Application for funding from SATS 2023

Project title: ComVis: a cost-effective end-to-end pipeline for processing camera trap pictures at NINA

Jane U. Jepsen (Senior researcher, Tromsø), **Jenny Stien** (Researcher, Tromsø), **Benjamin Cretois** (Senior engineer IT, Trondheim), **Francesco Frassinelli** (Senior engineer IT, Trondheim), **Signe Christensen-Dalsgaard** (Researcher, Trondheim), **Arild Landa** (Senior researcher, Bergen), **Steven Guidos** (PhD student, Bergen), **John Odden** (Senior researcher, Oslo/*Scandcam*), **Sunniva Bahlk** (Senior technician, Oslo/*Scandcam*), **Lars Rød-Eriksen** (researcher, Trondheim/*National Arctic Fox monitoring*), **Nina Eide** (Senior researcher, Trondheim/*National Arctic Fox monitoring*), **Arnaud Tarroux** (Researcher Tromsø/*Et grønnere NINA*)

Project idea

State-of-the-art

Camera traps are increasingly used to collect data addressing key questions in ecology and nature management. In an appropriate study-design, they provide a non-invasive way of collecting data to infer estimates of e.g. species' distribution, abundance and phenology [e.g. 1, 2-4]. Camera traps reduce observer effects on the target species and provide an unmatched capacity for monitoring biodiversity at spatial-temporal scales unachievable by human observation [5-8]. In addition, they drastically reduce the invasive human footprint in the wider environment, requiring limited maintenance other than deployment and retrieval. Sorting and labelling the large number of pictures collected by camera traps is nevertheless challenging. Both motion and time lapse triggered camera traps return numerous images of no interest and reliable filtering often constitutes the most labour- and financially costly task in collating the picture data. Thereafter, detected objects must be correctly labelled before they are ready for analysis. Recent developments in machine learning (ML), and more specifically deep learning, have made it possible to automatize such workflows, reducing the labour costs. The state-of-the-art classifier *MegaDetector* has been successfully used to detect animals, humans and vehicles in camera trap pictures in a wide diversity of environments [9] and can be used to create a pipeline for both discarding empty pictures and help with labelling the animals [10]. Currently at NINA, no automatic pipeline for sorting camera trap pictures exists and implementing this would greatly reduce the time and financial resources spent on sorting images. Moreover, the recent experience gained in using High Performance Computing (HPC) clusters to run complex machine learning algorithms (i.e. supercomputers; see documentation on SATS Machine Learning and Associated Technologies at NINA; [11]) will benefit this project as the workflow can be sped up to a degree that was not reachable previously.

Knowledge needs

We have identified three primary areas of development which are both timely and essential to address to strengthen NINAs competitive ability and resource use within the field of camera-based monitoring (see *Strategic impacts*). These are i) Development of a streamlined time and energy efficient processing pipeline from raw images to species level detections; ii) Improved standardization of protocols for labelling of images and training of ML models; and iii) Better handling of human bycatch. Given NINA's contribution to cutting edge national and international research, the value of camera trap studies is recognised and their use is expanding. However, there is a lack of standard guidelines on how to manage camera trap pictures, hindering efforts to provide an automated solution. This project will develop a generic pipeline which can be easily used and adapted to a wide range of projects. It uses *MegaDetector* as a base model, with tailored classifiers to fill the steps in the pipeline not encompassed in *MegaDetector*, and includes advocating the use standard protocols and software for labelling data [12]. Given the scale of camera trap deployment in NINA's research, people will inevitably be unsuspectingly caught on camera as human bycatch [13]. These pictures need to be handled accordingly to prevent a breach in privacy (Lover og regler | Datatilsynet). This may involve deletion of unwanted bycatch [13], or anonymization, in cases where camera traps are used to address questions of human disturbance on wildlife [14]. Currently, within NINA, the majority of camera-–based monitoring programs rely on human observers to identify and delete human bycatch in image collections prior to image processing. There is therefore an urgent need for efficient, automated procedures to detect and appropriately handle such images.

