



## **The Eastern Transportation Coalition:**

Hurricane Pilot Results Web Meeting: States' Experience with Real-time Pilot Data

January 28, 2021

### ***Question and Answer Summary***

#### ***Goals and Objectives***

**Q: Habte Kassa (Georgia DOT):** From the states shown, which state has the most connected vehicle fleet?

A: Stan Young (National Renewable Energy Laboratory): Fleets in this sense are all passenger (light-duty fleet) from a particular original equipment manufacturer (OEM).

A: Jaap van den Hoek (Wejo Ltd) and Wander Boesjes (Moonshadow Mobile, Inc.): Florida has the most connected vehicles.

**Q: Lev Pinelis (Transurban):** What was the data latency and frequency?

A: Jaap van den Hoek (Wejo Ltd): 3-second capture rate, between 30-60 seconds latency from the vehicle to Moonshadow in this case.

#### ***Overview of Connected Vehicle Data***

**Q: Vaishali Shah (AEM Corporation):** Any personally identifiable information (PII) concerns with public agency access?

A: Jaap van den Hoek (Wejo Ltd): No PII concerns as the data is fully consented to, anonymized, and checked again before it goes out to users either public or private.

#### ***Accuracy of Real-Time Volume Data***

**Q: Krishnan Viswanathan:** Given these are sample data from certain OEMs, how was it expanded? Does it rely on count data?

A: Jaap van den Hoek (Wejo Ltd): Kaveh will talk about the uplift factor now which will explain the process.

**Q: Josh Roll (Oregon DOT):** Do "uplift factors" equal expansion factors?

A: Stan Young (National Renewable Energy Laboratory): Yes - an uplift factor of 0.10 (or 10%) equates to an expansion factor of 10x.

**Q: Lev Pinellas (Transurban):** What are the sizes of the segments?

A: Eimar Boesjes (Moonshadow Mobile Inc.): These are Open street maps (OSM) segments, usually from intersection to intersection or off-ramp.

A: Jaap van den Hoek (Wejo Ltd): Open street maps (OSM) IDs vary in length, for the POC we only used functional road classes (FRC) 1, 2, and 3.

**C: Matthew Glasser (Georgia DOT):** I see the same patterns in regular probe data: 10 PM -5 AM show higher probe penetration volatility.



## The Eastern Transportation Coalition:

Hurricane Pilot Results Web Meeting: States' Experience with Real-time Pilot Data

January 28, 2021

### *Question and Answer Summary*

**Q: Vaishali Shah (AEM Corporation):** Would uplift factors remain robust/consistent when vehicle make/market penetrations vary?

A: Jaap van den Hoek (Wejo Ltd): Vehicles population/penetration vary per state but with connected vehicle data (CVD) you can be sure it's a genuine vehicle reporting not e.g., cell phones.

**Q: Lev Pinelis (Transurban):** What was the average estimated penetration rate?

A: Stan Young – during the demo we estimated penetration rates of 3%. The calibration work by Kaveh confirmed that generally, it was from 2% to 5% depending on the region and road class.

**Q: Autumn Young (Florida DOT):** What type of traffic volumes provided by the DOTs were used? Perhaps yearly annual average daily traffic (AADT) estimation or 24/7 data collection sites?

A: Stan Young (National Renewable Energy Laboratory): The calibration data from the states came from continuous count stations. Peter Carnes reached out to the participating states to get the corresponding data from all the continuous count stations in June of 2020.

**Q: Shawn Turner (Texas AM Trans Inst):** Did you quantify "workable volume estimates" in terms of a calculated error rate via cross-validation?

A: Stan Young (National Renewable Energy Laboratory): In this context, the 'cross-validation' term comes from machine learning and AI. We use the more traditional statistics. I invite you to discuss regression analysis with Kaveh. I know the University of Maryland is going to move towards a more in-depth evaluation that gets all the statistical summaries that you want to see. We tried to boil it up to the very top level here and make a judgment on whether there was enough evidence to support real-time volume estimates.

**Q: Autumn Young (Florida DOT):** On segments without continuous count stations, did you eliminate the segment or apply the conflation factor consistently?

A: Stan Young (National Renewable Energy Laboratory): Everything that you saw from Kaveh was based on calibration data from continuous count stations. Our goal was to turn that around and hand a table off to Moonshadow to say "here are the uplift factors that you would need to apply by state, by road type, by the time of day, and by the day of the week". We never got to the point of creating that full table, but we got to the point of assessing whether that table could be created and could be consistent enough to be workable. Moonshadow created the actual table in their software, but I don't believe we ever got to the handoff point. They used a straight uplift factor of three percent based on some preliminary estimates.



