Unclouding Pollution Maps

Ioannis Konstantinidis

February 21, 2014

The http://HoustonCleanAirNetwork.com team

A philosopher, a mathematician, and an atmospheric scientist walk into a bar ...

The http://HoustonCleanAirNetwork.com team

A philosopher, a mathematician, and an atmospheric scientist walk into a bar ...
where they meet a computer scientist and an environmental justice advocate to talk about ozone pollution.

The http://HoustonCleanAirNetwork.com team

A philosopher, a mathematician, and an atmospheric scientist walk into a bar ...
where they meet a computer scientist and an environmental justice advocate to talk about ozone pollution.

Jessica Crowley, Barry Lefer, Mark Huang, Ashik Khatri, Peggy Lindner, John Naruk, Ioannis Pavlidis, Dan Price, Matt Tejada, Ilyas Uyanik

The http://HoustonCleanAirNetwork.com team

A philosopher, a mathematician, and an atmospheric scientist walk into a bar ...
where they meet a computer scientist and an environmental justice advocate to talk about ozone pollution.

Jessica Crowley, Barry Lefer, Mark Huang, Ashik Khatri, Peggy Lindner, John Naruk, Ioannis Pavlidis, Dan Price, Matt Tejada, Ilyas Uyanik

Special thanks to our major sponsors:
The Houston Endowment, the American Lung Association, and the University of Houston.

The issue with ground-level ozone (O_3)

- Ground-level ozone is not emitted directly into the air, but forms through a reaction of nitrogen oxides (NO_x) and volatile organic compounds (VOC) in the presence of sunlight.

The issue with ground-level ozone (O_3)

- Ground-level ozone is not emitted directly into the air, but forms through a reaction of nitrogen oxides (NO_x) and volatile organic compounds (VOC) in the presence of sunlight.
- Major man-made sources of NO_x and VOC:
  - emissions from industrial facilities and electric utilities
  - motor vehicle exhaust
  - gasoline vapors
  - chemical solvents

The issue with ground-level ozone (O_3)

- Ground-level ozone is not emitted directly into the air, but forms through a reaction of nitrogen oxides (NO_x) and volatile organic compounds (VOC) in the presence of sunlight.
- Major man-made sources of NO_x and VOC:
  - emissions from industrial facilities and electric utilities
  - motor vehicle exhaust
  - gasoline vapors
  - chemical solvents
- O_3 is a highly reactive gas, and the main component of smog. When inhaled, it damages the lung membrane, decreases lung capacity, and causes inflammation.

The issue with ground-level ozone (O_3)

- Ground-level ozone is not emitted directly into the air, but forms through a reaction of nitrogen oxides (NO_x) and volatile organic compounds (VOC) in the presence of sunlight.
- Major man-made sources of NO_x and VOC:
  - emissions from industrial facilities and electric utilities
  - motor vehicle exhaust
  - gasoline vapors
  - chemical solvents
- O_3 is a highly reactive gas, and the main component of smog. When inhaled, it damages the lung membrane, decreases lung capacity, and causes inflammation.
- It is regulated by the EPA, which sets standards for acceptable exposure.

The standards

The EPA has developed an Air Quality Index (AQI) for ozone.

- To compute the AQI, the measured concentrations of ozone (in parts per billion, or ppb), are averaged over an eight hour period.

The standards

The EPA has developed an Air Quality Index (AQI) for ozone.

- To compute the AQI, the measured concentrations of ozone (in parts per billion, or ppb), are averaged over an eight hour period.
- Each 8-hr period is then classified as good, moderate, unhealthy for sensitive groups, unhealthy, very unhealthy, or hazardous, according to a series of threshold AQI values.

The standards

The EPA has developed an Air Quality Index (AQI) for ozone.

- To compute the AQI, the measured concentrations of ozone (in parts per billion, or ppb), are averaged over an eight hour period.
- Each 8-hr period is then classified as good, moderate, unhealthy for sensitive groups, unhealthy, very unhealthy, or hazardous, according to a series of threshold AQI values.
- Following the current rule, the threshold for moderate to unhealthy for sensitive groups is 75ppb.

The standards

The EPA has developed an Air Quality Index (AQI) for ozone.

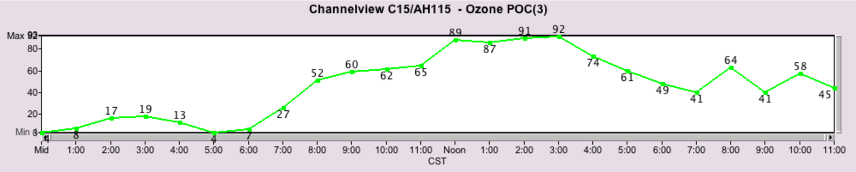
- To compute the AQI, the measured concentrations of ozone (in parts per billion, or ppb), are averaged over an eight hour period.
- Each 8-hr period is then classified as good, moderate, unhealthy for sensitive groups, unhealthy, very unhealthy, or hazardous, according to a series of threshold AQI values.
- Following the current rule, the threshold for moderate to unhealthy for sensitive groups is 75ppb.
- Attaining compliance to the EPA standard requires that this threshold is exceeded no more than 4 days a year.

