On the Link(s) Between *"D"* and *"A"* in Mobile *D*ata *A*nalytics

Goce Trajcevski

Dept. of EECS
Northwestern University
Evanston, Il, USA
goce@eecs.northwestern.edu

***Abstract*— This paper addresses the issue of (meta) management of the efficiency in the realm of applications relying on Mobile Data Analytics (MODA). We postulate that considering the efficient execution of the tasks of gathering, storing/retrieving and querying spatio-temporal data separately from the efficiency of executing control and decision-making (or, for that matter, even model-selection) algorithms may not be the most beneficial avenue in MODA settings. We argue that an important component is the management of the dynamics of the different trade-off(s) and, more importantly, linking them when orchestrating the processes of the data management and analytics. Towards that end, we present the DNA[2] (Data's Natural Associations with Analytics) hypothesis for coupling the data and control aspects in MODA contexts. We discuss in detail the role of *uncertainty*, along with some aspects of declarative specifications for merging the reactive and the pro-active behavior in MODA.**

I. Introduction

Miniaturization of computing and sensing devices, along with the advances in networking and communications, have provided a technological foundation for generating extremely large volumes of location-in-time data. According to a McKensay report from 2011, the size of location data of smart phone users is in the order of Peta-Bytes per year – not including location data inferred from cell-tower triangulation because such locations are relatively imprecise: *"Including those data would have increased our estimate of the size of personal location data 400-fold, given that the nearest cell-tower signals are determined every seven seconds"* [1].

The management of *(location,time)* information of mobile entities is essential for a variety of applications domains, ranging from navigation and efficient traffic management, through emergency/disaster rescue management, environmental monitoring, fly-through visualization, and various military applications (e.g., radar data, troops tracking) [2]. Essentially, every application requiring some form of Location Based Services (LBS) [3] needs efficient techniques for storage, retrieval and query processing of spatio-temporal data – topics studied in the field of Moving Objects Databases (MOD) [4]. Intelligent use of such data could have significant impact – e.g., it is estimated that by 2020, more than 70% of mobile phones will have GPS capability, up from 20% in 2010. In addition, the number of automobiles equipped with dashboard GPS devices will continue to grow. The potential global value of smart routing [5], [6] in the form of time and fuel savings could be about $500 billion by 2020. This is the equivalent of saving drivers 20 billion hours on the road, or 10 to 15 hours every year for each traveller, and about $150 billion on fuel consumption[1].

In addition to the sheer *(location,time)* values, many applications related to routing – e.g., optimizing the multi-modal fleet management [7], [8], [9] may require values from other types of sensors to be correlated with location data. As an extreme example, the U.S. Xpress gathers 900 to 970 data elements of various engine/component reading [10] that are used to plan the load and servicing regimes of its trucks fleet.

From a complementary (and broader) perspective, trends such as Big Data Analytics and Embedded Predictive Analytics have become the "grail-quest" of many enterprises seeking to optimize their profit, and have spurred the development of a plethora of Business Intelligence tools [11]. Regardless of the platform/technology and paradigm used, most of the approaches do have some common stages and, more importantly, have *feedback(s)* between certain stages. For example:
(1) The modules for storing prediction-models and "feeding" them into the OLAP modules are used to adjust the mining models and couple them with real-time observations. However, based on the data in the modules that monitor the effectiveness of the prediction(s), the dynamic behavior may be changed to adhere to a different model. This, in turn, necessitates a feedback into the module managing the prediction-models.
(2) The monitoring module typically provides a feedback into the OLAP module for the purpose of analyzing the reasons for bad prediction KPIs (Key Performance Indicators).

Given current trends in Big Data Analytics, the evolution of Mobile Data Management and MOD, the central issue addressed in this paper is: *what are the possible roles of spatio-temporal data management in MODA (Mobile Data Analytics) contexts*? Namely, in the recent years:

- Efficient processing of various spatio-temporal queries (e.g., range, (k) nearest neighbor, skyline), for both historic trajectories and streaming moving objects data has been addressed in many existing works works [4], [12], [13], [14], [15], in both "crisp" as well as uncertainty-aware settings [16], [17], [18], [19], [20], [21].

[1]Which, in addition, translates into an estimated reduction in carbon dioxide emissions of 380 million tonnes, or more than 5% a year [1].

- Efficient clustering and mining techniques for trajectories data have been proposed [22], [23], [24], [25], [26], [27], along with approaches for warehousing spatio-temporal data [28], [29], [30], [31], [32].
- Exploitation of cloud environments, along with MapReduce — based query processing has also been addressed more recently for spatio-temporal data [33], [34], [35], [36], [37].

So, one question that naturally arises is what are the links between the *D*ata part and *A*nalytics part in MODA? Could there be that the "D" part is only a service-provider for the "A" part? In the rest of this paper, we try to argue that this need not be the case. After a preliminary discussion in Section 2, which serves the purpose of motivating the thoughts behind the hypothesis of this paper, in Section 3 we will focus on two scenarios involving large-scale spatio-temporal datasets *and* decision processes. In Section 4, we present the main postulates of our hypothesis, called DNA^2 (Data's Natural Associations with Analytics) in MODA contexts and we identify some research challenges. We conclude the paper in Section 5 with a couple of observations regarding the applicability of the proposed hypothesis.

II. Small-Scale Observations

We now present two scenarios: one from the domain of mother-nature but extensively used in the domains of image processing and vision, and a correlated one from the topic of tracking in Wireless Sensor Networks (WSN).

