Exploration vs. Exploitation in Airborne Wind Energy Systems via Information-Directed Sampling Control

Guillaume Goujard[1], Patrick Keyantuo[1], Mathilde Badoual[1], Scott J. Moura[1]

Abstract— **Airborne Wind Energy systems (AWEs) are an emerging wind generation technology. They differ from conventional turbines in that they are attached to the ground by a tether and can evolve from low to high altitudes (approx. 1km). Informed altitude control of AWEs is key to track favorable wind speed and maximize power output in a time-varying and partially-observable environment. Leveraging recent advances in Multi-Armed Bandit problems, we recursively estimate the wind profile distribution and use the residuals to fit the noise covariance in an online fashion. This filtering approach paves the way for the computation of (i) the distribution of the wind-output given past observations and (ii) the expected reduction in entropy in the optimum distribution with respect to the potential future altitude set-point. We implement an *Information Directed Sampling* controller that minimizes the ratio of squared-regret per bit of information gained about the optimum. We finally compare our controller with different baseline controllers using real-world data.**

I. Introduction

A. Background

Decarbonizing the energy supply relies on the emergence and development of carbon-free generation sources that are competitive in terms of both cost and performance. While wind energy is already one of the most prevalent renewable resources in the United States [1], conventional turbines suffer from low capacity factor due to variability in wind speed. Airborne wind energy systems (AWEs) differ from conventional wind turbines by employing both a lifting body (kite, rigid wing, or aerostat) and adjustable length tethers, offering intriguing advantages. AWEs can harvest wind at higher altitudes [2] and across a range of operating altitudes. At a first approximation, wind speed increases monotonically with height [3] [4]. It follows that, AWEs can achieve a higher capacity factor compared with conventional stationary turbines by operating at higher altitudes, or by making adjustments to track favorable wind speeds across a wide range of altitudes. Higher altitude wind patterns - besides being stronger - are generally more temporally consistent and less turbulent [5] [6].

B. Literature Review

Past research on AWE altitude control focused on methods that find and stabilize at an optimal operating altitude [7] [8]. The notion of optimality is, however, difficult to characterize. No meteorological model exists to accurately forecast the whole wind shear profile at high altitudes for the short-term time horizons most relevant to AWE controls.

Characterizing the distribution of the maximum wind energy output is challenging. The wind profile is relatively high dimensional, and at each time step the observation is partial. The reason for this is that instrumentation for monitoring wind speeds is co-situated with the turbine itself; thus the wind speed is only measured at the current operating altitude. The challenge of estimating a distribution of hidden variables can be found in many different scientific fields. A traditional way to address this issue is to use model reduction, filter theory or Gaussian processes. Application of those methods are especially common in meteorology [9] [10] [11]. We particularly recommend the readings of both [12]–where an extended Kalman Filter is used to estimate the wind speed distribution and fit the model parameters – and [13] – where a Kalman Filter estimates states corresponding to a projection of wind speed on carefully designed basis functions. These fundamental ideas of low-dimensional state estimation of a spatio-temporal system under sparse observations will be further illustrated in this work.

AWE altitude control differs from these works in that there is a causal relationship between action and observation. Moving the AWEs to a specific altitude harvests wind power at the new location, contributing directly to the objective function (exploitation) that we seek to maximize. Yet, control actions also elicit information about the state estimate (exploration). This is the dual role of control as described by P. R. Kumar and P. Varaiya [14]. Assuming a structure in the wind speed evolution, our problem can be framed as a stochastic *restless* multi-armed bandit where the states (wind speeds) of all arms (altitudes) can change at each step according to a known stochastic transition function. Restless bandits are notoriously intractable. To overcome these challenges, we leverage the problem's specific structure and use Information-Directed Sampling, which is known to perform well in these instances [15].

Recent efforts have recognized the stochastic nature of the spatio-temporal evolution of the wind profile. Bin-Karim et al. developed a model predictive controller (MPC) [16] and Bafandeh et al. [17] used a Lyapunov-based extremum seeking, using surrogate power deficit as the objective. Both approaches rely on the statistical accuracy of the underlying forecasting model trained offline under full observability (assuming knowledge of the whole profile), but operating under sparse observation. As reckoned by Baheri et al. in [18] and [19], the statistical properties of the wind shear should be learned online to avoid a long and costly period

[1] Guillaume Goujard, Patrick Keyantuo, Mathilde Badoual and Scott J. Moura are with Department of Civil and Environmental Engineering, University of California, Berkeley, California, 94720, USA. {goujard,pkeyantuo,smoura}@berkeley.edu

of data collection. To address this shortcoming, they designed Gaussian Process forecasting models and used Bayesian Optimization to deal with the trade-off between exploitation and exploration. The first potential limit to this method, as Dunn et al. noted in [20], is that the power production function is nonlinear in the wind-speed. Hence, there is little reason that the power output would be normally distributed. Another unaddressed problem is that the power output is not simply a function of the wind speed, but also depends on the altitude adjustment. Third, the expected improvement is a greedy controller which focuses its sampling effort near the estimated optimum. It has been proven to perform poorly in a best-arm identification problem [21].

In previous work [22], we assumed that the wind profile follows a vector auto-regressive process. This approach highlights that the forecast model is difficult to fit in practice, as one needs to estimate the past lagged profiles from sparse observations to forecast future profiles. Finally, Dunn et al. examined the impact of sensor configurations on power output for different controllers. This work revealed that partial observability can significantly degrade performance, due to poor forecasting accuracy [20] [23].

