

Airplanes Aloft as a Sensor Network for Wind Forecasting

Ashish Kapoor[†], Zachary Horvitz[‡], Spencer Laube[‡] and Eric Horvitz[†]

[†]Microsoft Research, Redmond, WA, 98052 USA

[‡]Seattle Academy of Arts and Sciences, Seattle, WA 98122 USA

Abstract—We explore the feasibility of using commercial aircraft as sensors for observing weather phenomena at a continental scale. We focus specifically on the problem of wind forecasting and explore the use of machine learning and inference methods to harness air and ground speeds reported by aircraft at different locations and altitudes. We validate the learned predictive model with a field study where we release an instrumented high-altitude balloon and compare the predicted trajectory with the sensed winds. The experiments show the promise of using airplane in flight as a large-scale sensor network. Beyond making predictions, we explore the guidance of sensing with value-of-information analyses, where we consider uncertainties and needs of sets of routes and maximize information value in light of the costs of acquiring data from airplanes. The methods can be used to select ideal subsets of planes to serve as sensors and also to evaluate the value of requesting shifts in trajectories of flights for sensing.

I. INTRODUCTION

The efficiency of air transport is sensitive to winds encountered over the course of a flight. Pilots typically travel at a constant recommended airspeed for their craft based on engines, altitude, and weight [15]. Given a fixed airspeed, tail winds and headwinds increase and decrease the groundspeed of planes, respectively. Pilots, in collaboration with air traffic control, and in accordance with airline policies, establish and modify flight plans to avoid turbulence and to seek favorable winds. Typically, flight plans are filed at the outset of a flight and largely followed over the course of a trip.

To date, wind forecasts available for flight planning in the US have been limited to the use of the relatively coarse information provided by the *Winds Aloft* program, administered by the US National Oceanic and Atmospheric Administration (NOAA)[10]. The *Winds Aloft* forecasts for the continental US is informed by data collected from recurrent releases of weather balloons and radar. Linear interpolation is employed in the weather models to combine information from the available measurements.

We focus in this paper on the promise of leveraging airplanes as a large-scale and flexible sensor network for identifying wind speeds on a continental scale. We first discuss how we can access data from planes in flight to sense winds in regions of interest. Then, we present a model for combining reports from multiple aircraft to infer wind speeds. We show how we can update a learned model of the spatial correlation of winds with observations from individual planes. We also consider methods for guiding the collection of data with value of information (VOI). VOI can be used to identify subsets of planes that promise to provide the most valuable



Fig. 1. Visualization of data on all flying aircraft over the continental US. Such dense coverage enables real-time sensing and inference of weather phenomena on a large scale.

wind information at any moment in the context of current uncertainties. VOI can weigh the benefits of acquiring data from additional airplanes versus the costs of that access. Such guidance can be valuable when costs are associated with access and in situations where additional effort and communications can lead to more detailed reports, VOI also can be used to prioritize requests for modifications of flight plans when the value of sensing winds in regions that are not scheduled for coverage is identified to be high.

We first present background and key conceptual ideas. Then, we explore the promise of the approach with a field study using an instrumented high-altitude balloon. We compare the accuracy of predictions of the balloon’s trajectory using the NOAA model with the use of a finer-grained model that leverages data from airplanes in flight.

II. BACKGROUND AND RELATED WORK

A great deal of effort has been invested in understanding atmospheric phenomena, including efforts to construct planetary-scale wind models. We will not review the multitude of research efforts in this realm. Rather, we focus on the NOAA *Winds Aloft* service [10] in current use and the information that NOAA provides on winds, which we will refer to simply as the NOAA data.

Commercial airplanes cruise between altitudes of 23000 and 41000 feet. Winds at these heights can vary between 30 and 120 knots [6]. NOAA measures the wind for the *Winds Aloft* program at 176 weather stations over the US. The data is collected by lofting high-altitude weather balloons from

these sites every 6 hours. The balloons measure wind speeds and magnitudes at a set of altitudes from the mean sea level to 53000 feet. The set of measurements from adjacent wind stations, and other data on winds from radar observations, are combined via ad hoc rules, such as a weighted average, to provide estimates of the winds over wider regions [10], [15]. We introduce an alternate model for fusing multiple observations and compare the forecasts it provides with those available via the widely used approach in Section IV.

Pilot voice reports have been used in weather models for over 50 years. Over the last 20 years efforts have been undertaken to employ data from commercial aircraft for finer-grained modeling of weather phenomenon [3]. Proprietary data from commercial aircraft have been available to government agencies for weather prediction for nearly twenty years. The Aircraft Meteorological Data Relay (AMDAR) program [1] centers on making available meteorological data from aircraft. To date, much of the effort on AMDAR has centered on building out an infrastructure to collect and disseminate the data. For example, investments have been made to equip fleets of aircraft with special sensors versus using in-the-wild sensing, as we propose in this work. Efforts have also leveraged AMDAR data in qualitative analyses and in simple linear models [2].

We differentiate our work from prior efforts by focusing on combining information from publicly available flight tracking data that requires no special instrumentation and infrastructure. We also introduce a Gaussian Processes model for predicting winds from multiple measurements. We use the methodology to combine data sensed from a network of aircraft in a statistically efficient and mathematically rigorous framework. One of the key benefits of the proposed method is that it allows for better predictions by mining available data without requiring expensive hardware installations and programs that would fund such new hardware and communications infrastructure for sharing the data.

Gaussian processes (GP) [23] are powerful probabilistic modeling techniques that have been used with considerable success in machine learning. Per the focus here, GP have been used to model natural phenomenon, including spatial interdependencies [8], [12]. As an example, they have been applied in modeling wind energy and power forecasting [7], [16]. While this prior research is a direct application of GP, the work does not apply directly to our scenario. In our problem setting, rather than directly observe winds, we can only observe the aircraft ground speed. We need to extend the GP-based techniques to incorporate data that is auxiliary to the phenomenon being modeled.

Automated information acquisition with consideration of costs or of an overall sensing budget is an active area of research in the sensor network community [4], [19], [24]. Efforts center on computing the expected *value of information* (VOI) [13], [14] that can be realized when deploying sensors. Gaussian Processes have been shown to be particularly useful in prior work on VOI [20], [21]. Similar methods have been explored in *active* machine learning [11], [17], [18], [22], where the goal is to build predictive models under a budget. Our work extends the prior art by applying VOI to the scenario where data can be requested from commercial aircraft based on the current uncertainties of winds, the value of knowing more

about specific regions and the planned flight path of aircraft. The value of knowing more about different regions can be computed based on utilitarian considerations [19], taking into account the potential influences of inferences about winds on the plans employed by a larger community of airplanes.

