Transfer Learning for Data-Driven Background Modelling: Optimal Transport v. Classifier Extrapolation

Tudor Manole
Carnegie Mellon University

Joint work with: Patrick Bryant, John Alison, Mikael Kuusela and Larry Wasserman

PhyStat--Systematics, November 2021

Data-Driven Background Modelling

Setting: Observe collider events $X_1, X_2, \ldots$ arising from a Poisson point process with intensity function:

$$\lambda(x) = B(x) + \mu S(x).$$

- $S(\cdot)$ is the **known** intensity function of signal events;
- $B(\cdot)$ is the **unknown** intensity function of background events;
- μ is the signal strength modifier.

Data-Driven Background Modelling

Setting: Observe collider events $X_1, X_2, \ldots$ arising from a Poisson point process with intensity function:

$$\lambda(x) = B(x) + \mu S(x).$$

- $S(\cdot)$ is the **known** intensity function of signal events;
- $B(\cdot)$ is the **unknown** intensity function of background events;
- μ is the signal strength modifier.

Goal: Test the hypothesis

$$H_0 : \mu = 0, \quad H_1 : \mu > 0.$$

- Can perform likelihood-based inference for μ once $B(\cdot)$ is given.

Data-Driven Background Modelling

Setting: Observe collider events $X_1, X_2, \ldots$ arising from a Poisson point process with intensity function:

$$\lambda(x) = B(x) + \mu S(x).$$

- $S(\cdot)$ is the **known** intensity function of signal events;
- $B(\cdot)$ is the **unknown** intensity function of background events;
- μ is the signal strength modifier.

Goal: Test the hypothesis

$$H_0 : \mu = 0, \quad H_1 : \mu > 0.$$

- Can perform likelihood-based inference for μ once $B(\cdot)$ is given.

This Talk: Data-driven estimation of the background $B(\cdot)$.

- Necessary to impose some modelling assumptions.

Data-Driven Background Modelling

Setting: Observe collider events $X_1, X_2, \ldots$ arising from a Poisson point process with intensity function:

$$\lambda(x) = B(x) + \mu S(x).$$

- $S(\cdot)$ is the **known** intensity function of signal events;
- $B(\cdot)$ is the **unknown** intensity function of background events;
- μ is the signal strength modifier.

Goal: Test the hypothesis

$$H_0 : \mu = 0, \quad H_1 : \mu > 0.$$

- Can perform likelihood-based inference for μ once $B(\cdot)$ is given.

This Talk: Data-driven estimation of the background $B(\cdot)$.

- Necessary to impose some modelling assumptions.
- Significant source of systematic uncertainties.

Data-Driven Background Modelling

Setting: Observe collider events $X_1, X_2, \ldots$ arising from a Poisson point process with intensity function:

$$\lambda(x) = B(x) + \mu S(x).$$

- $S(\cdot)$ is the **known** intensity function of signal events;
- $B(\cdot)$ is the **unknown** intensity function of background events;
- μ is the signal strength modifier.

Goal: Test the hypothesis

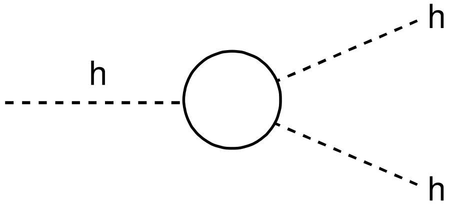
$$H_0 : \mu = 0, \quad H_1 : \mu > 0.$$

- Can perform likelihood-based inference for μ once $B(\cdot)$ is given.

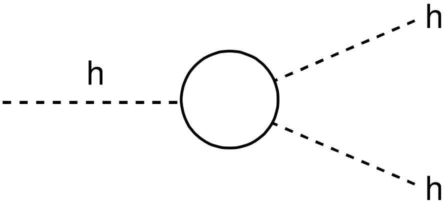
This Talk: Data-driven estimation of the background $B(\cdot)$.

- Necessary to impose some modelling assumptions.
- Significant source of systematic uncertainties.
- Our focus will be on providing cross-checks for estimating B.

Motivating Example: di-Higgs Background Modelling

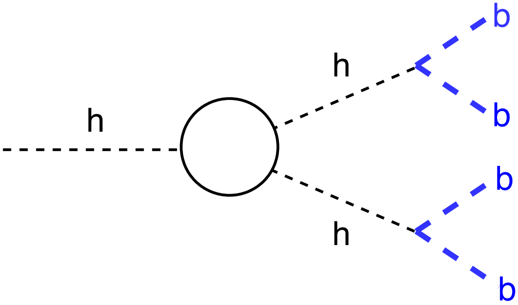


Motivating Example: di-Higgs Background Modelling



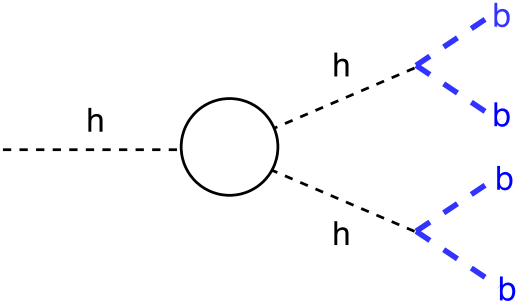
- $H \to HH$ is only expected to arise in about one per trillion events.

