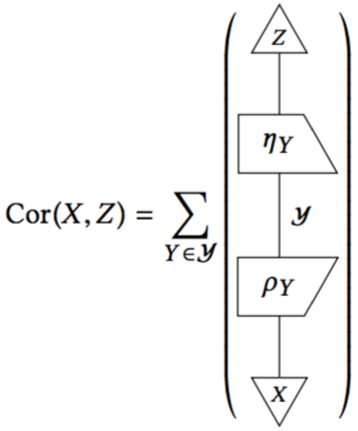


Indicator Frameworks

Joshua Tan (Oxford), Christine Kendrick (City of Portland),
Abhishek Dubey (Vanderbilt), and Sokwoo Rhee (NIST)

SCOPE 2017, Carnegie Mellon University

April 21, 2017

Overarching question: how do we synthesize models with data?

Towards a science of measuring systems

- Cities are **cyber-physical systems**
- Cities are **systems of systems**
- These are *mathematical descriptions*. They do not measure anything, per se.
- Other mathematical descriptions:
    - Network approaches
    - Economic models
    - Game theory

Towards a science of measuring cities

- Problem: the **models** fail to describe the world perfectly.
- Problem: the **data** doesn’t either.*

* cities are **complex**: there isn’t enough raw data
to describe all the interactions

Indicator frameworks

- They are simple.
- They are concrete.
- They are already being used.
- Step 1: give a **mathematical semantics** for indicator frameworks.
- Step 2: test whether indicator frameworks can be upgraded to synthesize **models** with **data**.

This Talk

- Background and Motivation
- Operational Indicator Frameworks
- Abstract Indicator Frameworks
- Causal Diagrams
- Future Work

Operational indicator frameworks

- An indicator is a column of numeric data values. They are typically to used to measure inputs, immediate outcomes, and long-term impacts of city projects. We assume that all indicators are time-varying.

```
id,id_wasp,id_secret,frame_type,frame_number,sensor,value,timestamp,raw,parser_type
44637,city1,408414489,128,132,noise,50,"2016-08-10 06:00:29",noraw,0
44679,city1,408414489,128,138,noise,52,"2016-08-10 06:02:24",noraw,0
44742,city1,408414489,128,143,noise,51,"2016-08-10 06:04:00",noraw,0
44777,city1,408414489,128,149,noise,55,"2016-08-10 06:05:55",noraw,0
44819,city1,408414489,128,152,noise,60,"2016-08-10 06:06:53",noraw,0
44875,city1,408414489,128,160,noise,62,"2016-08-10 06:09:27",noraw,0
```

- An operational indicator framework is just a list of indicators, sometimes organized hierarchically.

Congestion	% in hours	Increase in overall travel times when compared to free flow situation (uncongested situation
Public transport use	#/cap/year	Annual number of public transport trips per capita
Net migration	#/1000	Rate of population change due to migration per 1000 inhabitants
Population Dependency Ratio	#/100	Number of economically dependent persons (net consumers) per 100 economically active persons (net producers),
International Events Hold	#/100.000	The number of international events per 100.000 inhabitants
Tourism intensity	nights/100.000	Number of tourist nights per year per 100.000 inhabitants