III. AIRPLANES AS WEATHER SENSORS

We describe the construction of a predictive model for wind by coupling NOAA winds data with measurements derived from planes in flight. As mentioned, Winds Aloft provides forecasts of wind direction and speeds at different altitudes for specific locations. Data from airplanes are used to enhance NOAA predictions, which in turn already incorporate observations from balloons.

Formally, we represent the static wind station sites reported in the Winds Aloft report as $\mathbf{S}_L = \{\mathbf{s}_1, \dots, \mathbf{s}_{N_s}\}$, with forecast winds as observations $\mathbf{T}_L = \{\mathbf{t}_1, \dots, \mathbf{t}_{N_s}\}$. The Winds Aloft data provides historical forecasts for a predetermined set of geographical locations at many different altitudes, which we use as training data on winds. We note that the wind observations $\mathbf{t} = [t^x, t^y]$ are two-dimensional vectors, where the t^x and t^y denote projection of the wind vector on geographic 90° (towards east) and 360° (towards north) heading.

We also access data from airplanes in flight. Our goal is to incorporate publicly available flight data from all aircraft in flight. Specifically, we look at information available on public flight tracking websites and gather plane positions $\mathbf{S}_A = \{\mathbf{s}_1, \dots, \mathbf{s}_{N_a}\}$ and observed ground velocities $\mathbf{V}_A = \{\mathbf{v}_1, \dots, \mathbf{v}_{N_a}\}$. These observed velocities are highly informative about the latent winds at the aircraft locations. We shall use a two-dimensional vector representation for observations drawn from aircraft, $\mathbf{v} = [v^x, v^y]$, where the components are defined exactly as for the wind observations in the NOAA data, \mathbf{t} . We shall provide evidence via an experimental study that using the observations about wind derived from aircraft can significantly increase the predictive performance from the NOAA baseline.

Given the two sources of information about winds, we are interested in estimating wind observation \mathbf{t}_* for any test site \mathbf{s}_* . Taking a Bayesian perspective, we are most interested in the distribution, $p(\mathbf{t}_* | \mathbf{S}, \mathbf{T}_L, \mathbf{V}_A)$. Here we have combined all the sites $\mathbf{S} = \mathbf{S}_L \cup \mathbf{S}_A \cup \mathbf{s}_*$ for simplicity of presentation. We model this desired posterior distribution via a GP formulation as described below.

A. Gaussian Processes for Wind Modeling

GP is an appealing probabilistic approach to modeling natural spatial phenomena such as traffic, wind flows, and sea currents. The method provides a representation and efficient updating procedure for conditioning phenomena at one location based on observations at other locations. Gaussian Process regression (GPR) is closely related to kernel machines such as Support Vector Machines (SVM) [9] and has been well-explored in machine learning. One advantage of modeling phenomena with Gaussian Processes is that it provides probabilistic estimates of predictions, and thus is well-suited for the computation of the expected value of collecting additional information.

Figure 2 shows the graphical model corresponding to the proposed approach. The model shows the observed quantities

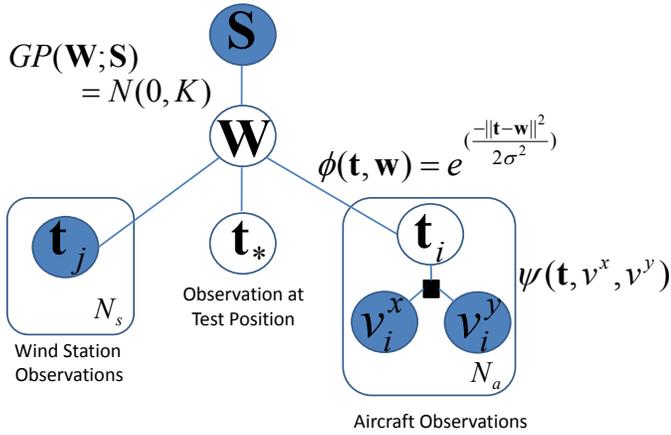


Fig. 2. A probabilistic graphical model for wind prediction, where shaded nodes represent observed variables. Physical sites are denoted by \mathbf{S} , which induce a Gaussian Process prior over the true wind velocities \mathbf{W} . The true velocity \mathbf{w} at each site is related to the observations \mathbf{t} via the potential $\phi(\mathbf{w}, \mathbf{t})$. The observations \mathbf{t} are only available for the wind stations (left box). The target observations \mathbf{t} for the aircraft are latent and only the ground speed components v^x and v^y corresponding to that aircraft are known, which in turn are related via the potential function $\psi(\mathbf{t}, v^x, v^y)$ (right box). The goal is to predict wind \mathbf{t}_* observed at the test site.

as shaded nodes, i.e., the collection of sites \mathbf{S} , observed winds at wind stations \mathbf{T}_L and the observed over the ground velocities $\mathbf{V}_A = [\mathbf{v}_i]_{i=1}^{N_A}$ for the aircraft. The rest of the unshaded nodes are latent variables. The boxes represent repetitions: N_s instances corresponding to the different wind stations and N_a instances for the separate aircraft.

The core idea behind using GP is to construct a probabilistic regression model for predicting wind velocity under constraints of spatial smoothness. That is, similar winds are observed at nearby or similar sites. Intuitively, the observed wind speed \mathbf{t} for any site s can be considered to be a noisy observation of a true wind velocity \mathbf{w} that is latent and arises due to a Gaussian process. The spatial smoothness represented by the prior is combined with observed data, either obtained at wind stations or aircraft observations, via likelihood models. Intuitively, when we infer the wind \mathbf{t}_* for the test site s_* , we probabilistically combine the smoothness constraint and the information obtained by observations \mathbf{T}_L and \mathbf{V}_A at the wind stations and the aircraft respectively. For the wind stations, the probabilistic relationship between the hidden true wind \mathbf{w} and the noisy observation \mathbf{t} is defined via a Gaussian likelihood model $\phi(\mathbf{t}, \mathbf{w})$.

