nature

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41586-018-0533-0

In the format provided by the authors and unedited.

Glider soaring via reinforcement learning in the field

Gautam Reddy[1,5], Jerome Wong-Ng[1,5], Antonio Celani[2], Terrence J. Sejnowski[3,4] & Massimo Vergassola[1]*

[1]Department of Physics, University of California, San Diego, La Jolla, CA, USA. [2]The Abdus Salam International Center for Theoretical Physics, Trieste, Italy. [3]The Salk Institute for Biological Studies, La Jolla, CA, USA. [4]Division of Biological Sciences, University of California, San Diego, La Jolla, CA, USA. [5]These authors contributed equally: Gautam Reddy, Jerome Wong-Ng.
*e-mail: massimo@physics.ucsd.edu

Supplementary Information for 'Soaring through reinforcement learning in the field'

Gautam Reddy, Jerome Wong Ng, Antonio Celani, Terrence J. Sejnowski, Massimo Vergassola

1 On-board estimation of the navigational cues

For a given desired pitch ϕ_d and desired bank angle μ_d, the flight controller implements a feedback control system such that:

$$\frac{d\phi}{dt} = \frac{\phi_d - \phi}{\tau}, \tag{1}$$

$$\frac{d\mu}{dt} = \frac{\mu_d - \mu}{\tau}, \tag{2}$$

where ϕ is the pitch (Extended Data Figure 2), μ is the bank angle and τ is a user-set time scale of control. ϕ_d is set fixed during flight and can be used to indirectly modulate the angle of attack, α, which determines the airspeed V and sink rate w.r.t air of the glider ($-\boldsymbol{v}_z$). When α is large, the glider has a low airspeed and a low sink rate, while at small α the glider is faster but also sinks more rapidly. The glider's glide polar curve, relating the sink rate and airspeed for different values of α at equilibrium, depends on μ, the lift coefficient $C_L(\alpha)$ and the drag coefficient $C_D(\alpha)$ as

$$\frac{-\boldsymbol{v}_z}{V} = \frac{C_D(\alpha)}{C_L(\alpha)\cos\mu}. \tag{3}$$

The ratio on the left hand side is called the glide angle γ (Extended Data Figure 2), where

$$\gamma \equiv -\boldsymbol{v}_z/V = \alpha - \phi. \tag{4}$$

The lift and drag coefficients are determined by the geometry of the plane; their form can be derived in certain simplified situations [1, 2]. Equations (3) and (4) together relate the measurable quantities, ϕ and the bank angle μ, to α at equilibrium. Actions of increasing, decreasing or keeping the same bank angle are taken in time steps of t_a by changing μ_d such that μ increases linearly from μ_i to μ_f in t_a:

$$\mu_d(t) = \mu_i + (\mu_f - \mu_i)\frac{t+\tau}{t_a}. \tag{5}$$

1.1 Estimation of the vertical wind acceleration

The vertical wind acceleration $\boldsymbol{a}_z$ is defined as:

$$\boldsymbol{a}_z \equiv \frac{d\boldsymbol{w}_z}{dt} = \frac{d}{dt}\left(\boldsymbol{u}_z - \boldsymbol{v}_z\right), \tag{6}$$

where $\boldsymbol{u}$ and $\boldsymbol{v}$ are the velocities of the glider w.r.t the ground and air respectively, and $\boldsymbol{w}$ is the wind velocity. Here, we have used the relation

$$\boldsymbol{w} = \boldsymbol{u} - \boldsymbol{v}. \tag{7}$$

An estimate of $\boldsymbol{u}$ is obtained in a straightforward manner from the EKF, which uses the GPS and barometer readings to form the estimate. On the other hand, the measurement of $\boldsymbol{v}_z$ is confounded by various aerodynamic effects that significantly affect it on relevant time scales of a few seconds. Artificial accelerations introduced due to these effects impair accurate estimation of the wind acceleration and thus alter the perceived state during decision-making and learning. Two effects significantly influence variations in $\boldsymbol{v}_z$: (1) Sustained pitch oscillations with a period of few seconds and a varying amplitude, and (2) Angle of attack variations, which occur in order to compensate for the imbalance of lift and weight while rolling. A full analysis of dynamic stability involves a set of four coupled differential equations [2] and is further complicated by the feedback control mechanism. We instead provide qualitative arguments and validate them using our data.