Attractiveness and competitiveness indicators, from CITYkeys, 2016

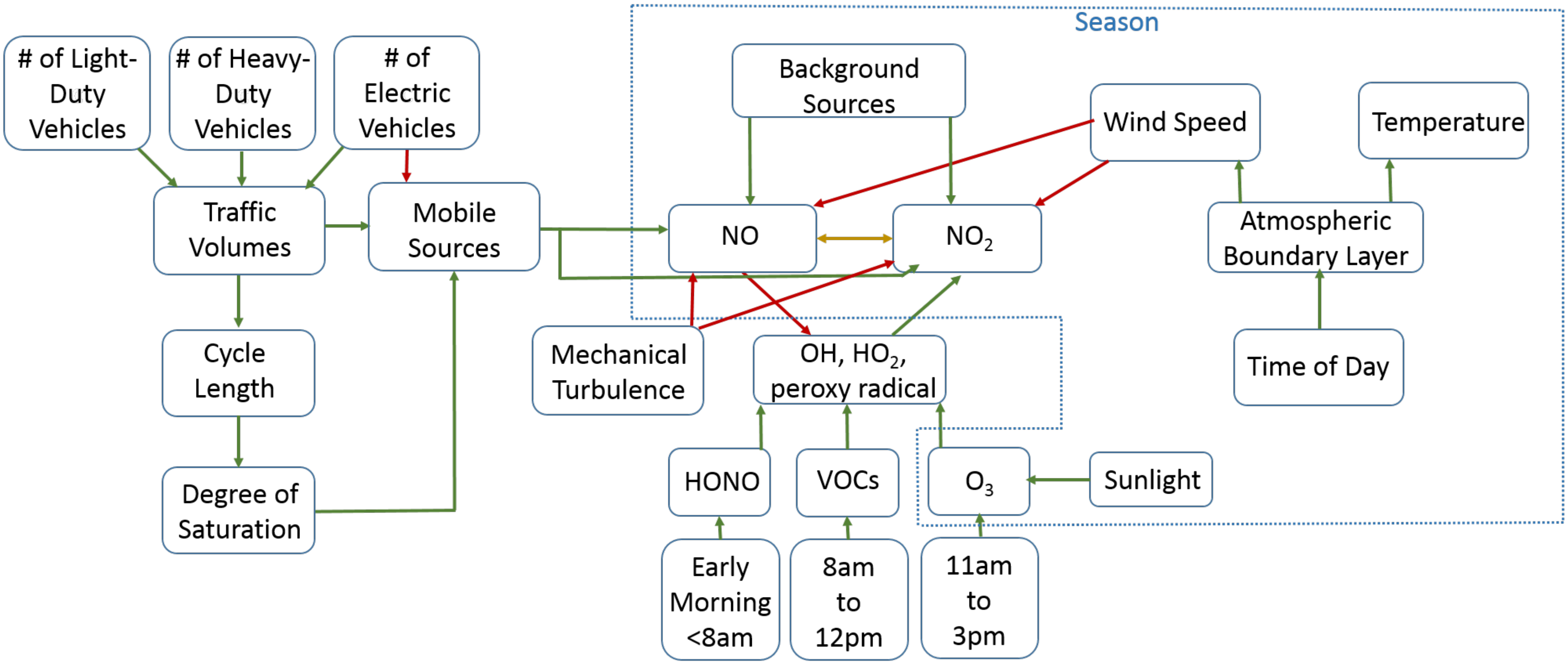
Topic	Day	Week	Month	QTR
311 CALL CENTER PERFORMANCE	0.92	0.88	0.89	0.9
GRAFFITI ON-TIME %	1.0	0.89	0.94	1.06
MISSED TRASH ON-TIME %	1.14	1.19	1.2	1.19
PARKS MAINTENANCE ON-TIME %	1.15	1.1	1.09	1.05
POTHOLE ON-TIME %	0.83	0.97	0.77	0.83
SIGN INSTALLATION ON-TIME %	0.83	1.01	1.14	1.07
SIGNAL REPAIR ON-TIME %	1.04	1.11	1.18	1.14
STREETLIGHT ON-TIME %	0.2	0.59	0.57	0.72
TREE MAINTENANCE ON-TIME %	1.14	1.18	1.19	1.19
ON-TIME PERMIT REVIEWS	0.98	0.97	0.91	0.99
LIBRARY USERS	1.44	1.15	1.27	1.29

Selection of indicators from CityScore, Mayor’s Office, City of Boston, 2017

How do we construct operational indicator frameworks?

- Expert input, Likert scales, but also…
- We can refine old indicators by “cleaning” the data.
- We can form combine sets of indicators into new indicators.
- We can compute or infer correlations between indicators, then describe proxies.

Adding **structure**



Abstract indicator frameworks, v1

- An abstract indicator framework is composed of:
  1. A **R**-valued matrix whose columns represent indicators and rows represent **data**
  2. An inner product operation between indicators, understood as their sample **correlation**
- The set of all abstract indicator frameworks forms something called a **category**

Mathematical background: category theory

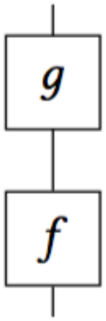
- Category theory was originally invented in the 1960s to integrate different aspects of mathematics, especially topology and algebra. It is now being tested by a variety of different agencies, like NIST and DARPA, as a language for modeling and integrating large, heterogeneous **systems**, e.g. Nextgen or CASCADE.
- A *category* **C** is a collection of objects, called the objects of C, along with a collection of maps, called the morphisms of C, satisfying certain properties.
- A *functor* F : **C** -> **D** between two categories is a map taking objects of C to objects of D, and morphisms of C to morphisms of D in a compatible way.