A. Foveation

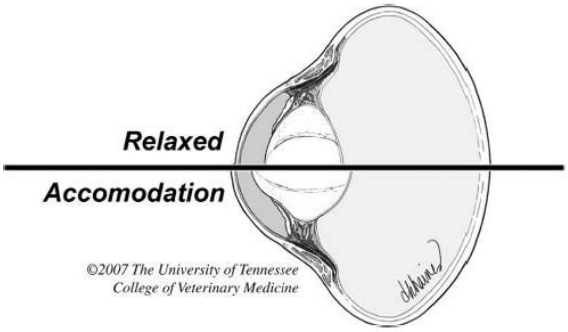

Fig. 1. Image Foveations

Eagles are one of the species that have the highest visual acuity and coordination with the muscular activities of all the animals. For instance, the Bald Eagles can focus on a swimming fish from hundreds of meters in the sky and have a fast (and successful) dive of 200mph to capture it [38]. Interestingly, unlike humans, eagles have two fovea in each eye, thus being able to focus on tracking a prey and simultaneously tracking a rival bird/animal, and adjusting their actions accordingly. In addition to the binocular vision – and more important for the issues addressed in this work – eagles have the ability to *adapt/accomodate* the muscular focusing of the fovea [39] (cf. Figure 1). Such accommodations are based on the detection of a particular event of interest – e.g., if the distance of a rival bird is decreasing faster than the distance to the prey, the peripheral selectivity is augmented.

The foveation phenomenon has been extensively used in computer vision (active vision and visualization) and image databases retrieval. For example, to compensate for a lack of peripheral vision, varying the points of interest in different foveated images may be used [40]. Similarly, a sequence of foveated images from a single high-resolution one, can yield better bandwidth utilization [41], [42]. In the context of this work, an important question is

Q1: *What is the analogue of foveation in MODA settings*?

B. Tracking in Sensor Networks

Tracking of moving entities is considered to be one of the canonical problems in Wireless Sensor Networks (WSN) research [43], [44], [45], [46]. For the purpose of this paper, one particular facet of the tracking problem illustrates a small-scale trade-offs revealing a possible link between "D" and "A" in this context.

Typically, during tracking, the sensor nodes are organized in (hierarchial) clusters [47], [48] according to their geographical positions. A designated sensor node is elected as the *tracking principal* of each cluster, acting as a temporal data fusion center and coordinator of the tracking process which involves both *localization* [49], [50] and *monitoring* data like speed, acceleration and moving direction. Tracking principles combine together such measurements [45], [51], but another important task of an on-duty tracking principal is the *selection the next principal*, to be handed off the target tracking information [52], [53]. To cater to the stringent energy constraints and provide certain quality of service in WSN, the selection of the next tracking principal has two main (complementary) aspects:
(1) Maximize the extent of trajectory coverage per principal.
(2) Minimize the number of changes/hand-offs between principals.

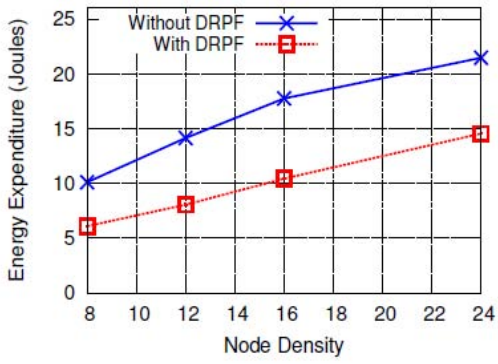

Fig. 2. Energy Savings in Tracking

Clearly, these tasks require some form of a *prediction*, based on the (near-past) history of the tracked object's motion, for which some form of Kalman filters – most often, particle filters [54] – are used. In our recent work [55] we postulated that a simple spatio-temporal location query under a

dead-reckoning policy (cf. [56]) would provide a comparable accuracy of the tracking while saving a substantial amount of energy by avoiding the computation-heavy calculations involved in particle filters. Figure 2 from [55] shows that the energy savings when using the *where_at* spatio-temporal query in the prediction could be over 50% when compared with the use of particle filters. The trade-offs (cf. [55]) were less than 10% of a location-estimation error, providing further energy savings due to reduced communication. Based on the above observation, we have a variation that can be phrased as:

Q2: *Could there be a coupling of spatio-temporal queries and control algorithms that would provide trade-offs between: acceptable (preferably bounded) error vs. overall efficiency of the process of (combining the data with the) control-based prediction*?

III. Large-Scale Mobile Data and Analytics

We now proceed with a discussion of a domain that can be considered as a classic: traffic management. The main reason for selecting this particular domain is that it brings about a few issues that are at the core of the hypothesis stated in Section IV. Namely, efficient traffic management certainly involves *large quantities* of moving objects' locations-in-time data. Complementary to it, a *decision-making* process is involved that is based on some analytics tool (e.g., game theory or control theory). Lastly, there are ever-present constraints: e.g., if determining the ideal regime for changing the traffic signal (optimizing, say, the average fuel consumption) takes too long, it may simply be invalid for the state of the traffic at which it is finally applied. Hence, one may wish to trade a sub-optimal control behavior, at the expense of timeliness.

The importance of the Automatic Traffic Management (ATM) has been recognized since the early 1900's - for instance, Ghiglieri's automatic traffic signals with red and green lights were introduced in California in 1917, and the first computer-based synchronized traffic lights in the US emerged in the 1970's. In the 1990's, the Federal Highway Administration (FHWA) established the Adaptive Control Software (ACS) research program [57], [58]. In 2004 the US Department of Transportation (US-DOT) launched the Vehicle Infrastructure Integration (VII) initiative [59] with a vision of deploying a communications infrastructure on the roadways and in all production vehicles. Methodologies, tools and systems addressing different aspects of the problem of efficient traffic management abound. However, existing best-practice approaches [60], [61], [62], [63] rely on *models* which base their decision on past observations augmented with recent traffic-related information (e.g., traffic flow) in *limited spatial regions*. This, in turn, prohibits the full exploitation of the knowledge of different spatio-temporal correlations among traffic data patterns, thereby limiting the spectrum of control algorithms that can be used in a network context. Typically, ATM system receives data from various sources: roadside vehicle identification sensors, inductive loop-detectors below the road surface, drivers who voluntarily provide location-in-time information via on-board GPS equipment, vehicular ad-hoc network, video detectors over roads, etc. [60]. Such data is used for: (1) traffic control performed by adaptive traffic signals and phase timings [64], [65], [66]; and (2) traffic advisory, ranging from roadside message signs and radio, through pre-trip and en-route guidance and navigation (e.g., Navteq [67], Roadnet [68]).