The longitudinal dynamic modes of the plane include short period oscillations and the phugoid. Short period oscillations are largely angle of attack variations, and the oscillations are usually heavily damped. Phugoid oscillations of longer period are less damped and are accompanied by oscillations of pitch at almost constant angle of attack. Using a reduced-order model of longitudinal stability [3], the time period of the phugoid oscillations can be estimated from the airspeed V as $\sqrt{2}\pi V/g \approx 3.5$s (here g is gravity and $V \approx 8$ m/s), which is consistent with the time period seen in experiments (Extended Data Figure 3a). Phugoidal oscillations are sustained due to constant perturbations to the pitch-wise moment when rolling. The amplitude and phase of the oscillations is determined by the magnitude and sign of $\dot{\mu}$ respectively. The amplitude is $\propto \mu^2$ and can be $> 5^\circ$ at bank angles of 30°. From (4), we see that pitch oscillations of a five degrees (~ 0.1 radian) at an airspeed $V = 8$m/s can give rise to a change in $\boldsymbol{v}_z$ of ~ 0.8m/s, which is of the same magnitude as the sink rate, and thus constitutes a significant contribution.

The lift-weight imbalance while rolling is compensated by a change in angle of attack. Suppose that a plane in equilibrium at bank angle μ_0, airspeed V_0 and angle of attack α_0 rolls to μ, V and α respectively. In equilibrium at μ_0 and μ, by balancing the forces along the vertical axis we get $L(\alpha_0)\cos\mu_0 = W = L(\alpha)\cos\mu$, where W is the weight of the glider. Here, the dependence of the lift on the angle of attack is emphasized (the contributions due to a non-zero glide angle are small and ignored here). Since $L(\alpha) \propto V^2 C_L(\alpha)$, this yields the relation $V^2 C_L(\alpha)\cos\mu = V_0^2 C_L(\alpha_0)\cos\mu_0$. Airspeed measurements in our experiments show that the change in V is negligible (Extended Data Figure 4a), and does not compensate for the change in lift. Instead, the change in lift is largely balanced by a change in the angle of attack, so that:

$$\frac{C_L(\alpha)}{C_L(\alpha_0)} \approx \frac{\cos\mu_0}{\cos\mu}. \tag{8}$$

Below the stall angle, the lift coefficient is approximately a linear function $C_L(\alpha) = A(\alpha - \alpha_i)$, where α_i is usually negative and its value depends on the geometry and the angle of incidence of the wing. We thus obtain

$$\frac{\alpha - \alpha_i}{\alpha_0 - \alpha_i} \approx \frac{\cos\mu_0}{\cos\mu}, \tag{9}$$

$$\Delta\alpha \approx (\alpha_0 - \alpha_i)\left(\frac{\cos\mu_0}{\cos\mu} - 1\right), \tag{10}$$

where $\Delta\alpha \equiv \alpha - \alpha_0$.

Suppose a plane which is steady at zero bank angle has an angle of attack α_0, pitch ϕ_0 and vertical velocity w.r.t air of $\boldsymbol{v}_{z,0}$. The deviation of $\boldsymbol{v}_z$ from $\boldsymbol{v}_{z,0}$ for a particular bank angle at a given instant is (from (4))

$$\Delta\boldsymbol{v}_z = -\Delta V\,(\alpha_0 - \phi_0) - V\,(\Delta\alpha - \Delta\phi)\,. \tag{11}$$

Here $\Delta\alpha$ is assumed to depend on the instantaneous bank angle as given in equation (10), which is justified by our arguments that the longitudinal oscillations are phugoidal i.e., the angle of attack is not influenced. Since the change in V is small, the first term can be ignored and the second term can

now be used as an approximation for the instantaneous $\boldsymbol{v}_z$ (up to a constant term) given the current bank angle and pitch, which are obtained from measurements. The constant term is ignored since our interest is in the derivative of $\boldsymbol{v}_z$.