Accessing data about winds from the airplanes in flight requires an extra step because the public reports from the planes provide only the ground velocities of the aircraft. We identify winds from ground velocity reports via the potential function $\psi(\mathbf{t}, v^x, v^y)$ that relates the ground velocity of aircraft to the encountered winds. In summary, the proposed model ties the site locations \mathbf{S} , the true wind \mathbf{W} , the noisy wind observations \mathbf{t} , and the aircraft ground velocities \mathbf{v} by inducing

the following distribution:

$$p(\mathbf{T}, \mathbf{V}_A, \mathbf{W} | \mathbf{S}) \propto GP(\mathbf{W}; \mathbf{S}) \times \prod_{i \in L \cup A \cup *} \phi(\mathbf{t}_i, \mathbf{w}_i) \prod_{j \in A} \psi(\mathbf{t}_j, v_j^x, v_j^y).$$

Here $\mathbf{T} = \mathbf{T}_L \cup \mathbf{T}_A \cup \mathbf{t}_*$. The potential ϕ applies to all the sites (wind station L , airplane A and the test site $*$), while ψ only effects the aircraft sites A . In the absence of airplane observations, the model simply reduces to a GPR [23] methodology. We describe all the three components below in detail:

Smoothness Constraints via the GP Prior: The smoothness constraint is imposed using a GP prior that defines the probabilistic relationship between the physical sites \mathbf{S} and the random variables \mathbf{W} . The GP prior $GP(\mathbf{W}; \mathbf{S})$ assigns higher probability to the wind observations that respect the similarity between the data points. Intuitively, the assumption is that similar data points should have the same wind; the similarity between two sites s_i and s_j is defined via a kernel $k(s_i, s_j)$. Probabilistic constraints are imposed on the collection \mathbf{W} of true wind vectors over all the sites. In particular, these random variables are assumed to be jointly Gaussian and the covariance between winds \mathbf{w}_i and \mathbf{w}_j is typically specified using a kernel function applied to s_i and s_j . Formally, $GP(\mathbf{W}; \mathbf{S}) \sim \mathcal{N}(0, \mathbf{K})$ where \mathbf{K} is a $(N_s + N_a + 1)$ -by- $(N_s + N_a + 1)$ kernel matrix with $K_{ij} = k(s_i, s_j)$, and $N_s + N_a + 1$ reflects the N_s wind sites, N_a airplane observations and one test site. We note that it is critical to have a good kernel function $k(\cdot, \cdot)$ that correctly represents underlying similarity for the prediction purposes. We learn this similarity function in this work via maximizing model evidence [23], which we discuss in detail in Section III-B.

Gaussian Noise Likelihood Model: The likelihood models the probabilistic relationships between the observed winds \mathbf{t} and the latent true winds \mathbf{w} . GPR assumes an additive Gaussian noise model. Consequently the relation between \mathbf{t} and \mathbf{w} is given by:

$$\phi(\mathbf{t}, \mathbf{w}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\|\mathbf{t}-\mathbf{w}\|^2}{2\sigma^2}}, \quad (1)$$

where σ is the noise model variance. We point out that, while there might be minor deviations in airspeed, for example due to turbulence, such deviations are easily handled by using the above noise model. Some airplanes may deviate from their filed air speeds. However, the majority of planes follow their flight plans. Robustness to small deviations in subsets of planes comes via sensing from large numbers of flying sensors. It is feasible to further enhance the proposed approach by incorporating such deviations explicitly via more expressive and complex noise models (e.g. by assuming different σ for each aircraft), which we defer to future work.

Incorporating Airplane Groundspeed: The flight tracking websites provide both the actual observed groundspeeds of planes and the intended airspeed. We seek to compute the winds that the aircraft is encountering, given both the ground-speed observation, $\mathbf{v} = (v^x, v^y)$, and the airspeed, as expressed in the flightplan. The airspeed is the fixed cruising velocity which the aircraft flies relative to the air mass it is flying through. When the encountered wind is zero, the magnitude of

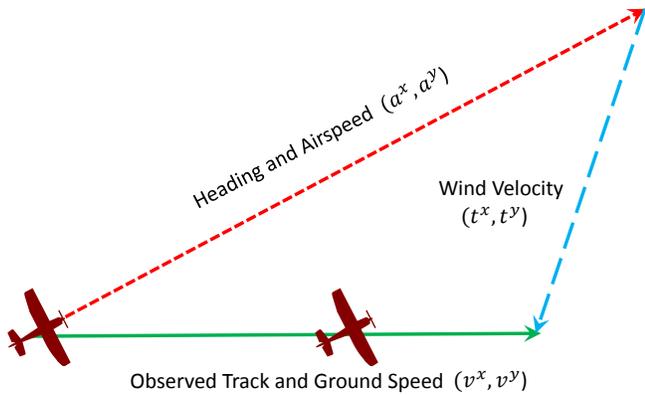


Fig. 3. Graphical representation of the wind triangle. The aircraft maintains an assigned heading at a reported constant and known airspeed (a^x, a^y) , while the observed wind (t^x, t^y) changes the observed course and the ground speed (v^x, v^y) via a vector addition. We use a Cartesian representation for velocities where the x and the y components denote projection of the vector on easterly and northerly headings respectively.

the ground speed ($\|\mathbf{v}\| = \sqrt{(v^x)^2 + (v^y)^2}$) should exactly match the airspeed.

Figure 3 shows how winds influence an airplane and the observed groundspeed. The aircraft points at a specific heading and flies with a constant airspeed with respect to the air mass. This is shown as the red dotted arrow, with two-dimensional vector (a^x, a^y) denoting the intended heading and the airspeed magnitude. The encountered wind $\mathbf{t} = (t^x, t^y)$ modifies the aircraft path resulting in the groundspeed velocity vector $\mathbf{v} = (v^x, v^y)$. While the flight plan does not give an intended heading, it does provide us with the filed airspeed. So we know the magnitude $\|\mathbf{a}\|$. We can show that the following relationship holds between the observed velocity \mathbf{v} and the winds encountered \mathbf{t} :

$$(v^x - t^x)^2 + (v^y - t^y)^2 = \|\mathbf{a}\|^2. \quad (2)$$

This equation arises from vector addition. We induce the following potential function that respects the above constraining property given the magnitude of the cruising airspeed $\|\mathbf{a}\|$:

$$\psi(\mathbf{t}, v^x, v^y) = e^{-\beta(\|\mathbf{v}-\mathbf{t}\|-\|\mathbf{a}\|)^2}. \quad (3)$$

This potential function assigns high likelihood to the wind assignments \mathbf{t} that respect the constraints in Equation (2). This likelihood falls exponentially as the winds begin to deviate from the constraints according to fall-off parameter β . We note that the potential ψ provides a natural way to incorporate information about the prevailing winds by considering the vector geometry of the natural process. However, a key challenge with this formulation is that the overall potential function corresponding to the graphical model has non-Gaussian terms making the inference non-trivial.

