

Conversational Figures: A Conversational Question-Answering Dataset Grounded in Scientific Figures and Text

Anirudh Sundar, William Gay, Antonia Rabisheva, Larry Heck

Goal

Create a **Virtual Research Assistant** that

- Amplifies human research
- Is capable of **contextual dialogue**
- Interprets document-grounded figures
- Supports **conversational Q-A**

Method

This work introduces a new dataset
CONVERSATIONAL FIGURES (cFIGS)

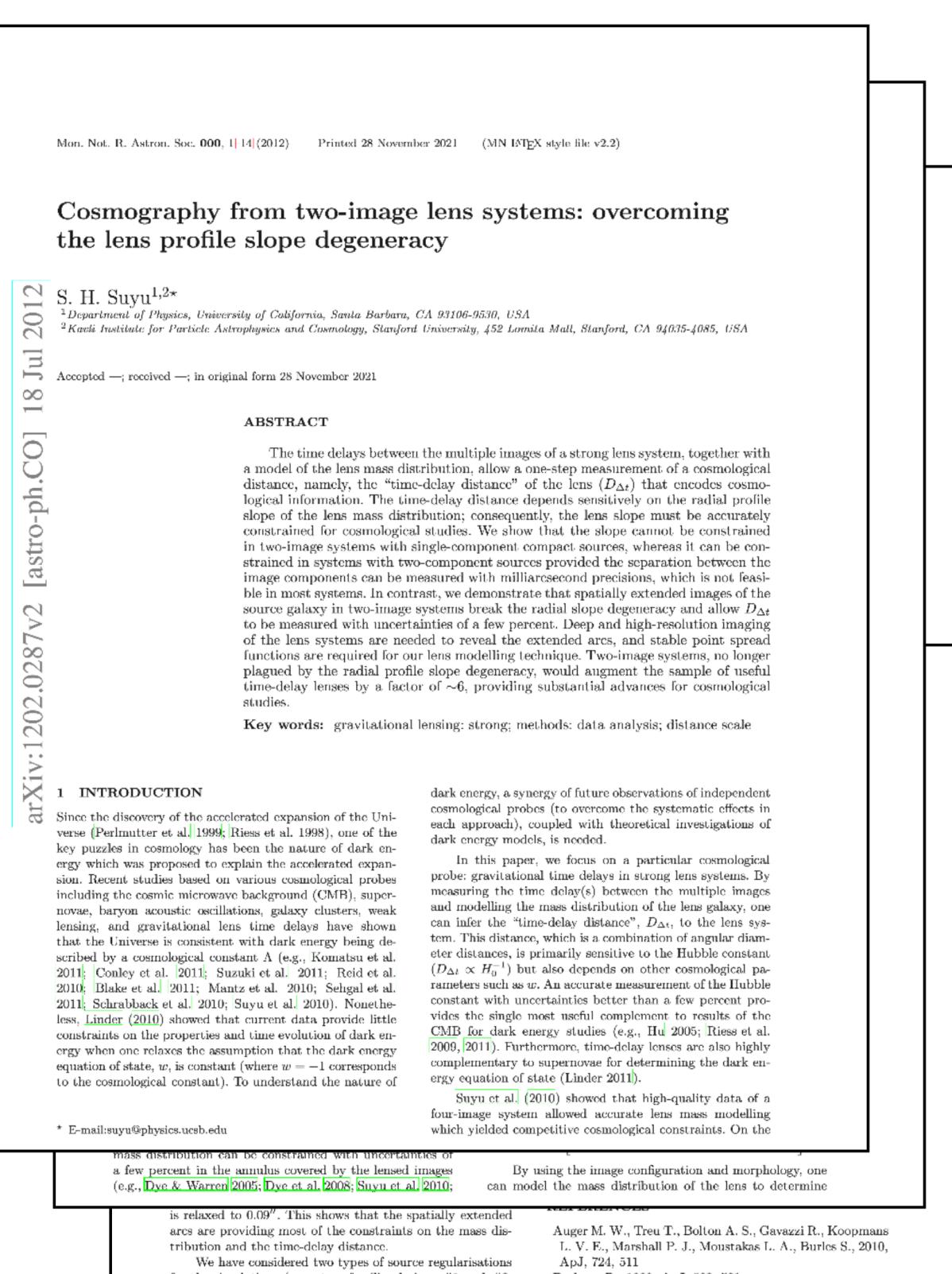
- Multi-turn
- Image-grounded
- Conversational question-answer pairs
- From scientific documents

