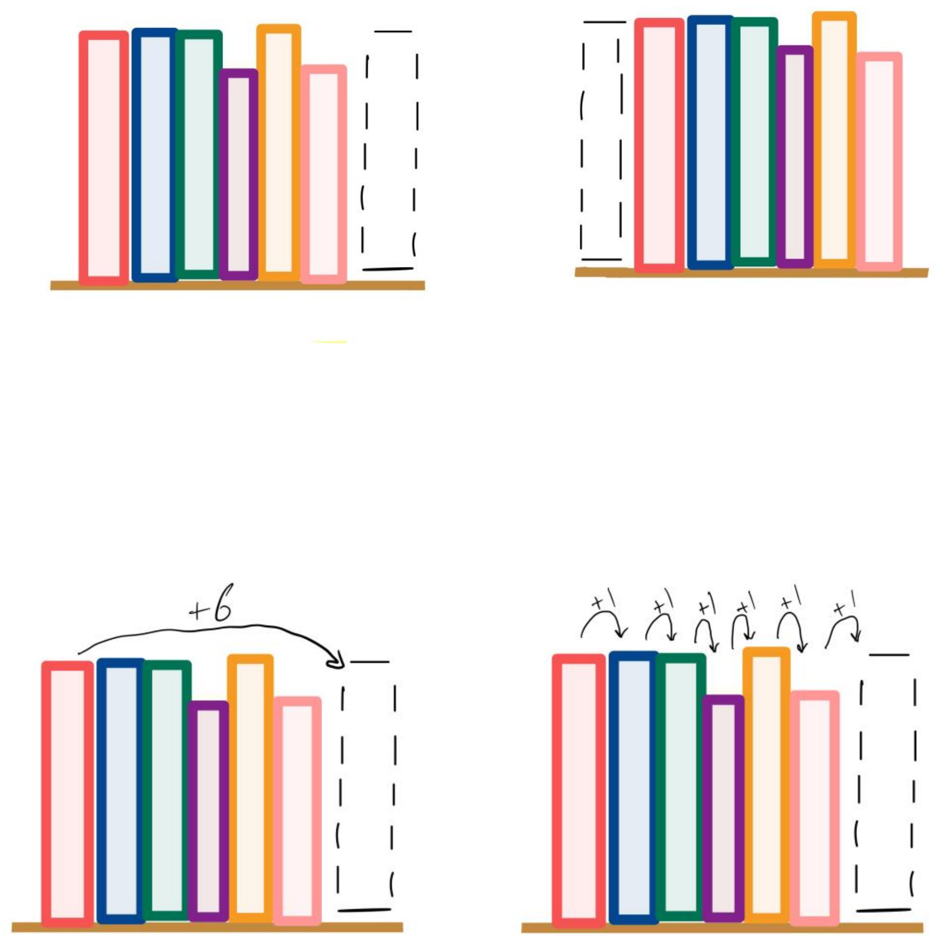
Unbalanced Optimal Transport: Exploring its use for Spatial Forecast Verification

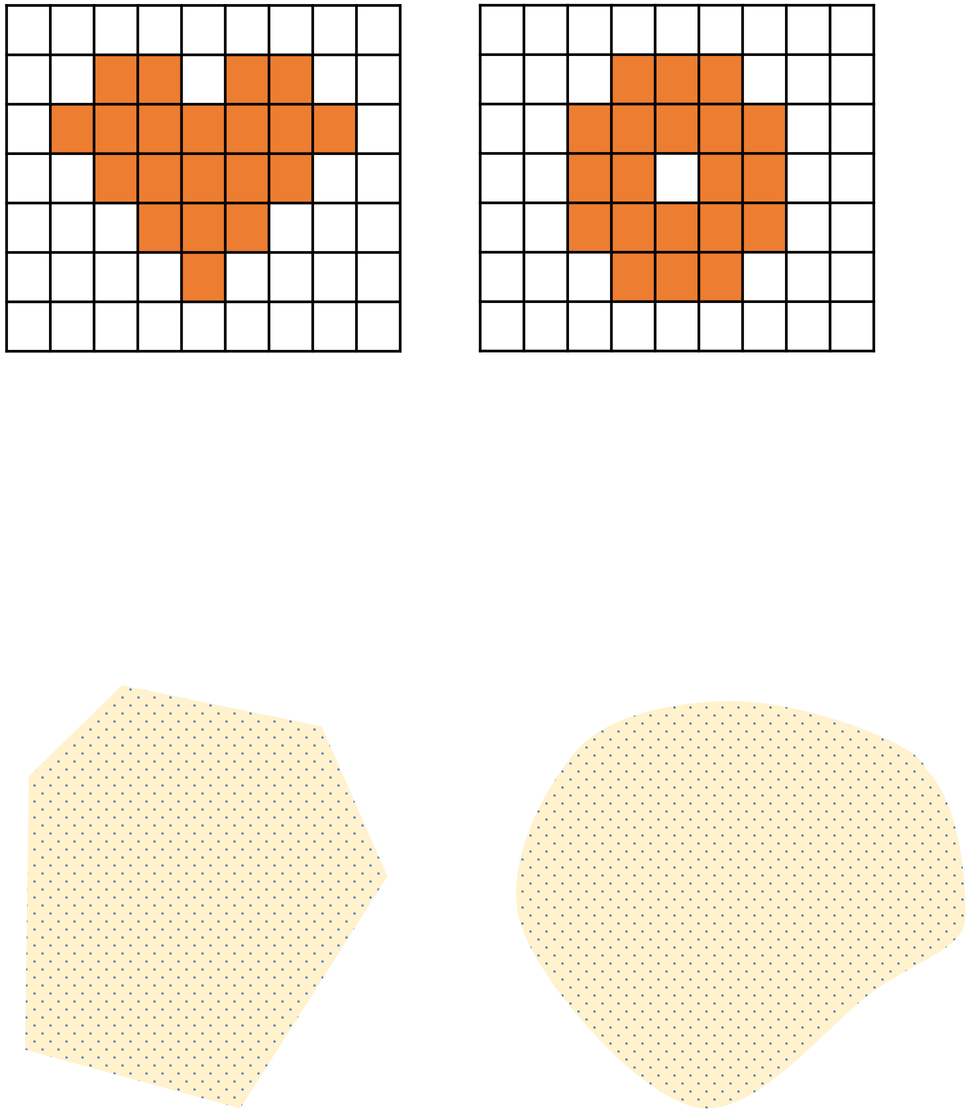
Jacob Francis

Supervisors; Professor Colin Cotter (Imperial College London), Dr Marion Mittermaier (MET Office UK)

Optimal Transport: Intuition



Optimal Transport: Intuition



Optimal Transport (Wasserstein Distance)

Given measures $\alpha \in \mathcal{P}(\mathcal{X})$, $\beta \in \mathcal{P}(\mathcal{Y})$, the optimal plan is given by $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ such that;

$$\text{Primal:} \quad \pi^* \in argmin_{\pi \in \mathcal{U}(\alpha,\beta)} \int_{\mathcal{X} \times \mathcal{Y}} c(x,y) d\pi(x,y) \tag{1}$$

$$\text{Dual:} \quad f^*, g^* \in argmax \int_{\mathcal{X}} f(x) d\alpha(x) + \int_{\mathcal{Y}} g(y) \beta(y), \tag{2}$$

for $f(x) \in \mathcal{C}(X), g(y) \in \mathcal{C}(Y) : f(x) + g(y) \leq c(x,y)$ and where $\mathcal{U}$ is the space of joint probability measures with marginals α, β.