Motivating Example: di-Higgs Background Modelling



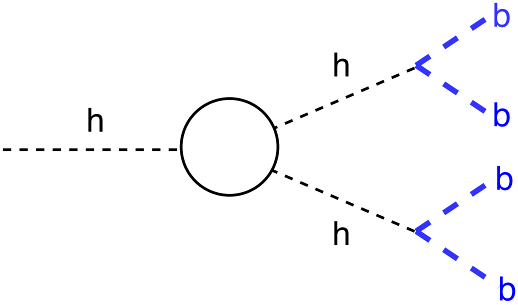
- $H \to HH$ is only expected to arise in about one per trillion events.
- $HH \to 4b$ decay channel accounts for over 30% of di-Higgs events.

Motivating Example: di-Higgs Background Modelling



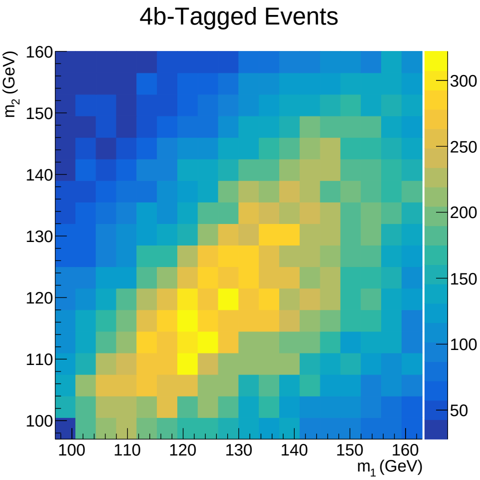
- $H \to HH$ is only expected to arise in about one per trillion events.
- $HH \to 4b$ decay channel accounts for over 30% of di-Higgs events.
  - **Signal events**: $H \to HH \to 4b$.
  - **Background events**: Any other process leading to 4 b-tagged jets.

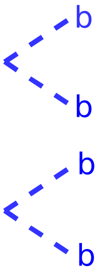
Motivating Example: di-Higgs Background Modelling



- $H \to HH$ is only expected to arise in about one per trillion events.
- $HH \to 4b$ decay channel accounts for over 30% of di-Higgs events.
  - **Signal events**: $H \to HH \to 4b$.
  - **Background events**: Any other process leading to 4 b-tagged jets.
- 4b background is very large and cannot be simulated with high fidelity $\Longrightarrow$ needs to be estimated from data.

Control and Signal Regions



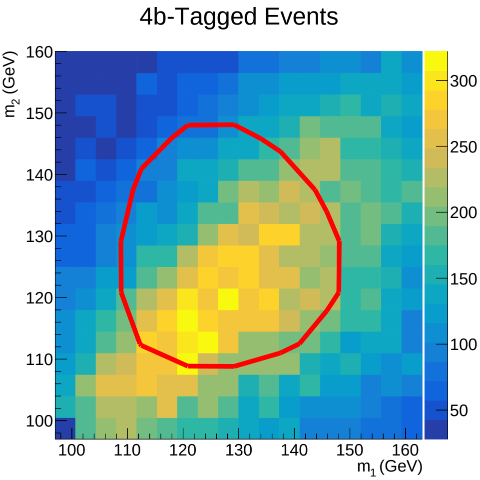


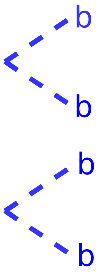
Dijet Invariant Mass m_1

Dijet Invariant Mass m_2

[1] ATLAS Collaboration (2019) "Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". Journal of High Energy Physics, 30.

Control and Signal Regions



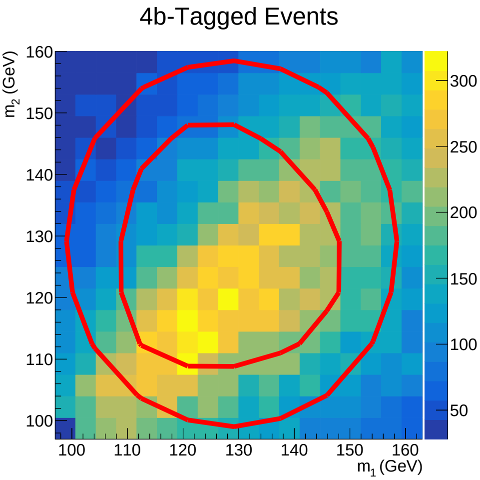


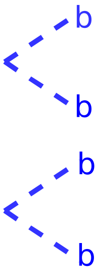
Dijet Invariant Mass m_1

Dijet Invariant Mass m_2

[2] ATLAS Collaboration (2019) ″Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector″. Journal of High Energy Physics, 30.