The standards

The EPA has developed an Air Quality Index (AQI) for ozone.

- To compute the AQI, the measured concentrations of ozone (in parts per billion, or ppb), are averaged over an eight hour period.
- Each 8-hr period is then classified as good, moderate, unhealthy for sensitive groups, unhealthy, very unhealthy, or hazardous, according to a series of threshold AQI values.
- Following the current rule, the threshold for moderate to unhealthy for sensitive groups is 75ppb.
- Attaining compliance to the EPA standard requires that this threshold is exceeded no more than 4 days a year.
- There is a separate standard based on 1-hr averages that applies to areas which fail to comply with the 8-hr standard.

Daily summary plot from an ambient air monitoring station in the Houston area for June 26, 2012;

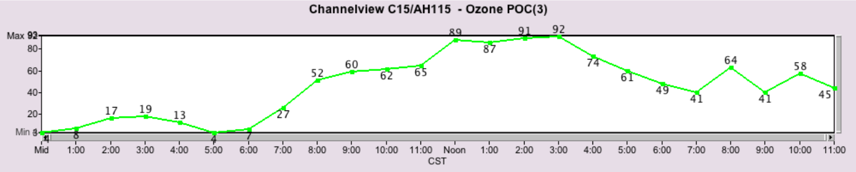


1-hr averages of ground-level ozone concentrations in parts per billion (ppb).

- The 8-hr averages exceeded 75 ppb for three time periods (those starting at 9am, 10am, and 11am).

Houston, we have a problem

Daily summary plot from an ambient air monitoring station in the Houston area for June 26, 2012;

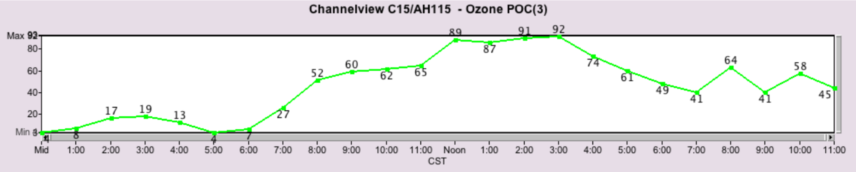


1-hr averages of ground-level ozone concentrations in parts per billion (ppb).

- The 8-hr averages exceeded 75 ppb for three time periods (those starting at 9am, 10am, and 11am).
- This is *not* atypical. In fact, the Houston area is not projected to meet the standard for years to come.

Houston, we have a problem

Daily summary plot from an ambient air monitoring station in the Houston area for June 26, 2012;

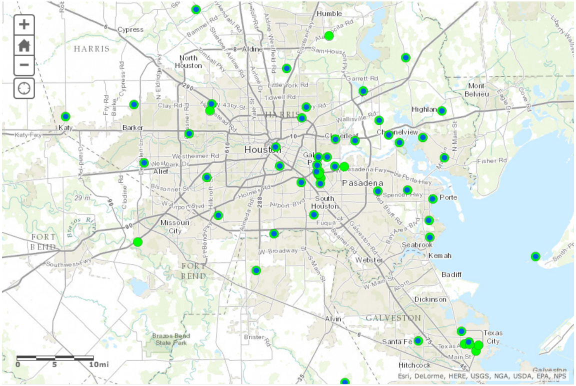


1-hr averages of ground-level ozone concentrations in parts per billion (ppb).

- The 8-hr averages exceeded 75 ppb for three time periods (those starting at 9am, 10am, and 11am).
- This is *not* atypical. In fact, the Houston area is not projected to meet the standard for years to come.
- Until the standard is met, how can Houstonians stay informed about *current* ozone conditions in their daily lives?

Houston, we have a problem ... and it is not lack of data

On the monitoring side, the Texas Commission on Environmental Quality (TCEQ) maintains a network of 45 stations in the greater Houston area, collecting measurements every five minutes.



The problem is what we do with the data: clouding it

The TCEQ *retroactively* makes the data they collect available on the internet, reporting only the 1-hr average for the previous hour.

www.tceq.state.tx.us/cgi-bin/compliance/monops/daily_summary.pl

This web page provides the most current hourly averaged data available. Our convention for time-tagging data is the beginning of each hour. For example, values shown for the noon hour are based on measurements taken from noon to 1:00 p.m. The noon average will not be calculated until after 1:00 p.m. The noon average will then be available on our external server from 1:15 p.m. to 1:30 p.m. This results in an apparent one-hour time lag in the data. We also present our data in Local Standard Time for each measuring site. For most of Texas this is Central Standard Time. During Daylight Savings, this introduces another apparent one-hour time lag in the data.

Use the controls below to select a different date or time format and to control cell highlighting based on measured nitrogen dioxide or PM-2.5 levels. Click on the Generate Report button once you have made your selections. Click on the Plot Data button once the tabular report has been generated to open a separate window containing data plots.