$$c(x,y) = \frac{1}{2}[(x_1 - y_1)^2 + (x_2 - y_2)^2]$$

Plans

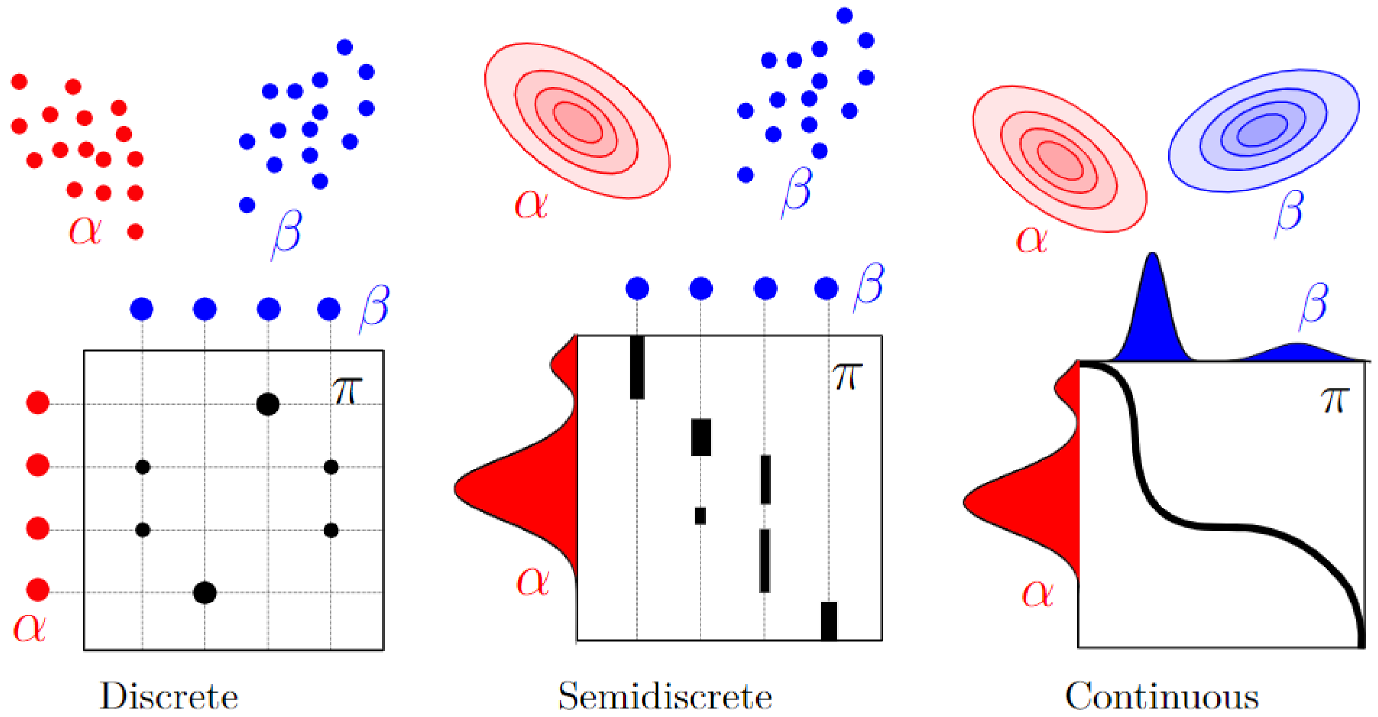
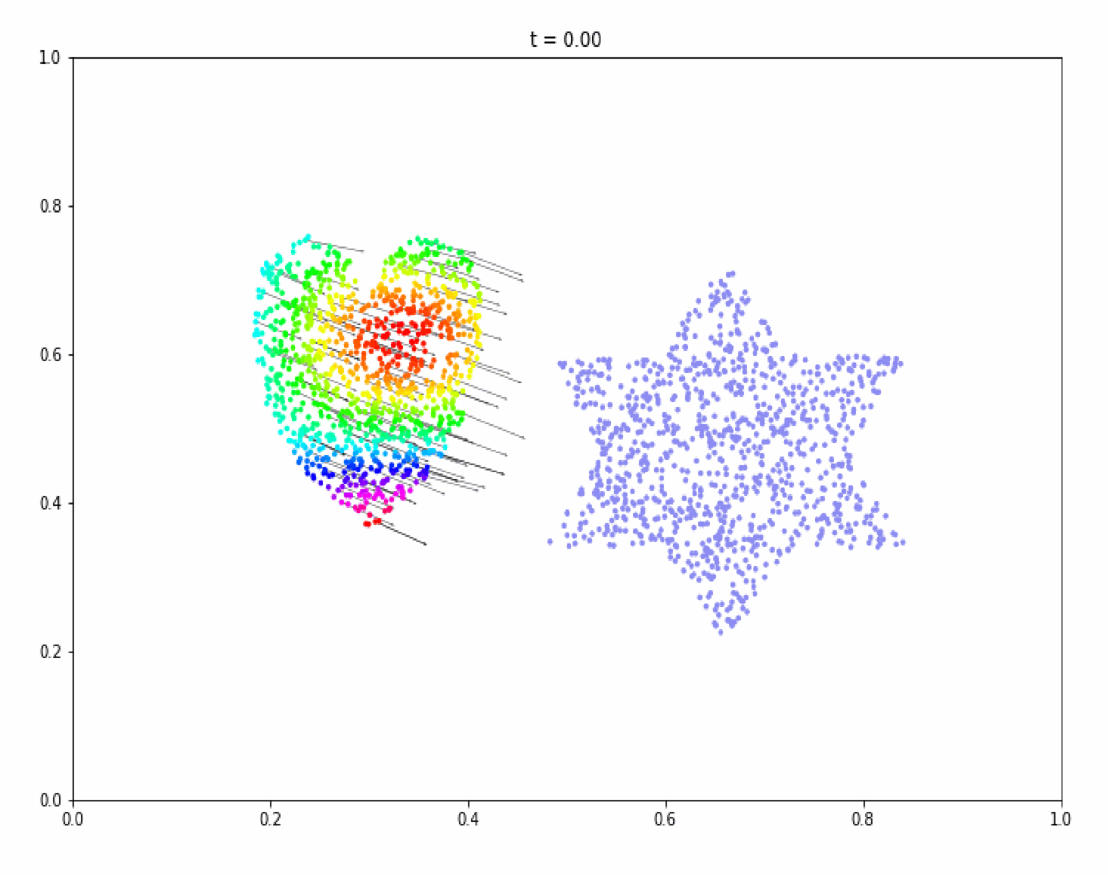


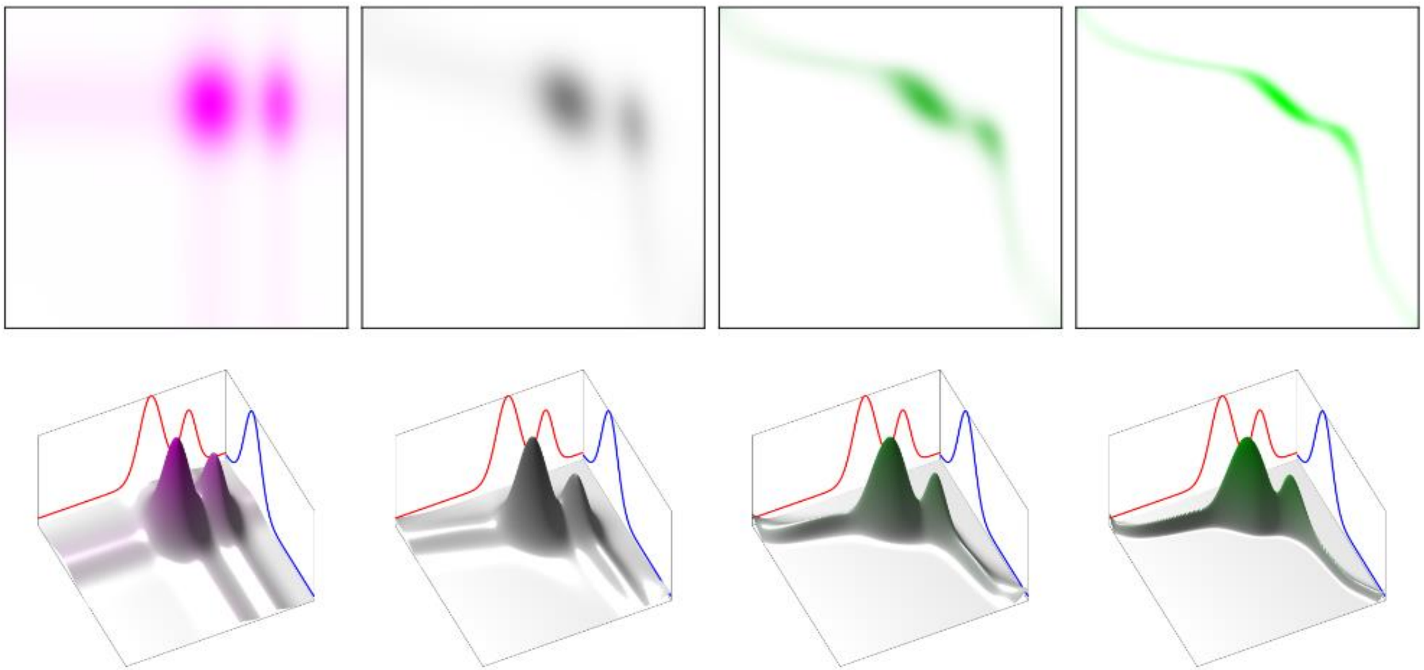
Figure 2.5: Schematic viewed of input measures (α, β) and couplings $\mathcal{U}(\alpha, \beta)$ encountered in the three main scenarios for Kantorovich OT. Chapter 5 is dedicated to the semidiscrete setup.

[Peyré and Cuturi, 2019]

Optimal Transport: Intuition



Entropic Plans



$\varepsilon = 10$ $\varepsilon = 1$ $\varepsilon = 10^{-1}$ $\varepsilon = 10^{-2}$

Figure 4.2: Impact of ε on the couplings between two 1-D densities, illustrating Proposition 4.1. Top row: between two 1-D densities. Bottom row: between two 2-D discrete empirical densities with the same number $n = m$ of points (only entries of the optimal $(\mathbf{P}_{i,j})_{i,j}$ above a small threshold are displayed as segments between x_i and y_j).

[Peyré and Cuturi, 2019]

Optimal Transport: Unbalanced

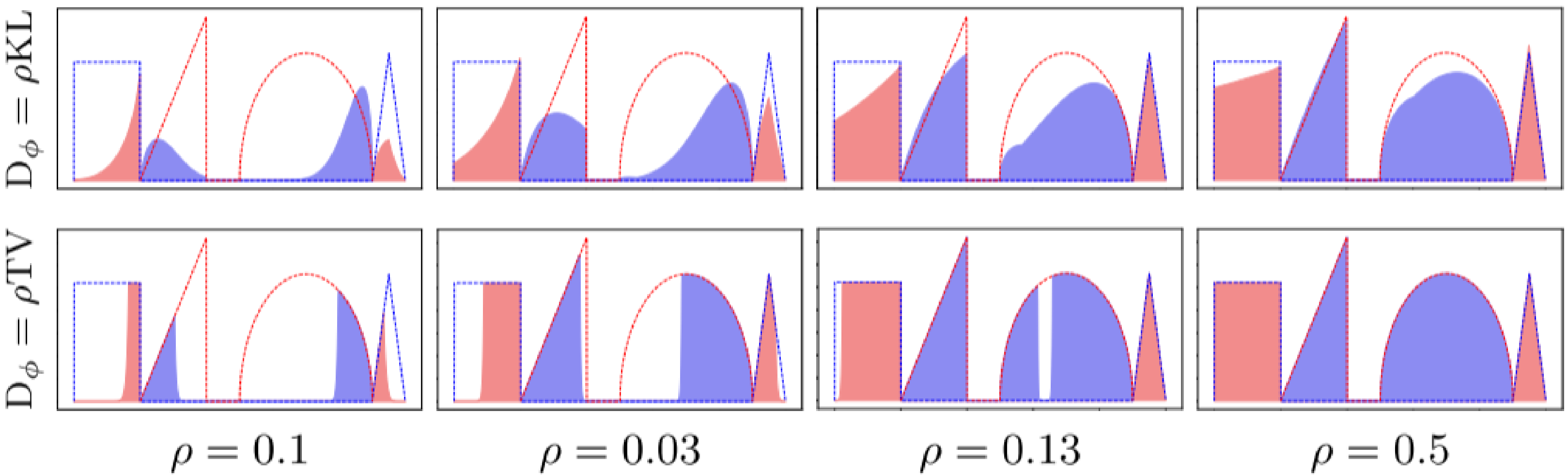


Fig. 3: Display of the impact of the reach parameter ρ on the marginals (π_1, π_2) given the same inputs (α, β) from Figure 2. First line corresponds to $\rho\mathrm{KL}$ and the second to $\rho\mathrm{TV}$.

