Spatial estimation of EV energy demand based on aggregated measurements

Andres Ferragut, Emiliano Espindola

Universidad ORT Uruguay

INFORMS APS 2023 – Nancy, France – June 2023

Introduction

- Electrical vehicle (EV) adoption is currently growing exponentially.
- Less carbon emissions, noise and other efficiency benefits.

Problems:

- We need to build the charging infrastructure to replace gas stations.
- Charging is power and energy intensive for the network, the grid must cope with the enlarged demand.

Introduction

- Electrical vehicle (EV) adoption is currently growing exponentially.
- Less carbon emissions, noise and other efficiency benefits.

Problems:

- We need to build the charging infrastructure to replace gas stations.
- Charging is power and energy intensive for the network, the grid must cope with the enlarged demand.

We need good spatial estimates of energy demand!

Outline of the talk

Problem description

Radial basis functions approach

A Poisson parametric model

Final remarks

The Problem

- We need an spatial estimate of energy demand in order to upgrade the distribution network.
- Currently, we do not have measurements of this demand due to low EV penetration.

Idea:

Use current gas consumption, measured at gas stations, converted to energy.

The Problem

- We need an spatial estimate of energy demand in order to upgrade the distribution network.
- Currently, we do not have measurements of this demand due to low EV penetration.

Idea:

Use current gas consumption, measured at gas stations, converted to energy.

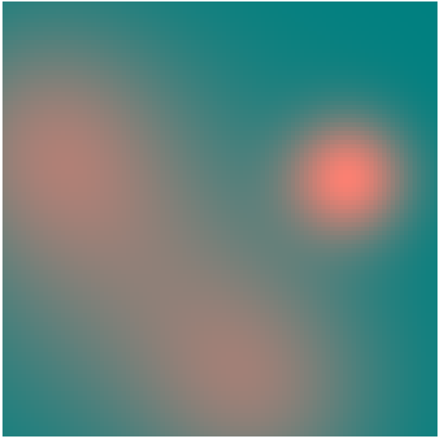
Challenge:

These measurements are concentrated at the gas stations. How to interpolate them?

Energy density

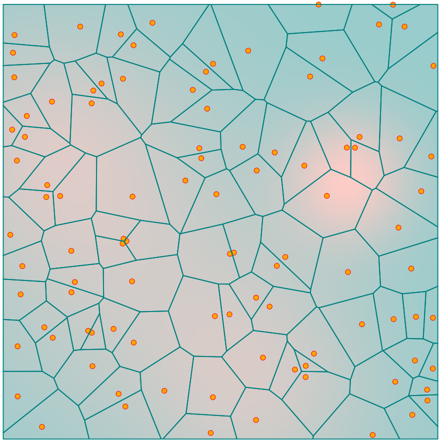
- We have an unkonwn energy density $g(x)$ (in energy/km^2) over a region $\mathcal{X}$.
- Represents amount of energy demand coming from a small ball around x.

Baseline density



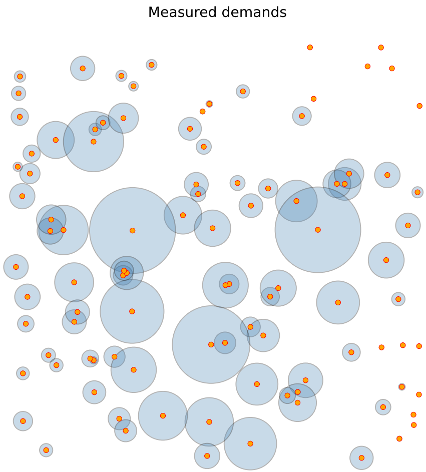
Measurement points

- We cannot sample from this density!
- All we have is some measurement points distributed over $\mathcal{X}$.
- What we can measure is the total demand coming from a cell around our measurement point.



Our dataset

- Each demand is measured at sites s_i.
- We have access to $y_i = \int_{V_i} g(x)dx$, where V_i is the Voronoi cell of site i.
- The size of the circle represents measured demand at the sites.