C. Contribution

Our work seeks to complement previous efforts. We define a persistent forecasting model (as in [20]) which assumes little on the wind profile evolution law. This defines a vectorial stochastic process governing the wind profile evolution as in [22]. To address the problem of estimating the whole profile given sparse observation, we reduce the system dimensions by (i) projecting onto a 4-dimensional subspace, and (ii) leveraging a Gaussian process structure on the wind profile evolution similar to [18] to develop an online learning algorithm for the process covariance function. Finally, (iii) the Kalman Filter (KF) recursive equations -by forecasting the impact of an altitude adjustment on the state estimates- allow us to finely balance regret minimization and information gain on maximizing wind power output. We finally illustrate the methodology via simulations.

II. Problem Formulation

A. Wind dynamics

Denote as $w(t,h)$ the wind *speed* at time t and altitude $h \in \bar{H} = (0,H)$; and $w_t\colon h \to w(t,h)$ is the wind *profile* belonging to $H = L^2(\bar{H},\mathbb{R})$ of square integrable functions. We assume the existence of a Partial Differential Equation governing the evolution of w_t,

$$\frac{\mathrm{d}w_t}{\mathrm{d}t} = \mathbf{f}(w_t,\boldsymbol{\lambda}) \tag{1}$$

The mis-specification of the law $\mathbf{f},\boldsymbol{\lambda}$ governing the evolution is the first source of uncertainty: the *model error*.

Using Euler's forward method, finite elements and assuming a linear PDE, one can show the previous PDE reduces to a simple persistent dynamical system with state $x_t \in \mathbb{R}^p$ and random noise ϵ_t, where x_t are the coefficients of w_t projected onto a p-dimensional basis of $L^2(\bar{H},\mathbb{R})$ [13].

$$x_{t+1} = x_t + \epsilon_t$$

There exists an observation function $\phi : \mathbb{R} \to \mathbb{R}$ from which we can approximately recover the wind profile form, which further introduces a second source of uncertainty: the *observation error*.

$$w_t(h) = \phi(h)^\top x_t + \gamma_t(h)$$

Finally, due to the problem structure we only observe one wind speed at one altitude per time step. This defines a *partially-observable* Markov model, which induces a third source of uncertainty. The hidden vector state x_t is estimated over a probability space $(\Omega,\Sigma,\mathbb{P})$, where the uncertainty arises from the *state error*, *observation error* and *partial observability*.

By designing a proper controller, we can influence the *partial observability* to reduce the entropy of the state estimate's distribution.

B. AWE Controller

At time t, the decision maker sequentially chooses actions u_t from a finite action set $\bar{U}$ (of the accessible altitudes) and observes the corresponding outcomes (wind speed) $(\mathbf{Y}_{t,u_t})_{t\in\mathbb{N}}$. With our previously defined notations:

$$\mathbf{Y}_{t,u_t} = \mathbf{w}_t(u_t)$$

The agent associates a reward $(y,u) \to R(y,u-u_{t-1})$ which is the wind power output for an outcome y and action u. The function is fixed and known [7] as,

$$R(y,\Delta u) = \underbrace{c_1 \cdot \min\left(y, V_{\text{rated}}\right)^3}_{\text{Generated energy}} - \underbrace{\left(c_2 y^2 + c_3 y^2|\Delta u|\right)}_{\text{Energy lost}} \tag{2}$$

where $\Delta u = u - u_t$ is the adjustment with respect to the past altitude. The reward function accounts for both the energy generated or lost by adjusting altitude. In section VI, this will be seen as a penalty for exploring. Parameters c_1, c_2, c_3, V_{rated} are constants imposed by the manufacturer. Figure 1 shows the power output with regards to wind speed for different altitude changes Δu. It is nonlinear, non-convex and non-monotonic. Note, that we should not specifically track the greatest wind speeds, but rather wind speed regions close to the rated speed.

Uncertainty about the vector state $\mathbf{x_t}$ induces uncertainty about the true optimal action $\mathbf{u_t^*}$, which we denote by: $\mathbf{u_t^*} \in \operatorname{argmax}_u(\mathbb{E}[\mathbf{R}_{t,u}|\mathbf{x_t}])$. Thus, we define the T–period regret of the sequence of actions $u_1,\cdots,u_T$ as the random variable:

$$\mathbf{Regret}(T) = \sum_{t=1}^{T} \mathbf{R}_{t,\mathbf{u_t^*}} - \mathbf{R}_{t,\mathbf{u_t}}$$

In this work we are interested in minimizing the Bayesian notion of regret $\mathbf{Regret}(T)$ conditioned to available information. It is important to stress that the expectation is taken over the randomness of action $\mathbf{u_t}$, over the system state $\mathbf{x_t}$, and outcomes $\mathbf{Y_t}$. Finally, we define the filtration (or history) of observations and actions as $\mathcal{F}_t = \sigma((\mathbf{Y_k},\mathbf{U_k}) \mid k \leq t)$ which is a sigma algebra of $\mathcal{F}$. To simplify notations we denote: $\mathbb{E}_t[\cdot] := \mathbb{E}[\cdot|\mathcal{F}_t]$.

Our final objective is to select a sequence of policies $(\pi_t)_{t\in\mathbb{N}}$ over action space: $\pi(\cdot) := \mathbb{P}_t(u = \cdot)$ that minimizes $\mathbb{E}[\mathbf{Regret}(\pi,T)]$. This is addressed in Section VI.