[Séjourné et al, 2019]

Kullback-Leiber (KL)	$\rho \log(\frac{d\gamma}{d\lambda}) - 1, \gamma + \rho 1, \lambda$
Total Variation (TV)	$\rho \lvert \frac{d\gamma}{d\lambda} - 1 \rvert, \lambda$

ICP Geometric Cases

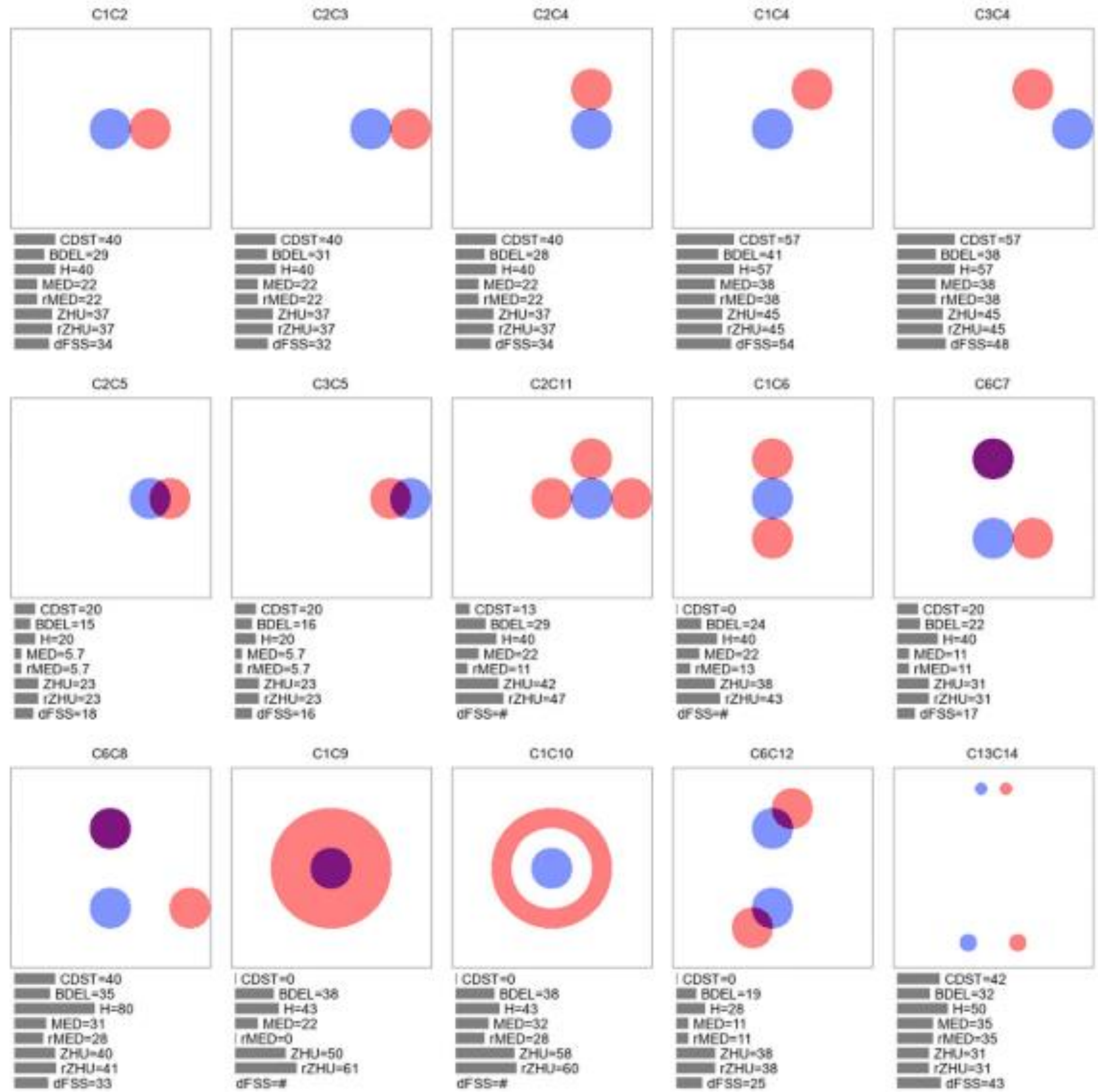


FIG. 6. As in Fig. 4, but for circular comparisons.

[Gilleland et al, 2020]

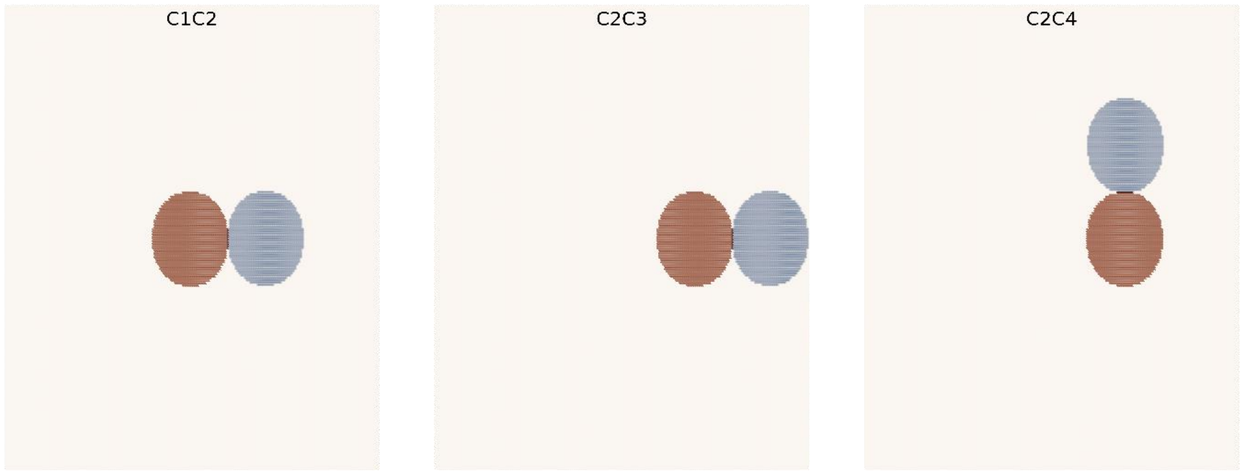
Summary of OT scores

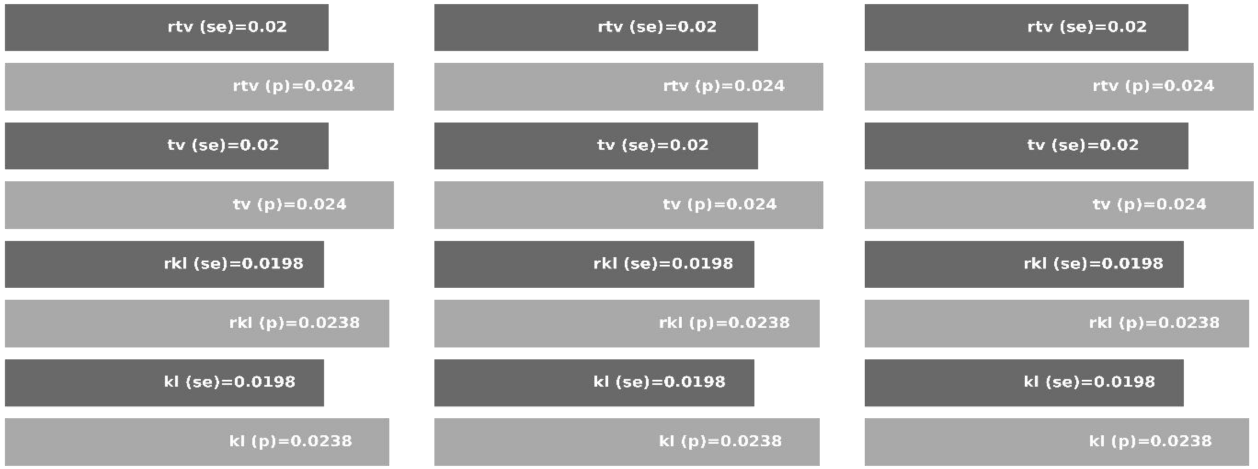
These are then accompanied by reverse version of each.