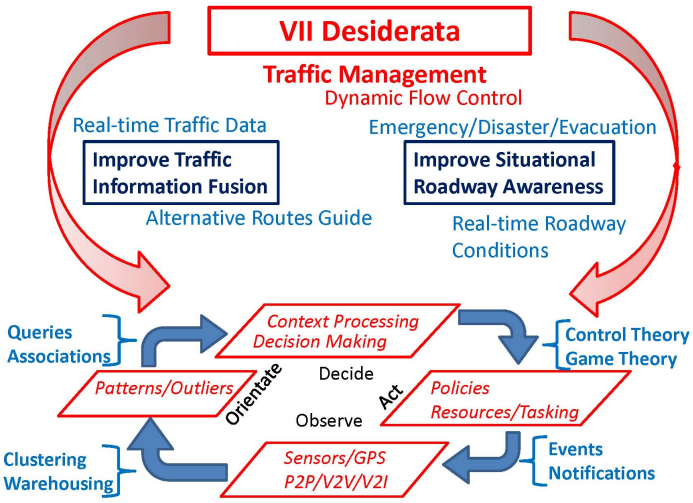

Fig. 3. Information Fusion and Decisions

A plethora of projects, prototypes and software tools exist (e.g., OPAC, RHODES, RTACLL, SIDRA [60], [65]) deployed for controlling traffic lights at intersections, generally aiming to optimize different traffic-related-metrics: average delay at signals, average trip-time, etc.. A generic diagram, illustrating the basic *predictive* paradigm for adaptive traffic management is illustrated in Figure 3, (cf. [60]). Under these settings the real-time information is used for prediction of the demands – e.g., duration of the green-light interval over a given horizon. Another paradigm used in traffic management is the *reactive one*, where the real time detection of pre-defined events triggers the execution of corresponding actions that implement a particular optimization policy. However, both paradigms are mostly *model-based*, without fully exploiting the wealth of historic information that can be fused at different levels of granularity and its use in distributed control algorithms.

Vehicle to Vehicle (V2V) and Vehicle to Infrastructure (V2I) paradigms [61], [62] view vehicles as a system that distributes the control, obtaining data by sensing the environment and communicating the individual parameters with the traffic-peers [69] in order to improve the operational efficiency in terms of fuel consumption and CO/HC emission [63]. Assistance can range from drivers-advisory to automatic control of vehicles (e.g., speed, distance from the other vehicles, etc). Several V2V/V2I control-frameworks and architectures have been proposed (e.g., PATH [70], Dolphin [71]), and actual projects have developed: – CarTALK, aiming at organizing the vehicles in an ad-hoc network for cooperative driver-assistance [72]; – SAFESPOT, aiming at improving safety and advanced detection of dangerous situations [73].

Despite the wide variety of centralized and distributed

algorithms for optimizing different parameters of relevance for efficient traffic management, e.g., such as signal control [74] and road pricing [75], there is a lack of intelligent triggering mechanisms that will enable balancing the trade-off between local vs. global control-decisions – and, most importantly for this work: their coupling with the data gathering process. As an observation – V2V/V2I paradigms, have advantages in terms of data-freshness and low communication latency, but they have drawbacks in limited spatio-temporal horizon of validity. For instance, reacting to an event notifying of a traffic congestion, may cause a re-routing action that, within short time may bring the vehicle in a region with yet another traffic congestion. The discussion of the issues related to the problem of efficient traffic management lead to the following questions:

Q3: *Could there exist some unified view of the mobile data and control/analytics in settings in which there are trade-offs among constraints at various levels of granularity and across different dimensions:*

- Quality assurances – e.g.:
  – yield of the desired optimization function such as average travel time.
  – horizon of validity for a selected control-model.
- Limited response time – e.g., in terms of changing the model used to control the traffic signals – within acceptable bounds from a desired level of quality assurances.
- Minimizing the use of other resources – e.g., bandwidth, energy, etc.

Q4: *Are there existing results in the field of MOD that could contribute towards the desiderata in* **Q3** *– and, if so, what are the directions to be taken?*

IV. The DNA^2 of MODA

We now discuss the main hypothesis of this paper, provide some guidance towards addressing the questions **Q1**–**Q4** and try to identify research avenues along those lines.

To better motivate the discussion, consider the scenario shown in Figure 4, illustrating the roadmap of Chicagoland. Assume that the objective is to minimize the number average number of "stop-and-go" patterns for the vehicles in the vicinity of the Midway airport. To achieve such objective, there may be several applicable control-models for traffic lights in the near-by streets, and the selection of a particular model – or, for that matter, tuning the parameters within a given model – is performed based on detection of certain patterns in the traffic data. Specifically, consider the following request:

Req1: *When the average speed of flocks containing at least 10 vehicles, between County Line Rd. exit and Pulaski Rd. along I-55 inbound has declined by $>$ 30%, if the average speed on the road-segments that have entry to I-55 has not decreased by $>$ 10%, then decrease the duration of the green-light interval on the nearest intersection to the ramp by 10%. Subsequently, when the density of vehicles on the streets within 2 miles along I-55 has increased by $>$ 10%, if the average speed within 1 mile from the loop has dropped by $>$ 40%, decrease by 15% the duration of the green-light interval on every intersection within 1 mile along I-55.*

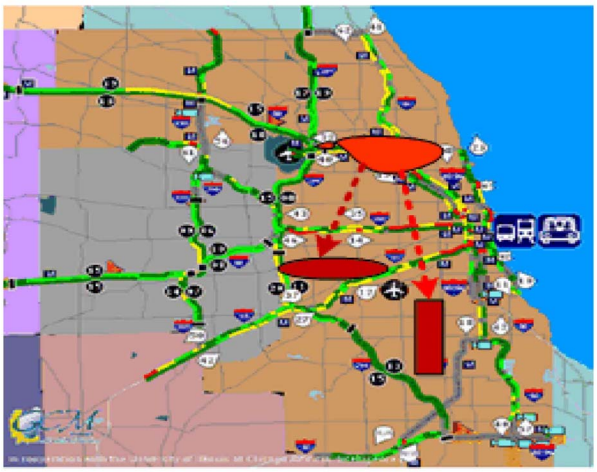

Fig. 4. Spatio-Temporal Correlations

Req1 is a verbalized example of applying the concept of *evolving triggers* – i.e., a trigger forking *children-triggers* based on the evolution of the environment [76] – to traffic management settings. Specifically, after detecting a particular event(i.e., average speed of flocks $\geq$ 10 vehicles...) and condition (i.e., average speed on the road-segments along I-55...): (1) Not only is the action executed "...decrease the duration of the green-light interval on the nearest intersection to the ramp by 10%", but also (2) another trigger is spawned. For the purpose of this section, there are a few important observations:

- Intuitively, one would subscribe the following statements to the *Action* part of a given trigger:
  – "...decrease the duration of the green-light interval on the nearest intersection to the ramp by 10%." and
  – "...decrease by 15% the duration of the green-light interval"
  However, in reality, the decision about "10%" and "15%" is likely to be generated as an output from some control module, based on the current state (and historic ones), for a given objective function(s).
- One could argue that the estimation of the "average speed of platoons..." should be part of a state-description and tasked to the control module. However, based on the discussion in Section II – we postulate that it may be more efficient to have them processed as spatio-temporal queries (over streaming data).
- It is likely that even the values of the parameters for the respective events and conditions in the triggers may be generated as outputs of the (observations of the) state of a given control module.

The above observations bring the following question which, in a sense, subsumes the ones identified so far:

Q5: *Which responsibilities should belong to data management*

and which ones belong to control/analytics in MODA?

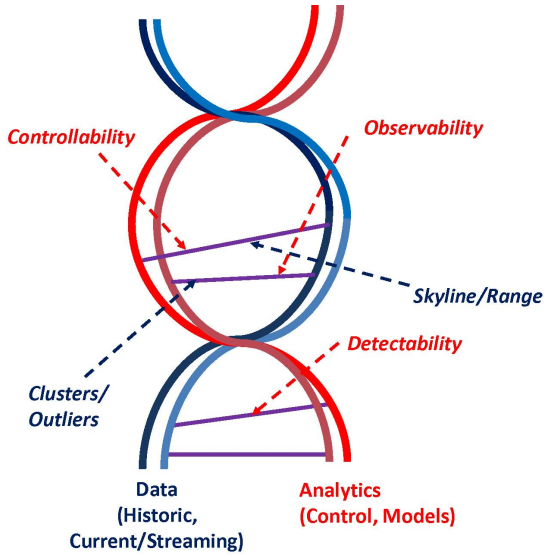

Fig. 5. The DNA^2 Hypothesis' Model

We believe that a possible answer to **Q5** is offered by, what we call, the DNA^2 hypothesis. As illustrated in Figure 5, the hypothesis postulates that for every activity in the control/analytics module, there exists some data-management activity that could improve its computational efficiency. Clearly, different applications will have certain context-dependent properties – to say the least, the technologies involved will be one determining factor. However, the hypothesis postulates that the impact of the "A" part on the roles and responsibilities of the "D" part – and vice-versa – should not be ignored. The main question is how to categorize such "mutual impacts"? Figure 5 illustrates this by showing a few different labels associated with the "D" part and the "A" part along the links between corresponding strains. One specific example that we address as "postulate" is related to the role of the uncertainty in the DNA^2. Namely, the extent of "foveation" (cf. **Q1**) in the collaboration between "D" and "A" parts will always involve some degree of uncertainty. Its unique nature in the context of the DNA^2 is addressed in more detail the sequel.

A. Uncertainty and Controllability/Observability

Consider the following (generic) equation[2]:

$$\overline{\mathbf{x}}(k+1) = \mathbf{A}\overline{\mathbf{x}}(k) + \mathbf{B}\overline{\mathbf{u}}(k)$$

Typically, it specifies the transition of a state of a discrete control system at time-instant $(k+1)$, based on the state at k and the impact of input-values at that time ($\overline{\mathbf{u}}(k)$). $\mathbf{A}$ is an $n \times n$ matrix, the coefficient of which may be determined based on some (e.g., optimality) desideratum, and $\mathbf{B}$ is an $n \times r$ matrix capturing the "weight" of different input values [77].

[2]While we are fully aware that the equation is overly-simplistic for traffic management scenarios, its purpose is solely to provide a basis for the subsequent observations/arguments.

Now, consider the question:

$\mathbf{Q}_c$: *Is the given system controllable*?

"Controllability" denotes a desirable property of traversing the entire configuration space of a given system or, in other words, ability of an external input to move the internal state of a system from any initial state to any other final state [78]. The mathematical tool for determining the controllability is to check whether the matrix $\mathcal{C} = [\mathbf{A}\,\mathbf{AB}\,\mathbf{A^2B}\ldots\,\mathbf{A^{n-1}B}]$ has a $rank(\mathcal{C}) = n$. However, the problem is that the computational complexity of determining the rank of a matrix is at least $\Omega(|\mathcal{C}|)$, where $|\mathcal{C}|$ is the number of non-zero elements in $\mathcal{C}$ [79]. Note that this could be a computationally expensive operation: imagine if every recorded data value (e.g., from every single road sensor or every camera) needs to be considered in the state-description. Hence, we have the following observation:

O1: *If one is to apply data reduction, what would be the impact on the correctness/error of the controllability test*?

In [80] we demonstrated that applying trajectories' simplification method in MOD settings provides a guaranteed bound on the error of the answers to some popular queries (e.g., range and nearest-neighbor). Subsequently, in [81] we correlated the bound with the one obtained in streaming-settings (i.e., the reduction is applied to *(location,time)* data as it arrives in the server. These are but small examples that indicate that it may be possible that in more general settings, applying data reduction techniques [82], [83] may yield benefits in the efficiency of the analytics/control modules execution.

The mathematical-dual of controllability is the *observability* – a measure for how well internal states of a system can be inferred by knowledge of its external outputs [77]. In other words, the system is said to be observable if it is possible to determine the behavior of the entire system based solely on its outputs. In the context of traffic management, this would entail that using road-side sensors, one would be able to judge whether a particular operating regime of the traffic signals satisfies the desired control law(s)[3]. Similarly to controllability, the test for observability again involves determining the rank of a corresponding (large-size) matrix. In a similar spirit to **O1**, we have the following observation:

$\mathbf{O1}_d$: *If one is to monitor "clusters" of trajectories instead of tracing individual object's trajectories, what would be the impact on the correctness/error of the observability test*?