B. Learning and Inference

First, we describe the inference procedure to infer observed winds at the test site \mathbf{s}_* assuming that the kernel function $k(\cdot, \cdot)$ is known. Given the NOAA data \mathbf{T}_L and aircraft observations \mathbf{V}_A , our goal is to infer $p(\mathbf{t}_*|\mathbf{S}, \mathbf{T}_L, \mathbf{V}_A)$. Specifically, we

wish to compute:

$$p(\mathbf{t}_*|\mathbf{S}, \mathbf{T}_L, \mathbf{V}_A) \propto \int_{\mathbf{W}} p(\mathbf{t}_*|\mathbf{W})p(\mathbf{W}|\mathbf{S}, \mathbf{T}_L, \mathbf{V}_A). \quad (4)$$

The above decomposition is feasible due to the structure of the graphical model where, given \mathbf{W} , the test random variable \mathbf{t}_* is independent of everything else. We note that, in the above equation, the term $p(\mathbf{t}_*|\mathbf{W})$ is of a Gaussian form but the term $p(\mathbf{W}|\mathbf{S}, \mathbf{T}_L, \mathbf{V}_A)$ take a non-standard form as it comprises of Gaussian and non-Gaussian terms:

$$p(\mathbf{W}|\mathbf{S}, \mathbf{T}_L, \mathbf{V}_A) \propto GP(\mathbf{W}; \mathbf{S}) \prod_{i \in L} \phi(\mathbf{t}_i, \mathbf{w}_i) \times \prod_{j \in A} \int_{\mathbf{t}_j} \phi(\mathbf{t}_j, \mathbf{w}_j) \psi(\mathbf{t}_j, v_j^x, v_j^y).$$

This equation probabilistically combines the smoothness constraints $GP(\mathbf{W}; \mathbf{S})$ imposed via the GP prior and the information provided in the observations. Note that computing this posterior requires marginalization over unobserved \mathbf{t}_j at airplane locations—where only ground velocity \mathbf{v}_j of aircraft are available. The terms that correspond to aircraft observations make exact inference infeasible due to their non-Gaussian form. We propose to address this problem via the Laplace approximation [5]:

$$p(\mathbf{W}|\mathbf{S}, \mathbf{T}_L, \mathbf{V}_A) \approx N(\bar{\mathbf{W}}, \Sigma_{\mathbf{W}}).$$

Here, $\bar{\mathbf{W}}$ is the mode of the posterior $p(\mathbf{W}|\mathbf{S}, \mathbf{T}_L, \mathbf{V}_A)$ computed via gradient descent and $\Sigma_{\mathbf{W}}$ is the Hessian matrix evaluated at the mode. We believe that it is reasonable to use a Laplace approximation; the strong spatial regularity of winds in conjunction with tight dependence on the observed ground velocities encourages the discovery of unique solutions. Once the Laplace approximation is estimated, new predictions can easily be made by noting that the Gaussian noise model links \mathbf{t}_* to \mathbf{w}_* at any test site. Consequently, the predictive distribution over \mathbf{t}_* is also a Gaussian and given by: $p(\mathbf{t}_*|\mathbf{S}, \mathbf{T}_L, \mathbf{V}_A) \sim \mathcal{N}(\bar{\mathbf{w}}_*, \Sigma_{\mathbf{w}_*} + \sigma^2 I)$.

Learning the Kernel Function: The performance of GP-based prediction depends on the kernel selected to capture the similarity among sites. Learning in a GP framework is equivalent to choosing the correct kernel and is often performed by maximizing the marginal likelihood or the model evidence (see [23]). Evidence is a numerical quantity and signifies how well a model fits the given data. By comparing the evidence corresponding to the different models, we can choose the one most suitable for the task.

The kernel function is the backbone of the graphical model presented in this work. Intuitively, the kernel function $K_{ij} = k(\mathbf{s}_i, \mathbf{s}_j)$ should adequately capture the notion of similarity between the two sites. We propose to represent this kernel function via a linear combination of simple geography-based similarity functions. In particular we induce several radial basis (RBF) kernels [23] based on longitude, latitude, altitude, and time of observation at many different values of the scale parameter. Formally, we generate multiple kernels $\mathbf{K}^{(1)}, \dots, \mathbf{K}^{(k)}$ based on simple geographical variables and then seek a linear combination of these base kernels, such that

the resulting kernel \mathbf{K} has a good predictive power:

$$\mathbf{K} = \sum_{i=1}^k \alpha_i \mathbf{K}^{(i)}, \quad (5)$$

Here $\alpha = \{\alpha_1, \dots, \alpha_k\}$ are the weight parameters solved by optimizing the following objective function:

$$\begin{aligned} & \arg \min_{\alpha} -\log p_{\alpha}(\mathbf{T}_L|\mathbf{S}) \\ & \text{subject to: } \alpha_i \geq 0 \text{ for } i \in \{0, \dots, k\}. \end{aligned}$$

Note that, in the above equation, we only have used data \mathbf{T}_L collected at the wind stations. This is due to the fact that the log evidence $\log p_{\alpha}(\mathbf{T}_L|\mathbf{S})$ term in the above equation can be written as a closed form equation:

$$\begin{aligned} \log p_{\alpha}(\mathbf{T}_L|\mathbf{S}) = & -\frac{1}{2} \mathbf{T}_L^T (\sigma^2 \mathbf{I} + \mathbf{K}_{LL})^{-1} \mathbf{T}_L - \\ & \frac{1}{2} \log |\sigma^2 \mathbf{I} + \mathbf{K}_{LL}| - \text{Const}. \end{aligned}$$

The non-negativity constraints on α ensure that the resulting \mathbf{K} is positive-semidefinite and can be used in the GP formulation. The proposed objective is a non-linear program and can be solved using any gradient-descent based procedure. In our implementation, we use a gradient descent procedure based on the projected BFGS method, using a simple line search. Once the parameters α are found, then the resulting linear combination of kernels (\mathbf{K}) can be used for prediction.

C. Decisions about Observations

While the flight information for airborne aircraft is publicly available, there may still be a significant cost associated with querying an aircraft. For example commercial websites such as www.flightaware.com charge according to the number and the type of information that is requested. Flightplan information, together with a live update, is the most expensive type of query. We focus here on methods for making decisions about which plane observations to use, given the informational benefits and the costs of the information with an eye to maximizing utility.

We take a decision-theoretic perspective and employ value of information (VOI) to guide the gathering of information. Formally, we seek to select a set of site locations \mathbf{S}_Q , from the set of all of the sites \mathbf{S}_A where the aircraft are currently present, that provide maximum gain per unit cost:

$$\mathbf{S}_Q = \arg \max_{\mathbf{S}_Q \in \mathbf{S}_A} \frac{\text{Gain}(\mathbf{S}_Q)}{\text{Cost}(\mathbf{S}_Q)}. \quad (6)$$

Given the observations at the wind station, we can make inferences over true winds \mathbf{w} at sites of interest \mathbf{S}_I . These inferences also provide uncertainties in predictions that enables us to compute the expected value of information [13], [14]. We shall define $\text{Gain}(\mathbf{S}_Q)$ as reduction in uncertainty:

$$\text{Gain}(\mathbf{S}_Q) = H(\mathbf{S}_I - \mathbf{S}_Q|\mathbf{S}_L) - H(\mathbf{S}_I - \mathbf{S}_Q|\mathbf{S}_L \cup \mathbf{S}_Q). \quad (7)$$

Here $H(\cdot)$ denotes entropy over the true winds \mathbf{w} and the formulation attempts to characterize the information contained in the set \mathbf{S}_Q about the rest of the sites of interest (denoted as $\mathbf{S}_I - \mathbf{S}_Q$). Specifically, this equation captures how much reduction in entropy occurs when additional observations at sites \mathbf{S}_Q are added to the pool. At the additional sites \mathbf{S}_Q ,

rather than observing the true winds \mathbf{w} , we can only access the aircraft ground velocities \mathbf{v} . Nonetheless, in order to simplify the computation, we proceed to reason about the most informative set by assuming that we can observe true winds at these sites. This assumption is reasonable because the process noise variance is very small compared to the wind velocities, and consequently inferring winds via the aircraft velocities will be relatively accurate.