CAMS 15 Channelview C15/AH115 **Select a different site**

Month: Day: Year: Time Format:

June 26 2012 12 Hour (AM/PM) Generate Report Plot Data

Nitrogen Dioxide Highlights: ☑Moderate ☑Unhealthy for Sensitives ☑Unhealthy ☑Very Unhealthy ☑Hazardous

☑Green underline for validated data

The table below contains hourly averages for all the pollutants and meteorological conditions measured at Channelview C15/AH115 for **Tuesday, June 26, 2012**. All times shown are in CST.

Parameter Measured	Morning												Afternoon											
	Mid	1:00	2:00	3:00	4:00	5:00	6:00	7:00	8:00	9:00	10:00	11:00	Noon	1:00	2:00	3:00	4:00	5:00	6:00	7:00	8:00	9:00	10:0
Nitric Oxide	1.9	1.2	1.3	1.3	1.0	**20.7**	11.1	3.7	0.4	0.6	0.6	0.4	0.8	0.8	1.0	1.2	1.6	1.5	2.0	1.8	1.0	0.8	0.4
Nitrogen Dioxide	20.4	19.9	13.1	9.1	11.9	25.8	25.5	19.5	11.0	6.0	***5.7***	5.9	11.0	10.6	13.3	14.3	22.5	26.8	**33.5**	32.6	12.9	19.1	13.7
Oxides of Nitrogen	22.8	21.6	14.7	10.5	13.1	**47.5**	37.4	23.6	11.6	6.7	***6.4***	6.4	12.0	11.7	14.5	15.9	24.6	28.8	36.3	35.1	14.2	20.3	14.4
Ozone	***4***	8	17	19	13	4	7	27	52	60	62	65	89	87	91	**92**	74	61	49	41	64	41	58
Wind Speed	1.8	3.1	2.7	***1.7***	2.2	1.8	2.4	2.6	3.3	3.5	4.0	5.9	7.1	8.5	**9.2**	8.3	7.9	7.0	5.2	3.0	3.2	2.2	2.3

The problem is what we do with the data: clouding it

The TCEQ *retroactively* makes the data they collect available on the internet, reporting only the 1-hr average for the previous hour.

www.tceq.state.tx.us/cgi-bin/compliance/monops/daily_summary.pl

This web page provides the most current hourly averaged data available. Our convention for time-tagging data is the beginning of each hour. For example, values shown for the noon hour are based on measurements taken from noon to 1:00 p.m. The noon average will not be calculated until after 1:00 p.m. The noon average will then be available on our external server from 1:15 p.m. to 1:30 p.m. This results in an apparent one-hour time lag in the data. We also present our data in Local Standard Time for each measuring site. For most of Texas this is Central Standard Time. During Daylight Savings, this introduces another apparent one-hour time lag in the data.

Use the controls below to select a different date or time format and to control cell highlighting based on measured nitrogen dioxide or PM-2.5 levels. Click on the Generate Report button once you have made your selections. Click on the Plot Data button once the tabular report has been generated to open a separate window containing data plots.

CAMS 15 Channelview C15/AH115 **Select a different site**

Month: Day: Year: Time Format:

June 26 2012 12 Hour (AM/PM) Generate Report Plot Data

Nitrogen Dioxide Highlights: ☑Moderate ☑Unhealthy for Sensitives ☑Unhealthy ☑Very Unhealthy ☑Hazardous

☑Green underline for validated data

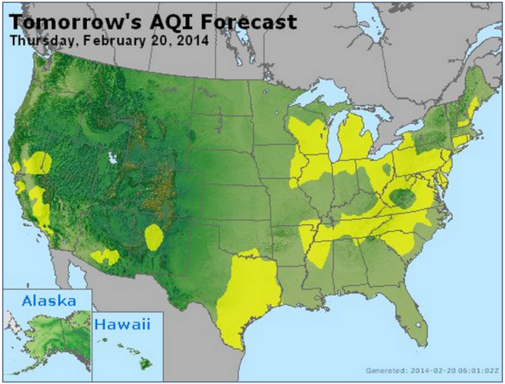
The table below contains hourly averages for all the pollutants and meteorological conditions measured at Channelview C15/AH115 for **Tuesday, June 26, 2012**. All times shown are in CST.

Parameter Measured	Morning												Afternoon										
	Mid	1:00	2:00	3:00	4:00	5:00	6:00	7:00	8:00	9:00	10:00	11:00	Noon	1:00	2:00	3:00	4:00	5:00	6:00	7:00	8:00	9:00	10:0
Nitric Oxide	1.9	1.2	1.3	1.3	1.0	**20.7**	11.1	3.7	0.4	0.6	0.6	0.4	0.8	0.8	1.0	1.2	1.6	1.5	2.0	1.8	1.0	0.8	0.4
Nitrogen Dioxide	20.4	19.9	13.1	9.1	11.9	25.8	25.5	19.5	11.0	6.0	***5.7***	5.9	11.0	10.6	13.3	14.3	22.5	26.8	**33.5**	32.6	12.9	19.1	13.7
Oxides of Nitrogen	22.8	21.6	14.7	10.5	13.1	**47.5**	37.4	23.6	11.6	6.7	***6.4***	6.4	12.0	11.7	14.5	15.9	24.6	28.8	36.3	35.1	14.2	20.3	14.4
Ozone	***4***	8	17	19	13	4	7	27	52	60	62	65	89	87	91	**92**	74	61	49	41	64	41	58
Wind Speed	1.8	3.1	2.7	***1.7***	2.2	1.8	2.4	2.6	3.3	3.5	4.0	5.9	7.1	8.5	**9.2**	8.3	7.9	7.0	5.2	3.0	3.2	2.2	2.3

but they do not produce forecasts or location-specific estimates.