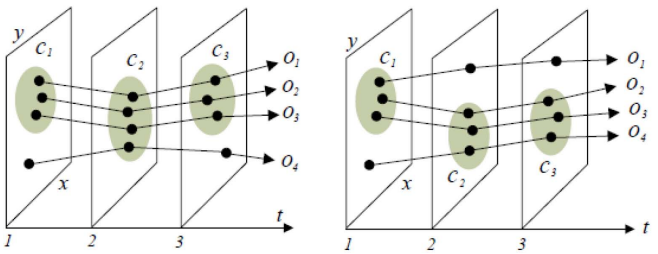

Fig. 6. Moving Clusters and Convoys

[3]Otherwise, the regime may need to be changed so that the overall traffic "behavior" is brought to another state.

As indicated in Section I, several recent works have addressed the problem of clustering trajectories [22], [23], [24], [26], [27] (see [25] for a recent survey). Each of them differs in the definition of the distance/proximity which, in turn, determine the boundaries of the respective clusters. As an example, Figure 6 illustrates two different clustering policies: moving clusters from [26] and convoys from [24]. Assuming that one could couple the distance and the density requirements in a manner that satisfies particular application requirements (cf. **Q3**), the argument that we make is in support of the observation $\mathbf{O1}_d$. Namely, the size of the observability matrix could be significantly reduced, thereby speeding up its rank computation. We note that there is a weaker notion of observability – namely, *detectability*. A system is considered detectable iff all its unobservable states are stable. Arguably, this may be a more appropriate concept to be used in the context of traffic management, bringing the inference of "semantic features" in the overall picture [84]. However, such investigation is beyond the scope of this paper.

B. The Flip-side of Uncertainty

A we mentioned in Section I, the impact of the location uncertainty – e.g., due to imprecision of the positioning technologies – on the various kinds of spatio-temporal queries has been subject of quite a few works over the past decade (see [21] for a survey). However, in the context of this paper, there is another perspective of the uncertainty and its role in spatio-temporal data management in MODA context. Namely, the sources of uncertainty will no longer be restricted to the imprecision of the positioning devices – they will also be dictated by the restrictions originating in the "A" part of a particular application. This, in some sense, provides an affirmative answer to **Q3**.

One "coupling" of heterogeneous uncertainties was presented in [85]: the work proposed a model to capture the impact of the *location uncertainty* on the *travel time uncertainty* in the context of probabilistic NN-queries on road networks. MODA context is natural ground for requirements which bind together temporal and/or quality constraints with spatio-temporal uncertain queries (cf. **Q3**), by far extending the link provided in [85]. This is especially true in applications like traffic management. Specifically, consider the settings of evaluating a particular control-model used for orchestrating the traffic signals. In order to determine whether the model is still valid, one is likely to rely on approaches that detect anomalies/outliers in the motion of the tracked entities in real time with respect to the historic data [86], [87]. However, applying data streams against an entire MOD for detecting such outliers is likely to be computationally prohibitive and incur unacceptable delays for the control/analytics modules. Clearly, some type of clusters of trajectories – thereby implying a coarser representation – will have to be used [29], [30]. Moreover, instead of checking for outliers using individual streams, some grouping may need to be performed "on the fly" [88], [89] and used against an existing OLAP engine. The extent of the uncertainty introduced will be dictated by the temporal-limitations or other quality criteria of the control/analytics module (cf. **Q3**). However, in the lieu of **Q2**, there is one more observation that we would like to discuss in this section. Consider distributed settings, where the spatial extents of governance of a particular server executing a control/analytics module may vary. While it is certainly desirable to focus local activities of MOD servers to cater to the geographically corresponding control/analytics modules, it may pay out to add an "ephemeral" uncertainty in the following sense:

- Let the geographical MOD servers additionally collaborate on selected collections of (moving) range or skyline queries.
- Use the answers to such queries within acceptable bounds of uncertainty to determine when to start detecting causality among spatio-temporal patterns across different locations [90].

The overhead of the collaboration among MOD servers for queries not directly associated with the control laws will introduce an extra degree of uncertainty – e.g., due to additional source of time-constraints. However, in a longer run, the answers to such may generate a prediction that will reduce the degree of uncertainty in terms of validity of a particular control law.

V. Instead of Conclusions

We believe to have touched – not quite even scratched – the surface of something that could generate a productive research in the alley of spatio-temporal data management – coupling it with the control/analytics modules in MODA settings. The intention of the DNA^2 hypothesis is to highlight both the potential and necessity of such couplings. We focused mostly on the aspect of the uncertainty and its "dual" roles in the context of the DNA^2 hypothesis. However, there are certain settings of MODA that may shift the balance of foveation along links between the strands in the DNA^2 in complementary spirits.

Some examples of such setting come from the domains of biomass monitoring and climate change studies [36], [91]. In addition to the peta-bytes of data describing complex evolving shapes, these domains offer extremely large collection of models of a given evolution – where the actual control aspect is an ongoing scientific challenge. As pointed in [36], the expected volume of climate data is expected to reach 350 Petabytes by 2030 (cf. [92]) – off of which half is contributed by the models, slightly less than a half by remote sensing data, and the rest from in-situ sensors. However, the in-situ sensors are likely to become denser in the future, providing even larger datasets.

A domain that is likely to generate unique perspectives towards identifying appropriate control laws, is the one of visual analytics [93], [94], [95], [96]. Visualization has always been a powerful tool to enable a productive communication between experts from different domains. It is our belief that visualizing the different impacts of control strategies on spatio-temporal phenomena is likely to yield criteria that can greatly

improve the selection of appropriate models to be employed in real-time settings.

Last, but not the least, we believe that a particularly interesting aspect of MODA is the issue of *declarative coupling* between the "D" part and the "A" part. Namely, in Section IV we mentioned the concept of evolving triggers [76] that has already been used to increase the energy-efficiency in terms of juggling the push vs. pull mode between different queries in WSNs. Projects such as TinyDB [97] have provided another demonstration that it is possible to have a high-level syntax that enables a collaboration between sensing (i.e., data management) and some basic control/actuation. Is it possible to build on such paradigm so that more complicated control-laws can become part of the high-level specifications? To what extent can an end user be provided with a degree of transparency regarding the details of the technologies involved – e.g., different activities in overlay networks [98]? Will there be possibilities to provide categorization of the different "strands" based on the capabilities to declarative specify their features and influence the foveation between the "D" and the "A"? We hope that some of these answers will be answered affirmatively in the near future.