The cost component of the VOI function in Equation 6 is straightforward. We assume a simple additive model where the cost reflects the number of selected points. That is, $\text{Cost}(\mathbf{S}_Q) = |\mathbf{S}_Q|$. Again this is a reasonable cost model and aligned with many commercial services that charge based upon the number of queries issued.

It is known that determining \mathbf{S}_Q by solving Equation 6 is computationally intractable for a large sets. However, a greedy solution to this problem results in a tight approximation to the ideal solution in settings where the sub-modularity property holds [21]. Formally, we start with an initial set of active sites $\mathcal{A} = \mathbf{S}_L$, and use a greedy selection procedure to add new sites s that maximize the one-step reduction in uncertainty $\Delta H(s)$. This procedure continues until a budget is exhausted or until the incremental cost of an additional observation exceeds an estimate of the increase in value associated with the reduction in uncertainty. Formally, assuming that we directly observe the true winds at the selected sites, we can show that the information-theoretic gain can be written as [21]:

$$\Delta H(s) = \log \frac{\mathbf{K}_{ss} - \mathbf{K}_{s\mathcal{A}} \mathbf{K}_{\mathcal{A}\mathcal{A}}^{-1} \mathbf{K}_{\mathcal{A}s}}{\mathbf{K}_{ss} - \mathbf{K}_{s\bar{\mathcal{A}}} \mathbf{K}_{\bar{\mathcal{A}}\bar{\mathcal{A}}}^{-1} \mathbf{K}_{\bar{\mathcal{A}}s}}. \quad (8)$$

Here $\bar{\mathcal{A}}$ denotes $\mathbf{S}_I - \{\mathcal{A} \cup s\}$. Intuitively, the heuristic aims to select a data point that has the highest uncertainty given the sensed sites and is also the most central among the sites of interest \mathbf{S}_I .

We note that the work can be extended to several other utility-theoretic considerations of using airplanes aloft as sensors. In one direction of work, we can formulate a community-sensing model [19], where we consider the value of increasing the accuracies of wind forecasts at different locations and altitudes given the cumulative value of such data over a population of concurrent or forthcoming flights. For example, we can consider the workload of flightplans and compute the expected gains in value with having more accurate arrival times for multiple aircraft based in the reduction of error in these predictions for the set of craft under consideration. We can also consider in a utility function the opportunities and value of making interventions in flight plans of multiple aircraft in flight so as to reduce head winds or to gain access strong tailwinds for the larger community of aircraft. We note that the above analysis only considers static sites at which airplanes are present. An interesting future extension to this formulation includes the prospect of dynamically re-routing subsets of airplanes in order to make important observations for its own flight or as part of a larger community-sensing solution. The cost of a change in flightplan, per changes required in the trajectory and altitude of one or more aircraft may be balanced by gains in the increased net value to the overall system.

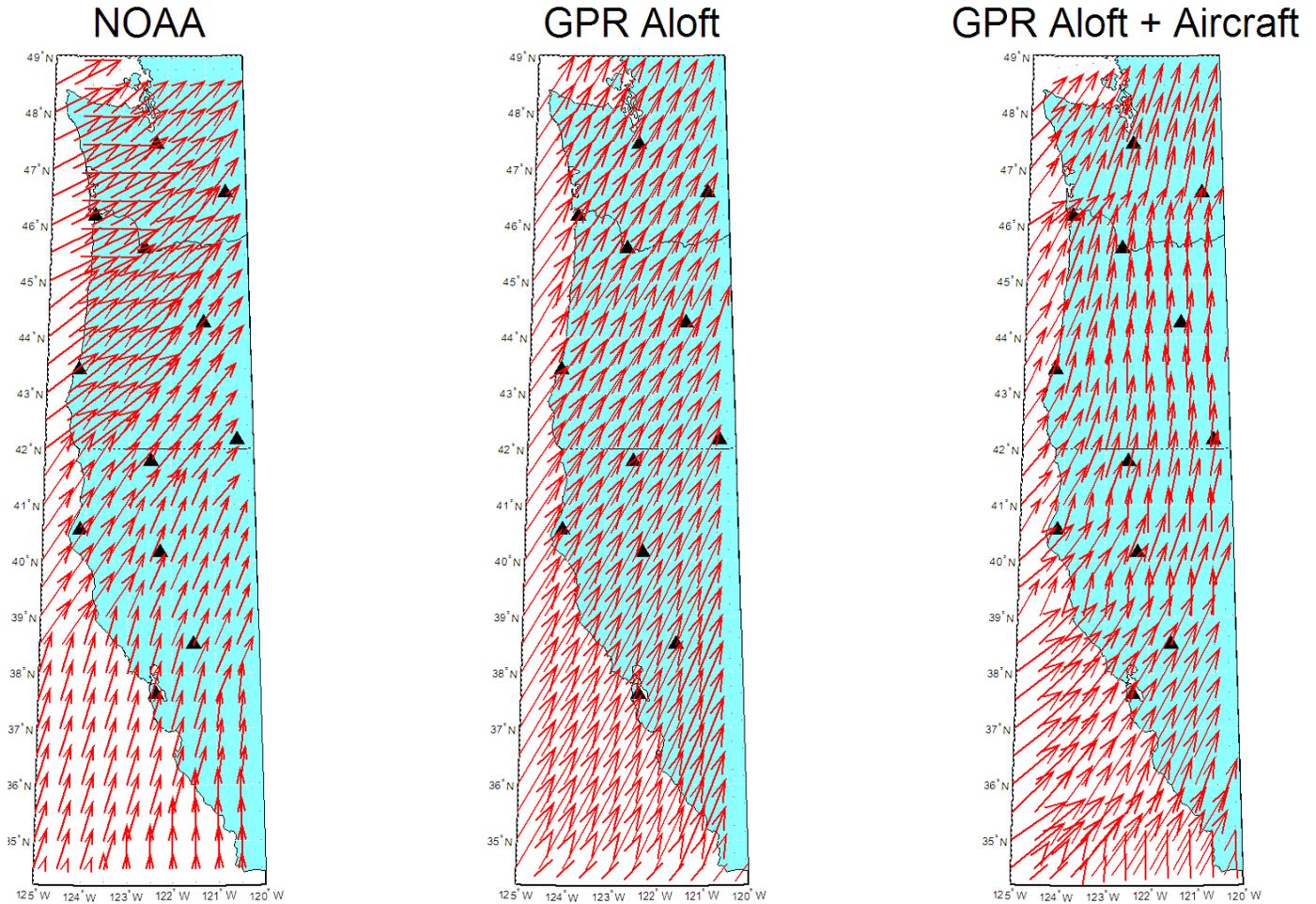


Fig. 4. Comparison of estimated winds at 30000 feet during the second balloon launch. The figure shows forecast made with the NOAA Winds Aloft data (left), Gaussian process estimation using wind stations only (middle), and the proposed model where both the aircraft data and the Winds Aloft data are used (right). Significant differences are seen in the predictions made with the Winds Aloft data.