Control and Signal Regions



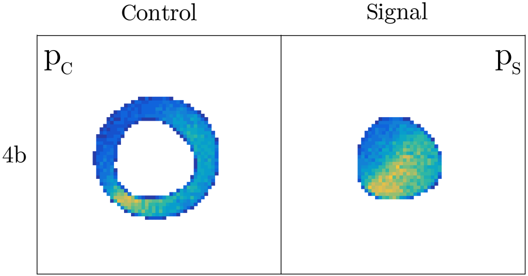


Dijet Invariant Mass m_1

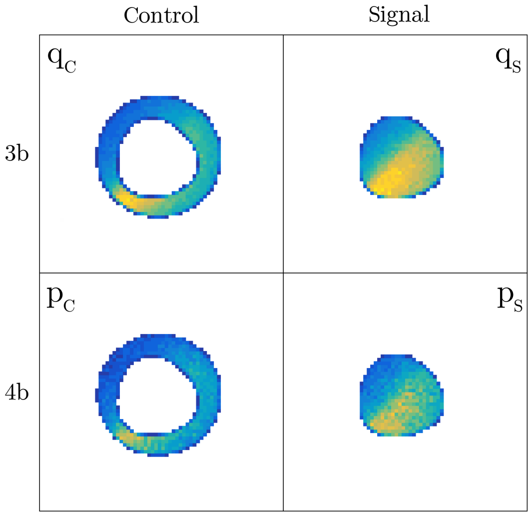
Dijet Invariant Mass m_2

[3] ATLAS Collaboration (2019) "Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". Journal of High Energy Physics, 30.

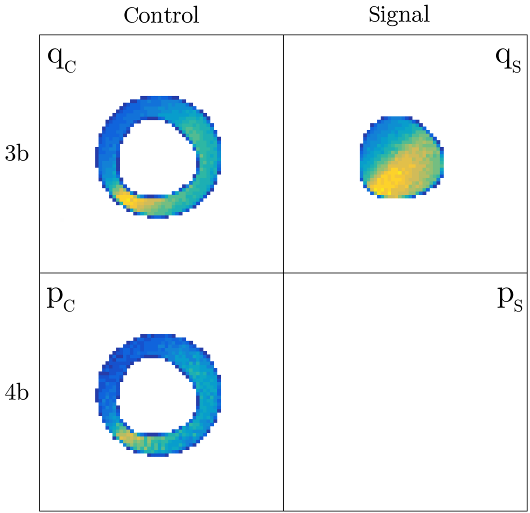
Control and Signal Regions



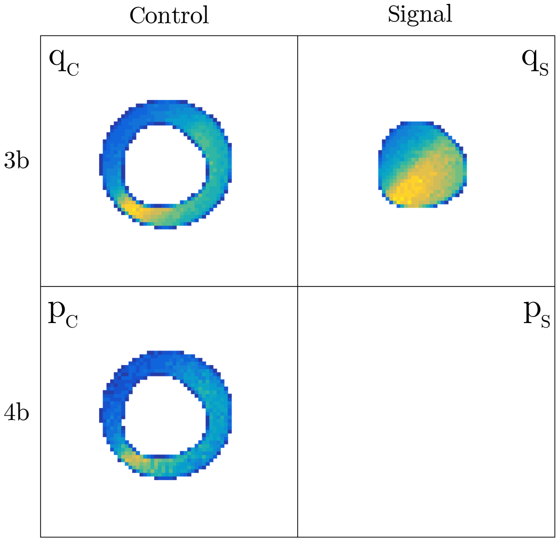
Introduce a Sample of Three-Tagged Collider Events



Introduce a Sample of Three-Tagged Collider Events



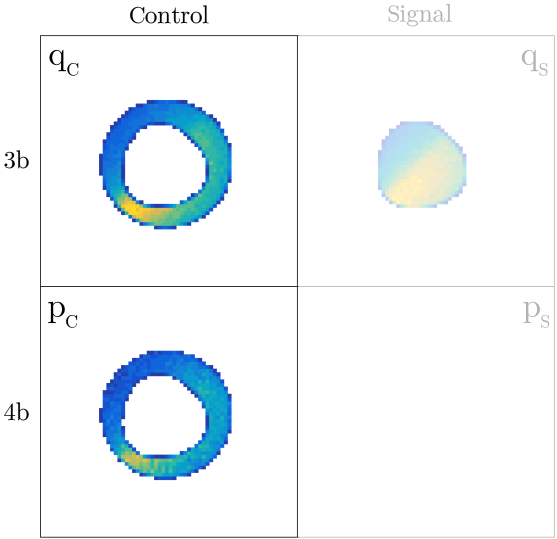
Existing Approach for Background Estimation: Density Ratio Extrapolation