Cost	Description
Primal KL cost	Entropic regularised unbalanced bias cost with Kullback-Leiber unbalanced regularisation
Primal TV cost	Entropic regularised unbalanced bias cost with Total variation unbalanced regularisation
Sinkhorn Divergence KL cost	Debiased primal KL cost. Including extra terms to post optimisation remove noise
Sinkhorn Divergence TV cost	Debiased primal TV cost. Including extra terms to post optimisation remove noise

Relative boundary

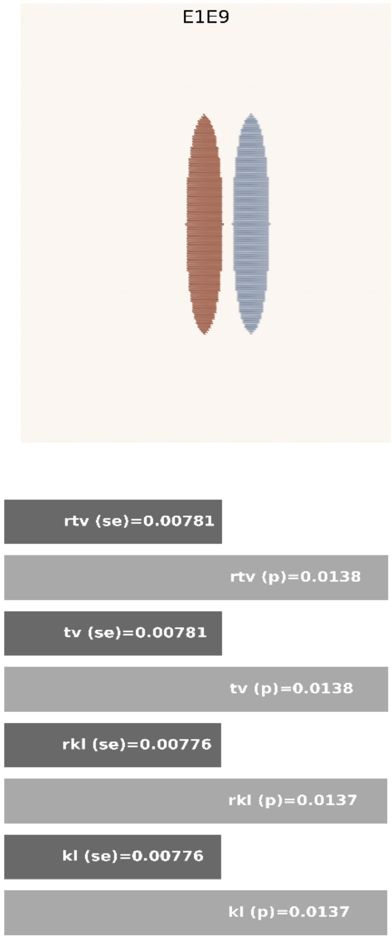
Key:
tv = total variation cost
kl = Kullback-Leiber cost
(se) = Sinkhorn Divergence cost
(p) = primal cost
r__ = reverse cost

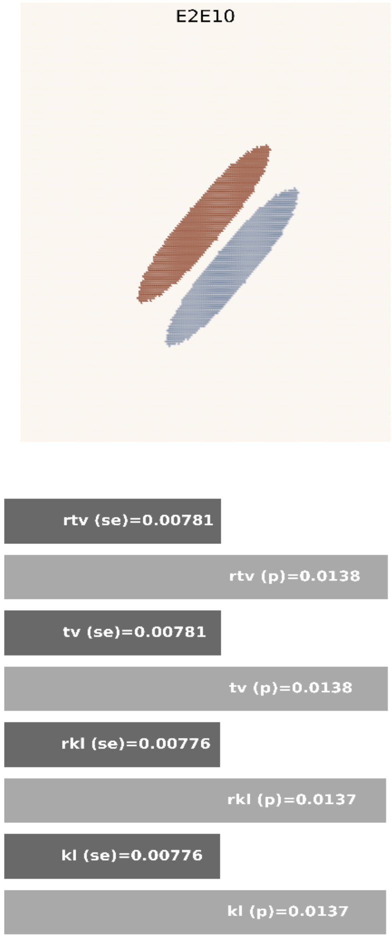


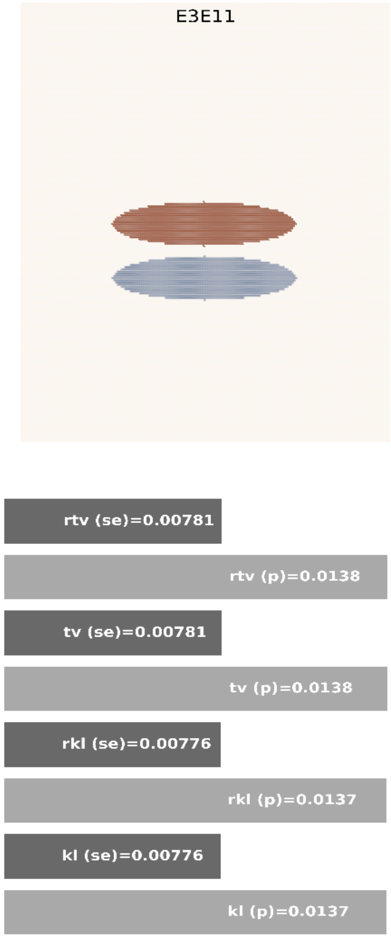


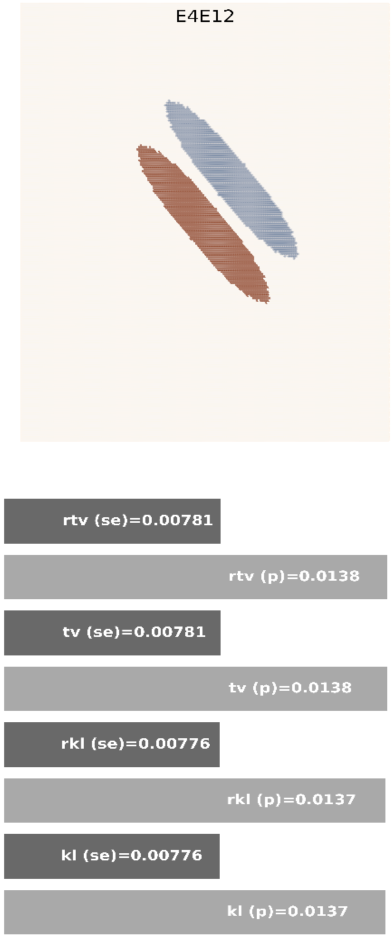
Relative boundary (rotation)

Key:
tv = total variation cost
kl = Kullback-Leiber cost
(se) = Sinkhorn Divergence cost
(p) = primal cost
r__ = reverse cost