Novelty and ambition

The overall ambition of the project is to initiate a more efficient use of state-of-art machine learning tools, that will streamline the processing of camera-based monitoring data in NINA. We will achieve this by providing an end-to-end pipeline for processing of camera trap data based on a combination of an established model (*MegaDetector*) and customized classifiers trained on data from NINAs monitoring programs. In doing so we will i) Contribute to standardization of image data processing across projects; ii) Reduce time consumption and costs of implementing camera-based monitoring design in current and future projects; and iii) Provide a case for how NINA can measure the footprint of image processing, so that an optimal route for minimal energy consumption can be developed. We believe the potential gain for NINA as an institute is substantial. For example, a recent test of a *MegaDetector* based pipeline indicated that processing speed increased by 500% compared to manual classification, and time in manual steps decreased more than 8-fold [14].

Strategic impact

Societal impact of the project

Evidence of the potential of artificial intelligence (AI) technologies to affect Sustainable Development Goals (SDGs) in both positive and negative terms is a topic of much attention in the literature [15, 16], including AI as an enabler in reaching environmental SDGs. Use of computer vision to streamline environmental monitoring is part of this package, and serves foremost to reduce the environmental foot print [17] of field-based activities, while simultaneously permitting data to be collected at temporal and spatial scales unachievable by means of human observers alone. However, camera-based monitoring also holds an environmental cost both on the production and deployment side (e.g. production of sensors and batteries, deployment and maintenance of sensors in the field), and on the processing side (e.g. energy consumption of running the algorithms). ML workloads have high energy requirements, and there is a growing literature on 'best practices' of designing workloads in order to minimize the carbon footprint, including a recent suggestion that ML based research papers should include emission statistics in order to foster competition on carbon efficiency in addition to model performance [18]. We will implement a monitor of the energy consumption of our proposed pipeline and provide an example of how such consumption budgets may be presented in ML based research publications from NINA.

Strategic contribution of the project to NINA

Machine learning and computer vision is rapidly transforming the impact of camera based monitoring on ecology, as it permits a drastic increase in the amounts and types of data which can be processed, and hence the types of ecological questions which can be addressed [19]. This is also influencing the market in which NINA operates, and the expectations by which we are met by stakeholders and funding bodies. A good example is the new national monitoring program for small rodents in Norway currently in planning which is envisaged to be entirely camera based [20], and build on both trap design and image classifiers developed in the small rodent module in COAT [21]. Other good examples are national level monitoring of carnivores such as lynx and arctic foxes which rely heavily on camera-based techniques. NINA needs to be at the forefront in handling, processing and analysis of image-based monitoring data and be able to incorporate such designs in projects at short notice. This pilot project contributes directly to lowering the threshold for using ML tools. We build on a current SATS project which address the infrastructure available within and outside NINA [11] and we present a generic pipeline for analysis which NINA researchers can tailor to their own needs. By pioneering a discussion of energy efficiency of ML workflows we contribute directly to an operationalization of NINAs environmental goals which states that NINA shall *map our environmental footprint and implement measures that reduce this* (NINAs mission statement 2020-2024).

Scientific impact of the project

The distillation of relevant ecological data, and ultimately ecological understanding, from distributed sensors (such as camera traps) is lagging behind the growing use and hence the accumulation of raw measurements (such as images). Lack of automatization, inefficient processing workflows, but also an insufficient merger between expertise from the two domains of ecology and computer science, are among the prime reasons for this [9]. We believe NINA is in an excellent position to make advances in this field drawing on a close integration between computer technical and ecological expertise and a portfolio of projects from a wide range of ecological contexts which implement camera traps. We argue that the immediate impact of this pilot will be mostly within institution in the form of more cost effective and integrated workflows, but believe that the project is also in line with advancements currently being made internationally [9, 10, 14]. This should secure that advances made within this project, also have an impact at an international level.