Existing Approach for Background Estimation: Density Ratio Extrapolation

- Let $\hat{\psi}$ be an estimator for:

$$\psi(x) = \frac{p_C(x)}{q_C(x)}, \quad x \in \mathrm{CR}.$$



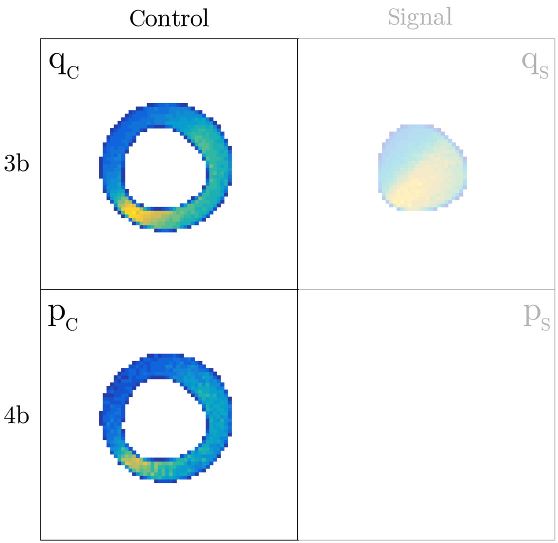
Existing Approach for Background Estimation: Density Ratio Extrapolation

- Let $\hat{\psi}$ be an estimator for:

$$\psi(x) = \frac{p_C(x)}{q_C(x)}, \quad x \in \mathrm{CR}.$$

- Assuming $\hat{\psi}$ is defined beyond the control region, we may hope for:

$$\hat{\psi}(y) \approx \frac{p_S(y)}{q_S(y)}, \quad y \in \mathrm{SR}.$$



Existing Approach for Background Estimation: Density Ratio Extrapolation

- Let $\hat{\psi}$ be an estimator for:

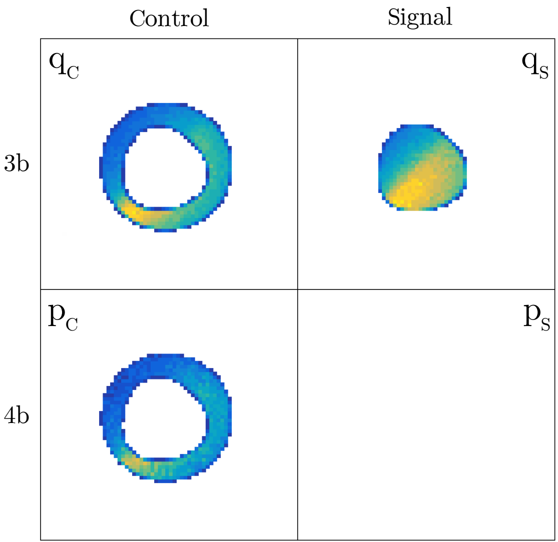
$$\psi(x) = \frac{p_C(x)}{q_C(x)}, \quad x \in \mathrm{CR}.$$

- Assuming $\hat{\psi}$ is defined beyond the control region, we may hope for:

$$\hat{\psi}(y) \approx \frac{p_S(y)}{q_S(y)}, \quad y \in \mathrm{SR}.$$

- Final estimator for p_S:

$$\hat{p}_S(y) \propto \hat{\psi}(y)\hat{q}_S(y).$$



Existing Approach for Background Estimation: Density Ratio Extrapolation

- Let $\hat{\psi}$ be an estimator for:

$$\psi(x) = \frac{p_C(x)}{q_C(x)}, \quad x \in \mathrm{CR}.$$

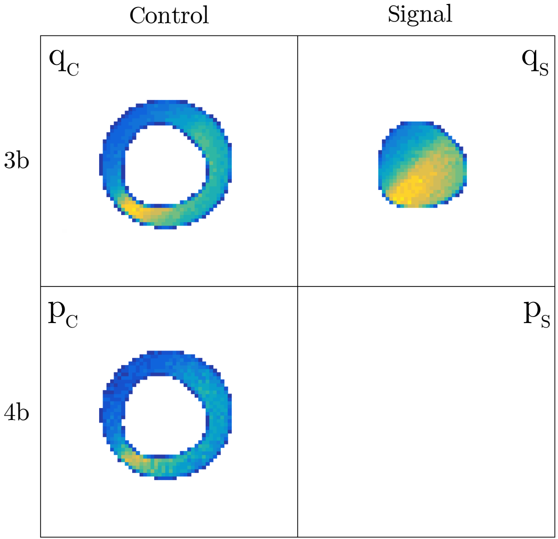
- Assuming $\hat{\psi}$ is defined beyond the control region, we may hope for:

$$\hat{\psi}(y) \approx \frac{p_S(y)}{q_S(y)}, \quad y \in \mathrm{SR}.$$

- Final estimator for p_S:

$$\hat{p}_S(y) \propto \hat{\psi}(y)\hat{q}_S(y).$$

- $\hat{\psi}$ is often estimated using a classifier for discriminating 3b and 4b.



Implicit Modelling Assumptions

Assumption 1 (Informal): The following density ratio is "sufficiently" smooth over $\mathrm{CR} \cup \mathrm{SR}$

$$\psi(x) = \begin{cases} \frac{p_C(x)}{q_C(x)}, & x \in \mathrm{CR}, \\ \frac{p_S(x)}{q_S(x)}, & x \in \mathrm{SR}. \end{cases}$$

- The extent to which Assumption 1 fails is a systematic effect, which is difficult to quantify.