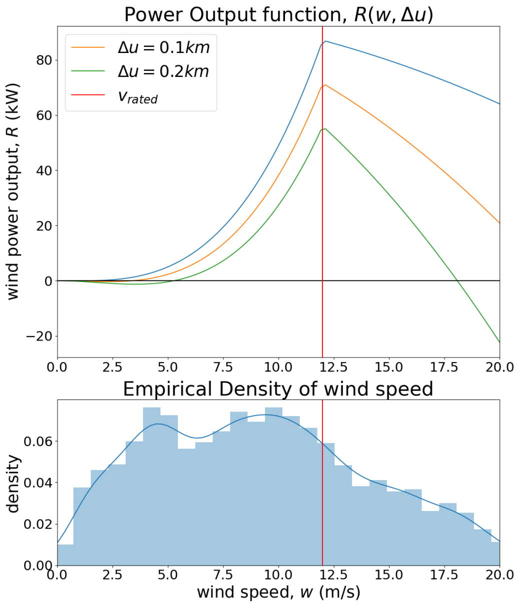

Fig. 1: Power output function of wind-speed and altitude adjustment $\Delta u \in \{0, 0.1, 0.2\}$ & Empirical density of wind speed. High wind speed is likely and induce lower power output. Altitude adjustments Δu come with an increasing cost with wind speed (in $w^2 \cdot |\Delta u|$).

C. *Offline setup*

We consider the data set from [24], which was collected from an experimental campaign at Cape Henlopen State Park in Lewes, Delaware. In this work, the data used consists of wind speed measurements recorded by a 915-Mhz wind profiler between July 1, 2014 and August 31, 2014. The profiler records wind speeds in $\Delta h = 50$ meter altitude increments from 150 m to 1000 m, in 30 minute intervals.

The wind profile is sampled at constant altitudes $\mathcal{H} = \{i\Delta h\}_{i \leq n}$ such that $y_t \in \mathbb{R}^n$ with $n = 18$, and $(y_t)_i = w_t(i\Delta h)$. We use this dataset to both select the functional basis and set up the simulation environment to evaluate our online controller.

III. Model order Reduction

We aim to select p basis functions with $p < n$, $\Phi = (\phi_i)_{i \leq p}$ with $\phi_i \in \mathcal{L}^2(\bar{H}, \mathbb{R})$, so that the projected wind profile is accurate and its dynamics are simple to track. Practically, the choice of Φ will decompose w_t into a linear combination of elements of Φ multiplied by coefficient vector (the system state) $x \in \mathbb{R}^p$, i.e.

$$w_t(h) = \sum_{i=1}^{p} x_{t,i}\phi_i(h) + \gamma_t(h) \tag{3}$$

Fitting the coefficients is accomplished by least squares:

$$\begin{aligned} x_t &= \text{argmin}_x \sum_{j=1}^{n} \left(w_t(j\Delta h) - \sum_{i=1}^{p} x_i \phi_i(j\Delta h) \right)^2 \\ &= \text{argmin}_x (y_t - B(\phi)^\top x)^2 \end{aligned}$$

Hence, we can write that:

$$x_t = \left(B(\phi)^\top B(\phi)\right)^{-1} B(\phi) y_t, \text{ where } [B(\phi)]_{i,j} = \phi_i(j\Delta h)$$

In section IV, we will formalize a persistent dynamical system as a Linear Gauss-Markov model such that:

$$\begin{cases} x_{t+1} &= x_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, Q) \\ y_t &= B(\phi)^\top x_t + \nu_t, \quad \nu_t \sim \mathcal{N}(0, R) \end{cases} \tag{4}$$

Ideally, we seek a functional basis which (i) provides inertia, (ii) has good model accuracy and (iii) has good observability. Our methodology is the following: we restrict the family of functions to polynomials of degree smaller than n:

$$\Phi = (\Phi_i)_{i \leq n} = \left\{ \phi_i(h; \alpha) := \sum_{k=0}^{i} \alpha_i (H - h)^i; \ i < n \right\}$$

A high inertia system (i) evolves slowly with time i.e. $\mathbb{E}[(x_{t+1} - x_t)(x_{t+1} - x_t)^\top] = ||Q||_F$ is small. A projection is accurate (ii) if its reconstruction error is small: $\mathbb{E}[(y_t - y_t^r)(y_t - y_t^r)^\top] = ||R||_F$. Finally, the system has good observability (iii) if observing y gives high information content about the state. That is, the mutual information between X and Y is high: $I(X;Y) = H(Y) - H(Y|X)$. Both quantities can be estimated with the data and written as a function of R and $Cov(Y_i, Y_j)$.

Upon selecting a polynomial order $p = 4$, we obtain a covariance error matrix $\hat{R}$ such that $\gamma_t \sim \mathcal{N}(0, \hat{R})$. Note that we could build this approximation with some general notion of the wind profile and hence this offline step does not require collecting data in an actual test bed. We denote $B(\phi)$ as Φ in the following.

IV. Kalman filter formulation of forecast model

The challenge in building a forecast model lies in the complex dynamics of the wind shear's physics. This can be highlighted by plotting the standard deviation of the profile which is heteroskedastic (see Fig. 2).