In order to measure the variations in $\boldsymbol{v}_z$ in response to the glider's turn, we first observe that $\langle \boldsymbol{u}_z \rangle = \langle \boldsymbol{v}_z \rangle$ from (7) since $\langle \boldsymbol{w}_z \rangle = 0$. We compute $\langle \boldsymbol{u}_z \rangle$ (and thus $\langle \boldsymbol{v}_z \rangle$) by averaging the change in $\boldsymbol{u}_z$ (measured in the field) over hundreds of specific bank angle transitions. We verify that changes in $\boldsymbol{v}_z$ over 13 possible bank angle transitions are indeed captured by (11) (Extended Data Figure 3b). Note that there is only one free parameter, $\alpha_0 - \alpha_i$, which is fit. The vertical wind velocity $\boldsymbol{w}_z$ is then estimated from (7).

The vertical wind acceleration $\boldsymbol{a}_z$ is smoothed using an exponential smoothing kernel with time scale σ_a. An exponential filter of time scale σ acts on an input x to give the smoothed output $\tilde{x}$ as,

$$\tilde{x}(t) = \int_{-\infty}^{t} x(s) e^{-\frac{t-s}{\sigma}} \sigma^{-1} ds, \tag{12}$$

where the tilde is hereafter used to denote quantities smoothed by an exponential filter. Substituting $\boldsymbol{a}_z = \frac{d\boldsymbol{w}_z}{dt}$ for x and integrating by parts, we get

$$\tilde{\boldsymbol{a}}_z = \frac{\boldsymbol{w}_z - \tilde{\boldsymbol{w}}_z}{\sigma_a}, \tag{13}$$

In our implementation, another layer of exponential smoothing of smoothing time scale $\sigma_a' \ll \sigma_a$ is applied in order to average over sensory noise. As a consistency check, we verify that $\tilde{\boldsymbol{a}}_z$ is unbiased for different bank angle transitions (Extended Data Figure 4b).

1.2 Estimation of vertical wind velocity gradients across the wings

Spatial gradients in the vertical wind velocity induce rolling moments on the plane, which we estimate using the deviation of the measured bank angle from the expected bank angle. The total rolling moment of the plane has contributions from three sources – (1) The feedback control of the plane, which acts according to equation (2), (2) The spatial gradients in the wind including turbulent fluctuations, and (3) Roll moments created due to various aerodynamic effects, which we detail below.

The latter two contributions create a dynamical effect that perturb the evolution of the bank angle from equation (2). The modified evolution of the bank angle is modeled as

$$\frac{d\mu}{dt} = \frac{\mu_d - \mu}{\tau} + \omega(t) + \omega_{\text{aero}}(t), \tag{14}$$

where $\omega(t)$ and $\omega_{\text{aero}}(t)$ are contributions to the roll-wise angular velocity due to the wind and aerodynamic effects respectively. We empirically find four major contributions to $\omega_{\text{aero}}(t)$ - (1) the dihedral effect, which is a stabilizing moment due to the effects of sideslip on a dihedral wing geometry, (2) the overbanking effect, which is a destabilizing moment that occurs during turns with small radii, (3) Trim effects, which create a constant moment due to asymmetric lift on the two wings, and (4) a loss of rolling moment generated by the ailerons while rolling at low airspeeds.

Expanding the dihedral and overbanking effects around $\mu = 0$, their contributions to $\omega_{\text{aero}}(t)$ can be modeled with two terms of the form $-\mu/T_{\text{dih}}$ and $-\mu/T_{\text{ob}}$ respectively, with $T_{\text{dih}} > 0$ and $T_{\text{ob}} < 0$. The value of T_{dih} depends on the geometry of the wing and the airframe, whereas T_{ob} depends on the radius of the turns at μ. The radius of a turn at bank angle μ and airspeed V is given by

$$R = \frac{V^2}{g \tan \mu}, \tag{15}$$

For $V = 8$m/s and $\mu = 30°$, the radius is ~ 10m. For wingspans of a few meters, typical of model sailplanes, the effect can be significant. The trim effect appears as a constant bias $-b$. The effective loss