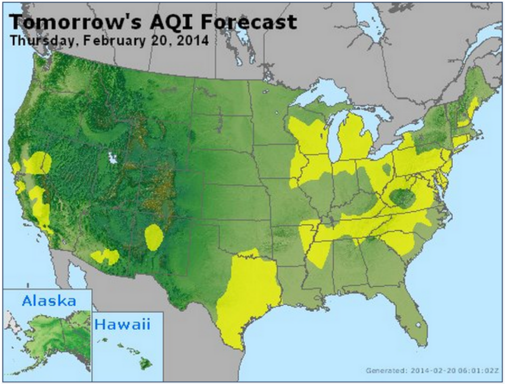
The problem is what we do with the data: mapping it

The EPA produces daily forecast maps for the AQI:



The problem is what we do with the data: mapping it

The EPA produces daily forecast maps for the AQI:



but they do not capture the dynamics of ozone pollution, since they use coarse scales for time and location grids.

Unclouding the map: real-time risk

The EPA and TCEQ are bound by the regulatory context, which is based on retrospective analysis, and only report data accordingly.

Unclouding the map: real-time risk

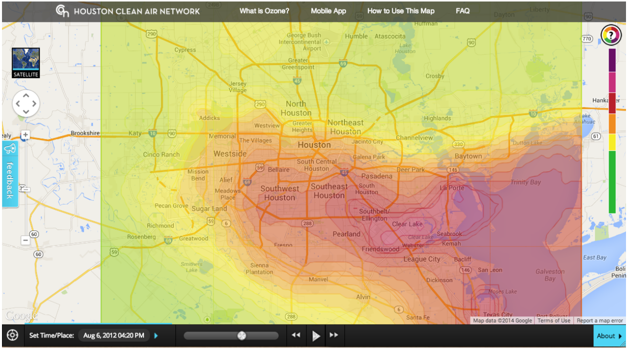
The EPA and TCEQ are bound by the regulatory context, which is based on retrospective analysis, and only report data accordingly.

Our task was to build mobile apps and a website that provide maps and individualized estimates of current ozone density, using the existing measurement framework.

Unclouding the map: real-time risk

The EPA and TCEQ are bound by the regulatory context, which is based on retrospective analysis, and only report data accordingly.

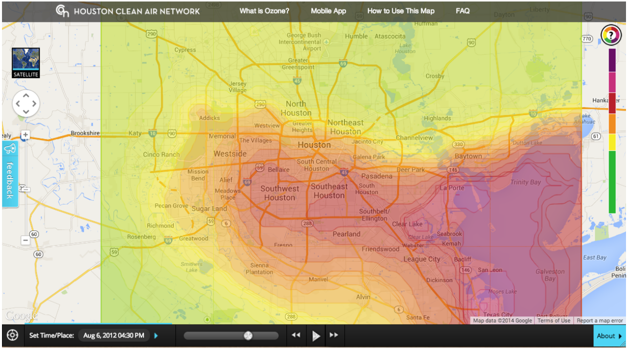
Our task was to build mobile apps and a website that provide maps and individualized estimates of current ozone density, using the existing measurement framework.



Unclouding the map: real-time risk

The EPA and TCEQ are bound by the regulatory context, which is based on retrospective analysis, and only report data accordingly.

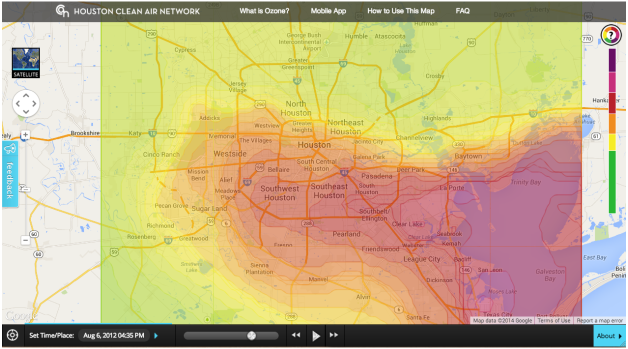
Our task was to build mobile apps and a website that provide maps and individualized estimates of current ozone density, using the existing measurement framework.



Unclouding the map: real-time risk

The EPA and TCEQ are bound by the regulatory context, which is based on retrospective analysis, and only report data accordingly.