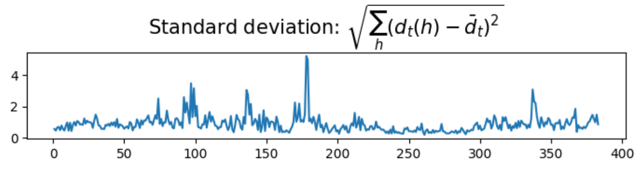

Fig. 2: Wind Profile Evolution $d_t = w_t - w_{t-1}$ presents a Heteroskedastic variance

An alternative to a complex-to-fit and estimate-dependent forecasting model is to formulate a persistent model following (4). Persistent models are known to perform relatively well [20]. In this case, the computational effort involves estimating the current profile and the required prior knowledge of the model is minimal. Note that the persistent state equation (4) is linear in the state dynamics, as is the observation process. A significant difference with a plain vanilla Kalman Filter is that the observation function is directly controllable since it relates to the altitude.

$$\begin{cases} x_{t+1} &= x_t + \epsilon_t \\ y_t &= \phi(u_t)^\top x_t + \nu_t \end{cases} \tag{5}$$

Recall the central assumptions of the Kalman Filter (KF):

1) The random variables $x_0, \omega_0, \nu_0, \cdots, \omega_T, \nu_T$ are jointly Gaussian and mutually independent,
2) Noise ϵ_t is independent, centered and normally distributed with $\mathbb{E}[\epsilon_t \epsilon_t^\top] = Q_t$. We later develop an online adaptive learning algorithm for Q_t,
3) Noise ν_t is independent, centered and normally distributed with $\mathbb{E}[\nu_t \nu_t^\top] = \phi(u_t)^\top R \phi(u_t)$. Matrix R can be estimated offline using the procedure explained in Section III, and corresponds to the distribution of approximation errors.

The objective of a filter is to obtain the state density x_T given past observations up to a certain time $\mathcal{F}_t$. Since all random quantities are Gaussian and the state and observation equations are linear, the filter optimally blends new information by simply applying a linear feedback to the prior state. We encourage interested readers to consult [14] for more information. The recursive KF equations are summarized as follows:

1) The state x_t is normally distributed and given by density $p_t(x_t)$, defined by its first two moments: $x_{t|t} = \mathbb{E}_t[x_t]$ and $\Sigma_{t|t} = \mathbb{E}_t[(x_t - x_{t|t})(x_t - x_{t|t})^T]$.
2) **Prediction Step** Given $x_{t|t}$, $\Sigma_{t|t}$, the forward (or prediction) equations apply model dynamics in (5):

$$x_{t+1|t} = \mathbb{E}_t[x_{t+1|t}] = x_{t|t}$$
$$\Sigma_{t+1|t} = \Sigma_{t|t} + Q_t$$

3) **Analysis Step** To ease notation denote ϕ_{t+1} as $\phi(u_{t+1})$. Given $x_{t+1|t}$, $\Sigma_{t+1|t}$, note that $\mathcal{F}_{t+1} = \{y_{t+1}\} \cup \mathcal{F}_t$ and hence using Bayes formula $x_{t+1}|\mathcal{F}_{t+1} = (x_{t+1}|\mathcal{F}_t)|(y_{t+1}|\mathcal{F}_t)$. We know:

$$\begin{bmatrix} x_{t+1|t} \\ y_{t+1|t} \end{bmatrix} = \mathcal{N}\left(\begin{bmatrix} x_{t|t} \\ \phi_{t+1}^\top x_{t|t} \end{bmatrix}, \begin{bmatrix} \Sigma_{t+1|t} & \Sigma_{t+1|t}\phi_{t+1} \\ \phi_{t+1}^\top \Sigma_{t+1|t} & \phi_{t+1}^\top(\Sigma_{t+1|t} + R)\phi_{t+1} \end{bmatrix} \right)$$

Setting the Kalman Gain as:

$$L_t = \Sigma_{t+1|t}\phi_{t+1}\left[\phi_{t+1}^\top(\Sigma_{t+1|t} + R)\phi_{t+1}\right]^{-1}$$

After collecting y_{t+1}, the measurement update reads:

$$x_{t+1|t+1} = x_{t|t} + L_t\left(y_{t+1} - \phi(u_{t+1})^\top x_{t|t}\right)$$
$$\Sigma_{t+1|t+1} = \Sigma_{t+1|t} - L_t \phi(u_{t+1})^\top \Sigma_{t+1|t}$$

The measurement update indicates how the density of the state estimate will change given the next altitude u_{t+1}. Importantly, note that the covariance update depends on u_{t+1}. We can leverage this structure to select the altitude that decreases the state entropy at the optimum altitude. This particular fact will be used to develop our controller.

V. Open-Loop online estimation of parameters

A. *Parametric Process Covariance*

As illustrated in Fig. 2, the process covariance Q_t should be considered as time-varying and hence needs to be learned. To address, we propose to (i) continuously estimate the covariance over a sample of s past observations, and (ii) reduce the number of free parameters to decrease the estimate's variance. For (ii), we consider a geo-physical structure for the process w_t. This follows previous work on this matter [18]. Denote $d_t = w_t - w_{t-1}$ as the process noise of the functional wind speed process. We can then assume that d_t is a Gaussian process with respect to space, i.e.,

$$d_t(z) \sim \mathcal{GP}(0, k(z, z'))$$

where $k(z, z') = \sigma_0^2 \exp(-\frac{1}{2}\frac{(z-z'))^2}{\sigma_1^2}) + \sigma_2^2$ is its covariance function. We chose the Squared Exponential (SE) kernel plus a constant kernel for its overall quality of fit (see Fig. 3). The only notable difference between the sample and kernel-based covariance is the higher uncertainty that reigns at the upper altitude layers. Effectively, it boils down to parameterize the covariance matrix by three parameters $\theta = (\sigma_0^2, \sigma_1^2, \sigma_2^2)$. In particular $d_t(h_i)$ has covariance K_θ where $[K_\theta]_{i,j} = k(h_i, h_j)$. Hence, after projection Q_t is itself parametrized by θ:

$$Q_{t,\theta} = \Phi^\top K_\theta \Phi$$

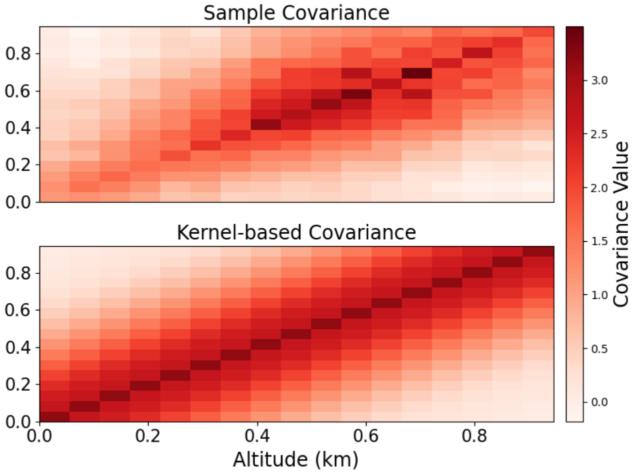

Fig. 3: Sample $\hat{K}$ and Kernel-based covariance $K_{\hat{\theta}}$. The 3-parameters kernel is a reasonable estimation.

B. *Fixed Horizon Expectation Maximization Algorithm*

Our state space model is a Hidden Markov Chain. The log likelihood of a sequence of state and observation from $t = 0$ to $t = T$ conveniently has a separable form. We also note that only the transitions from x_{t-1} to x_t depend on θ.

$$p(x_{0:T}, y_{0:T}; \theta) = p(x_0) \prod_{t=1}^{T} p(x_t | x_{t-1}; \theta) p(y_t | x_t)$$

$$\log p(x_{0:T}, y_{0:T}; \theta) = \sum_{t=1}^{T} \log\left(p(x_k | x_{k-1}; \theta)\right) + C$$

One traditionally estimates the covariance matrix of a Kalman Filter offline through an expected maximization (EM) type of algorithm. Given a set of observations, the algorithm iteratively generates a sequence of parameter estimates (θ^c) such that the log-likelihood of the observations is non-decreasing over the iterations. Let θ' be a parameter that we assume governs the evolution of x. Since $L(\theta)$ is not random, and thus $\mathbb{E}_{\theta'}[L(\theta)] = L(\theta)$, we have:

$$L(\theta) = \log p_\theta(y_{1:T}) = \log p_\theta(y_{1:T}, x_{1:T}) - \log p_\theta(y_{1:T}|x_{1:T})$$

$$\mathbb{E}_{\theta'}[L(\theta)] = \mathbb{E}_{\theta'}[\log p_\theta(y_{1:T}, x_{1:T})] - \mathbb{E}_{\theta'}[\log p_\theta(y_{1:T}|x_{1:T})]$$

$$L(\theta) = Q(\theta;\theta') - H(\theta,\theta')$$

Iteratively increasing $\theta^{(c+1)} = \text{argmax}_\theta \; Q(\theta, \theta^{(c)})$ leads to increasing $(L(\theta^{(c)}))$ (see [25]). Hence the EM algorithm converges to a local maxima of the log-likelihood of the model. We now specify the expectation step and drop the conditions $\theta', y_{1:T}$ on the expectation, for ease of reading.

$$Q_T(\theta,\theta') = \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_T[\log p(x_t|x_{t-1};\theta)]$$

$$= -\frac{1}{2T}\sum_{t=1}^{T} \underbrace{\mathbb{E}_T[\epsilon_t^\top Q_\theta^{-1}\epsilon_t]}_{c_t(\theta)} - \frac{1}{2}\log\det(Q_\theta)$$

Using the Rauch–Tung–Striebel smoother, we can determine: $x_{t|T}$, $\epsilon_{t|T}$, $\Sigma_{t|T}$, $\Sigma^t_{t-1|T} = \mathbb{E}_T[\epsilon_t\epsilon_t^\top]$, and

$$c_t(\theta) = \epsilon_{t|T}^\top Q_\theta^{-1}\epsilon_{t|T} + \text{Tr}\big(Q_\theta^{-1}(\Sigma_{t|T} + \Sigma_{t-1|T} - 2\cdot\Sigma^t_{t-1|T})\big)$$

Interested readers can refer to [26]. Practically, we use $s = 50$ past samples and the Scikit-Learn Gaussian Process kernels to minimize the log-likelihood using the box-contrained Broyden–Fletcher–Goldfarb–Shanno algorithm (L-BFGS-B) to return our sequence of estimates, $(\theta^*_{t-s:t})_t$.

$$\theta^*_{T-s:T} = \text{argmin}_\theta \; -\frac{1}{2s}\sum_{t=T-s}^{T} c_t(\theta) - \frac{1}{2}\log\det(Q_\theta)$$

VI. Closed-loop Information Directed Sampling

From $x_{t|t}, \Sigma_{t|t}$, we can obtain a non-stationary distribution of the wind-profile at altitude u by applying:

$$w_t(u)|\mathcal{F}_t \sim \mathcal{N}(\phi(u)^\top x_{t|t} \;,\; \phi(u)^\top\Sigma_{t|t}\phi(u))$$

Our objective is to select a sequence of policies $(\pi_t)_{t\in\mathbb{N}}$ that minimizes $\mathbb{E}[\textbf{Regret}(\pi, T)]$. We consider policies from a set of binomial distributions, with n denoting the number of altitude bins and parameters $p_i = (\frac{i}{n})_{i\le n}$ so that, for each altitude bin i, the distribution is centered on $i = p_i \cdot n$,

$$D = \Big(Binom(p_i, n)\Big)_{i\le n} \qquad \text{(Set of Randomized Policies)}$$