of rolling moment at low airspeeds is modeled as an additional term of the form $-\frac{\mu_d-\mu}{T_{\rm roll}}$ that opposes changes in the bank angle towards the desired bank angle. In summary, an unbiased estimation of the torque requires a calibration of three parameters related to the aerodynamics of the glider - $T_s^{-1} \equiv T_{\rm dih}^{-1} + T_{\rm ob}^{-1}$, $T_{\rm roll}$ and b. The full equation for the evolution of the bank angle is now written as:

$$\frac{d\mu}{dt} = \frac{\mu_d - \mu}{\tau} - \frac{\mu}{T_s} - \frac{\mu_d - \mu}{T_{\rm roll}} - b + \omega(t), \tag{16}$$

The three parameters are measured by making repeated transitions between bank angles of $0°, \pm 15°, \pm 30°$ by increasing, decreasing the bank angle by $15°$ or keeping same angle over intervals of 3 seconds. Averaging the bank angle in this interval over many such transitions yields the evolution of the bank angle without the wind contribution in (16) (Extended Data Figure 5a). The averaged (16) can be integrated exactly to get an analytical form for the bank angle. For linear transitions from μ_i to μ_f in a time interval t_a, plugging (5) into the averaged (16) and integrating leads to

$$\begin{aligned}\mu(t) = \frac{\tau' t \Delta\mu}{\tau'' t_a} + \left(\tau''^{-1}\mu_i - b\right)\tau' e^{-t/\tau'} + \\ + \left(\frac{\tau\Delta\mu}{\tau'' t_a} - b - \frac{\tau'\Delta\mu}{\tau'' t_a}\right)\tau'\left(1 - e^{-t/\tau'}\right),\end{aligned} \tag{17}$$

where we have defined $\tau'^{-1} = \tau^{-1} - T_{\rm roll}^{-1} + T_c^{-1}$, $\tau''^{-1} = \tau'^{-1} - T_c^{-1}$ and $\Delta\mu = \mu_f - \mu_i$. The three parameters are then obtained by fitting the predicted curves from the above equation to the 13 experimentally obtained bank angle transition curves, as shown in Extended Data Figure 5a.

The roll-wise torque is smoothed over a time scale σ_ω using (12) to obtain the equation for the smoothed torque $\tilde{\omega}$:

$$\tilde{\omega}(t) = \frac{\mu - \tilde{\mu}}{\sigma_\omega} + \frac{\tilde{\mu}}{T_s} - (\tilde{\mu}_d - \tilde{\mu})\left(\frac{1}{\tau} - \frac{1}{T_{\rm roll}}\right) + b, \tag{18}$$

where we again use the tilde to denote quantities smoothed over the time scale σ_ω using (12). As in the case of $\boldsymbol{a}_z$, another layer of smoothing of time scale σ'_w is applied. We find that the bias b can change across different flights of the same glider. The bias is estimated on-board before the soaring algorithm is activated by exponentially averaging the torque uncorrected for bias over a time scale of two minutes. Finally, we verify that the estimated ω for different bank angle transitions is indeed unbiased, as shown in Extended Data Figure 5b.

2 Reward shaping and policy invariance

Reinforcement learning algorithms [5] are typically posed in the framework of a Markov Decision Process (MDP). In an MDP, an agent traverses a state space by taking actions while receiving associated rewards. A transition matrix, denoted by $T(s'|s,a)$, gives the probability of transitioning to a particular state s' given the agent's current state s and its current action a and encodes the statistics of the environment and its interactions with the agent. The reward function $R(s,a)$ defines the expected reward given when action a is taken in state s. The agent's control over actions is represented by its policy $\pi(a|s)$, which is the probability that the agent takes action a at state s. The expected discounted sum of future rewards for a particular state-action pair (s,a) is given by the Q function, which is written here in a recursive form:

$$Q_\pi(s,a) = R(s,a) + \gamma \sum_{s',a'} T(s'|s,a)\pi(a'|s')Q_\pi(s',a'). \tag{19}$$

Here γ $(0 \le \gamma < 1)$ is the discount factor, which determines the time scale of future rewards the agent cares about, and the subscript is used to highlight that the Q values depend on the policy π.