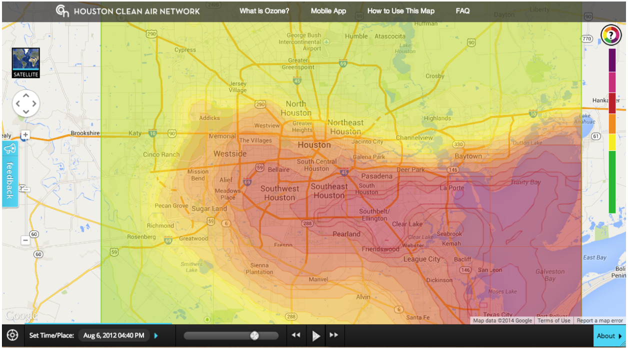
Our task was to build mobile apps and a website that provide maps and individualized estimates of current ozone density, using the existing measurement framework.



Unclouding the map: real-time risk

The EPA and TCEQ are bound by the regulatory context, which is based on retrospective analysis, and only report data accordingly.

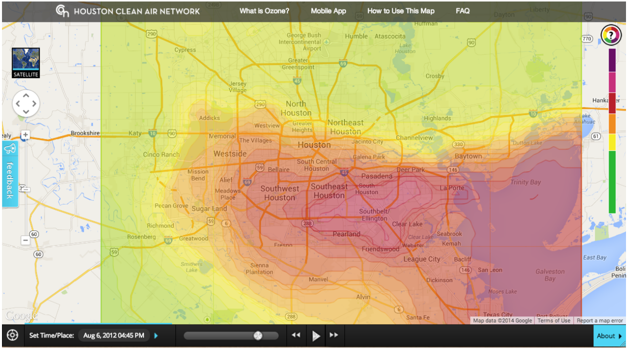
Our task was to build mobile apps and a website that provide maps and individualized estimates of current ozone density, using the existing measurement framework.



Unclouding the map: real-time risk

The EPA and TCEQ are bound by the regulatory context, which is based on retrospective analysis, and only report data accordingly.

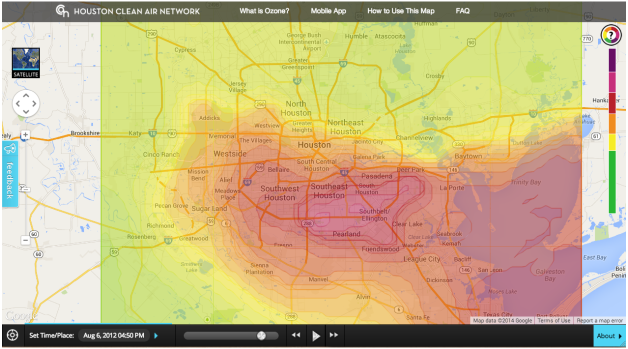
Our task was to build mobile apps and a website that provide maps and individualized estimates of current ozone density, using the existing measurement framework.



Unclouding the map: real-time risk

The EPA and TCEQ are bound by the regulatory context, which is based on retrospective analysis, and only report data accordingly.

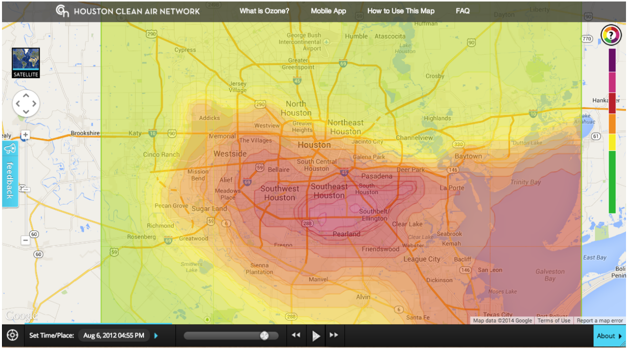
Our task was to build mobile apps and a website that provide maps and individualized estimates of current ozone density, using the existing measurement framework.



Unclouding the map: real-time risk

The EPA and TCEQ are bound by the regulatory context, which is based on retrospective analysis, and only report data accordingly.

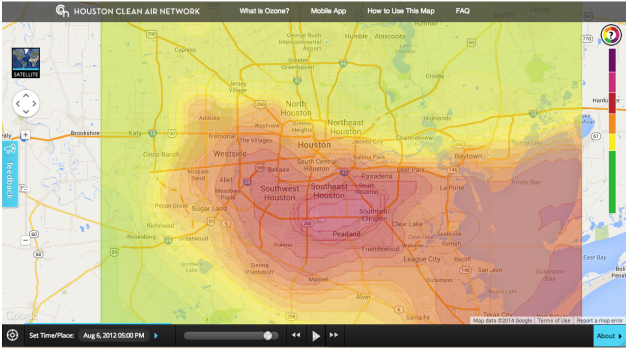
Our task was to build mobile apps and a website that provide maps and individualized estimates of current ozone density, using the existing measurement framework.



Unclouding the map: real-time risk

The EPA and TCEQ are bound by the regulatory context, which is based on retrospective analysis, and only report data accordingly.