Let $\Delta_t(u) = \mathbb{E}_t[R_{t,U^*} - R_{t,u}]$ be the regret of taking action u and $g_t(u) := I_t(U^*; Y_{t,u})$ be the information gain about the optimum. $\mathbf{U}^*$ is the optimal altitude (and random) for gross power output, i.e. $R(y, 0)$ in (2). Note $\mathbf{U}^*$ is independent of altitude adjustment Δu. Its distribution is:

$$\alpha_t(u) = \mathbb{P}_t(U^* = u) = \mathbb{P}_t\big(\cap_{a\ne u}\{R(\mathbf{Y}_{t,u}, 0) \ge R(\mathbf{Y}_{t,a}, 0)\}\big)$$

$g_t(u)$ is also known as the mutual information between the wind speed at u and the maximum $\mathbf{U}^*$. The information directed sampling algorithm selects the probability which minimizes the information ratio:

$$\pi_t^{IDS} = \text{argmin}_{\pi\in D}\left\{\Psi_t(\pi) = \frac{\Delta_t(\pi)^2}{g_t(\pi)}\right\} \quad \text{(IDS Policy)} \tag{6}$$

The controller minimizes the squared regret incurred per-bit of information acquired about the optimum [15]. This policy has sub-linear regret growth, and further work could devise a precise bound for our specific persistent model.

A. Estimation of Regret

As mentioned in the introduction, the distribution of the power output $R(\mathbf{Y})$ even under normally distributed wind speed has no closed-form distribution. We must rely on Monte Carlo simulations to estimate the regret. Hence, we take M $(x^m)_{m\le M}$ samples from $\mathcal{N}(x_{t|t}, \Sigma_{t|t})$ and record their rewards (power output) $(r^m)_m$ given the last altitude $u_{t-1} \in \mathcal{F}_t$ and net optimizer u^*_m.

$$\Delta_t(u) = \mathbb{E}_t[R_{t,U^*}] - \mathbb{E}_t[R(Y_{t,u}, u - u_{t-1})]$$

$$\approx \frac{1}{M}\sum_{m=1}^{M} R(y^m_{t,u^*_m}, u^*_m - u_{t-1}) - \mathbb{E}_t[R(Y_{t,u}, u - u_{t-1})]$$

The second term is a tractable integral with respect to a one-dimensional normal density.

B. Estimation of the information gain function

We choose U^*, the *gross optimum*, as the optimal altitude for gross power $R(y, \Delta u = 0)$ instead of net power $R(y, \Delta u = u - u_{t-1})$. This resolves the following issue. IDS optimizes an instantaneous objective, which would penalize exploration via the third term in (2) without considering the potential long-term reward. Using gross power $R(y, 0)$ side-steps this issue. The information gain is defined as the expected reduction in entropy $H(\alpha)$ of the posterior distribution of U^* for observing $Y_{t,u}$:

$$g_t(u) = \mathbb{E}_t\big[H(\alpha_t) - H(\alpha_{t+1})|u_t = u\big]$$

To estimate the first and second term, we again use Monte Carlo simulations to approximate α_t. Then, as explained in Section IV, taking action u will update the covariance function: $\Sigma_{t+1|t+1}(u) = \Sigma_{t+1|t} - L_t\phi(u)^\top\Sigma_{t+1|t}$. Hence, we can sample M scenarios from $\mathcal{N}(x_{t|t}, \Sigma_{t+1|t+1}(u))$ and approximate $\alpha_{t+1}(u)$.

$$\alpha_{t/t+1}(u) \approx \frac{1}{M}\sum_{m=1}^{M} \mathbb{1}(u = \text{argmax}\; r_m^{t/t+1})$$

Finally, we select a policy π from D such that the information ratio is minimized: $\min_{\pi\in D}(\pi^\top\Delta_t)^2/(\pi^\top g_t)$.

VII. Results

To illustrate the KF-IDS controller, we simulate different baseline controllers over the same wind profile and compare various performance measures. As presented in Section II, the data comes from Cape Henlopen for days ranging from July 7, 2014 (time index 0) to July 11 (time index 96). The values for the power output function can be found in [7]:

Symbol	Value	Symbol	Value
h_{min}	0.15 km	Δt	30 min
h_{max}	1 km	c_1	0.0579 kW s^3/m^3
r_{max}	0.01 km/min	c_2	0.09 kW s^2/m^2
v_r	12 m s^{-1}	c_3	1.08 kW s^2/m^2 ·km

TABLE I: AWE Model Parameters

A. Baseline controllers

The most fundamental performance metric to evaluate our controller is power generation. As in previous works, we also consider the following baseline controllers:

1) *Omniscient Dynamic Programming (Omniscient)*: We recursively solve the DP equations yielding the trajectory that maximizes overall power generation. This provides the upper bound of energy generation.
2) *Omniscient Fixed Altitude (Optimal Fixed)*: The optimal fixed altitude.
3) *Lowest Fixed Altitude (Fixed 100m)*: This baseline represents conventional wind turbines and helps us measure the incremental gains of AWE systems.

We additionally consider the following controllers. For notations see Section VI.

4) *Regret Minimization (Greedy)*: We sample from the distribution π that minimizes expected squared regret $(\pi^\top \Delta)^2$.
5) *Information Gain Maximization (Info.)*: We sample from the distribution π that maximizes expected information gain $\pi^\top g$.
6) *Information-Directed Sampling (IDS)*: We sample from the distribution π that maximizes the information ratio $(\pi^\top \Delta)^2/\pi^\top g$.

B. Discussion

In the following simulations, we assume wind speed is measured at ground level (this is inexpensive), in addition to measuring wind speed at the AWE's controlled altitude.