IV. EXPERIMENTAL EVALUATION

We studied the promise of the proposed methodology via experiments centering on the launch and tracking of two helium-filled high-altitude balloons. We describe the apparatus, methodology, and results below.

A. Apparatus

The high-altitude balloons have the capability to take the payload over 100,000 feet. The payload consists of a GPS device, barometric and temperature sensors, and an onboard computer for recording time-encoded data from the sensors. The GPS device and service, purchased from Spot LLC, enabled tracking of the GPS status via satellite during the flight, which enabled us to retrieve the payload when it returned to the ground. The payload returned to the ground via a parachute which deployed automatically following balloon rupture at the balloon expansion limit of the balloon (with the lowering of atmospheric pressure), defining the maximum altitude reached. The real-time feed and higher-resolution onboard recording of flight data provided detail about the balloon trajectories and, thus, the true winds encountered during the flight. For each

balloon launch, the FAA was contacted and provided with the time of the launch.

The first balloon launch was a pilot study that was released on April 8, 2013 in the city of Orondo, Washington, USA at 47.72162N, -120.18508W. This balloon reached an altitude of 105,000 feet. The GPS track and the synchronized time information provided valuable data about the ascent profile, which in turn was used in the second experiment designed to verify the benefits of the proposed algorithm.

The second balloon was released with a similar payload on June 16, 2013 in Suncadia, Washington, USA at GPS coordinates 47.122348N and -121.010384W. This balloon reached an altitude of 95,000 feet.

The second launch provided a validation data set of the methods. We logged data from aircraft flying in the airspace over the states of Washington, Oregon, and California between 32° N and 49° N latitude and 114° W and 125° W longitude. The aircraft data were accessed via queries to www.flightaware.com and contained information about the over-the-ground velocity as well as the airspeed filed in flight plans for each airplane. The data was queried starting one

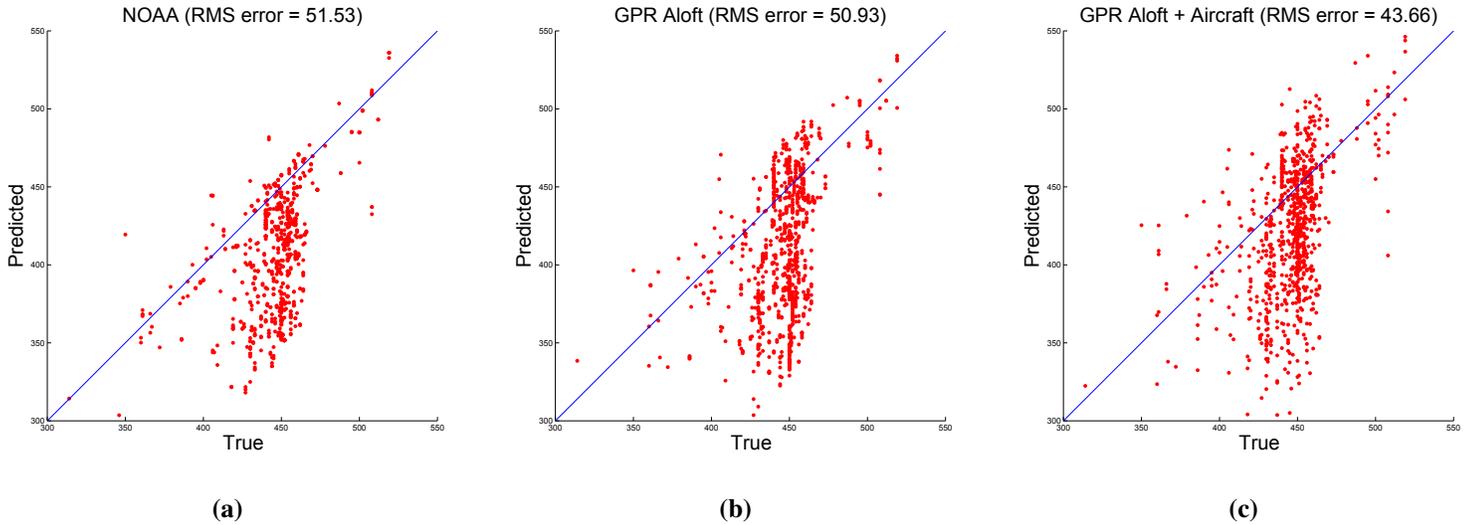


Fig. 5. Performance comparison in predicting ground speed of aircraft aloft using the leave-one-out evaluation strategy. The figure demonstrates scatter plots between the true ground speed and the ones estimated by (a) NOAA Winds Aloft forecast, (b) Gaussian process estimation using wind stations only and (c) the proposed model where both the aircraft data and the Winds Aloft data are used. Each point is (true, predicted) and corresponds to one aircraft aloft. The RMS error obtained by using the proposed model is significantly better than both the NOAA forecasts and the GP model based on only the wind stations data.

hour prior to the launch at a frequency of every five minutes, and this data logging continued until the end of the balloon flight. The wind model was constructed by using the data recorded only up to the time of the balloon launch. The aircraft data consisted of 1653 observations from 496 unique aircraft. Finally, a filtering step was performed that removed all the aircraft observations where the magnitude of difference between the filed airspeed and the observed groundspeed was either exactly zero or differed by more than 100 knots. This filtering step is necessary as often in flight reports incomplete data is automatically imputed and does not represent the true groundspeed.

B. Computational Details and Baselines

We explore two additional baselines beyond the proposed approach. In summary, the following three approaches are considered:

- **NOAA:** The first baseline is based on the NOAA Winds Aloft forecast. In order to predict the winds at a location (latitude, longitude, altitude), we simply consider the geographically nearest Winds Aloft station and linearly interpolate between the reported altitudes in the reports. Note that this is the methodology that is currently employed in most of the automated flight-planning software [15].
- **GP aloft:** The second baseline goes beyond the simple scheme using NOAA estimates by using Gaussian processes to build a predictive model. However, this predictive model, only considers the Winds Aloft data to build the wind predictions. The model is richer than NOAA as it allows for non-linear interpolation across longitude, latitude and altitude.
- **GP aloft + aircraft:** This is the scheme proposed in the paper that uses the aircraft data in conjunction

with the Winds Aloft forecast via the Gaussian process model.