Mathematical formulation

- In a region of space $\mathcal{X} \subset \mathbb{R}^d$, we are given:
  - A list of fixed sites $\{s_1, \ldots, s_m\}$, $s_i \in \mathcal{X}$.
  - A list of measurements $\{y_1, \ldots, y_m\}$, $y_i \geqslant 0$.
- Goal: construct an estimate $\hat{g}(x;\theta)$ of the spatial density such that:

$$\int_{V_i} g(x;\theta)\, dx \approx y_i \quad \forall i$$

where V_i is a cell associated with site s_i (e.g. the Voronoi cell).

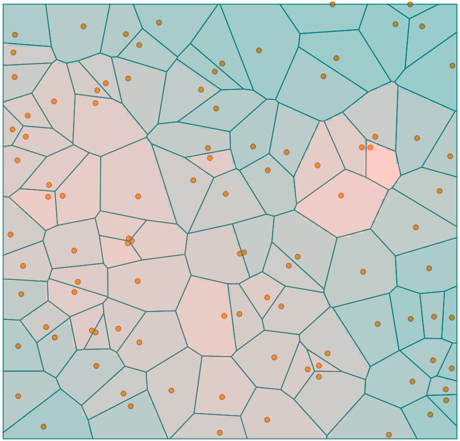
- First approach: histogram counts.

- Estimate

$$g_H(x) = \sum_i \frac{y_i}{\mathrm{m}(V_i)} \mathbf{1}_{V_i}(x)$$

- Non-smooth. Low interpolation properties. Not suitable for low-dimensional representation.

Histogram counts



Radial basis functions

To obtain a lower dimensional representation we use radial basis functions to estimate $g(x)$. Namely, our estimator has the form:

$$g_{RBF}(x;\theta) = \sum_{j=1}^{n} w_j e^{-\frac{||x-\mu_j||^2}{2\sigma_j^2}}$$

where $\theta = (\{w_j\}, \{\mu_j\}, \{\sigma_j^2\})$.

- $w_j \in \mathbb{R}^+$ are the weights,
- $\mu_j \in \mathbb{R}^d$ are the nodes and
- $\sigma_j^2 \in \mathbb{R}^+$ the bandwidths.

Least squares approach

- Since we have access to the cell measurements, it makes sense to consider the loss function:

$$L(\theta) = \frac{1}{2}\sum_{i=1}^{m}\left(\int_{V_i} g_{RBF}(x;\theta)dx - y_i\right)^2$$

- Therefore, the least squares estimator becomes:

$$\hat{\theta}_{LS} = \arg\min_{\theta} L(\theta)$$

- We now show an algorithm to compute this estimator.

Computing the weights

Consider first given the nodes μ_j and the bandwidths σ_j^2, we have:

$$\int_{V_i} g_{RBF}(x,\theta)\,dx = \sum_{j=1}^{n} w_j \int_{V_i} e^{-\frac{||x-\mu_j||^2}{2\sigma_j^2}}\,dx =: \sum_{j=1}^{n} a_{ij} w_j.$$

The loss becomes:

$$L(\theta) = \frac{1}{2}\sum_{i=1}^{m}\left(\sum_{j=1}^{n} a_{ij} w_j - y_i\right)^2 = \frac{1}{2}||Aw - y||^2$$

And thus we have the linear least squares problem:

$$\min \frac{1}{2}||Aw - y||^2, \quad s.t.\ w \geqslant 0.$$

It can be readily solved, typically the constraint is not active.

Estimating nodes and bandwidths

- To estimate μ_j and $\{\sigma_j^2\}$ we may use gradient descent. Note that:

$$\frac{\partial L}{\partial \theta_k} = \sum_{i=1}^{m} \left(\int_{V_i} g_{RBF}(x;\theta)\,dx - y_i \right) \left(\int_{V_i} \frac{\partial}{\partial \theta_k} g_{RBF}(x;\theta)\,dx \right)$$

- Moreover, due to the structure of the RBF functions:

$$\begin{aligned}
\frac{\partial}{\partial \mu_j} g_{RBF}(x;\theta) &= \left[\frac{x-\mu_j}{\sigma_j^2} \right] w_j e^{-\frac{||x-\mu_j||^2}{2\sigma_j^2}} \\
\frac{\partial}{\partial \sigma_j^2} g_{RBF}(x;\theta) &= \left[\frac{||x-\mu_j||^2}{2(\sigma_j^2)^2} \right] w_j e^{-\frac{||x-\mu_j||^2}{2\sigma_j^2}}
\end{aligned}$$