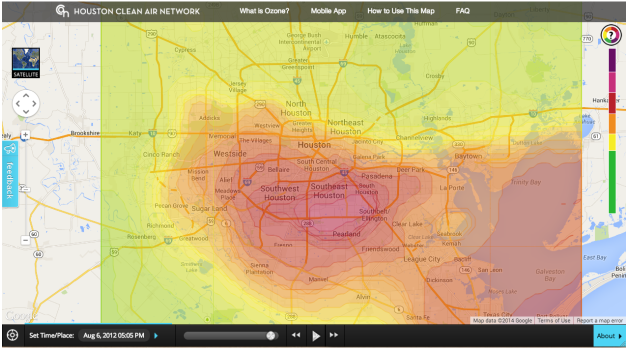
Our task was to build mobile apps and a website that provide maps and individualized estimates of current ozone density, using the existing measurement framework.



Unclouding the map: real-time risk

The EPA and TCEQ are bound by the regulatory context, which is based on retrospective analysis, and only report data accordingly.

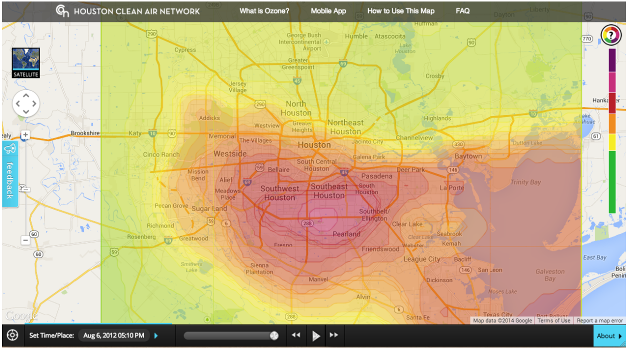
Our task was to build mobile apps and a website that provide maps and individualized estimates of current ozone density, using the existing measurement framework.



Unclouding the map: real-time risk

The EPA and TCEQ are bound by the regulatory context, which is based on retrospective analysis, and only report data accordingly.

Our task was to build mobile apps and a website that provide maps and individualized estimates of current ozone density, using the existing measurement framework.



Part II: the method

Assume an unknown

$$f : \mathbb{R}^D \rightarrow \mathbb{R} : \mathbf{x} \mapsto f(\mathbf{x})$$

Part II: the method

Assume an unknown

$$f : \mathbb{R}^D \to \mathbb{R} : \mathbf{x} \mapsto f(\mathbf{x})$$

and a set of N observations

$$\begin{bmatrix} \mathbf{X} \\ Y \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_N \\ y_1 & y_2 & \dots & y_N \end{bmatrix}$$

such that

$$y_n = f(\mathbf{x}_n), \; n = 1, \dots, N$$

Part II: the method

Assume an unknown

$$f : \mathbb{R}^D \to \mathbb{R} : \mathbf{x} \mapsto f(\mathbf{x})$$

and a set of N observations

$$\begin{bmatrix} \mathbf{X} \\ Y \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_N \\ y_1 & y_2 & \dots & y_N \end{bmatrix}$$

such that

$$y_n = f(\mathbf{x}_n), \ n = 1, \dots, N$$

Can we estimate f at a given $\mathbf{x}_\star \in \mathbb{R}^D$?

Part II: the method

Assume unknown

$$f : \mathbb{R}^D \rightarrow \mathbb{R} : \mathbf{x} \mapsto f(\mathbf{x})$$

AND

$$\phi : \mathbb{R}^D \rightarrow \mathbb{R}^P : \mathbf{x} \mapsto \phi(\mathbf{x})$$

Part II: the method

Assume unknown

$$f : \mathbb{R}^D \to \mathbb{R} : \mathbf{x} \mapsto f(\mathbf{x})$$

AND

$$\phi : \mathbb{R}^D \to \mathbb{R}^P : \mathbf{x} \mapsto \phi(\mathbf{x})$$

(ϕ stands for ϕeature vector)

Part II: the method

Assume unknown

$$f : \mathbb{R}^D \to \mathbb{R} : \mathbf{x} \mapsto f(\mathbf{x})$$

AND

$$\phi : \mathbb{R}^D \to \mathbb{R}^P : \mathbf{x} \mapsto \phi(\mathbf{x})$$

(ϕ stands for ϕeature vector)
($\mathbb{R}^P$ stands for Pheature space)

Part II: the method

Assume unknown

$$f : \mathbb{R}^D \to \mathbb{R} : \mathbf{x} \mapsto f(\mathbf{x})$$

AND

$$\phi : \mathbb{R}^D \to \mathbb{R}^P : \mathbf{x} \mapsto \phi(\mathbf{x})$$

Key assumption: ϕ must be invertible and known.