Georgia Institute of Technology

Preliminary Results

	ROUGE 1	ROUGE 2	ROUGE L
Image + Dialogue	0.248	0.095	0.173
Image + Caption + Dialogue	0.233	0.113	0.175
Image + Caption + Paragraph + Dialogue	0.280	0.124	0.199

Dataset Collection



Open
Review
.net

Official Review of Submission85 by Reviewer 5zr1

Strengths:

- 1.The paper adds optical flow-based landmark localisation for detection of electrodes from a sequence of X-ray images.
- 2.An algorithm to construct a dataset on-the-fly is proposed, which is able to adapt a pre-trained optical model to the type of X-ray image observed in the task.

Weaknesses:

- 1.The application of optical flow seems to be incomplete. The paper claims that it views the problem as a video landmark detection network, yet optical flow is used only for two consecutive frames, with the resulting localisation information not being used for further detections. The authors could explore using longer frame stacks, adding smoothing constraints to landmark locations over time, etc. As it is presented, the methodology seems to lack several possible explorations.
- 2.The authors mention Demouster et al. 2023 as the closest work to their paper, yet it is not compared/evaluated against the proposed version. Indeed, Demouster et al. 2023 also propose using temporal context to refine detections.

Questions for the authors:

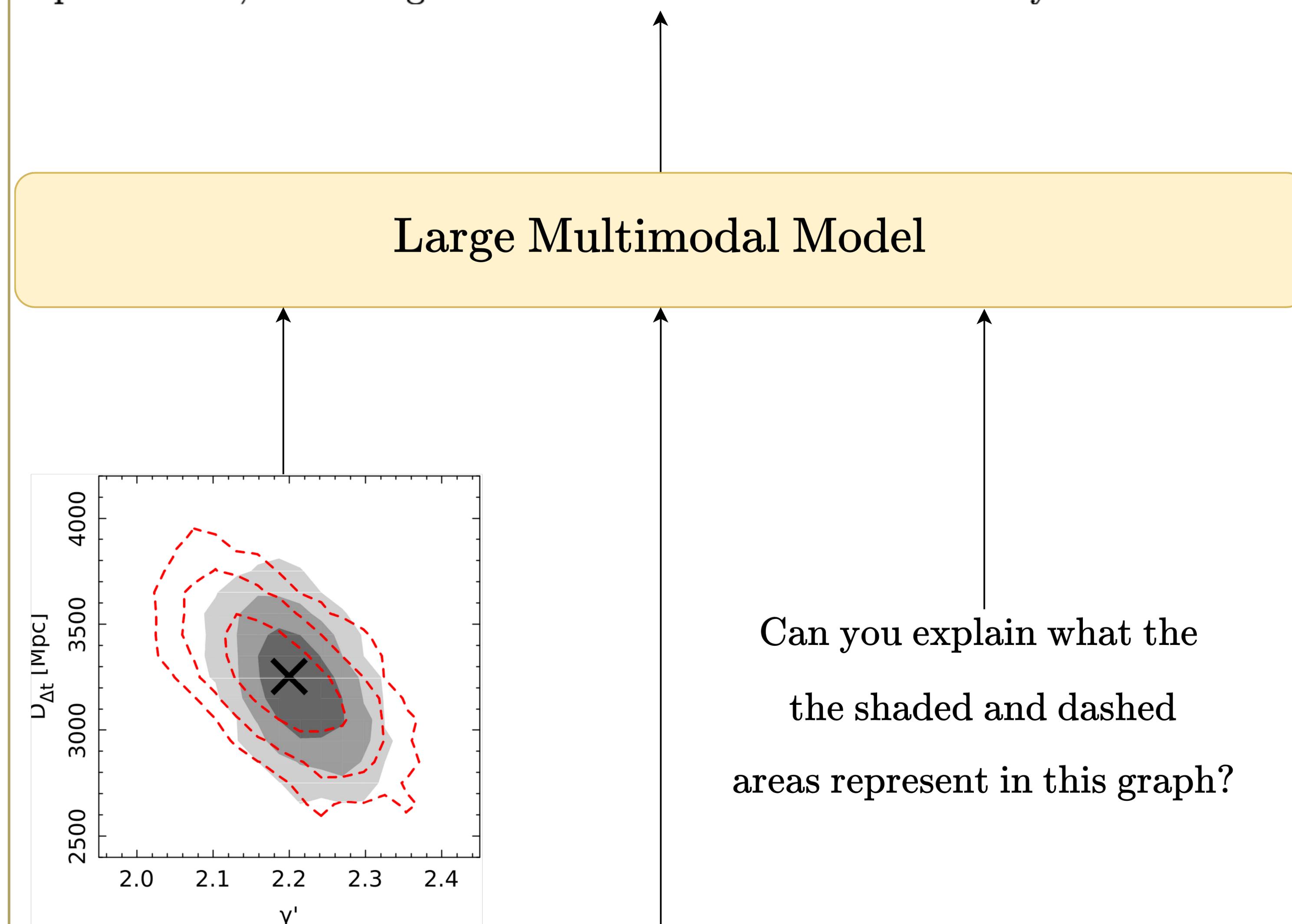
In Figure 9, could the authors clarify what the cross at the center of the image indicates?