ACKNOWLEDGMENT

Research supported in part by the National Science Foundation under grants CNS 0910952 and III 1213038.

Many thanks to Roberto Tamassia, Isabel Cruz and Alok Choudhary for the useful discussions and comments.

REFERENCES

[1] M. K. G. Institute, "Big data: The next frontier for innovation, competition, and productivity," 2011.

[2] E. Pitoura and G. Samaras, "Locating objects in mobile computing," *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, vol. 13, no. 4, 2001.

[3] J. Schiller and A. V. (editors), *Location-Based Services*. Morgan Kaufmann, 2004.

[4] R. H. Güting and M. Schneider, *Moving Objects Databases*. Morgan Kaufmann, 2005.

[5] C. Guo, M. Ma, B. Yang, C. S. Jensen, and M. Kaul, "Ecomark: Evaluating models of vehicular environmental impact," in *GIS*, 2012.

[6] S. Shang, R. Ding, B. Yuan, K. Xie, K. Zheng, and P. Kalnis, "User oriented trajectory search for trip recommendation," in *EDBT*, 2012.

[7] S. Lan, J.-P. Clarke, and C. Barnhart, "Planning for robust airline operations: Optimizing aircraft routings and flight departure times to minimize passenger disruptions," *Transportation Science*, vol. 40, no. 1, 2006.

[8] M. Christiansen, K. Fagerholt, and D. Ronen, "Ship routing and scheduling: Status and perspectives," vol. 38, no. 1, 2004.

[9] A. E. Rizzoli, N. Casagrande, A. V. Donatia, L. M. Gambardella, D. Lepori, R. Montemanni, P. Pina, and M. Zaffalon, "Planning and optimisation of vehicle routes for fuel oil distribution," in *International Congress on Modelling and Simulation (MODSIM)*, 2003.

[10] T. Leonard, "Delivering deeper insights with big data and real-time analytics," 2012.

[11] B. Franks, *Taming The Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics*. Wiley and SAS Business Series, 2012.

[12] R. H. Güting, T. Behr, and J. Xu, "Efficient *k*-nearest neighbor search on moving object trajectories," *VLDB Journal*, vol. 19, no. 5, pp. 687–714, 2010.

[13] K. Hose and A. Vlachou, "A survey of skyline processing in highly distributed environments," *VLDB J.*, vol. 21, no. 3, 2012.

[14] M. F. Mokbel and W. G. Aref, "SOLE: scalable on-line execution of continuous queries on spatio-temporal data streams," *VLDB Journal*, vol. 17, no. 5, pp. 971–995, 2008.

[15] Y. Zheng and X. Zhou, *Computing with Spatial Trajectories*. Springer, 2011.

[16] M. J. Atallah and Y. Qi, "Computing all skyline probabilities for uncertain data," in *PODS*, 2009, pp. 279–287.

[17] B. Kuijpers, R. Grimson, and W. Othman, "An analytic solution to the alibi query in the space-time prisms model for moving object data," *International Journal of Geographical Information Science*, vol. 25, no. 2, pp. 293–322, 2011.

[18] X. Lian and L. Chen, "Probabilistic ranked queries in uncertain databases," in *EDBT*, 2008, pp. 511–522.

[19] G. Trajcevski, A. Choudhary, O. Wolfson, Y. Li, and G. Li, "Uncertain range queries for necklaces," in *MDM*, 2010.

[20] Y. Tao, X. Xiao, and R. Cheng, "Range search on multidimensional uncertain data," *ACM Trans. Database Syst.*, vol. 32, no. 3, p. 15, 2007.

[21] G. Trajcevski, "Uncertainty in spatial trajectories," in *Computing with Spatial Trajectories*. Springer, 2011.

[22] Z. Fu, W. Hu, and T. Tan, "Similarity based vehicle trajectory clustering and anomaly detection," in *ICIP (2)*, 2005.

[23] J. Gudmundsson and M. J. van Kreveld, "Computing longest duration flocks in trajectory data," in *GIS*, 2006.

[24] H. Jeung, M. L. Yiu, X. Zhou, C. S. Jensen, and H. T. Shen, "Discovery of convoys in trajectory databases," *PVLDB*, vol. 1, no. 1, 2008.

[25] H. Jeung, M. L. Yiu, and C. S. Jensen, "Trajectory pattern mining," in *Computing with Spatial Trajectories*. Springer, 2011.

[26] P. Kalnis, N. Mamoulis, and S. Bakiras, "On discovering moving clusters in spatio-temporal data," in *SSTD*, 2005.

[27] J.-G. Lee, J. Han, X. Li, and H. Gonzalez, "*TraClass*: trajectory classification using hierarchical region-based and trajectory-based clustering," *PVLDB*, vol. 1, no. 1, 2008.

[28] F. Braz, S. Orlando, R. Orsini, A. Raffaetà, A. Roncato, and C. Silvestri, "Approximate aggregations in trajectory data warehouses," in *ICDE Workshops*, 2007, pp. 536–545.

[29] S. Campora, J. de Macedo, and L. Spinsanti, "St-toolkit: A framework for trajectory data warehousing," in *AGILE*, 2011.

[30] L. I. Gómez, S. A. Gómez, and A. A. Vaisman, "A generic data model and query language for spatiotemporal olap cube analysis," in *EDBT*, 2012.

[31] C. S. Jensen, T. B. Pedersen, and C. Thomsen, *Multidimensional Databases and Data Warehousing*. Morgan & Claypool, 2012.

[32] S. Orlando, R. Orsini, A. Raffaetà, A. Roncato, and C. Silvestri, "Spatio-temporal aggregations in trajectory data warehouses," in *DaWaK*, 2007, pp. 66–77.

[33] A. Akdogan, U. Demiryurek, F. B. Kashani, and C. Shahabi, "Voronoi-based geospatial query processing with mapreduce," in *CloudCom*, 2010.