Part II: the method

Assume unknown

$$f : \mathbb{R}^D \to \mathbb{R} : \mathbf{x} \mapsto f(\mathbf{x})$$

AND

$$\phi : \mathbb{R}^D \to \mathbb{R}^P : \mathbf{x} \mapsto \phi(\mathbf{x})$$

Key assumption: ϕ must be invertible and known.
Now replace $\mathbf{X}$ by $\mathbf{\Phi} = \phi(\mathbf{X})$ and consider the observations

$$\begin{bmatrix} \mathbf{\Phi} \\ Y \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \dots & \phi_N \\ y_1 & y_2 & \dots & y_N \end{bmatrix}$$

such that

$$y_n = f \circ \phi^{-1}(\phi_n),\ n = 1, \dots, N$$

Can we estimate $\tilde{f} = f \circ \phi^{-1}$ at a given $\phi_\star \in \mathbb{R}^P$?

Finite Frames!

$\mathbf{\Phi}$ is a (finite) frame for its span in $\mathbb{R}^P$. Let $\mathcal{H} = \text{span}\mathbf{\Phi}$
Obligatory definitions follow:

Finite Frames!

$\mathbf{\Phi}$ is a (finite) frame for its span in $\mathbb{R}^P$. Let $\mathcal{H} = \text{span}\mathbf{\Phi}$
Obligatory definitions follow:
- Analysis/Bessel

$$L : \mathcal{H} \to \mathbb{R}^N : \phi \mapsto \mathbf{\Phi}^T \phi = \{\langle \phi, \phi_n \rangle\}_{n=1}^N$$

Finite Frames!

$\mathbf{\Phi}$ is a (finite) frame for its span in $\mathbb{R}^P$. Let $\mathcal{H} = \text{span}\mathbf{\Phi}$
Obligatory definitions follow:
- Analysis/Bessel

$$L : \mathcal{H} \to \mathbb{R}^N : \phi \mapsto \mathbf{\Phi}^T \phi = \{\langle \phi, \phi_n \rangle\}_{n=1}^N$$

- Frame operator

$$S : \mathcal{H} \to \mathcal{H} : \phi \mapsto \mathbf{\Phi}\mathbf{\Phi}^T \phi = \sum_{n=1}^{N} \langle \phi, \phi_n \rangle \phi_n$$

Finite Frames!

$\mathbf{\Phi}$ is a (finite) frame for its span in $\mathbb{R}^P$. Let $\mathcal{H} = \text{span}\mathbf{\Phi}$
Obligatory definitions follow:

- Analysis/Bessel

$$L : \mathcal{H} \to \mathbb{R}^N : \phi \mapsto \mathbf{\Phi}^T \phi = \{\langle \phi, \phi_n \rangle\}_{n=1}^N$$

- Frame operator

$$S : \mathcal{H} \to \mathcal{H} : \phi \mapsto \mathbf{\Phi}\mathbf{\Phi}^T \phi = \sum_{n=1}^{N} \langle \phi, \phi_n \rangle \phi_n$$

- Gram matrix

$$G = \mathbf{\Phi}^T \mathbf{\Phi}$$

- Gram operator

$$\mathbb{R}^N \to \mathbb{R}^N : Y \mapsto \mathbf{\Phi}^T \mathbf{\Phi} Y$$

More frames

Lemma

If $\tilde{f}$ is a linear functional, i.e., $\tilde{f}(\phi_\star) = \phi_\star^T \alpha$, and $y_n = \tilde{f}(\phi_n)$, then

$$\alpha = (\mathbf{\Phi}\mathbf{\Phi}^T)^{-1}\mathbf{\Phi} Y = \mathbf{\Phi}(\mathbf{\Phi}^T\mathbf{\Phi})^{-1} Y$$

More frames

Lemma

If $\tilde{f}$ is a linear functional, i.e., $\tilde{f}(\phi_\star) = \phi_\star^T \alpha$, and $y_n = \tilde{f}(\phi_n)$, then

$$\alpha = (\mathbf{\Phi}\mathbf{\Phi}^T)^{-1}\mathbf{\Phi} Y = \mathbf{\Phi}(\mathbf{\Phi}^T\mathbf{\Phi})^{-1} Y$$

Proof.

Since $y_n = \tilde{f}(\phi_n)$, we have $Y = \mathbf{\Phi}^T\alpha$, so $L^\star(Y) = \mathbf{\Phi} Y = \mathbf{\Phi}\mathbf{\Phi}^T\alpha = S(\alpha)$
Hence,

$$\alpha = S^{-1}L^\star(Y) = (\mathbf{\Phi}\mathbf{\Phi}^T)^{-1}\mathbf{\Phi} Y$$

Note that,

$$L^\star G = L^\star(LL^\star) = (L^\star L)L^\star = SL^\star$$

$$S^{-1}(L^\star G)G^{-1} = S^{-1}(SL^\star)G^{-1}$$

$$S^{-1}L^\star = L^\star G^{-1}$$

□

An estimate

Corollary

If $\tilde{f}$ is a linear functional, i.e., $\tilde{f}(\phi_\star) = \phi_\star^T \alpha$, and

$$\phi_\star \in \mathcal{H} = span\mathbf{\Phi} \subset \mathbb{R}^P,$$

then

$$\tilde{f}(\phi_\star) = \phi_\star^T \mathbf{\Phi}(\mathbf{\Phi}^T \mathbf{\Phi})^{-1} Y \tag{1}$$

An estimate

Corollary

If $\tilde{f}$ is a linear functional, i.e., $\tilde{f}(\phi_\star) = \phi_\star^T \alpha$, and

$$\phi_\star \in \mathcal{H} = span\mathbf{\Phi} \subset \mathbb{R}^P,$$

then

$$\tilde{f}(\phi_\star) = \phi_\star^T \mathbf{\Phi}(\mathbf{\Phi}^T \mathbf{\Phi})^{-1} Y \tag{1}$$

- What if $\phi_\star = \phi(x_\star) \notin \mathcal{H}$?