Implicit Modelling Assumptions

Assumption 1 (Informal): The following density ratio is "sufficiently" smooth over CR ∪ SR

$$\psi(x) = \begin{cases} \frac{p_C(x)}{q_C(x)}, & x \in \mathrm{CR}, \\ \frac{p_S(x)}{q_S(x)}, & x \in \mathrm{SR}. \end{cases}$$

- The extent to which Assumption 1 fails is a systematic effect, which is difficult to quantify.
- We develop a method which makes the following **orthogonal** assumption.

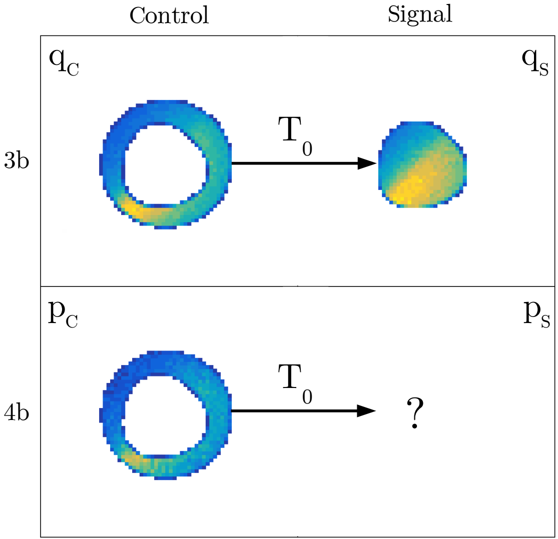
Assumption 2: There exists a map $T_0 : \mathrm{CR} \to \mathrm{SR}$ such that

$$3b: \quad X \sim q_C \implies T_0(X) \sim q_S,$$
$$4b: \quad Z \sim p_C \implies T_0(Z) \sim p_S.$$

Optimal Transport Maps

Assumption 2: There exists a map $T_0 : \mathrm{CR} \to \mathrm{SR}$ such that

$$3b: \quad X \sim q_C \Longrightarrow T_0(X) \sim q_S,$$
$$4b: \quad Z \sim p_C \Longrightarrow T_0(Z) \sim p_S.$$

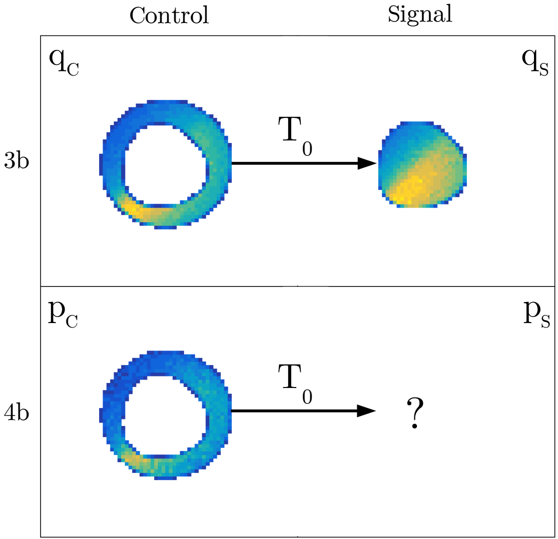


Optimal Transport Maps

Assumption 2: There exists a map $T_0 : \mathrm{CR} \to \mathrm{SR}$ such that

$$3b: \quad X \sim q_C \implies T_0(X) \sim q_S,$$
$$4b: \quad Z \sim p_C \implies T_0(Z) \sim p_S.$$

- There are many maps pushing q_C to q_S.



Optimal Transport Maps

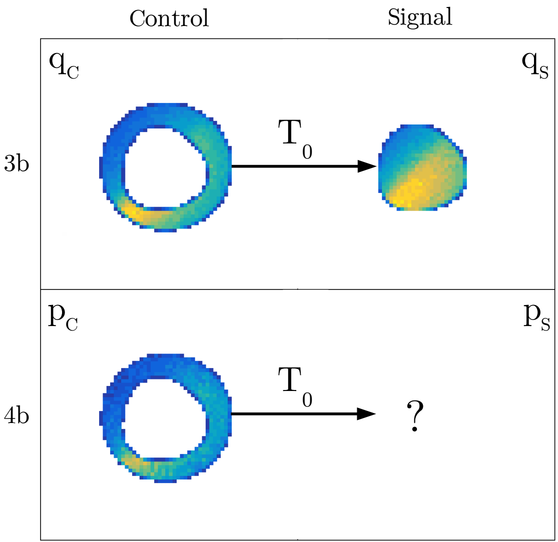
Assumption 2: There exists a map $T_0 : \mathrm{CR} \to \mathrm{SR}$ such that

$$3b: \quad X \sim q_C \Longrightarrow T_0(X) \sim q_S,$$
$$4b: \quad Z \sim p_C \Longrightarrow T_0(Z) \sim p_S.$$

- There are many maps pushing q_C to q_S.
- We will assume T_0 is the one which requires the least work for transforming $X \sim q_C$:

$$T_0 := \operatorname*{argmin}_{T: T(X) \sim q_S} \mathbb{E}\big[D(X, T(X))\big].$$

Here, D is a metric over collider events[4].