Response

In Figure 9, the cross at the center of the image represents the input values of the parameters used in the simulation

System Architecture

The shaded area represents the credible intervals for the parameters when using the correct PSF in simulation. In contrast, the dashed area represents the credible intervals obtained when an offset PSF is used. These intervals show the 68.3%, 95.4%, and 99.7% confidence levels on the estimated parameters, indicating where the true values are most likely to be found.



Can you explain what the shaded and dashed areas represent in this graph?

Figure 9. The effect of the PSF. Marginalised PDF of γ' and D_{dt} for Simulation 3 Realisation 3 modelled using the input TinyTim PSF (shaded) and a TinyTim PSF that is offset by $\sim 45''$ from the input PSF (dashed). The three areas show the 68.3 %, 95.4% and 99.7% credible regions. The cross marks the input values. With the imperfect (offset) PSF and the scaled pixel uncertainty in equation (26), the precision of the recovered D_{dt} is slightly degraded due to misfits near the bright AGN images.

Scan to be in the loop

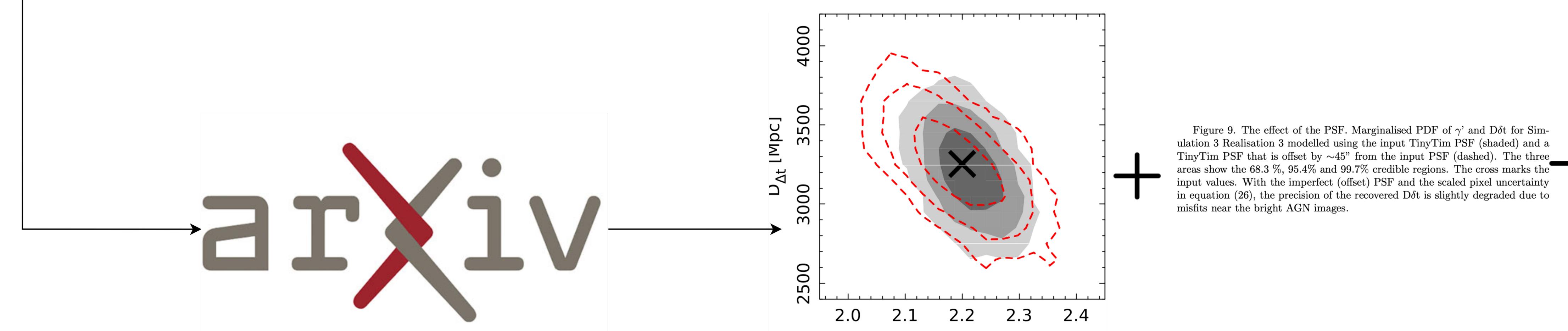


Figure 9. The effect of the PSF. Marginalised PDF of γ' and D_{dt} for Simulation 3 Realisation 3 modelled using the input TinyTim PSF (shaded) and a TinyTim PSF that is offset by $\sim 45''$ from the input PSF (dashed). The three areas show the 68.3 %, 95.4% and 99.7% credible regions. The cross marks the input values. With the imperfect (offset) PSF and the scaled pixel uncertainty in equation (26), the precision of the recovered D_{dt} is slightly degraded due to misfits near the bright AGN images.