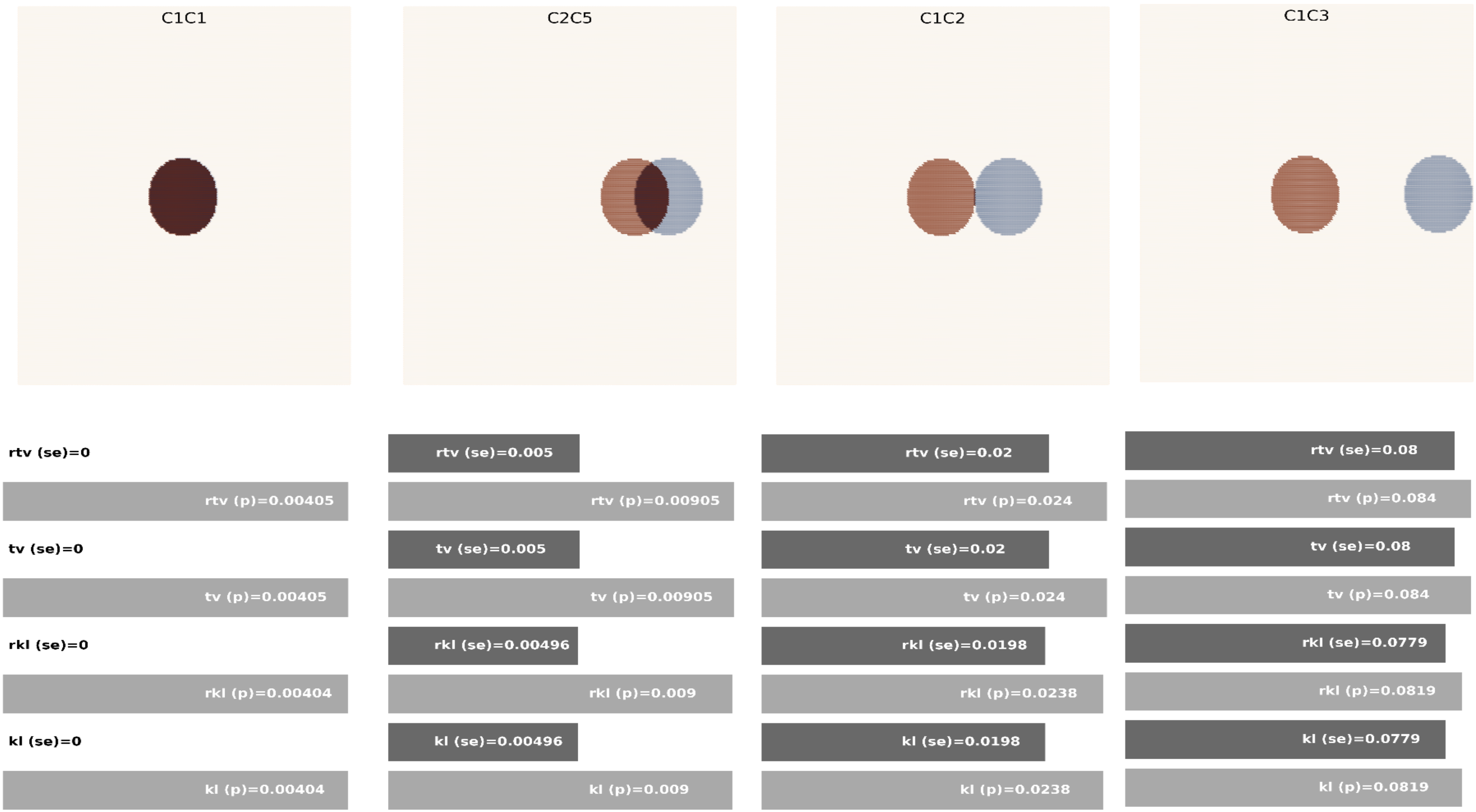




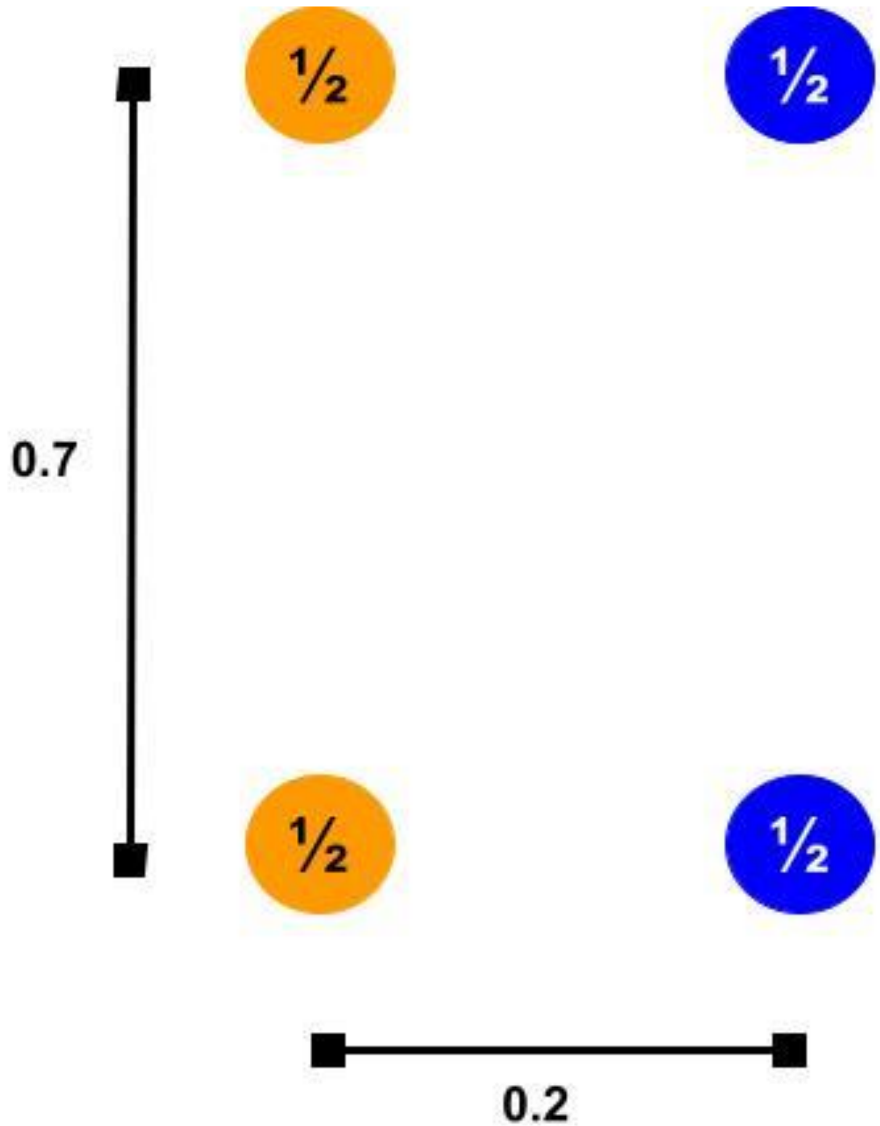


Translation

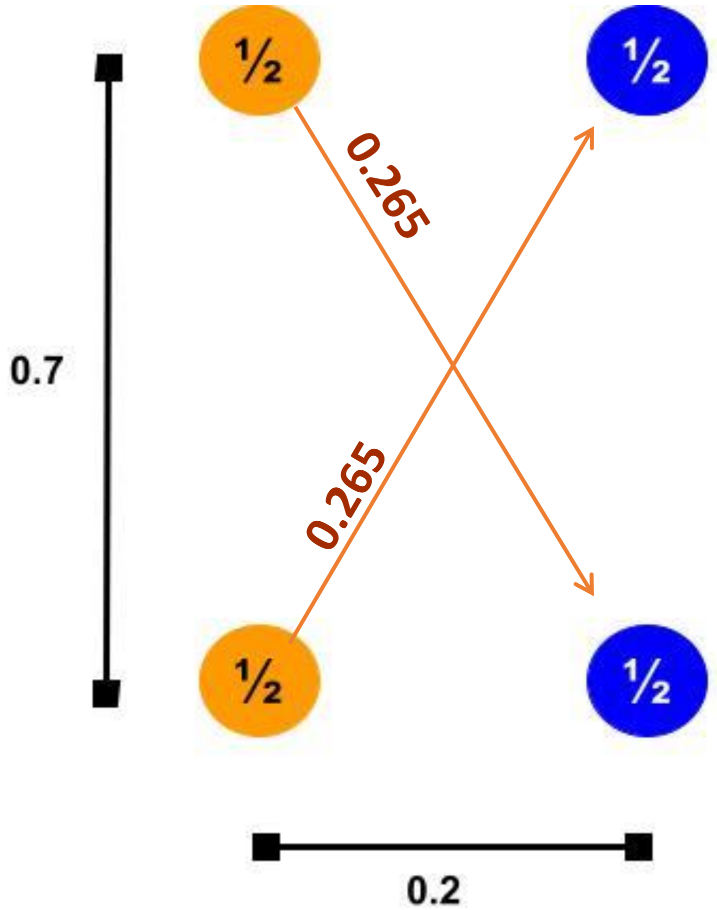
Key:
tv = total variation cost
kl = Kullback-Leiber cost
(se) = Sinkhorn Divergence cost
(p) = primal cost
r__ = reverse cost

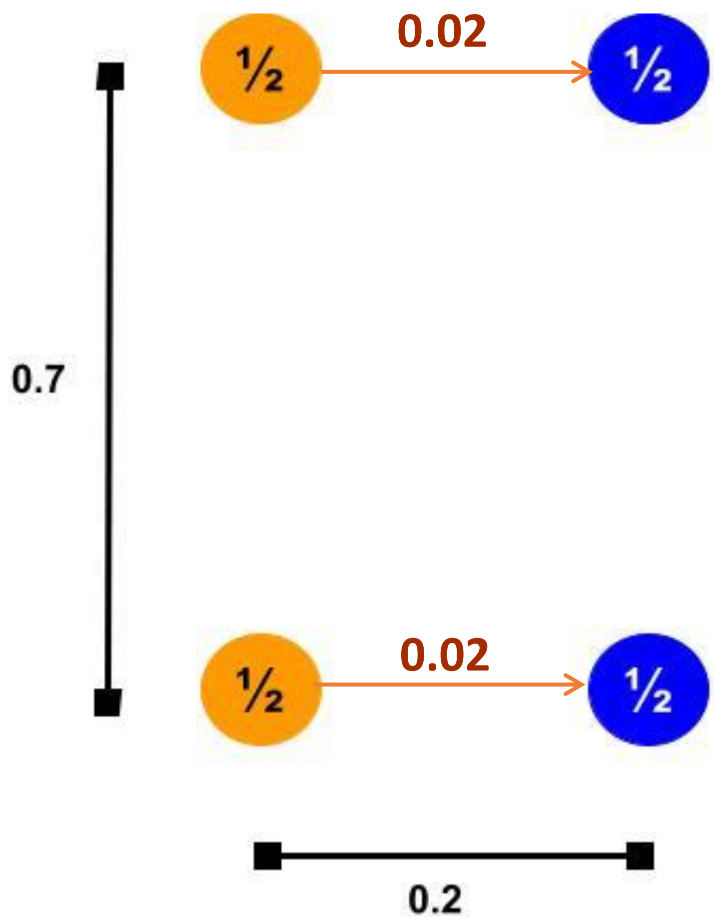


Simple case



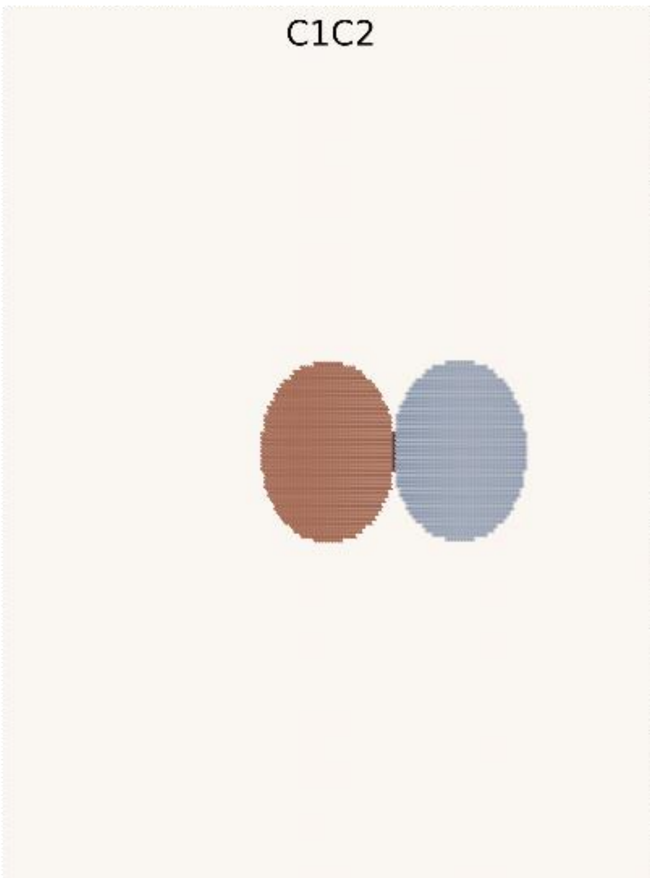
Simple case





Translation

Key:
tv = total variation cost
kl = Kullback-Leiber cost
(se) = Sinkhorn Divergence cost
(p) = primal cost
r__ = reverse cost