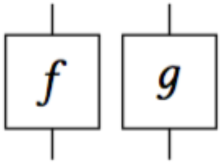
Definition 3.4. The category of $\mathbb{R}$-valued data tables, Data, is defined by the following data:

(1) objects $\mathcal{X} = (\mathcal{X}, \Omega_{\mathcal{X}}, \mathbb{I}_{\mathcal{X}})$ of Data are $m \times n$ tables of $\mathbb{R}$-valued data vectors whose rows are assigned an index key given by $\mathbb{I}_{\mathcal{X}} : \Omega_{\mathcal{X}} \rightarrow \mathbb{R}$ and whose columns, $B_{\mathcal{X}} = \{X_1, ..., X_n\}$, represent indicators
(2) morphisms $f : \mathcal{X} \rightarrow \mathcal{Y}$ are linear transformations of the column values of $\mathcal{X}$ by vector addition (of other columns in $\mathcal{X}$) and scalar multiplication
(3) the composition is just the matrix product
(4) the tensor product of $\mathcal{X} \otimes \mathcal{Y}$ is the integrated table of their data values over a table of linkages, $S \subset \Omega_{\mathcal{X}} \times \Omega_{\mathcal{Y}}$

Mathematical background: monoidal categories

- A category with a tensor operation ⊗ : **C** × **C** -> **C**, satisfying certain properties, is called a *monoidal category*.
- Essentially the tensor allows you to compare morphisms in parallel, while composition allows you to compare morphisms in series.





Abstract indicator frameworks, v2

- We want to abstract from the data management aspect. It’s the choice of the indicators that is important, not the individual rows of data underneath.
- Many of operations on indicators are purely statistical, e.g. correlation, so we would like to define them in a general context.
- We especially want to emphasize correlation, because it emphasizes relations *between* indicators rather than the indicators themselves.
- We want to set the stage for linking **models** to **data**.
- This motivates the following definition:

Definition 3.2. The category of random variables, Rand, is defined by the following data:

(1) objects are finite-dimensional Hilbert spaces

$$\mathcal{X} = L^2(\Omega_{\mathcal{X}}, \Sigma_{\mathcal{X}}, \mathbb{P}_{\mathcal{X}})$$

of square-integrable random variables (under the equivalence relation $X_1 \sim X_2$ if $\mathbb{P}_{\mathcal{X}}(X_1 = X_2) = 1$) with inner product $\langle X, Y\rangle = E(XY)$, defined over probability spaces $(\Omega_{\mathcal{X}}, \Sigma_{\mathcal{X}}, \mathbb{P}_{\mathcal{X}})$, with an associated basis $\mathcal{B}_{\mathcal{X}} = \{X_1, X_2, ..., X_n\} \cup$ $\mathbf{1}$, where $\mathbf{1}$ is the random variable with constant value 1.

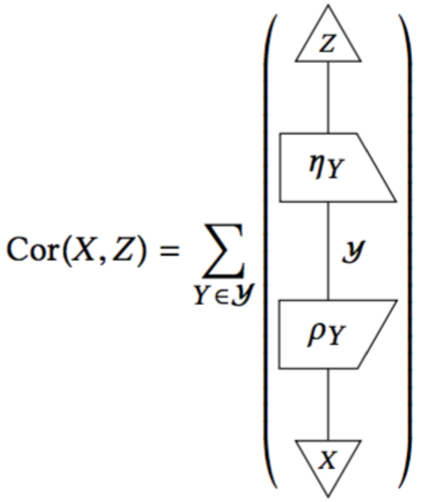
(2) morphisms $F : \mathcal{X} \to \mathcal{Y}$ are bounded linear operators

(3) the composition is the usual composition of bounded linear operators

(4) the tensor product of $\mathcal{X}$ and $\mathcal{Y}$ is the pushout over their joint support in $\Omega_{\mathcal{X}} \times \Omega_{\mathcal{Y}}$

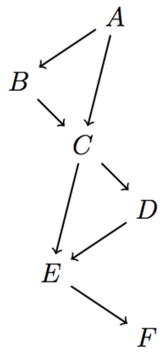
Defining new indicator frameworks in **Rand**

- Given two indicator frameworks **X** and **Z** in **Rand**, one can write "formulas" in **Rand** that describe new indicator frameworks, exactly analogous to how one defines mediating (or confounding) variables in statistics. In the diagrammatic calculus, these look something like this:



A simple model: causal diagrams

- Causal diagrams are directed acyclic graphs, e.g.