Computing the gradient

So in order to compute the gradient, we need to estimate the following moments of our current density estiamte:

$$\int_{V_i} g_{RBF}(x,\theta)\,dx, \quad \int_{V_i} \left[\frac{x-\mu_j}{\sigma_j^2}\right] g_{RBF}(x,\theta)\,dx, \quad \int_{V_i} \left[\frac{||x-\mu_j||^2}{2(\sigma_j^2)^2}\right] g_{RBF}(x,\theta)\,dx.$$

for each cell i.

Sample N uniformly distributed points in the region $\mathcal{X}$ and estimate:

$$\int_{V_i} g_{RBF}(x;\theta)dx \approx \frac{\mathrm{m}(\mathcal{X})}{N} \sum_{k=1}^{N} g_{RBF}(u_k;\theta)\mathbf{1}_{V_i}(u_k)$$

Estimating the moment integrals

Monte Carlo approach

Sample N uniformly distributed points in the region $\mathcal{X}$ and estimate:

$$\int_{V_i} g_{RBF}(x;\theta)dx \approx \frac{\mathrm{m}(\mathcal{X})}{N}\sum_{k=1}^{N} g_{RBF}(u_k;\theta)\mathbf{1}_{V_i}(u_k)$$

Two possible variants:

- Use a large N and fix the estimation points $\rightarrow$ slightly more bias, less variance, faster to compute.
- Resample a relatively small N on each step $\rightarrow$ less bias, high variance, amounts to Stochastic Gradient Descent.

Algorithm

Stochastic gradient descent version

Given a suitable initial condition $\theta^{(0)} = (\{w_j^{(0)}\}, \{\mu_j^{(0)}\}, \{\sigma_j^{2\,(0)}\})$, at each step k:

1. Sample N uniformly distributed random points in $\mathcal{X}$.
2. Estimate the moment integrals and compute the gradient $\nabla L(\theta^{(k)})$.
3. Perform a gradient step:

$$\mu_j \leftarrow \mu_j - \alpha_k \nabla L(\theta^{(k)})_{\mu_j}, \quad \sigma_j^2 \leftarrow \sigma_j^2 - \alpha_k \nabla L(\theta^{(k)})_{\sigma_j^2}.$$

   with step size $\alpha_k \sim O(1/k)$.
4. For the new nodes and bandwidths, recompute w_j using linear least squares.
5. Update $\theta^{(k+1)}$ and iterate until convergence.

Choosing the initial condition

We need a good first estimate $\theta^{(0)}$. We propose the following method:

Bootstrapping:
Fix the number of kernels n as an hyperparameter and do:

1. Given the sites $\{s_1, \ldots, s_m\}$ and the measurements $\{y_1, \ldots, y_m\}$, run weighted $k-$means with n clusters to optimize:

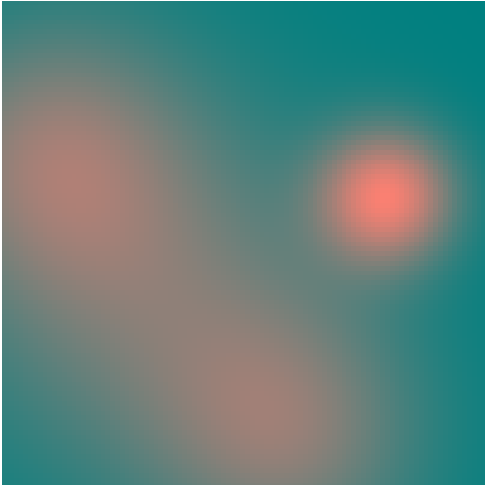
$$\min_{\mu_j} \sum_{j=1}^{n} \sum_{i \text{ closest to } \mu_j} y_i ||s_i - \mu_j||^2$$

2. Estimate the bandwidths σ_j^2 as the mean square distance of the allocated sites to node j.
3. Compute a first estimate of w_j by solving the linear least squares problem with the above initial estimates.