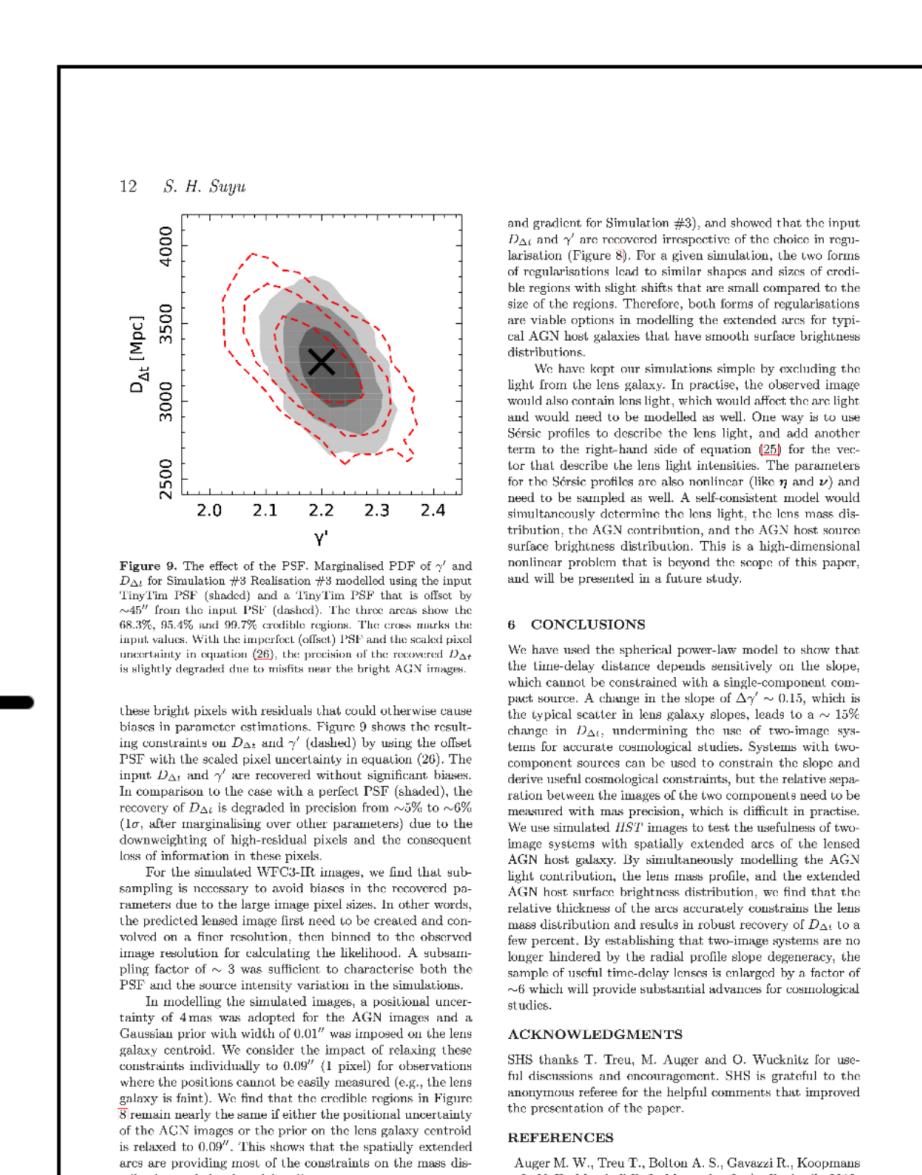


Figure 9. The effect of the PSF. Marginalised PDF of γ' and D_{dt} for Simulation 3 Realisation 3 modelled using the input TinyTim PSF (shaded) and a TinyTim PSF that is offset by $\sim 45''$ from the input PSF (dashed). The three areas show the 68.3 %, 95.4% and 99.7% credible regions. The cross marks the input values. With the imperfect (offset) PSF and the scaled pixel uncertainty in equation (26), the precision of the recovered D_{dt} is slightly degraded due to misfits near the bright AGN images.