[34] Q. Ma, B. Yang, W. Qian, and A. Zhou, "Query processing of massive trajectory data based on mapreduce," in *CloudDb*, 2009, pp. 9–16.

[35] R. Sugumaran, J. Burnett, and A. Blinkmann, "Big 3d spatial data processing using cloud computing environment," in *BIGSPATIAL Workshop (ACM GIS)*, 2012.

[36] R. R. Vatsavai, V. Chandola, S. Klasky, A. Ganguly, A. Stefanidis, and S. Shekhar, "Spatiltemporal data mining in the era of big spatial data: Algorithms and applications," in *BIGSPATIAL Workshop (ACM GIS)*, 2012.

[37] Y. Zhong, X. Zhu, and J. Fang, "Elastic and effective spatio-temporal query processing scheme on hadoop," in *BIGSPATIAL Workshop (ACM GIS)*, 2012.

[38] K. A. Akins, "A bat without qualities?" in *Consciousness: Psychological and Philosophical Essays*. Blackwell, 1993.

[39] M. P. Jones, K. E. Pierce, and D. Ward, "Avian vision: A review of form and function with special consideration to birds of prey," *Journal of Exotic Pet Medicine*, no. 2, 2007.

[40] E.-C. Chang and C.-K. Yap, "A wavelet approach to foveating images," in *Symposium on Computational Geometry*, 1997, pp. 397–399.

[41] S. Lee, A. C. Bovik, and Y. Y. Kim, "High quality, low delay foveated visual communications over mobile channels," *J. Visual Communication and Image Representation*, vol. 16, no. 2, pp. 180–211, 2005.

[42] Z. Wang, L. Lu, and A. C. Bovik, "Foveation scalable video coding with automatic fixation selection," *IEEE Transactions on Image Processing*, vol. 12, no. 2, pp. 243–254, 2003.

[43] J. A. Fuemmeler and V. V. Veeravalli, "Energy efficient multi-object tracking in sensor networks," *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 3742–3750, 2010.

[44] C. Gui and P. Mohapatra, "Power conservation and quality of surveillance in target tracking sensor networks," in *MOBICOM*, 2004, pp. 129–143.

[45] T. He, C. Huang, B. M. Blum, J. A. Stankovic, and T. F. Abdelzaher, "Range-free localization schemes for large scale sensor networks," in *MOBICOM*, 2003, pp. 81–95.

[46] H. Zhou, M. Taj, and A. Cavallaro, "Target detection and tracking with heterogeneous sensors," *IEEE Journal on Selected Topics in Signal Processing*, vol. 2, no. 4, pp. 503–513, 2008.

[47] W.-P. Chen, J. C. Hou, and L. Sha, "Dynamic clustering for acoustic target tracking in wireless sensor networks," *IEEE Trans. Mob. Comput.*, vol. 3, no. 3, pp. 258–271, 2004.

[48] S. Oh, S. Sastry, and L. Schenato, "A hierarchical multiple-target tracking algorithm for sensor networks," in *ICRA*, 2005.

[49] M. Rudafshani and S. Datta, "Localization in wireless sensor networks," in *IPSN*, 2007, pp. 51–60.

[50] Z. Zhong and T. He, "Achieving range-free localization beyond connectivity," in *SenSys*, 2009.

[51] O. Ghica, G. Trajcevski, F. Zhou, R. Tamassia, and P. Scheuermann, "Selecting tracking principals with epoch-awareness," in *Proceedings of the 18th ACM SIGSPATIAL GIS Conference*, 2010, pp. 222–231.

[52] Q. Ren, J. Li, and S. Cheng, "Target tracking under uncertainty in wireless sensor networks," in *MASS*, 2011, pp. 430–439.

[53] H. Wang, K. Yao, G. J. Pottie, and D. Estrin, "Entropy-based sensor selection heuristic for target localization," in *IPSN*, 2004, pp. 36–45.

[54] Z. Chen, "Bayesian filtering : From kalman filters to particle filters, and beyond," in *Statistics*, 2003, p. 169.

[55] F. Zhou, G. Trajcevski, O. Ghica, R. Tamassia, A. Khokhar, and P. Scheuermann, "Deflection aware tracking principal selection in active wireless sensor networks," *IEEE Trans. on Vehicular Technology*, vol. 61, no. 7, 2012.

[56] O. Wolfson, A. P. Sistla, S. Chamberlain, and Y. Yesha, "Updating and querying databases that track mobile units," *Distributed and Parallel Databases*, vol. 7, 1999.

[57] J. M. Sussman, "Intelligent vehicle highway systems: Challenge for the future," in *IEEE Micro*, 1993.

[58] M. Broucke and P. Varaya., "The automated highway system: A transportation technology for the 21st century," *Control Engineering Practice*, 1997.

[59] "US_DOT-VII," http://www.itsa.org/US_DOT_VII/c282/ITS_Resources/Library/US_DOT_VII.html.

[60] P. Mirchandani and F.-Y. Wang, "Rhodes to intelligent transportation systems," *IEEE Intelligent Systems*, 2005.

[61] L. D. Baskar, B. D. Schutter, and H. Hellendoorn, "Hierarchical traffic control and management with intelligent vehicles," in *IEEE Intelligent Vehicles Symposium*, 2007.

[62] R. Bishop, *Intelligent Vehicles Technology and Trends*. Artech House, 2005.

[63] A. Widodo, T. Hasegawa, and S. Tsugawa, "Vehicle fuel consumption and emission estimation in environment-adaptive driving with of without inter-vehicle communications," in *IEEE Intelligent Vehicles Symposium*, 2000.

[64] W.-C. Lin and C. Wang, "An enhanced 0-1 mixed-integer lp formulation for traffic signal control," in *IEEE Transactions on Intelligent Transportation Systems*, 2004.

[65] P. Mirchandani and K. Head, "A real-time traffic signal control system: Architecture, algorithms, and analyses," *Transportation Research Part C: Emerging technologies*, 2005.

[66] N. Zou and P. Mirchandani, "Analyses of vehicular delays and queues at intersections with adaptive and fixed timing control strategies," in *IEEE International Conference on Intelligent Transportation Systems*, 2005.