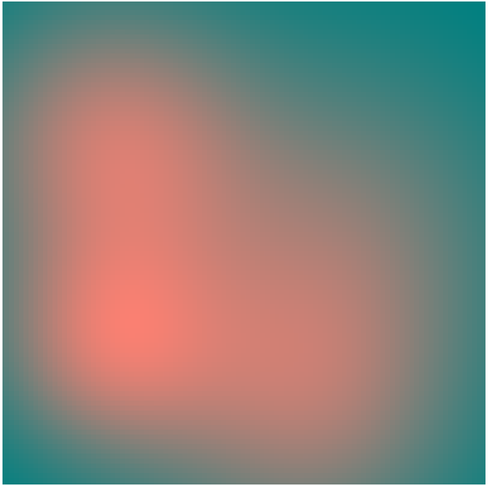
Example: reconstructing the original density

Initial condition

Baseline density

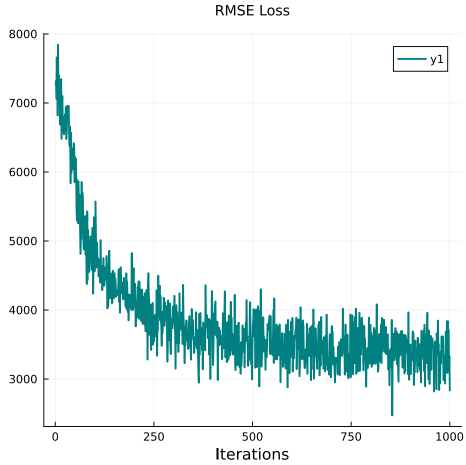


Initial density estimation



Example: reconstructing the original density

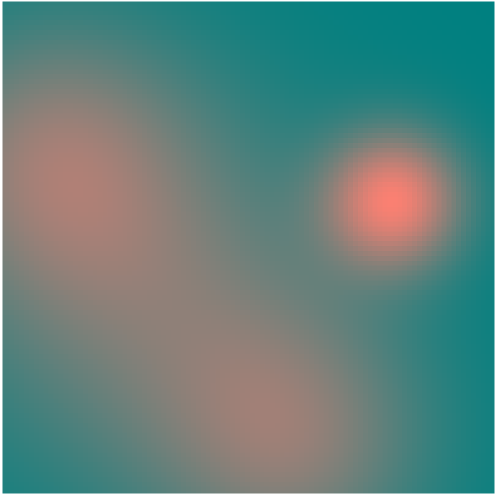
Root mean square loss evolution:



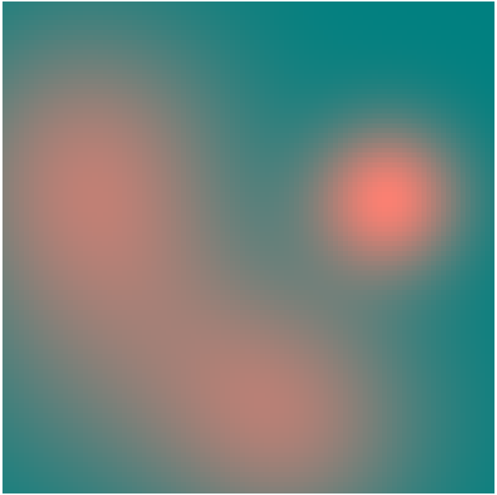
Example: reconstructing the original density

Final estimate:

Original density



Reconstructed density



Hidden Poisson process parametric model

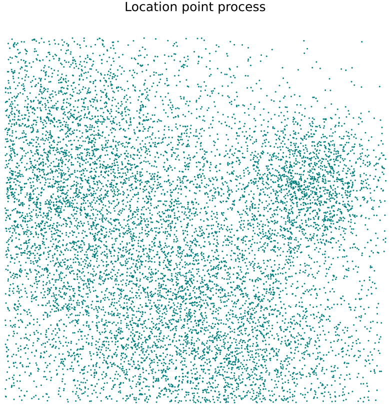
Assume now that demands come from a marked Poisson process

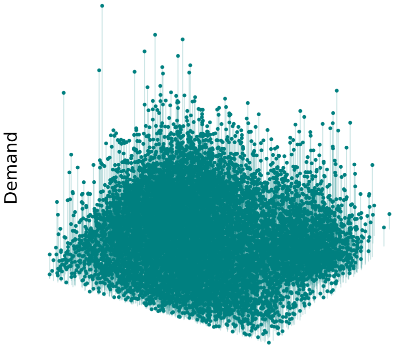
$$\Phi = \sum \delta_{(x_k, v_k)}$$

with:

- Spatial intensity $\Lambda(dx)$, $x \in \mathcal{X}$, modeled through an RBF density $\lambda(x;\theta)dx$.
- Individual marks (customer demands) exponentially distributed with rate ν.