rtv (se)=0.02

rtv (p)=0.024

tv (se)=0.02

tv (p)=0.024

rkl (se)=0.0198

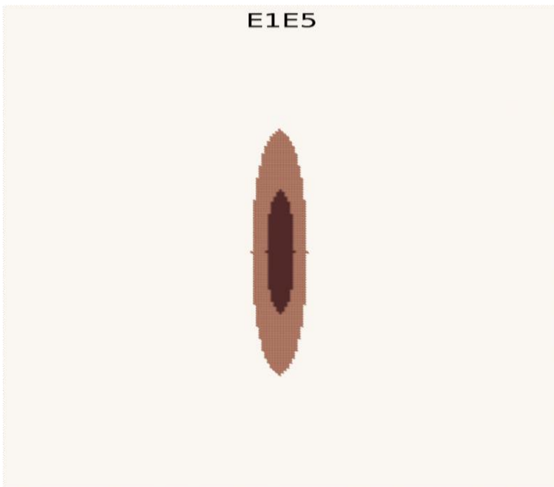
rkl (p)=0.0238

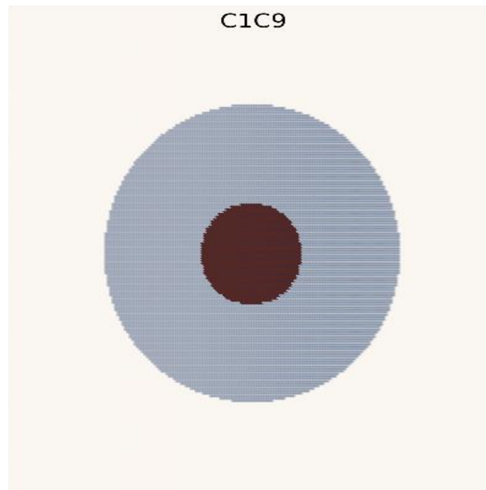
kl (se)=0.0198

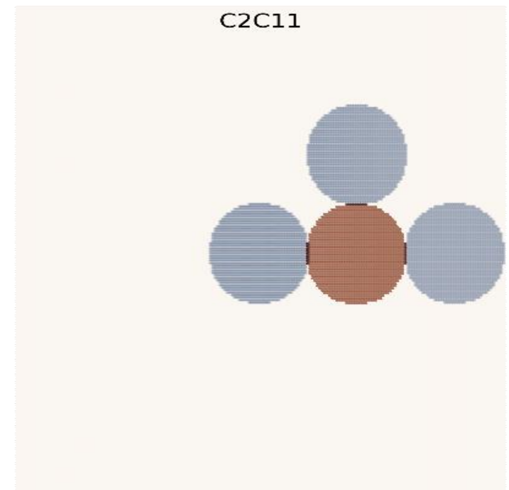
kl (p)=0.0238

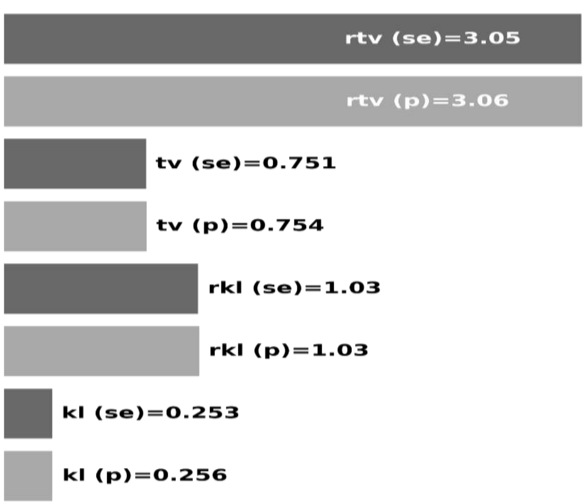
Reverse Costs (non-symmetric?)

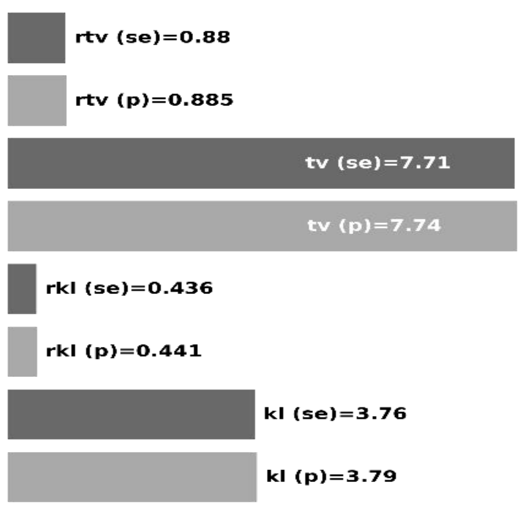
Key:
tv = total variation cost
kl = Kullback-Leiber cost
(se) = Sinkhorn Divergence cost
(p) = primal cost
r__ = reverse cost