[4] Komiske, P. T., Metodiev, E. M., and Thaler, J. (2019). Metric Space of Collider Events. Physical Review Letters, 123(4).

Optimal Transport Maps

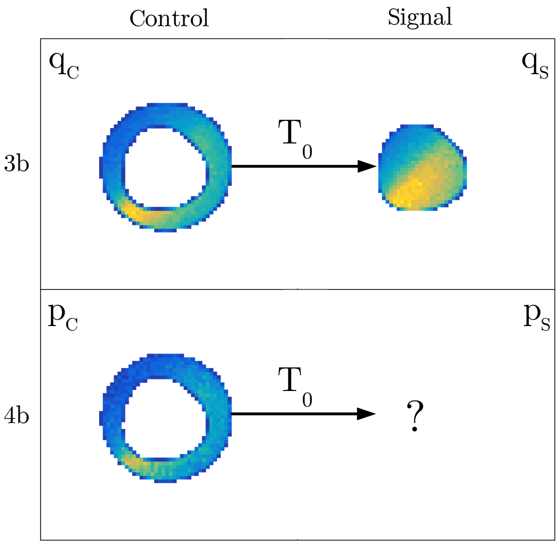
Assumption 2': The OT map from q_C to q_S also pushes p_C onto p_S, i.e.:

$$Z \sim p_C \Longrightarrow T_0(Z) \sim p_S.$$

- There are many maps pushing q_C to q_S.
- We will assume T_0 is the one which requires the least work for transforming $X \sim q_C$:

$$T_0 := \underset{T: T(X) \sim q_S}{\operatorname{argmin}} \mathbb{E}\big[D(X, T(X))\big].$$

Here, D is a metric over collider events[4].

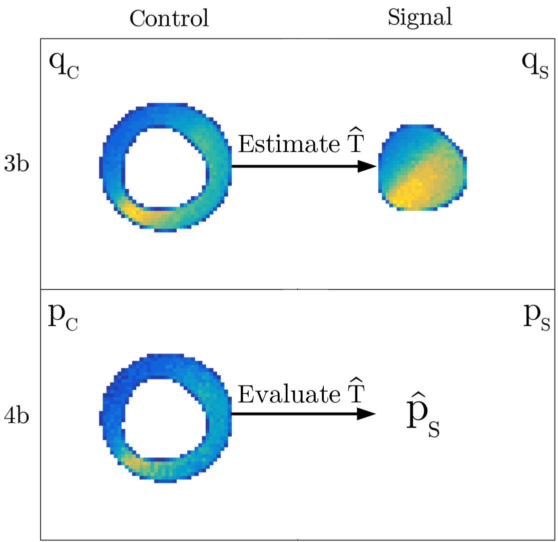


[4] Komiske, P. T., Metodiev, E. M., and Thaler, J. (2019). Metric Space of Collider Events. Physical Review Letters, 123(4).

Estimating p_S using Optimal Transport

Background Modelling Procedure:

- Form estimate $\hat{T}_n$ of T_0 using 3b data.
- Evaluate $\hat{T}_n$ on 4b data in CR.

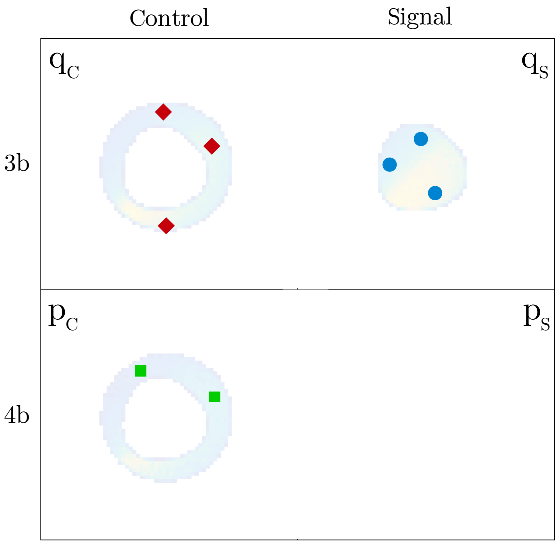


Estimating p_S using Optimal Transport

Background Modelling Procedure:

- Form estimate $\hat{T}_n$ of T_0 using 3b data.
- Evaluate $\hat{T}_n$ on 4b data in CR.

Estimators $\hat{T}_n$[5]:



[5] e.g.: arXiv:2107.12364; arXiv:2107.01718.

