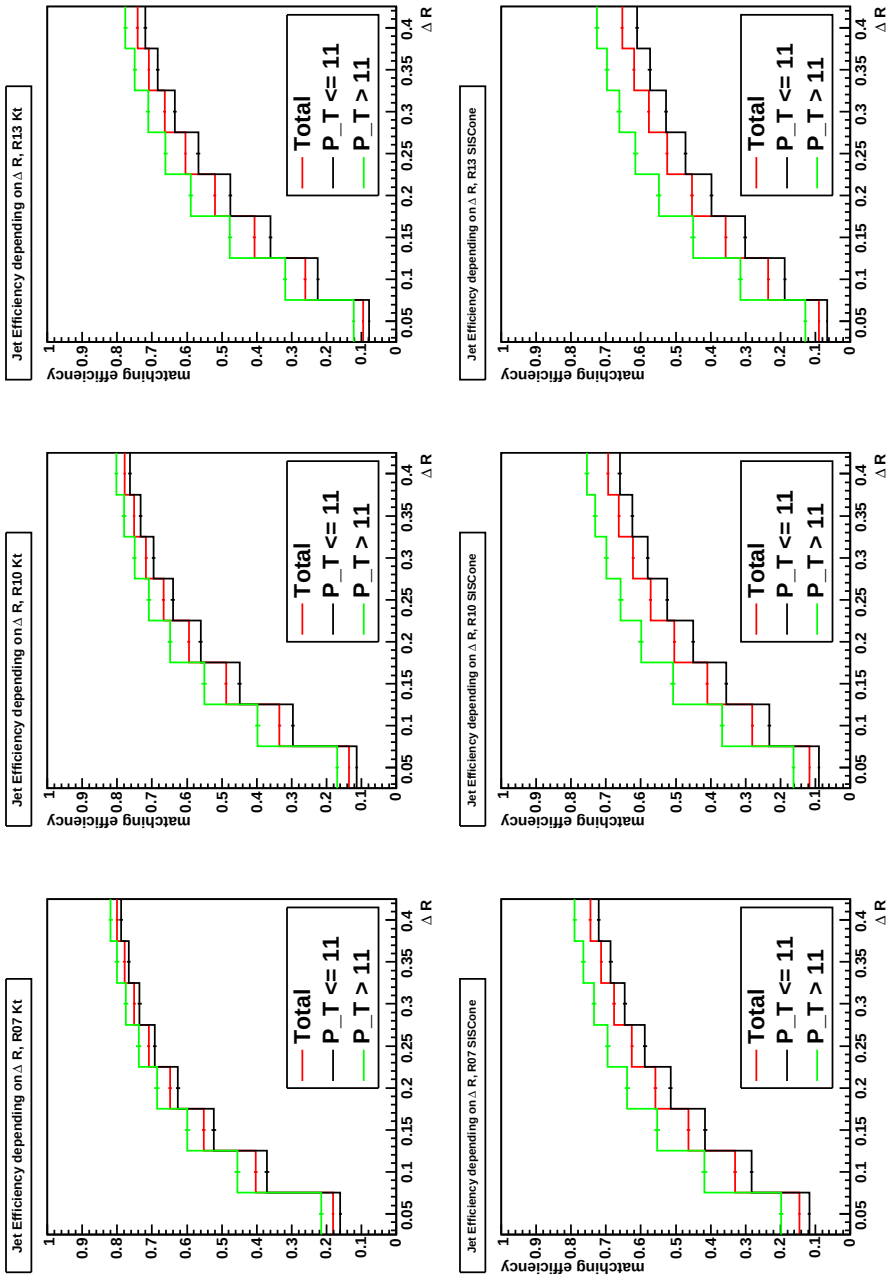


FIG 58. *$P_T$ dependance of the matching efficiency for the $k_T$ and SISCone*