An estimate

Corollary

If $\tilde{f}$ is a linear functional, i.e., $\tilde{f}(\phi_\star) = \phi_\star^T \alpha$, and

$$\phi_\star \in \mathcal{H} = span\mathbf{\Phi} \subset \mathbb{R}^P,$$

then

$$\tilde{f}(\phi_\star) = \phi_\star^T \mathbf{\Phi}(\mathbf{\Phi}^T\mathbf{\Phi})^{-1} Y \qquad (1)$$

- What if $\phi_\star = \phi(x_\star) \notin \mathcal{H}$?
- Replace $\mathbf{\Phi}^T\mathbf{\Phi}$ by $\mathbf{\Phi}^T\mathbf{\Phi} + \sigma^2\mathbb{I}$ in Eq (1) to find the expected value of the Bayesian estimation of $\tilde{f}$, given a zero-mean Gaussian prior for $\alpha \sim \mathcal{N}(0, \mathbb{I})$ and assuming additive errors in measurement that follow $\mathcal{N}(0, \sigma^2)$:

$$\tilde{f}(\phi_\star) = \phi_\star^T \mathbf{\Phi}(\mathbf{\Phi}^T\mathbf{\Phi} + \sigma^2\mathbb{I})^{-1} Y$$

From features to kernels

Remark: If we can extend the mapping

$$G : \{\mathbf{x}_n\}_{n=1}^N \times \{\mathbf{x}_n\}_{n=1}^N \to \mathbb{C}$$

$$(\mathbf{x}_n, \mathbf{x}_m) \mapsto \langle \boldsymbol{\phi}_n, \boldsymbol{\phi}_m \rangle$$

to a kernel

$$G : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{C}$$

then we can drop our key assumption; we don't need an explicit formula for the feature map ϕ, since we only need to compute $\langle \phi(\mathbf{x}_\star), \phi(\mathbf{x}_n) \rangle$ for Eq (1).

From features to kernels

Remark: If we can extend the mapping

$$G : \{\mathbf{x}_n\}_{n=1}^N \times \{\mathbf{x}_n\}_{n=1}^N \to \mathbb{C}$$

$$(\mathbf{x}_n, \mathbf{x}_m) \mapsto \langle \phi_n, \phi_m \rangle$$

to a kernel

$$G : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{C}$$

then we can drop our key assumption; we don't need an explicit formula for the feature map ϕ, since we only need to compute $\langle \phi(\mathbf{x}_\star), \phi(\mathbf{x}_n) \rangle$ for Eq (1).

A common assumption is homogeneity, i.e., that there exists h such that

$$G(\mathbf{x}_n, \mathbf{x}_m) = h(\|\mathbf{x}_n - \mathbf{x}_m\|)$$

From features to kernels

Remark: If we can extend the mapping

$$G : \{\mathbf{x}_n\}_{n=1}^N \times \{\mathbf{x}_n\}_{n=1}^N \to \mathbb{C}$$

$$(\mathbf{x}_n, \mathbf{x}_m) \mapsto \langle \phi_n, \phi_m \rangle$$

to a kernel

$$G : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{C}$$

then we can drop our key assumption; we don't need an explicit formula for the feature map ϕ, since we only need to compute $\langle \phi(\mathbf{x}_\star), \phi(\mathbf{x}_n) \rangle$ for Eq (1).
A common assumption is homogeneity, i.e., that there exists h such that

$$G(\mathbf{x}_n, \mathbf{x}_m) = h(\|\mathbf{x}_n - \mathbf{x}_m\|)$$

A common choice for h is a Gaussian, leading to Gaussian Process Regression

Part III

The Golden Jubilee

Mathematics Genealogy Project

- Home
- Search
- Extrema
- About MGP
- Links
- FAQs
- Posters
- Submit Data
- Mirrors

A service of the NDSU Department of Mathematics, in association with the American Mathematical Society.

Please email us with feedback.

John Joseph Benedetto

MathSciNet

Ph.D. University of Toronto 1964

Dissertation: *The Laplace Transform of Generalized Functions*

Advisor: H. Chandler (Horace) Davis

Students:
Click here to see the students listed in chronological order.

Name	School	Year	Descendants
Enrico Au-Yeung	University of Maryland College Park	2011	
George Benke	University of Maryland College Park	1971	
Erica Bernstein	University of Maryland College Park	1992	
Abdelkrim Bourouihiya	University of Maryland College Park	2006	
Somantika Datta	University of Maryland College Park	2007	
Kevin Duke	University of Maryland College Park	2012	