To train the glider, we choose the local vertical wind acceleration $\boldsymbol{a}_z$ as our reward function. The choice of $\boldsymbol{a}_z$ as an appropriate reward signal is motivated by observations made in simulations from [6]. In general, multiple reward functions can lead to the same policy, which opens the possibility for *reward shaping*, where a reward function modified from that of the underlying MDP is chosen in order to accelerate the learning process [7]. Reward shaping is particularly useful when the reward is delayed and learning is encumbered by the credit assignment problem. For the purpose of soaring, we aim to maximize the expected gain in height over a time interval of a few minutes. An intuitive choice for the reward function would then be the local vertical wind velocity $\boldsymbol{w}_z$, in which case the the RL algorithm maximizes the quantity $\langle \sum_{i=0}^{\infty} \boldsymbol{w}_z(t_i)\gamma^i \rangle$, where t_i is the time of the ith time step and the angular brackets denote expectation values. In the limit of $\gamma \to 1$, this quantity is the expected gain in height over a time interval $\sim (1-\gamma)^{-1}$. However, we find that the choice of $\boldsymbol{w}_z$ as the reward function fails to drive learning in the soaring problem, possibly because the velocities are strongly correlated across states and their temporal statistics fails to satisfy the Markovian assumption. To justify our choice of $\boldsymbol{a}_z$ as the reward function, we show here that a policy π that is optimal for an MDP with expected reward $R(s,a)$ is also optimal for the same MDP with reward $\propto \langle R(s',a') \rangle_{s,a,\pi} - \gamma R(s,a)$, where $\langle R(s',a') \rangle_{s,a,\pi}$ is the expected reward at the next time step given by

$$\langle R(s',a') \rangle_{s,a,\pi} = \sum_{s',a'} R(s',a') T(s'|s,a) \pi(a'|s'). \tag{20}$$

Intuitively, this implies that using a particular reward function for an MDP is equivalent to using any "derivative" of the reward function as its proxy, where the derivatives are defined in the discounted difference sense as above. We first observe that the policy induces a Markov chain on the MDP defined by the transition probabilities

$$T_\pi(s'|s) = \sum_a T(s'|s,a)\pi(a|s). \tag{21}$$

The key assumption we make here is that the induced Markov chain has a stationary distribution ρ_π given by

$$\rho_\pi(s') = \sum_s T_\pi(s'|s)\rho_\pi(s) \tag{22}$$

The expected sum of future rewards for policy π is then

$$\begin{aligned} \mathcal{V}_\pi &= \sum_{s,a} \rho_\pi(s)\pi(a|s) Q_\pi(s,a), \qquad (23) \\ &= \sum_{s,a} \rho_\pi(s)\pi(a|s) \left(R(s,a) + \gamma \sum_{s',a'} T(s'|s,a)\pi(a'|s') Q_\pi(s',a') \right), \\ &= \mathcal{R}_\pi + \gamma \sum_{s',a'} \rho_\pi(s')\pi(a'|s') Q_\pi(s',a'), \\ &= \mathcal{R}_\pi + \gamma \mathcal{V}_\pi, \end{aligned}$$

where we have defined the expected immediate reward, $\mathcal{R}_\pi = \sum_{s,a} \rho_\pi(s)\pi(a|s)R(s,a)$. The second step above follows from (19) whereas the third step uses the relation (22). We then have

$$\mathcal{R}_\pi = (1-\gamma)\, \mathcal{V}_\pi. \tag{24}$$

We wish to show that the expected sum of future rewards $\tilde{\mathcal{V}}_\pi$ for the MDP with reward function $\langle R(s',a') \rangle_{s,a,\pi} - \gamma R(s,a)$ is directly related to $\mathcal{V}_\pi$. The expected immediate reward $\tilde{\mathcal{R}}_\pi$ for this new

process is given by (from (20))

$$\begin{aligned}\tilde{\mathcal{R}}_\pi &= \sum_{s,a} \rho_\pi(s)\pi(a|s) \left(\sum_{s',a'} T(s'|s,a)\pi(a'|s')R(s',a') - \gamma R(s,a) \right), \\ &= (1-\gamma)\mathcal{R}_\pi,\end{aligned} \tag{25}$$