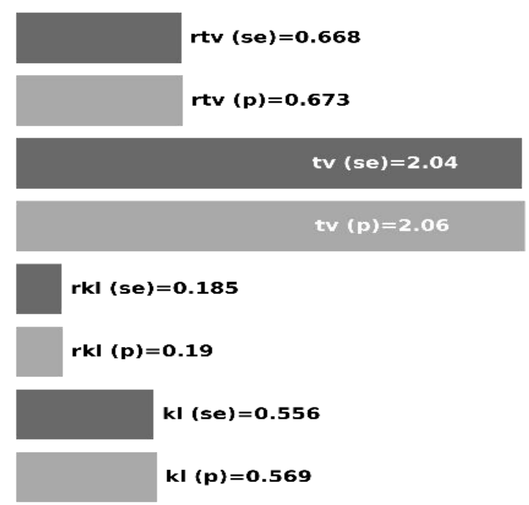




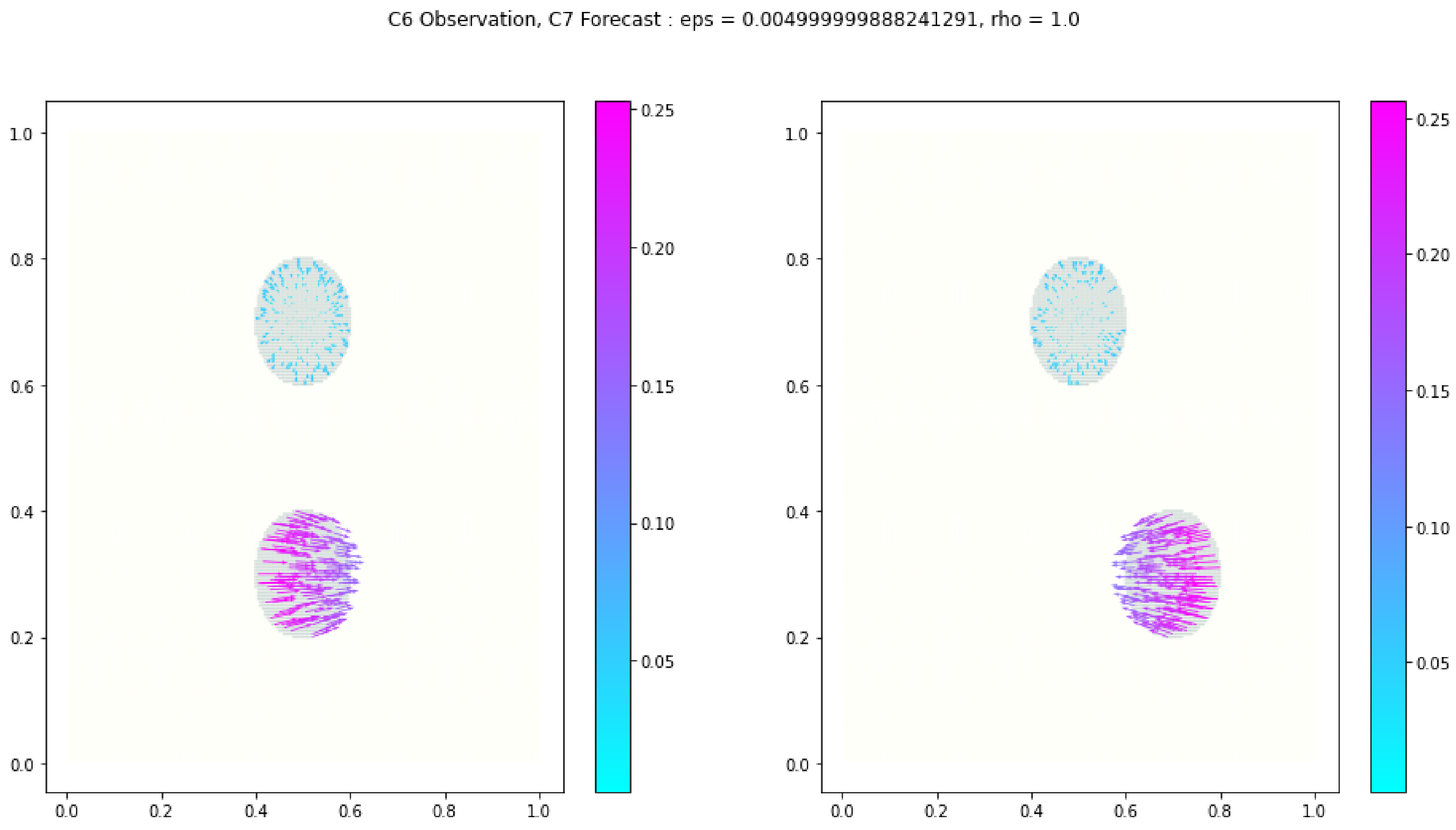




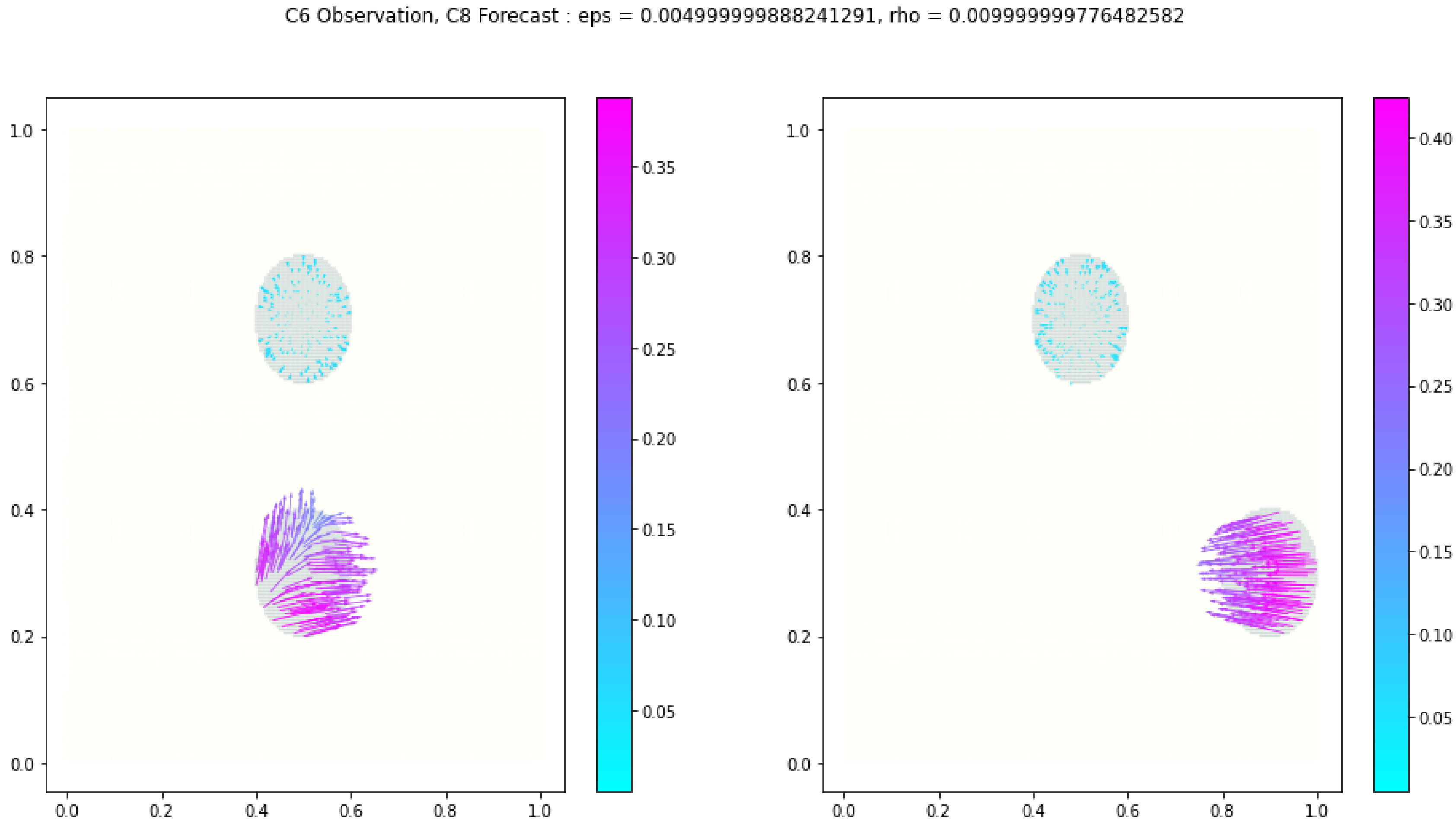




Transport Plan: rho=1

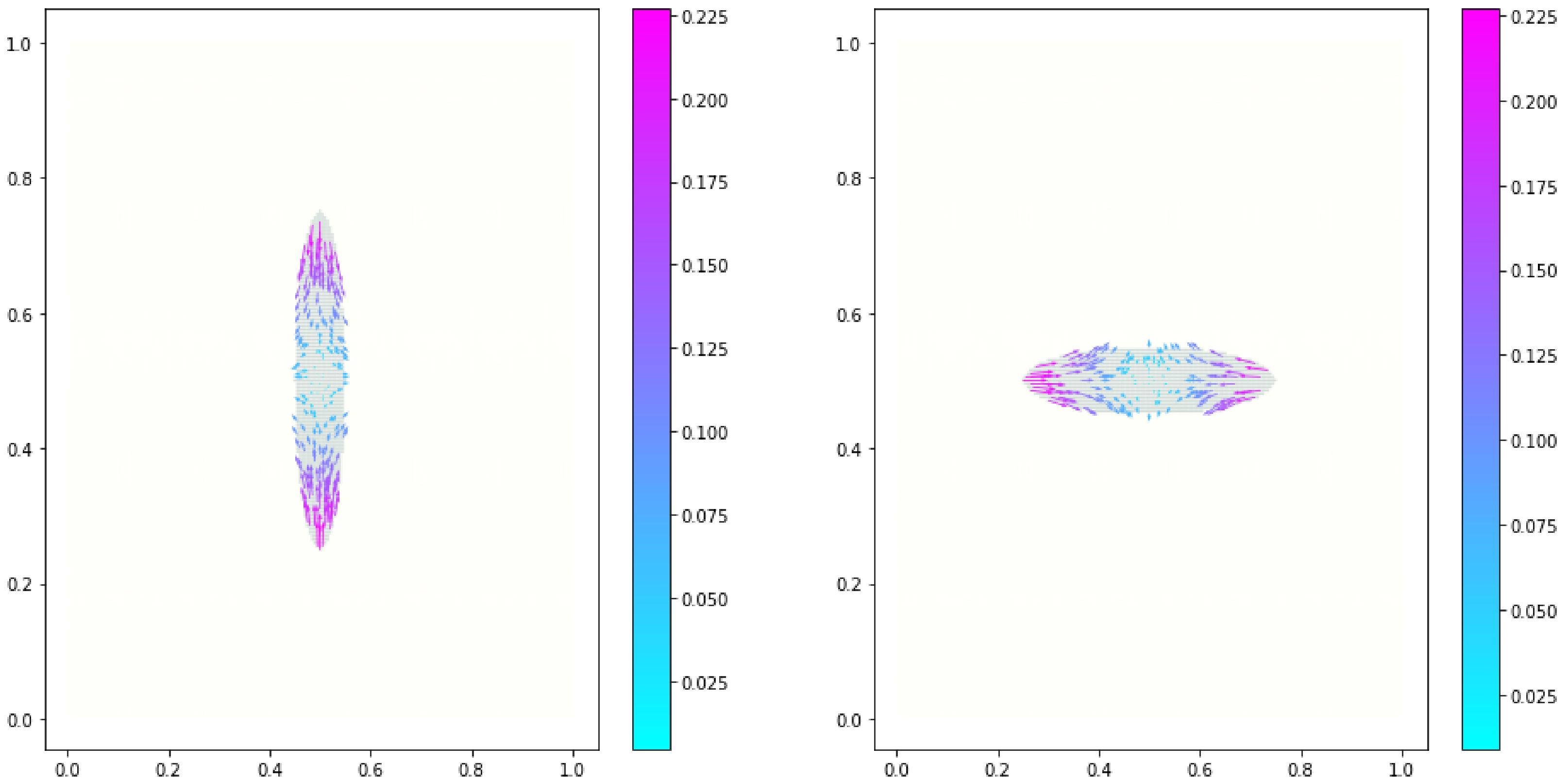


Transport Plan: rho=0.001



Transport Plan: rotation

E1 Observation, E2 Forecast : eps = 0.004999999888241291, rho = 1.0



An OT Metric and Toolbox

Questions to ask of the metric:

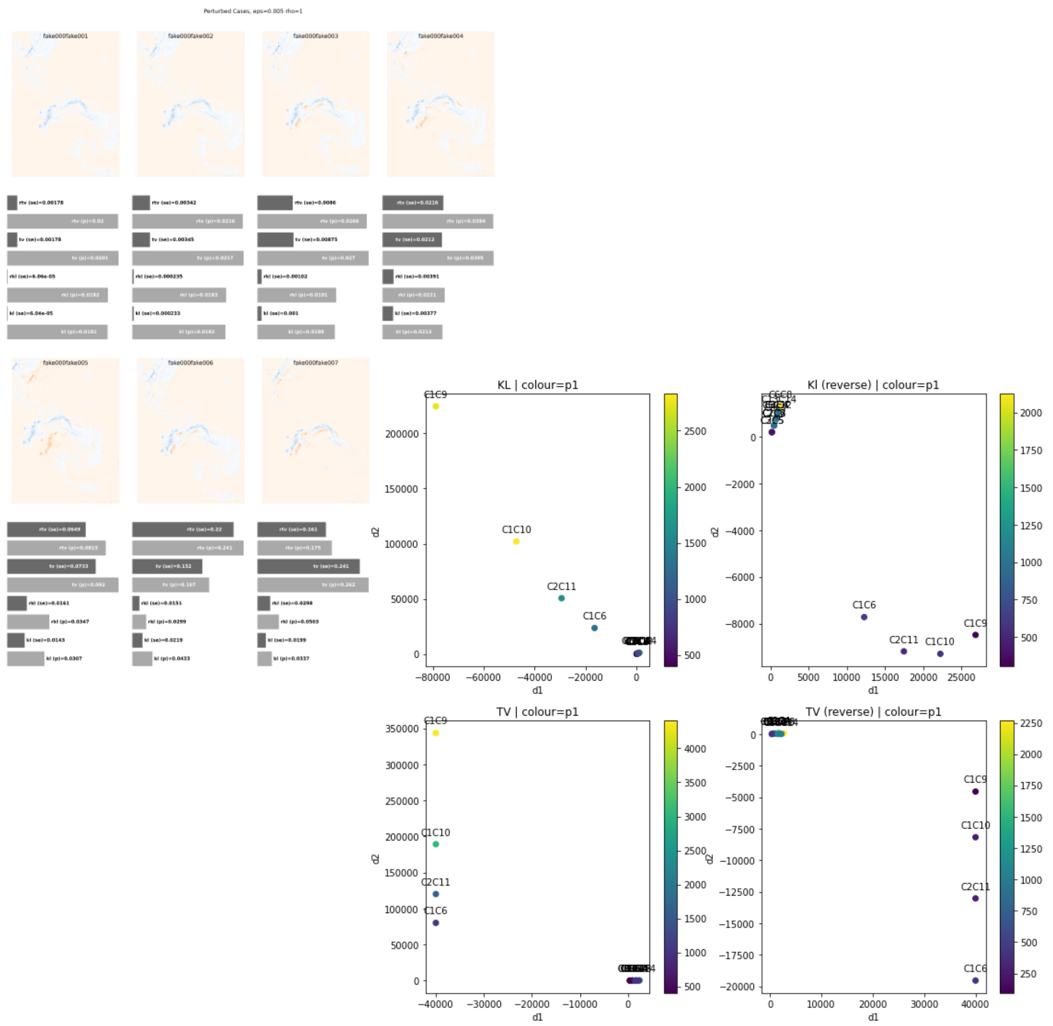
- ➢ Does the method inform **performance at different scales**?
- ➢ Does the method provide information on **location error**?
- ➢ Does the method provide information on **intensity errors and distributions**?
- ➢ Does the method provide information on **structural error**?
- ➢ Can the method give **traditional information**, hits, misses, false alarms etc. ?