- Each causal diagram can be used to construct a monoidal category, called a *causal theory* [Fong '13].
- Each graph, essentially, defines a procedure for computing mediating variables.

Abstract indicator frameworks, v3

- The causal diagram serves as a primitive **model** of a given context.
- The category **Rand** of random variables serves as a (still primitive) **semantics of how we use data**.
- The idea: define constraints in **Rand**, and thus indicator frameworks, by mapping the causal theory into **Rand**; this creates a **model** of the causal theory in **Rand**., i.e. an indicator framework based on the causal theory.

Definition 3.7. The category Ind of abstract indicator frameworks is defined by the following data:

1. an object I of Ind is a strong symmetric monoidal functor $C \to \mathsf{Rand}$ from a causal theory C to the category of random variables.
2. a morphism η between abstract indicator frameworks is a natural transformation of strong symmetric monoidal functors

Future Work

- More examples!
- Constructions besides mediating frameworks
- Improved semantics on **Rand**
- True “hybrid indicator frameworks” for CPS models

Thank you.

Example: Shot Spot in South Bend

- Courtesy of Santiago Garces, CIO of South Bend, Indiana
- Target indicator: reduce crime
- Target indicator: reduce gun crime and group-related activity
- Target indicator: target interventions at specific group members
- Means: Incorporate Shot Spot information with 911 dispatch calls. “Whenever, a shot incident is detected and a resident also calls 911 to inform of the incident, the dispatch is classified differently than when a signal is detected but not accompanied with a resident's call.”

Example: Shot Spot in South Bend

Indicators Used

1. **The total number of shootings involving a group member, compared to a 3 year rolling average;**
2. the ratio of group member involved shootings, compared to the total number of criminal assault shootings;
3. **Number of shooting incidents recorded both by Shot Spotter and residents**
4. **Number of direct interventions with group members, or close social relatives**
5. Number of call-ins (large meetings where notorious group members are presented with the opportunity to get involved with social services, or communicated the enforcement action alternative)
6. Number of enforcement actions, interventions directed at executing warrants and investigations against the most violent group
7. Ratio of shots where a resident called as a shot is detected by Shot Spotter (proxy for community trust and collaboration with the Police Department)
8. Percentage/ s-value of number of complaints relative to calls for service

Example: Shot Spot in South Bend

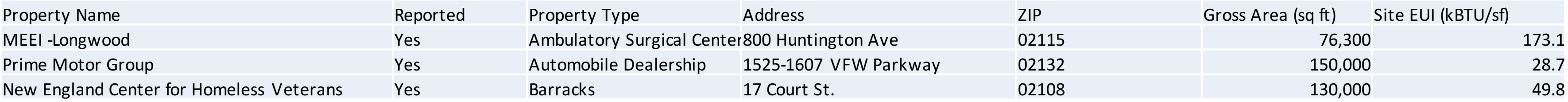
- Things we need to model:
  - Frameworks that draw on indicators from a number of different sources: ISO 37120 for overview, FBI uniform crime reporting program, 911 CAD system, Shot Spotter, internal databases for group activity and “interventions”
  - Comparisons of indicators as indicators, e.g. the total # of shootings compared to a 3 year rolling average
  - Logical operations on indicators, e.g. # of shootings recorded by Shot Spotter AND local residents, # of direct interventions with group members OR close relatives
  - Relationship between indicator and `sub-indicators’
  - Properties of indicators, like how difficult it is to supply that indicator

Example: Nextgen

- Courtesy of John Baez and Metron
- Nextgen: “Next Generation Air Transportation System”
- From DARPA: “The DoD and urban infrastructure capabilities are increasingly based on the integration and coordination of heterogeneous systems using System-of-Systems (SoS) architectures. However, it is difficult to model and currently impossible to systematically design such complex systems using state-of-the-art tools, leading to inferior performance, unexpected problems, and weak resilience. This inadequacy in design capability results from the complexity of interactions between system structures and behaviors across multiple time and length scales that cannot be adequately modeled using conventional approaches.”

Non-temporal indicators

- E.g., location-varying indicators like energy use per building:



Property Name	Reported	Property Type	Address	ZIP	Gross Area (sq ft)	Site EUI (kBTU/sf)
MEEI -Longwood	Yes	Ambulatory Surgical Center	800 Huntington Ave	02115	76,300	173.1
Prime Motor Group	Yes	Automobile Dealership	1525-1607 VFW Parkway	02132	150,000	28.7
New England Center for Homeless Veterans	Yes	Barracks	17 Court St.	02108	130,000	49.8