where the second step is derived in a fashion similar to the third and fourth steps in (23). Since relation (24) also holds for the new MDP, we have $\tilde{\mathcal{R}}_\pi = (1-\gamma)\,\tilde{\mathcal{V}}_\pi$, which yields (together with (24) and (25)) that

$$\tilde{\mathcal{V}}_\pi = (1-\gamma)\mathcal{V}_\pi. \tag{26}$$

The above relation holds for *any* policy. In particular, it holds for the optimal policy π^*, which maximizes $\tilde{\mathcal{V}}_\pi$, and therefore is also the policy that maximizes $\mathcal{V}_\pi$.

3 Noisy gradient sensing in the turbulent atmospheric boundary layer

The navigational cues $\boldsymbol{a}_z$ and ω measure the gradients in the vertical wind velocity along and perpendicular to the heading of the glider. Updrafts and downdrafts are relatively stable structures in a varying turbulent environment. Thermal detection through gradient sensing constitutes a discrimination problem of deciding whether a thermal is present or absent given recent $\boldsymbol{a}_z$ and ω values. In this section, we estimate the magnitude of 'noise' due to turbulence that unavoidably accompanies gradient sensing in the atmospheric boundary layer. Similar estimates of the noise due to the statistical properties of the surrounding physical environment have been made for the sensing of concentration gradients by motile bacteria finding nutrients via chemotaxis [8]. There, the noise arises due to the properties of diffusion; slow diffusion of the few ligands present in the local environment of the cell leads to repeated binding of the same ligands on the cell's receptors, resulting in a biased estimate of the local mean concentration. We consider, as in [8], the case of a 'perfect instrument', which perfectly measures the local vertical wind velocity at its location and across its wings with no accompanying measurement noise and aerodynamics-induced bias.

In the Methods section, we estimate the SNR using simple scaling arguments. Here, we validate our scaling arguments by explicitly computing the SNR for the case of ω estimation. The calculation for $\boldsymbol{a}_z$ estimation is similar and we omit it for the sake of conciseness. Note that the estimate below is still accurate only up to constant factors of order unity.

The instantaneous rate of rotation or 'torque' generated by the vertical component of the fluctuating velocity field $\boldsymbol{w}(\boldsymbol{r},t)$ is given by $\left(\boldsymbol{w}_z^+ - \boldsymbol{w}_z^-\right)/l$, where the $\pm$ superscripts denote the vertical wind velocities on the two wings and we have $\langle \boldsymbol{w}_z^\pm \rangle = 0$. We assume the torque is averaged over a time scale T using an exponential kernel, as in (12). We expect the dependence of the noise on l,V and T to remain invariant to the specific computation performed in integrating the torque; using an exponential kernel is convenient for simplifying the calculations. Suppose the glider moves with a fixed velocity $\boldsymbol{V}$ and the unit vector along the wings is $\hat{\boldsymbol{y}}$ (note that $\boldsymbol{V}$ and $\hat{\boldsymbol{y}}$ are perpendicular to each other). For ease of notation, suppose also that at the final time t the glider is at the origin. We have

$$l\tilde{\omega}(t) = \tilde{\boldsymbol{w}}_z^+(t) - \tilde{\boldsymbol{w}}_z^-(t), \tag{27}$$

where

$$\tilde{\boldsymbol{w}}_z^\pm(t) = \frac{1}{T}\int_{-\infty}^{t} \boldsymbol{w}_z\left(-\boldsymbol{V}(t-s) \pm \frac{l}{2}\hat{\boldsymbol{y}}, s\right) e^{-\frac{t-s}{T}}\, ds. \tag{28}$$

The variance $\delta\tilde{\omega}^2 = \langle\tilde{\omega}^2\rangle$ is

$$l^2\delta\tilde{\omega}^2 = \langle\tilde{\boldsymbol{w}}_z^{+2}\rangle + \langle\tilde{\boldsymbol{w}}_z^{-2}\rangle - 2\langle\tilde{\boldsymbol{w}}_z^{+}\tilde{\boldsymbol{w}}_z^{-}\rangle. \tag{29}$$