## The Eastern Transportation Coalition:

Hurricane Pilot Results Web Meeting: States' Experience with Real-time Pilot Data

January 28, 2021

### **Question and Answer Summary**

**Q:** **Matthew Glasser (Georgia DOT):** Did you have any randomly large volume days from any of the states? Did we have a good example of people evacuating an area and did the volume estimations hold?

A: Stan Young (National Renewable Energy Laboratory): Most of the anomalies that we saw in the calibration of what Kaveh looked at were hourly overnight periods, particularly when the volumes were extremely low. Kaveh can comment on that. The other question related to seeing changes in behavior relative to hurricanes? Thankfully during this period, there were no direct hurricane hits in the six-state region. There was a significant tropical storm in early September and we worked directly with Moonshadow and Wander to do some ad-hoc analysis that is included as part of the presentation.

**Q:** **Josh Roll (Oregon DOT):** I am also interested in the error when compared to the permanent count sites, doing some kind of cross-validation holdout out certain sites and applying the methods to other sites.

A: Stan Young (National Renewable Energy Laboratory): We could do that. Let's take that up with Kaveh. Cross-validation is a process where we develop a methodology based on a portion of the calibration data, say 80%, and then apply the methodology to the remaining 20% to see how accurate it is. Josh and Shawn, I'd encourage you to discuss it with Kaveh and see if we could do that quickly with the effort that we currently have in-house.

A: Kaveh Farokhi (University of Maryland CATT): The tight inter-quartile range indicates good cross-validation. At least 50% of cases in FRC1 segments were within +/-2% of the mean.

**Q:** **Patricia Hendren (The Eastern Transportation Coalition):** Amazing work. We are going to rename the "bleeding edge" to the "exciting edge". Can you clarify (similarities/differences) between this work with Wejo to the volume and turning movement (VTM) project?

A: Stan Young (National Renewable Energy Laboratory): This work is showing what is possible. The biggest difference is getting data in real-time and handling a large quantity of data. I think everyone who participated in this understood that this would come fast and furious, but our expectations of fast and furious versus what came was an order of magnitude of difference across the entire team. So, we had to learn to adjust quickly and we were able to do that. The other item is that the previous volume turning movement research work was always in a retrospective sense. We had the luxury of applying any and every data methodology we could as time was not constrained. Previously, teams resorted to machine learning, tree learning methodologies as well as AI to squeeze every ounce of accuracy that was possible. Going into this we knew we didn't have the resources in either funding or time to go to that level of detail, so we stepped back and did a more basic approach of estimating scaling factors. If the scaling factors worked,



## **The Eastern Transportation Coalition:**

Hurricane Pilot Results Web Meeting: States' Experience with Real-time Pilot Data

January 28, 2021

### ***Question and Answer Summary***

we had confidence that if we had enough time and resources, we could bring in more sophisticated methods. We all feel that we've obtained that.

**Q: Krishnan Viswanathan (Cambridge Systematics):** What was your source for vehicle occupancy factors to convert a vehicle to person trips?

A: Stan Young (National Renewable Energy Laboratory): Everything in this POC dealt with vehicle trips - we did not attempt to assess vehicle occupancy.

### ***Additional Questions***

**C:** Jaap van den Hoek (Wejo Ltd): Performance was addressed, and upload and the display were down to less than 3 minutes overall. Working with such a big data set required some tweaking and including the calculations for the volumes, the application performed well.

**C: Ben Jacobs (Rhode Island Division of Statewide Planning):** For origin data (OD) data, many of us have travel demand models of one kind or another for air quality purposes, and those use self-defined traffic analysis zone (TAZs). As this becomes more mature, integrating the data structures of the regional models TAZs may help avoid wheel spinning.

A: Stan Young (National Renewable Energy Laboratory): I fully agree. This data comes in with some built-in spatial aggregations. I believe zip code level and county level are two of them. Getting that data married as quickly as possible to additional spatial aggregations will add value to the DOTs and MPOs.

C: Jaap van den Hoek (Wejo Ltd): Yes, the CVD data is just as useful for vehicle miles traveled (VMT) and travel demand models, O/D, etc.

**C: Patricia Hendren (The Eastern Transportation Coalition):** Since our inception 25 years ago, the Coalition has been pushing the capabilities of data to increase our ability to keep people safe and improve efficiency. You can quote me here that the Coalition will continue to pursue the "volume" holy grail.

**Q: Autumn Young (Florida DOT):** Are there applications to the CVD data beyond volumes and O&D?

A: Jaap van den Hoek (Wejo Ltd): Yes! Autumn. Please contact Denise or Stan!

A: Eimar Boesjes (Moonshadow Mobile Inc.): Many-safety, corridor travel times, OD, greenhouse gas (GHG), and much more.

**C: Matthew Glasser (Georgia DOT):** Can all research be like this?