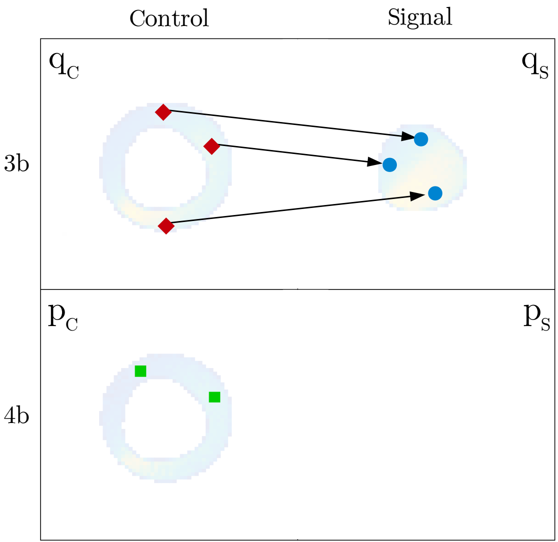
Estimating p_S using Optimal Transport

Background Modelling Procedure:

- Form estimate $\hat{T}_n$ of T_0 using 3b data.
- Evaluate $\hat{T}_n$ on 4b data in CR.

Estimators $\hat{T}_n$[5]:

- **"Map-then-Smooth"**: Compute the empirical OT map between 3b points, then extend it using linear smoothers (e.g. nearest neighbors).



[5] e.g.: arXiv:2107.12364; arXiv:2107.01718.

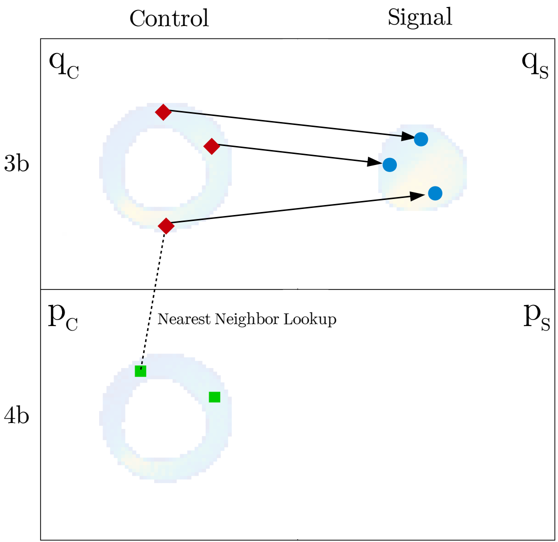
Estimating p_S using Optimal Transport

Background Modelling Procedure:

- Form estimate $\hat{T}_n$ of T_0 using 3b data.
- Evaluate $\hat{T}_n$ on 4b data in CR.

Estimators $\hat{T}_n$[5]:

- **"Map-then-Smooth"**: Compute the empirical OT map between 3b points, then extend it using linear smoothers (e.g. nearest neighbors).



[5] e.g.: arXiv:2107.12364; arXiv:2107.01718.

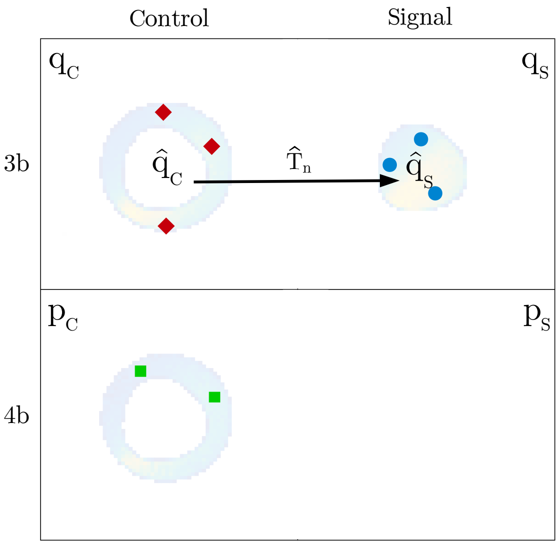
Estimating p_S using Optimal Transport

Background Modelling Procedure:

- Form estimate $\hat{T}_n$ of T_0 using 3b data.
- Evaluate $\hat{T}_n$ on 4b data in CR.

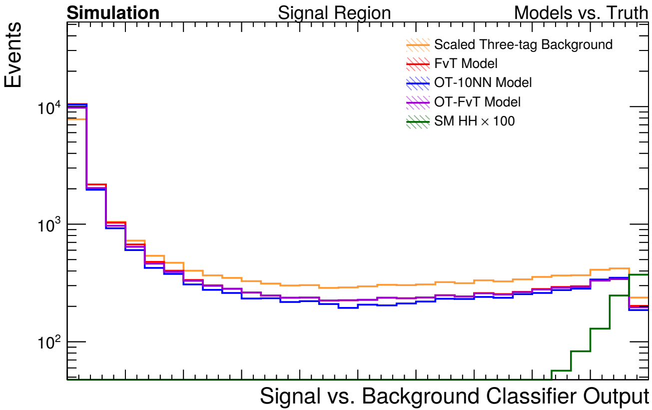
Estimators $\hat{T}_n$[5]:

- **"Map-then-Smooth"**: Compute the empirical OT map between 3b points, then extend it using linear smoothers (e.g. nearest neighbors).
- **"Smooth-then-Map"**: Compute the unique OT map between density estimators $\hat{q}_C$ and $\hat{q}_S$.
  - We again avoid density estimation by using 3b-vs-4b classification.