A realization of the process Φ:





Observation model

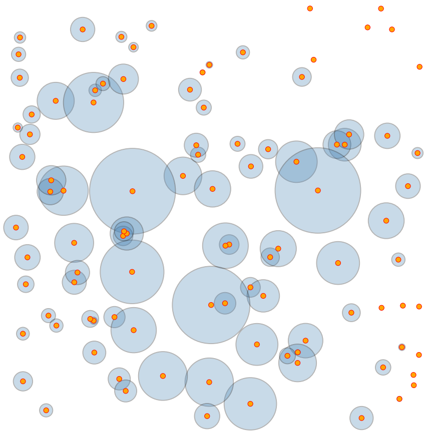
- We observe the total cell demands:

$$Y_i = \int_{V_i} v\,\Phi(dx, dv) = \sum_k v_k \mathbf{1}_{\{x_k \in V_i\}}$$

- For each cell:

$$\begin{aligned} N_i \sim & \quad \text{Poisson}(\lambda_i(\theta)) \\ Y_i = & \quad \textstyle\sum_{k=1}^{N_i} V_k \end{aligned}$$

with $V_k \sim \exp(\nu)$ and cells are independent.



Problem: the number of points acts as a *hidden variable.*

Maximum likelihood approach

- Ideally one would like to maximize $p(y;\theta)$. Difficult to compute.
- Expectation-Maximization approaches fail (a posteriori distribution also difficult).
- Consider maximizing the combined likelihood:

$$p(n,y) = \prod_{i=1}^{m} e^{-\lambda_i(\theta)} \frac{\lambda_i(\theta)^{n_i}}{n_i!} p(y_i \mid n_i), \quad \text{with } \lambda_i(\theta) = \int_{V_i} \lambda(x;\theta)\,dx$$

- Now, since given n_i, the demands are independent exponentials we have:

$$p(y_i \mid n_i) = \frac{1}{(n_i-1)!} \nu^{n_i} y_i^{n_i-1} e^{-\nu y_i}.$$

Maximizing the counts likelihood

Given an estimate of $\Lambda(dx)$ and ν, we can maximize each term over n_i, thus decoding the hidden variable:

$$\max_{n_i} e^{-\lambda_i(\theta)} \frac{\lambda_i(\theta)^{n_i}}{n_i!} \frac{1}{(n_i - 1)!} \nu^{n_i} y_i^{n_i - 1} e^{-\nu y_i}.$$

The maximum is attained for:

$$n_i^*(n_i^* + 1) = \lambda_i(\theta) \nu y_i$$

That is:

$$n_i^* \approx \sqrt{\lambda_i(\theta) \nu y_i}.$$

Estimating the RBF parameters

With the hidden variables estimated, the joint likelihood as a function of θ is:

$$\ell(\theta; n, y) = \sum_{i=1}^{m} -\lambda_i(\theta) + n_i \log(\lambda_i(\theta)) + n_i \log(\nu) - \nu y_i + (n_i - 1)\log(y_i) - \log(n_i!(n_i - 1)!).$$

The new estimate for ν follows immediately by differentiation:

$$\hat{\nu} = \frac{\sum_i n_i}{\sum_i y_i}.$$

As for the RBF parameters, we perform a gradient approach as before:

$$\frac{\partial \ell}{\partial \theta} = \sum_{i=1}^{m} \frac{\partial \ell}{\partial \lambda_i} \frac{\partial \lambda_i}{\partial \theta} = \sum_{i=1}^{m} \left[\frac{n_i}{\lambda_i(\theta)} - 1 \right] \frac{\partial \lambda_i(\theta)}{\partial \theta}.$$

The derivatives of $\lambda_i(\theta)$ are similar to the preceding ones.