We have $\langle\tilde{\boldsymbol{w}}_z^{+2}\rangle = \langle\tilde{\boldsymbol{w}}_z^{-2}\rangle$, where

$$\langle\tilde{\boldsymbol{w}}_z^{+2}\rangle = \frac{1}{T^2}\int_{-\infty}^{t}\int_{-\infty}^{t}\left\langle \boldsymbol{w}_z\left(-\boldsymbol{V}(t-s)+\frac{l}{2}\hat{\boldsymbol{y}},s\right)\boldsymbol{w}_z\left(-\boldsymbol{V}(t-s')+\frac{l}{2}\hat{\boldsymbol{y}},s'\right)\right\rangle e^{-\frac{t-(s+s')/2}{T/2}}dsds', \tag{30}$$

$$= \frac{w^2}{T^2}\int_{-\infty}^{t}\int_{-\infty}^{t}\left(1-\left(\frac{V|s-s'|}{L}\right)^{2/3}\right)e^{-\frac{t-(s+s')/2}{T/2}}dsds', \tag{31}$$

where L is the length scale of the ABL and w is the magnitude of vertical wind velocity fluctuations $\langle\boldsymbol{w}_z^2\rangle^{1/2}$. Here, we have used the two-point velocity correlation function in the turbulent regime, $\langle(\boldsymbol{w}(\boldsymbol{r})-\boldsymbol{w}(\boldsymbol{r'}))^2\rangle \sim |\boldsymbol{r}-\boldsymbol{r'}|^{2/3}$. Further, for V much larger than the velocity scale of the eddies w, any decorrelation of wind velocities is due to the glider's motion. We can then assume that the eddies are frozen in time, which allows us to approximate the spatio-temporal correlations in the equations above using only the spatial component of the two-point correlation function. The above integral is simplified by transforming variables to $p = (s+s')/2, q = (s-s')/2$:

$$\langle\tilde{\boldsymbol{w}}_z^{+2}\rangle = \frac{2w^2}{T^2}\int_{-\infty}^{\infty}\left(1-\left(\frac{2V|q|}{L}\right)^{2/3}\right)dq\int_{-\infty}^{t-|q|}e^{-\frac{t-p}{T/2}}dp \tag{32}$$

$$= \frac{2w^2}{T}\int_{0}^{\infty}\left(1-\left(\frac{2Vq}{L}\right)^{2/3}\right)e^{-\frac{2q}{T}}dq \tag{33}$$

$$= w^2\left(1-\Gamma(5/3)\left(\frac{VT}{L}\right)^{2/3}\right) \tag{34}$$

The calculation of the last term in the RHS of (29) follows in a similar manner. We have

$$\langle\tilde{\boldsymbol{w}}_z^{+}\tilde{\boldsymbol{w}}_z^{-}\rangle = \frac{w^2}{T^2}\int_{-\infty}^{t}\int_{-\infty}^{t}\left(1-\left(\frac{\left(l^2+V^2|s-s'|^2\right)^{1/2}}{L}\right)^{2/3}\right)e^{-\frac{t-(s+s')/2}{T/2}}dsds'. \tag{35}$$

Using the same transformation as above, and performing a straightforward calculation with $q' = 2q/T$, we get

$$\langle\tilde{\boldsymbol{w}}_z^{+}\tilde{\boldsymbol{w}}_z^{-}\rangle = w^2\left(1-\left(\frac{VT}{L}\right)^{2/3}\int_0^{\infty}\left(\alpha^2+q'^2\right)^{1/3}e^{-q'}dq'\right) \tag{36}$$

where we have substituted $\alpha = l/VT$. Combining (29), (32) and (36), we get

$$l^2\delta\tilde{\omega}^2 = 2w^2\left(\frac{VT}{L}\right)^{2/3}\left(\int_0^{\infty}\left(\alpha^2+q'^2\right)^{1/3}e^{-q'}dq' - \Gamma(5/3)\right) \tag{37}$$