The noise parameter σ^2 of the Gaussian noise model ϕ and the final kernel for both the GP based approaches were learnt via evidence maximization as described earlier. The base-level kernels consisted of individual RBF kernels over latitude, longitude, and altitude using eight different kernel widths exponentially ranging from 10^{-5} to 10^5 (i.e. $3 \times 11 = 33$ different kernels). The parameter β was simply set to 1.

C. Results

1) *Estimating Winds Aloft:* Figure 4 graphically shows wind predictions at 30,000 feet using the proposed approach and the two baselines. We observe that the estimated winds by the proposed method differs significantly from the NOAA generated forecasts. Specifically, the NOAA generated forecast shows more easterly winds when compared to the GP aloft + aircraft. We also observe that the GP aloft forecast is a smoother version of NOAA, which is not surprising considering the fact that GP aloft is a more refined interpolation scheme and does not employ aircraft data.

We can directly observe qualitative differences in the forecasts. However, we seek to evaluate correctness of the methods empirically. The evaluation is carried out by attempting to predict the true ground speed of the aircraft aloft. Specifically, we attempt to predict ground speeds of the flying aircraft given wind estimates from each of the wind models. Since the actual intended heading for the aircraft are unknown, we first reason about the actual heading of the aircraft by estimating the winds via the proposed methodology and then applying the vector operation $\mathbf{a} = \mathbf{v} - \mathbf{t}$. Next, we assume these \mathbf{a} to be given and then attempt to estimate the ground speeds via different wind models. For GP aloft + aircraft we use a leave-one-aircraft-out methodology for evaluation, where the predictive model is built on all but one test aircraft, and the procedure is repeated by considering each airplanes as a test case.

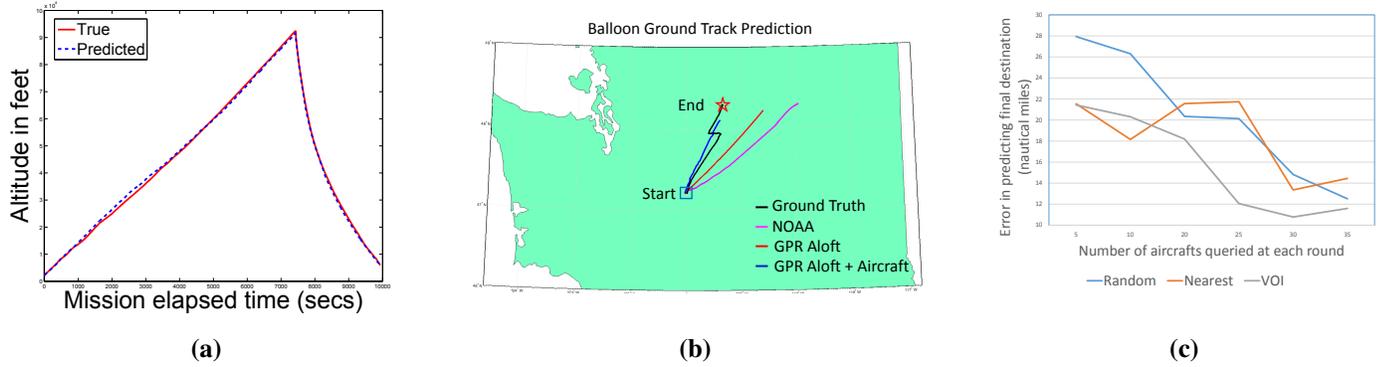


Fig. 6. (a) This figure demonstrates the accuracy in determining the ascent profile of the second balloon. (b) The path predicted using the three wind estimation models. The proposed model (blue) estimates the path far more accurately than the NOAA forecasts (magenta) and the GP aloft model (red) based only on the wind stations data. (c) Performance comparison of the VOI-based querying of aircraft data with random sampling and the nearest heuristic. The figure shows that significant accuracy in path prediction can be obtained for a fraction of cost if the decision-theoretic method proposed in the paper is used.

Figure 5 shows the scatter plots, where the reported ground-speed (x -axis) is plotted against the predicted values (y -axis). Good predictions lie close to the diagonal, reflecting high correlation between true and predicted, whereas points above and below the diagonal reflect over and under estimations respectively. We observe that the GP aloft + aircraft provides better root mean squared (RMS) error (43.66), than both the NOAA model (51.53) and GP aloft (50.93) baselines. Also, we note that the NOAA baseline consistently underestimates the ground speed and highlights the issues in estimation based in data scarcity. The results indicate the promise of using the proposed approach to estimated winds.

2) *Path Prediction:* We further verified the accuracy of the estimated winds by predicting the path of a high-altitude balloon. The balloon is carried by the wind, and accurate prediction of the path highly depends upon on the accuracy of the wind forecasts. As the winds change with altitude, it is important to model the ascent profile of the balloon, i.e. the vertical velocity of the balloon at any specific altitude. We obtain training data to model this ascent profile using an initial balloon flight. Specifically, from the training data we learn a simple regression function that predicts the vertical velocity of a balloon given its current altitude, ambient temperature and the atmospheric pressure. Figure 6(a) shows how well the vertical ascent profile is predicted by the model. The red line shows the true ascent profile for the second flight, while the blue line shows the predictions. We can see that the learned model is fairly accurate in predicting the ascent profile. Thus, we believe we can use the ascent information from the initial balloon launch to predict the ascent, as well as the trajectory of the second launch, if the wind estimations are accurate.

We combine the ascent profile with the wind predictions in order to predict the path of the second high-altitude balloon. Figure 6(b) shows the true (black) and the three predicted paths over the ground. The GP aloft+aircraft (blue) provides the best prediction, where the final recovery point was only predicted to be 11.59 miles away from the ground truth. The other baselines (NOAA in magenta, GP aloft in red), however, miss the final destination by 56.08 and 30.24 miles respectively. Note that these baselines show a more easterly path for the balloon, in line with the wind predictions shown in figure 4(a) and (b). The proposed model, however, corrects the error

associated with the use of the NOAA winds aloft report by effectively taking into account the data available from aircraft aloft. We note that there appears to be a sudden course change of the balloon around the time the balloon was expected to rupture. Our hypothesis is that this sudden shift is based in the aerodynamic effects of the rupture that can propel the balloon in unexpected ways. Beyond this sudden change, the ground trajectory is predicted well by the proposed approach.