Maximum likelihood algorithm

Given a suitable initial condition $\theta^{(0)} = (\{w_j^{(0)}\}, \{\mu_j^{(0)}\}, \{\sigma_j^{2\,(0)}\}, \nu^{(0)})$, at each step k, iterate until convergence:

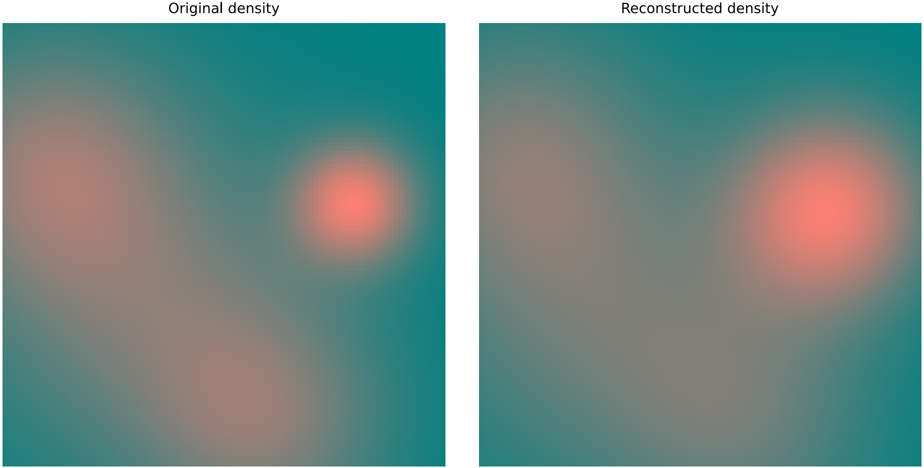
1. Sample N uniformly distributed random points in $\mathcal{X}$.
2. Estimate $\lambda_i(\theta)$ ant the moment integrals to compute its gradient.
3. Decode the hidden variables $n_i = \sqrt{\lambda_i(\theta^{(k)})\nu^{(k)} y_i}$
4. Perform a gradient step on nodes and bandwidths:

$$w_j^{(k+1)} = w_j^{(k+1)} + \alpha_k \nabla \ell(\theta^{(k)})_{w_j},$$
$$\mu_j^{(k+1)} = \mu_j^{(k)} + \alpha_k \nabla \ell(\theta^{(k)})_{\mu_j}, \quad (\sigma_j^2)^{(k+1)} \leftarrow (\sigma_j^2)^{(k)} + \alpha_k \nabla \ell(\theta^{(k)})_{\sigma_j^2}.$$

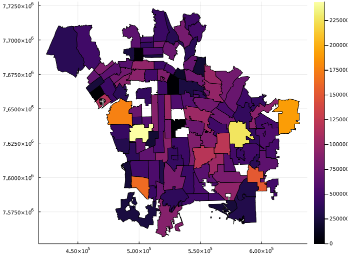
5. Update $\hat{\nu}^{(k+1)} = \frac{\sum_i n_i}{\sum_i y_i}$.

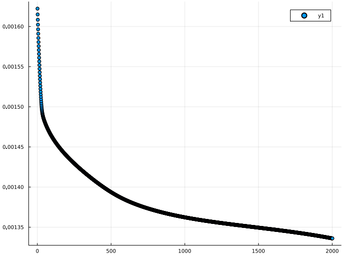
Example

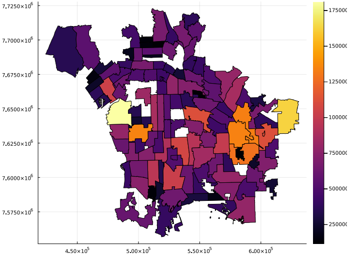
Maximum likelihhod reconstruction



Annual consumption of gas in LAX aggregated by ZIP code:







Take home messages...

- EVs are popping up, we have to prepare the infrastructure.
- We can estimate spatial energy demand based on current gas consumption measurements.
- We analyzed two different approaches with different properties.
- We would like to expand on the mathematical analysis, in particular the connection with stochastic gradient descent and more general transport measures and problems.

Thank you!

Andres Ferragut
ferragut@ort.edu.uy
http://aferragu.github.io