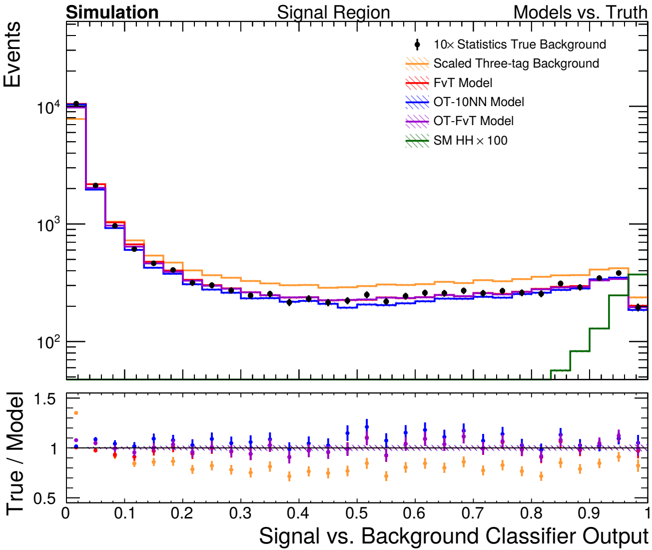


[5] e.g.: arXiv:2107.12364; arXiv:2107.01718.

Simulation Results



Simulation Results



Discussion

- **Main Message**: Optimal transport can provide a useful cross-check for existing background estimation methods, due its complementary modelling assumptions.

Discussion

- **Main Message**: Optimal transport can provide a useful cross-check for existing background estimation methods, due its complementary modelling assumptions.
- The extent to which these methods differ can potentially be used to quantify background modelling systematics.

Discussion

- **Main Message**: Optimal transport can provide a useful cross-check for existing background estimation methods, due its complementary modelling assumptions.
- The extent to which these methods differ can potentially be used to quantify background modelling systematics.
- Quantifying the uncertainty in estimating optimal transport maps also remains a challenging open problem.

Discussion

- **Main Message**: Optimal transport can provide a useful cross-check for existing background estimation methods, due its complementary modelling assumptions.
- The extent to which these methods differ can potentially be used to quantify background modelling systematics.
- Quantifying the uncertainty in estimating optimal transport maps also remains a challenging open problem.
- This is only one of many potential applications of the optimal transport toolbox in particle physics...

Thank You

Backup

Some Theoretical Properties

Theorem (M. et al., 2021; Informal). Under some technical conditions, if T_0 is Lipschitz, then the "map-then-smooth" estimator $\hat{T}_n$ based on one-nearest neighbor extrapolation satisfies (for $d \geq 5$),

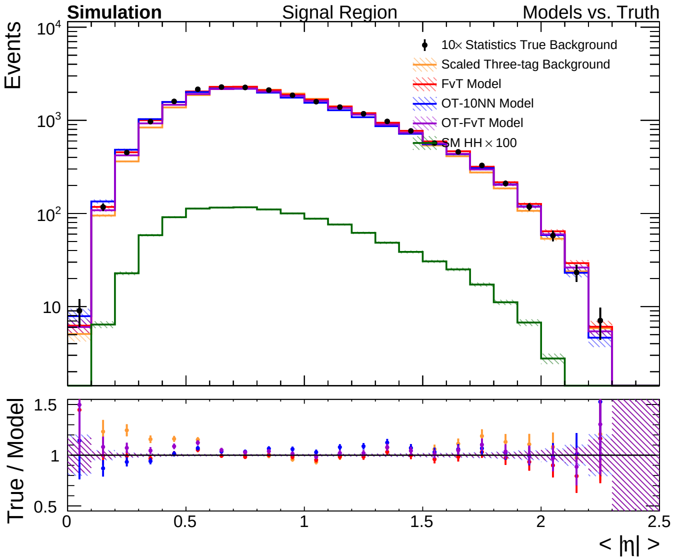
$$\mathbb{E}\|\hat{T}_n - T_0\|_{L^2}^2 \lesssim \left(\frac{\log n}{n}\right)^{\frac{2}{d}}.$$

- Convergence rates as fast as $1/n$ can be obtained for "smooth-then-map" estimators under stronger conditions.
- **Caveat**: This assumes $D = \|\cdot\|^2$, whereas we take D to be the metric of Komiske et al. (2019), itself arising from an optimal transport problem between event jets.

[6] Manole, T., Balakrishnan, S., Niles-Weed, J. and Wasserman, L. (2021). Plugin Estimation of Smooth Optimal Transport Maps. Preprint, arXiv:2107.12364.

[7] Komiske, P. T., Metodiev, E. M., and Thaler, J. (2019). Metric Space of Collider Events. Physical Review Letters, 123(4).

Further Simulation Results



Further Simulation Results