OT offers many possibilities, including;

- ✓ Geometric interpretation, with 7 potential terms to consider (3 dual, 4 primal)
- ✓ Provides approximate transport plan, showing where most transport is required.
- ✓ Parsimonious, only two parameters to tune. (Epsilon, Rho)
- ✓ Can work on two different meshed data
- ✓ Works on true intensity maps, rather than binary maps.

Future Work

- Create informative diagrams/plots
- Real data application
- Study the physical interpretation of rho
- Optimise the algorithm for large data sets and ensemble models
- **Work with stakeholders for better understanding of use cases, and information they want.**



Thank you

Feedback, thoughts and comments welcomed – jjf817@ic.ac.uk

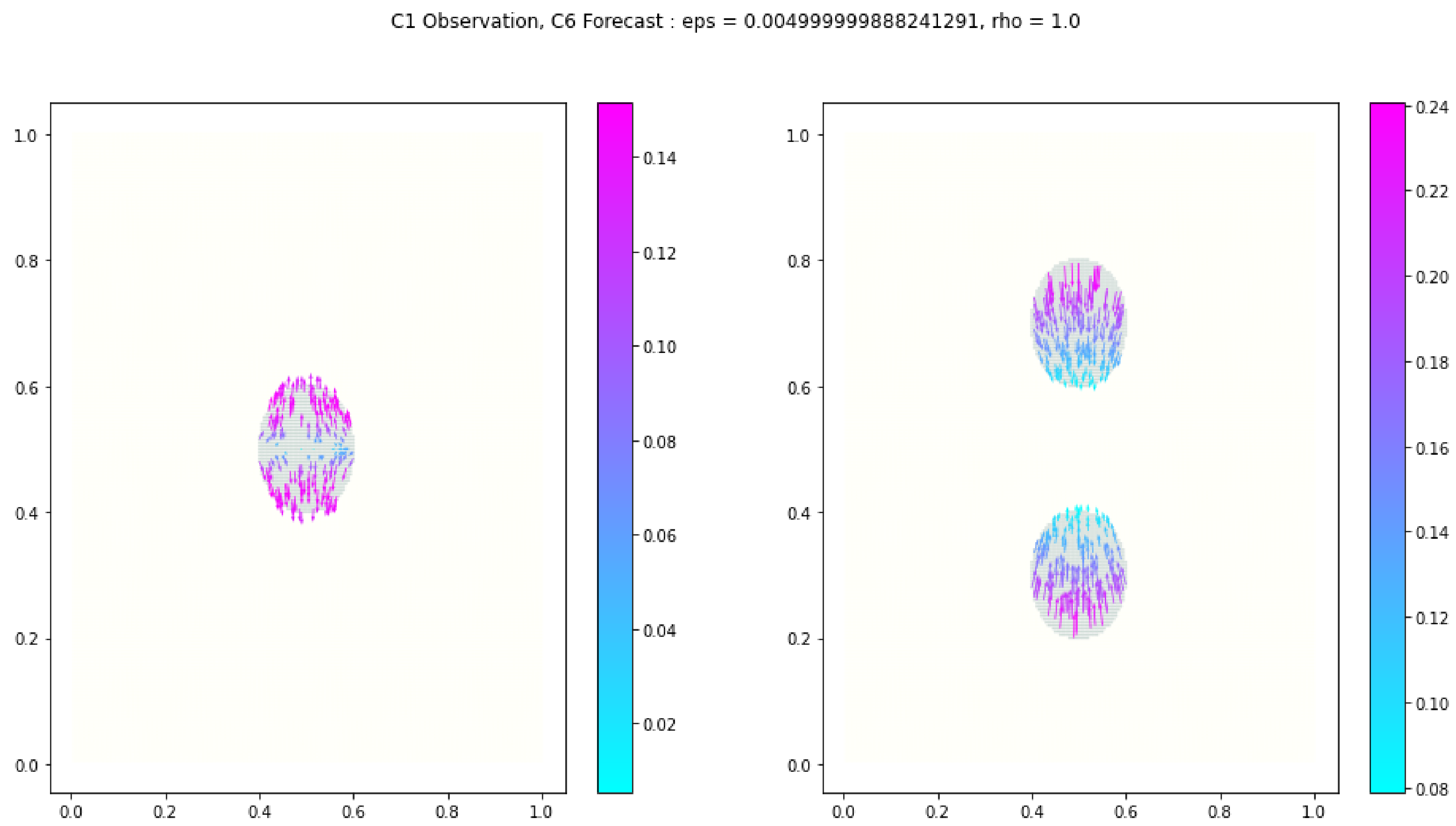
References:

Peyré, G, and Cuturi, M. "Computational optimal transport: With applications to data science." *Foundations and Trends in Machine Learning* 11.5-6 (2019): 355-607.

Séjourné, T., Feydy, J., Vialard, F. X., Trouvé, A., & Peyré, G. "Sinkhorn divergences for unbalanced optimal transport." *arXiv preprint arXiv:1910.12958* (2019).

Gilleland, E., Skok, G., Brown, B.G., Casati, B., Dorninger, M., Mittermaier, M.P., Roberts, N., Wilson, L.J., A Novel Set of Geometric Verification Test Fields with Application to Distance Measures. Monthly Weather Review 148, (2020) 1653–1673.

Geometric Example



Entropic Regularisation

Following work by Cuturi (2013) we add an entropic regularisation term corresponds to a Kullback Leibler penalty.

$$\pi^* \in argmin_{\pi} \langle c, \pi \rangle + \varepsilon KL(\pi|\alpha \otimes \beta) + \mathcal{D}_{\iota}(\pi_0|\alpha) + \mathcal{D}_{\iota}(\pi_1|\beta), \tag{1}$$

$$f^*, g^* \in argmax_{f,g} - \mathcal{D}_{\iota}^*(-f|\alpha) + -\mathcal{D}_{\iota}^*(-g|\beta) - \varepsilon \left\langle e^{\frac{f+g-c}{\varepsilon}} - 1, \alpha \otimes \beta \right\rangle \tag{2}$$

with divergence, $\mathcal{D}_{\iota}(\mu|\nu)$ is zero if $\mu = \nu$ and blows up otherwise and where $\mathcal{D}_{\iota}^*$ denotes convex conjugates. For this balanced setting we have that, $\mathcal{D}_{\iota}^*(\gamma|\lambda) = \gamma, \lambda$. We recover the explicit plan via;

$$\pi(x, y) = \alpha(x)\beta(y)e^{\frac{f(x)+g(y)-c(x,y)}{\varepsilon}}$$

Optimal Transport: Unbalanced

	Divergence	Transform
Balanced	$\mathcal{D}_\iota(\gamma\|\lambda)$	ψ, λ
Kullback-Leiber (KL)	$\rho\log(\frac{d\gamma}{d\lambda}) - 1, \gamma + \rho 1, \lambda$	$\rho e^{\frac{\psi}{\rho}} - 1, \lambda$
Total Variation (TV)	$\rho\|\frac{d\gamma}{d\lambda} - 1\|, \lambda$	$\max(-\rho, \psi), \lambda,\ \psi \leq \rho$
L^2-norm (L2)	$\frac{\rho}{2}\frac{d\gamma}{d\lambda}^2, \lambda$	$\frac{1}{2\rho}\psi^2, \lambda$

Table 1: Table displaying the different cases of divergences we utilise, along with their convex conjugate transform terms. Note here that the input measure, γ, represents the primal measure with reference measure, λ, ψ is a dual function and ρ some constant parameter.

$$\pi^* \in argmin_\pi c, \pi + \varepsilon KL(\pi|\alpha \otimes \beta) + \mathcal{D}_1(\pi_0|\alpha) + \mathcal{D}_2(\pi_1|\beta), \qquad (1)$$

$$f^*, g^* \in argmax_{f,g} - \mathcal{D}_1^*(-f|\alpha) + -\mathcal{D}_2^*(-g|\beta) - \varepsilon e^{\frac{f+g-c}{\varepsilon}} - 1, \alpha \otimes \beta, \qquad (2)$$