3) *Value-of-Information Studies:* We next evaluate the scheme to optimize querying airplanes under budget constraints. In our experiment above, we assumed the ability to query data from all the aircraft flying above WA, OR, and CA every five minutes. We investigate if it is possible to achieve similar performance when we limit the number of airplanes that can be queried every five minutes.

We compared the proposed VOI-based strategy, with two different strategies: (a) random sampling and (b) nearest airplanes to the area of interest. While the random sampling strategy simply picks up k airplanes at random, the nearest strategy picks k nearest aircraft to the launch location. Also, for the VOI based strategy, the set S_I of sites of interest is set to a 5×5 grid of points centered around initial launch location, where each step in the grid is either 0.01° change in latitude or longitude. Further, these grids are replicated for every 10000 feet of altitude up to 100,000 feet, resulting in total $25 \times 10 = 250$ points in S_I . Consequently, the goal of the VOI based method is to query the airplanes that provide maximal information with respect to winds at these points around the launch coordinates.

Figure 6(c) shows the error in estimating the coordinates of the final destination as the query budget is varied. We observe that the VOI-based strategy provides the most efficient sampling and outperforms both the random and nearest strategies. The VOI-based method allows better statistical efficiency than random, as the later scheme is completely ignorant of the sites of interest. The nearest neighbor strategy, while informed about sites of interest, still performs poorly due to the fact that it can often choose redundant aircraft (e.g. two aircraft very close to each other that are in the vicinity of the launch point). The VOI-based approach, however, chooses sites that are not only informative about the sites of interest but also non-redundant.

V. CONCLUSIONS

We presented a methodology for harnessing commercial aircraft as a large-scale mesh of sensors for forecasting winds on a continental scale. The proposed methodology leverages publicly available flight information without requiring special equipment on aircraft. We developed a probabilistic graphical model that combines data about the groundspeed and airspeed of planes at different locations and altitudes with publicly available NOAA Winds Aloft forecasts. We demonstrated the feasibility of the approach via real-world weather balloon experiments, where we released an instrumented high-altitude balloon and compared the predicted trajectory with the actual trajectory, based on sensed winds. Our results highlight the promise of the proposed Gaussian Process prediction methodology over the use of currently available NOAA wind forecasts. We described methods for harnessing the probabilistic model to provide guidance on sensor access and positioning via the computation of the expected value of information. We presented an approach to selecting ideal subsets of planes to serve as sensors under a budget constraint and showed that the optimized selection of observations can significantly reduce the number of queries needed to achieve a specified level of accuracy. The method can be applied with an eye to providing value to a population of participating airplanes via community sensing solutions that consider uncertainties in winds and needs over sets of routes. With such an approach, value of information can be used to guide sensing and routing in accordance with maximizing information value for multiple flights, in light of the costs of sensing. We believe that the proposed methodology and potential extensions are a promising approach to modeling a wide spectrum of weather phenomena via the statistically efficient integration of streams of data from aircraft aloft. We are excited about the potential to incorporate these ideas into aviation. Uses extend from optimizing flight plans for single planes to larger-scale community sensing services and optimization for the larger air transport industry. Such community sensing could guide the formulation and dynamic updating of flight plans that minimize flying times and fuel usage while providing updates about weather phenomenon of interest.

ACKNOWLEDGMENTS

We thank Michael Laube for his insights about commercial air transport and support of the high-altitude balloon studies.

REFERENCES

- [1] <http://amdar.noaa.gov/docs/bams/>.
- [2] <http://amdar.noaa.gov/docs/mamrosh-ams-98/>.
- [3] <http://amdar.noaa.org>.
- [4] Z. Abrams, A. Goel, and S. Plotkin. Set k-cover algorithms for energy efficient monitoring in wireless sensor networks. In *IPSN*, 2004.
- [5] A. Azevedo-Filho and R. Shachter. Laplace's method approximations for probabilistic inference in belief networks with continuous variables. In *Uncertainty in Artificial Intelligence*, 1994.
- [6] L. Boccia, P. Pace, G. Amendola, and G. D. Massa. Low multipath antennas for GNSS-based attitude determination systems applied to high-altitude platforms. In *GPS Solutions*, 2008.
- [7] N. Chen, Z. Qian, X. Meng, and I. T. Nabney. Short-term wind power forecasting using gaussian processes. In *IJCAI*, 2013.
- [8] N. A. C. Cressie. *Statistics for Spatial Data (revised edition)*. Wiley, 1993.
- [9] T. Evgeniou, M. Pontil, and T. Poggio. Regularization networks and support vector machines. *Advances in Computational Mathematics*, 13(1), 2000.
- [10] FAA. *Aviation Weather Services, Advisory Circular AC 00-45G Change 1*. Aviation Supplies and Academics, 2006.
- [11] Y. Freund, H. S. Seung, E. Shamir, and N. Tishby. Selective sampling using the query by committee algorithm. *Machine Learning*, 28(2-3), 1997.
- [12] R. Guhaniyogi, A. O. Finley, S. Banerjee, and A. E. Gelfand. Adaptive gaussian predictive process models for large spatial datasets. *Environmetrics*, 22(8), 2011.
- [13] E. Horvitz, J. Breese, and M. Henrion. Decision theory in expert systems and artificial intelligence. *International Journal of Approximate Reasoning*, 1988.
- [14] R. A. Howard. Information value theory. *IEEE Transactions on Systems Science and Cybernetics*, 1966.
- [15] Jeppsen. *Private Pilot Manual*. Sanderson, 2001.
- [16] X. Jiang, B. Dong, L. Xie, and L. Sweeney. Adaptive gaussian process for short-term wind speed forecasting. In *Proceedings of the 2010 conference on ECAI 2010: 19th European Conference on Artificial Intelligence*, 2010.
- [17] A. Kapoor, K. Grauman, R. Urtasun, and T. Darrell. Active learning with Gaussian Processes for object categorization. In *ICCV*, 2007.
- [18] A. Kapoor, E. Horvitz, and S. Basu. Selective supervision: Guiding supervised learning with decision-theoretic active learning. In *IJCAI*, 2007.
- [19] A. Krause, E. Horvitz, A. Kansal, and F. Zhao. Towards community sensing. In *IPSN*, 2008.
- [20] A. Krause, J. Leskovec, C. Guestrin, J. VanBriesen, and C. Faloutsos. Efficient sensor placement optimization for securing large water distribution networks. *Journal of Water Resources Planning and Management*, 136(6), 2008.
- [21] A. Krause, A. Singh, and C. Guestrin. Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, 9, 2008.
- [22] D. MacKay. Information-based objective functions for active data selection. *Neural Computation*, 4(4), 1992.
- [23] C. E. Rasmussen and C. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.
- [24] F. Zhao, J. Shin, and J. Reich. Information-driven dynamic sensor collaboration for tracking applications. *IEEE Signal Processing*, 19(2), 2002.