The integral above can be found in [10] and is expressed in terms of the Struve functions, $\boldsymbol{H}_\nu$, and the Bessel functions of 2nd kind, N_ν, to get

$$l^2\delta\tilde{\omega}^2 = 2w^2\left(\frac{VT}{L}\right)^{2/3}\left(\alpha^{5/6}2^{-1/6}\Gamma(1/2)\Gamma(4/3)\left(\boldsymbol{H}_{5/6}(\alpha)-N_{5/6}(\alpha)\right)-\Gamma(5/3)\right) \tag{38}$$

For $\alpha \ll 1$, the first terms of the series expansions of $\boldsymbol{H}_\nu$ and N_ν can be used to verify the scaling obtained from the above arguments. It is convenient to express N_ν in terms of the Bessel functions of

the first kind, J_ν, as $N_\nu(x) = (\cos(\nu\pi)J_\nu(x) - J_{-\nu}(x)) \,/\, \sin(\nu\pi)$ and expand $J_{\pm\nu}(x)$ for small x. After a straightforward but lengthy calculation involving Gamma function identities we arrive for $\alpha \ll 1$:

$$l^2\delta\tilde{\omega}^2 = w^2\left(\frac{VT}{L}\right)^{2/3}\alpha^{5/3}\frac{\sqrt{3}\Gamma(1/3)\Gamma(1/6)}{\Gamma(1/2)} \tag{39}$$

The mean vertical velocity difference for a glider travelling tangential to a thermal having profile $\boldsymbol{W}_z$ is $|l\hat{\boldsymbol{y}}.\frac{\partial \boldsymbol{W}_z}{\partial \boldsymbol{r}}| \sim lW/R$ where W is the strength of the thermal and R its size. The signal to noise ratio for $\alpha \ll 1$ is therefore

$$\frac{|l\hat{\boldsymbol{y}}.\partial \boldsymbol{W}_z/\partial \boldsymbol{r}|}{l\delta\tilde{\omega}} \sim \frac{WV^{1/2}T^{1/2}l^{1/6}L^{1/3}}{wR}. \tag{40}$$

Plugging in typical values: $W = 2$ m/s, $R = 50$m, $w = 0.5$ m/s, $l = 2$m, $V = 8$ m/s, $T = 3$s, $L = 1$ km, we obtain an SNR of ~ 4. A similar calculation can be performed for the accelerations. For a glider moving towards a thermal as above, using the arguments above and simple dimensional considerations, we have

$$\frac{|\boldsymbol{V}.\partial \boldsymbol{W}_z/\partial \boldsymbol{r}|}{\delta\tilde{\boldsymbol{a}_z}} \sim \frac{WV^{2/3}T^{2/3}L^{1/3}}{wR}. \tag{41}$$

References

[1] J. D. Anderson Jr., *Introduction to Flight*, McGraw-Hill Education, 7th edition, 2011.

[2] R. von Mises, *Theory of Flight*, Dover Publications, 1st edition, 1959.

[3] R. Stengel, *Flight Dynamics*, Princeton University Press, 1st edition, 2004.

[4] J. R. Garrat, *The Atmospheric Boundary Layer*, Cambridge Atmospheric and Space Science Series, Cambridge University Press, 1994.

[5] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, MIT Press, 1st edition, 1998.

[6] G. Reddy, A. Celani, T. Sejnowski, M. Vergassola, Learning to soar in turbulent environments, *Proc. Natl. Acad. Sci.*, 113(33):E4877-E4884, 2016.

[7] A. Y. Ng, D. Harada, S. J. Russell, Policy Invariance Under Reward Transformations: Theory and Application to Reward Shaping, *Proc. of the 16th International Conference on Machine Learning*, P278-287, 1999.

[8] H. C. Berg, E. M. Purcell, Physics of chemoreception, *Biophysical Journal*, 20(2): 193219, 1977.

[9] U. Frisch, *Turbulence: The Legacy of A. N. Kolmogorov*, Cambridge University Press, 1995.

[10] I. S. Gradshteyn, I. M. Ryzhik, *Table of Integrals, Series, and Products*, ed. D. Zwillinger, Academic Press, 8th edition, 2014.