1) Qualitative Analysis - Profile of the trajectories: Sampled trajectories for 5 policies are displayed in Fig. 4. The IDS and Greedy trajectories are similar. High wind speed periods (e.g. between 24 and 60 hours) have relatively fixed trajectories. This results from the objective function structure: the penalty for adjusting scales with squared wind speed. Hence, the squared regret for exploring (adjusting) will be dissuasive. On the other hand, during lower wind speed periods the incentive to explore becomes stronger (e.g. 60 - 85 hours). A penalty-unaware controller, such as the Information controller, constantly adjusts altitude to improve its knowledge of the wind power optimum.

2) Qualitative Analysis - KF Estimation Performance: In Fig. 5, the actual wind profile (upper subplot) is visually compared to the recovered profile from the KF estimates, formally defined as the sequence $(\Phi^\top x_{t|t})_t$ (lower subplot). When an IDS controller is used, the filter catches most of the wind profile patterns except when it is static (e.g. between 24 to 60 hours). During these periods, the IDS controller sacrifices recovery of the actual wind speed profile (exploration) for performance (exploitation).

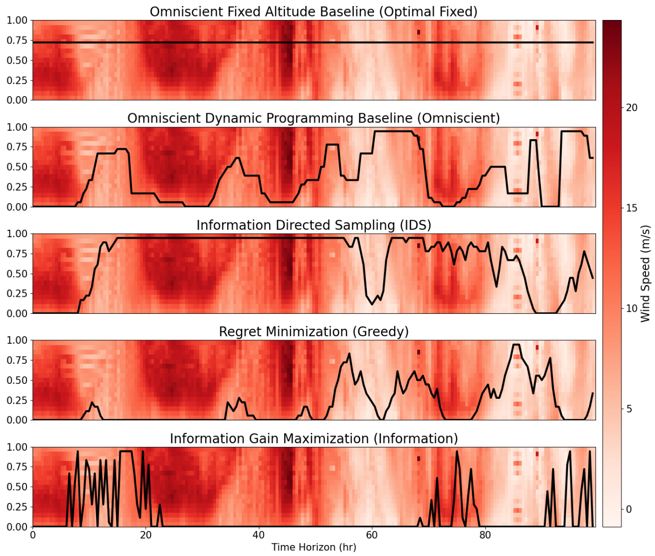

Fig. 4: Samples of controller trajectories for 5 different policies over the same wind-profile. IDS and Greedy presents some qualitative similarities due to the large penalty associated with adjusting altitudes with high-speed wind.

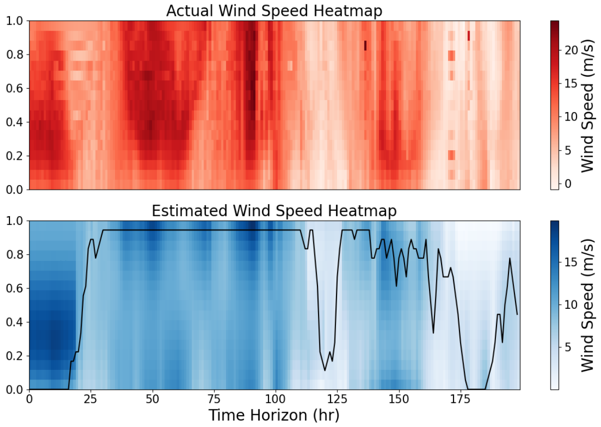

Fig. 5: Comparison of actual wind-profile (top) and KF estimates (below) when using an IDS controller. The KF successfully tracks the main trend, except when the cost to explore is overwhelming (between 40 and 60 hours).

3) Quantitative Analysis - filter and policies performance: The previous analyses are limited to sampled roll-outs and specific wind profiles. By reporting other performance measures in Table II, we illustrate that IDS can be superior to a fixed or regret-based policy in the long-term since its knowledge of the wind profile is enhanced. We examine 4 relevant performance measures: the first two relate to filter accuracy, whereas the last two relate to IDS performance.

A well-conditioned KF is paramount: its mean and covariance are fed to the controller objective function. One way to verify the KF is to track the average reconstruction error (see Section III). Additionally, the average filtered state entropy, defined as $\frac{1}{T}\sum_t ||\Sigma_{t|t}||_F$, reveals the effectiveness of exploration in the policy. As a result, the information controller excels – without surprise – in this task and the IDS controller follows in second.

The IDS controller seeks to improve its knowledge of the maximum output distribution α_t. The average likelihood of the gross optimum, $\frac{1}{T}\sum_t \alpha_t(u_t^*)$, measures the efficiency of the controller's exploration component. Finally, the output average regret: $\frac{1}{T}\sum_t \mathrm{Regret}(\pi_t, t)$ is the objective we ultimately seek to minimize. These last two performance measures illustrate the superiority of the IDS controller.

	Selected Baselines		KF Policies		
	DP	100m	IDS	Greedy	Info.
Reconstruction Error (m/s)	2.67	4.91	2.64	3.18	**1.67**
Froebenious Norm State Entropy	5.80	9.48	5.70	6.65	**4.34**
Likelihood of Maximum	—	0.00	**0.16**	0.14	0.09
Output Average Regret (kW)	—	13.65	**4.62**	13.82	52.32

TABLE II: Selected performance measures for different controllers. IDS presents a good tradeoff between observability of the profile (low reconstruction error, high likelihood of the maximum) and performance (least average regret)

VIII. Conclusion and Further Work

This work focuses on identifying and leveraging a specific structure of the wind profile evolution to design an Information-Directed Sampling controller based on the Kalman Filter equations. The policy blends exploitation and exploration into a single objective function.

We showed two important results based on real-world wind profile data: (i) The filter state estimates are reasonably accurate, even under sparse observations and with weak assumptions on the underlying evolution dynamics. (ii) Kalman Filtered-Information Directed Sampling improves net energy generation relative to other controllers.