[67] "Navteq," http://www.navteq.com.

[68] "Roadnet," http://www.upslogistcstech.com/pub/products/Roadnet/.

[69] U. Lee, E. Magistretti, K. B. Zhou, M. Gerla, P. Bellavista, and A. Corradi, "Mobeyes: Smart mobs for urban monitoring with vehicular sensor networks," *IEEE Wireless Communications*, vol. 13, no. 5, 2006.

[70] R. Horowitz and P. Varaiya, "Control design of an automated highway system," *IEEE Special Issue on Hybrid Systems*, vol. 88, no. 7, pp. 913–925, Jul 2000.

[71] S. Tsugawa, S. Kato, T. Matsui, H. Naganawa, and H. Fujii, "An architecture for cooperative driving of automated vehicles," in *IEEE Symposium on Intelligent Transportation*, 2000.

[72] D. Reichard, M. Miglietta, L. Moretti, P. Morsink, and W. Schulz, "Cartalk2000 - safe and comfortable driving based upon inter-vehicle communication," in *IEEE Symposium on Intelligent Vehicles*, 2002.

[73] "Safespot," http://wwww.safespot-eu.org.

[74] E. Cipriani and G. Fusco, "Solution procedures for the global optimization of signal settings and trafic assignment combined problem," in *Proc. ICTTS*, 2008.

[75] A. Hayrapetyan, E. Tardos, and T. Wexler, "A network pricing game for selfish traffic," *Distributed Computing*, 2005.

[76] G. Trajcevski, P. Scheuermann, O. C. Ghica, A. Hinze, and A. Voisard, "Evolving triggers for dynamic environments," in *Extending Database Technology (EDBT)*, 2006.

[77] E. D. Sontag, *Mathematical Control Theory: Deterministic Finite Dimensional Systems*. Springer, 1998.

[78] J. Klamka, *Controllability of dynamical systems*. Kluwer Academic, 1991.

[79] H. Y. Cheung, T. C. Kwok, and L. C. Lau, "Fast matrix rank algorithms and applications," in *STOC*, 2012, pp. 549–562.

[80] H. Cao, O. Wolfson, and G. Trajcevski, "Spatio-temporal data reduction with deterministic error bounds," *VLDB Journal*, vol. 15, no. 3, 2006.

[81] G. Trajcevski, H. Cao, P. Scheuermann, O. Wolfson, and D. Vaccaro, "On-line data reduction and the quality of history in moving objects databases," in *MobiDE*, 2006, pp. 19–26.

[82] J. Jestes, K. Yi, and F. Li, "Building wavelet histograms on large data in mapreduce," *PVLDB*, vol. 5, no. 2, pp. 109–120, 2011.

[83] G. Cormode and M. N. Garofalakis, "Sketching probabilistic data streams," in *SIGMOD Conference*, 2007, pp. 281–292.

[84] P. Wang, H. Wang, and W. Wang, "Finding semantics in time series," in *SIGMOD Conference*, 2011, pp. 385–396.

[85] G. Trajcevski, R. Tamassia, I. Cruz, P. Scheuermann, D. Hartglass, and C. Zamierowski, "Ranking continuous nearest neighbors for uncertain trajectories," *VLDB J.*, vol. 20, no. 5, pp. 767–791, 2011.

[86] Y. Bu, L. Chen, A. W.-C. Fu, and D. Liu, "Efficient anomaly monitoring over moving object trajectory streams," in *KDD*, 2009, pp. 159–168.

[87] D. Pokrajac, A. Lazarevic, and L. J. Latecki, "Incremental local outlier detection for data streams," in *CIDM*, 2007, pp. 504–515.

[88] T. Guo, Z. Yan, and K. Aberer, "An adaptive approach for online segmentation of multi-dimensional mobile data," in *MobiDE*, 2012, pp. 7–14.

[89] Z. Yan, N. Giatrakos, V. Katsikaros, N. Pelekis, and Y. Theodoridis, "Setrastream: Semantic-aware trajectory construction over streaming movement data," in *SSTD*, 2011, pp. 367–385.

[90] W. Liu, Y. Zheng, S. Chawla, J. Yuan, and X. Xing, "Discovering spatio-temporal causal interactions in traffic data streams," in *KDD*, 2011, pp. 1010–1018.

[91] S. Shekhar, V. Gunturi, M. R. Evans, and K. Yang, "Spatial big-data challenges intersecting mobility and cloud computing," in *MobiDE*, 2012, pp. 1–6.

[92] J. T. Overpeck, G. A. Meeh, S. Bony, and D. R. Easterling, "Climate data challenges in the 21st century," *Science*, vol. 331, no. 6018, 2011.

[93] A. Raffaetà, L. Leonardi, G. Marketos, G. L. Andrienko, N. V. Andrienko, E. Frentzos, N. Giatrakos, S. Orlando, N. Pelekis, A. Roncato, and C. Silvestri, "Visual mobility analysis using t-warehouse," *IJDWM*, vol. 7, no. 1, pp. 1–23, 2011.

[94] N. V. Andrienko, G. L. Andrienko, H. Stange, T. Liebig, and D. Hecker, "Visual analytics for understanding spatial situations from episodic movement data," *KI*, vol. 26, no. 3, pp. 241–251, 2012.

[95] M. A. Sakr, G. L. Andrienko, T. Behr, N. V. Andrienko, R. H. Güting, and C. Hurter, "Exploring spatiotemporal patterns by integrating visual analytics with a moving objects database system," in *GIS*, 2011, pp. 505–508.

[96] G. L. Andrienko, N. V. Andrienko, and S. Wrobel, "Visual analytics tools for analysis of movement data," *SIGKDD Explorations*, vol. 9, no. 2, 2007.

[97] S. Madden, M. Franklin, J. Hellerstein, and W. Hong, "Tinydb: An acquisitional query processing system for sensor networks," *ACM TODS*, vol. 30, no. 1, 2005.

[98] S. Girdzijauskas, A. Datta, and K. Aberer, "Structured overlay for heterogeneous environments: Design and evaluation of oscar," *TAAS*, vol. 5, no. 1, 2010.