Future work can include the following: (i) Quantify the performance loss for using a persistent model versus a more complex forecast model. (ii) Synthesize statistically accurate scenarios of wind speed profiles to rank controllers and evaluate the learning algorithm for the time-varying covariance matrix. (iii) Relax the KF hypotheses and consider alternative nonlinear filters. (iv) Leverage the stochastic process structure to derive tighter bounds on regret growth.

References

[1] M. Francis, "Renewables became the second-most prevalent u.s. electricity source in 2020." [Online]. Available: https://www.eia.gov/todayinenergy/detail.php?id=48896#

[2] M. Diehl, "Airborne wind energy: Basic concepts and physical foundations," in *Airborne wind energy*. Springer, 2013, pp. 3–22.

[5] C. L. Archer, L. Delle Monache, and D. L. Rife, "Airborne wind energy: Optimal locations and variability," *Renewable Energy*, vol. 64, pp. 180–186, 2014.

[3] C. L. Archer and K. Caldeira, "Global assessment of high-altitude wind power," *Energies*, vol. 2, no. 2, pp. 307–319, 2009.

[4] L. Fagiano, M. Milanese, and D. Piga, "High-altitude wind power generation," *IEEE Transactions on Energy Conversion*, vol. 25, no. 1, pp. 168–180, 2009.

[6] A. Cherubini, A. Papini, R. Vertechy, and M. Fontana, "Airborne wind energy systems: A review of the technologies," *Renewable and Sustainable Energy Reviews*, vol. 51, pp. 1461–1476, 2015.

[7] A. Bafandeh and C. Vermillion, "Real-time altitude optimization of airborne wind energy systems using lyapunov-based switched extremum seeking control," in *2016 American Control Conference (ACC)*. IEEE, 2016, pp. 4990–4995.

[8] C. Vermillion, T. Grunnagle, R. Lim, and I. Kolmanovsky, "Model-based plant design and hierarchical control of a prototype lighter-than-air wind energy system, with experimental flight test results," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 2, pp. 531–542, 2013.

[9] C. Penland and P. D. Sardeshmukh, "The optimal growth of tropical sea surface temperature anomalies," *Journal of climate*, vol. 8, no. 8, pp. 1999–2024, 1995.

[10] C. Penland, "Random forcing and forecasting using principal oscillation pattern analysis," *Monthly Weather Review*, vol. 117, no. 10, pp. 2165–2185, 1989.

[11] H. Mena and L. Pfurtscheller, "An efficient spde approach for el niño," *Applied Mathematics and Computation*, vol. 352, pp. 146–156, 2019.

[12] E. B. Iversen, J. M. Morales, J. K. Møller, and H. Madsen, "Short-term probabilistic forecasting of wind speed using stochastic differential equations," *International Journal of Forecasting*, vol. 32, no. 3, pp. 981–990, 2016.

[13] K. Scerri, M. Dewar, and V. Kadirkamanathan, "Estimation and model selection for an ide-based spatio-temporal model," *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 482–492, 2008.

[14] P. R. Kumar and P. Varaiya, *Stochastic systems: Estimation, identification, and adaptive control*. SIAM, 2015.

[15] D. Russo and B. Van Roy, "Learning to optimize via information-directed sampling," *Operations Research*, vol. 66, no. 1, pp. 230–252, 2018.

[16] S. Bin-Karim, A. Bafandeh, and C. Vermillion, "Spatio-temporal optimization through model predictive control: A case study in airborne wind energy," in *2016 IEEE 55th Conference on Decision and Control (CDC)*. IEEE, 2016, pp. 4239–4244.

[17] A. Bafandeh, "Hierarchical control strategies for spatiotemporally varying systems with application to airborne wind energy," Ph.D. dissertation, The University of North Carolina at Charlotte, 2018.

[18] A. Baheri and C. Vermillion, "Altitude optimization of airborne wind energy systems: A bayesian optimization approach," in *2017 American Control Conference (ACC)*. IEEE, 2017, pp. 1365–1370.

[19] A. Baheri, S. Bin-Karim, A. Bafandeh, and C. Vermillion, "Real-time control using bayesian optimization: A case study in airborne wind energy systems," *Control Engineering Practice*, vol. 69, pp. 131–140, 2017.

[20] L. N. Dunn, C. Vermillion, F. K. Chow, and S. J. Moura, "On wind speed sensor configurations and altitude control in airborne wind energy systems," in *2019 American Control Conference (ACC)*. IEEE, 2019, pp. 2197–2202.

[21] C. Qin, D. Klabjan, and D. Russo, "Improving the expected improvement algorithm," *arXiv preprint arXiv:1705.10033*, 2017.

[22] P. Keyantuo, L. N. Dunn, B. Haydon, C. Vermillion, F. K. Chow, and S. J. Moura, "A vector auto-regression based forecast of wind speeds in airborne wind energy systems," in *2021 IEEE Conference on Control Technology and Applications (CCTA)*, 2021.

[23] L. N. Dunn, *Data-Driven Decision Analysis in Electric Power Systems*. University of California, Berkeley, 2020.

[24] C. L. Archer, "Wind profiler at cape henlopen." [Online]. Available: https://www.ceoe.udel.edu/our-people/profiles/carcher/fsmw

[25] S. Gibson and B. Ninness, "Robust maximum-likelihood estimation of multivariable dynamic systems," *Automatica*, vol. 41, no. 10, pp. 1667–1682, 2005.

[26] S. Särkkä, *Bayesian filtering and smoothing*. Cambridge University Press, 2013, no. 3.