Implementation

Proposed activities

The project is a methodological pilot study of 12 months duration. Central to the project is the identification of a set of case studies (Table 1). The cases represent typical examples of how camera-based monitoring is used within NINA's research and monitoring projects. They have been selected to represent a range of habitat (terrestrial, coastal, marine), taxa (mammals, birds) and image acquisition modes (passive/active camera traps, motion trigger/timelapse) relevant to camera trap activities currently in NINA, as well as in the perceived methodological challenges involved in detecting animal objects in the images. An example set of images will be drawn from each of these case studies and serve as input data for training and testing the image classifiers developed during this project.

Table 1. *Summary of the four cases selected as examples of camera-based monitoring designs employed in NINA. Image acquisition design is broadly characterized as either* active *or* passive *depending on the presence of bait which may actively attract animals to the trap, and* motion *or* timelapse *depending on whether or not the camera use a motion trigger, a timelapse trigger or both*.

Case	Target species	Bycatch species	Habitat	Image acquisition	Contact
Case A: Forest ungulates	reindeer, moose	hare, roe deer, sheep, fox, terrestrial birds	birch forest	passive, motion/timelapse	JUJ
Case B: Large predators	wolverine, white-tailed eagle, golden eagle	raven, crow, stoat, weasel, pine marten, terrestrial birds	birch forest, alpine	active, motion/timelapse	JS
Case C: Predation on kittiwake nests	kittiwake, white-tailed eagle, raven, crow	peregrine falcon, otter	bird cliff	passive, motion/timelapse	SCD
Case D: Mesopredators	otter, mink	red fox, small rodents, terrestrial birds	riverine/coastal	passive, motion	SG

The activities in the project are organized in four consecutive stages (Fig. 1), where stages 1-3 are focused on methodological developments and stage 4 on documentation and reproducibility of the developments. We summarize each stage briefly below and refer to Table 2 for specific deliverables.

Stage 1: Image quality filtering: All camera-based monitoring designs must handle a certain proportion of images of such low quality that further processing is impossible because the target object would not be visible even if it was within the camera frame (Fig. 1; 'unusable' vs 'usable' images). Unusable images need to be removed from the image sample prior to processing and this is currently a manual, time-consuming process. For human observers scoring usability can be extremely challenging and requires calibration between observers to achieve a consistent result across observers. Further the 'optimal' threshold will depend on the intended use of the image data. An automatic classifier however, offers much more flexibility and can in theory be trained on multiple thresholds. During stage 1 we will develop an automatic quality filter, trained using images from all four cases, and evaluated based on its ability to classify images into two categories; 'usable' and 'unusable' images for each case. We will involve all project participants in a discussion of quality threshold definitions with the aim of providing guidelines on how to arrive at rigorous definitions of the quality categories. The performance of the classifier will be assessed relative to a manual classifications made by multiple human observers. The sample of automatically filtered 'usable' images from each case will constitute the input data for stage 2.

Stage 2. Image target filtering with *MegaDetector*: Removing empty camera trap pictures and annotating the remaining images with species present is the most time-consuming step if done manually but can be automatized using existing tools. *MegaDetector*, a machine learning model whose task is to detect animals on camera trap pictures, is now widely used in the conservation community and could provide a cost-efficient alternative to manual tagging. In stage 2 (Fig. 1) we will construct an end-to-end pipeline using *MegaDetector* for automatic detection of species of camera trap pictures for 1) local processing of less demanding tasks (NINAs external server) and 2) for remote processing when high computational resources are required (SAGA/NIRD). Because *MegaDetector* has mainly been trained with camera traps collected in African and US national parks we will also assess the performance of *MegaDetector* on the camera trap pictures collected in the Norwegian landscape. We will manually label 1,000 camera trap pictures collected from each case (Table 1) and compute the precision and recall of *MegaDetector*.

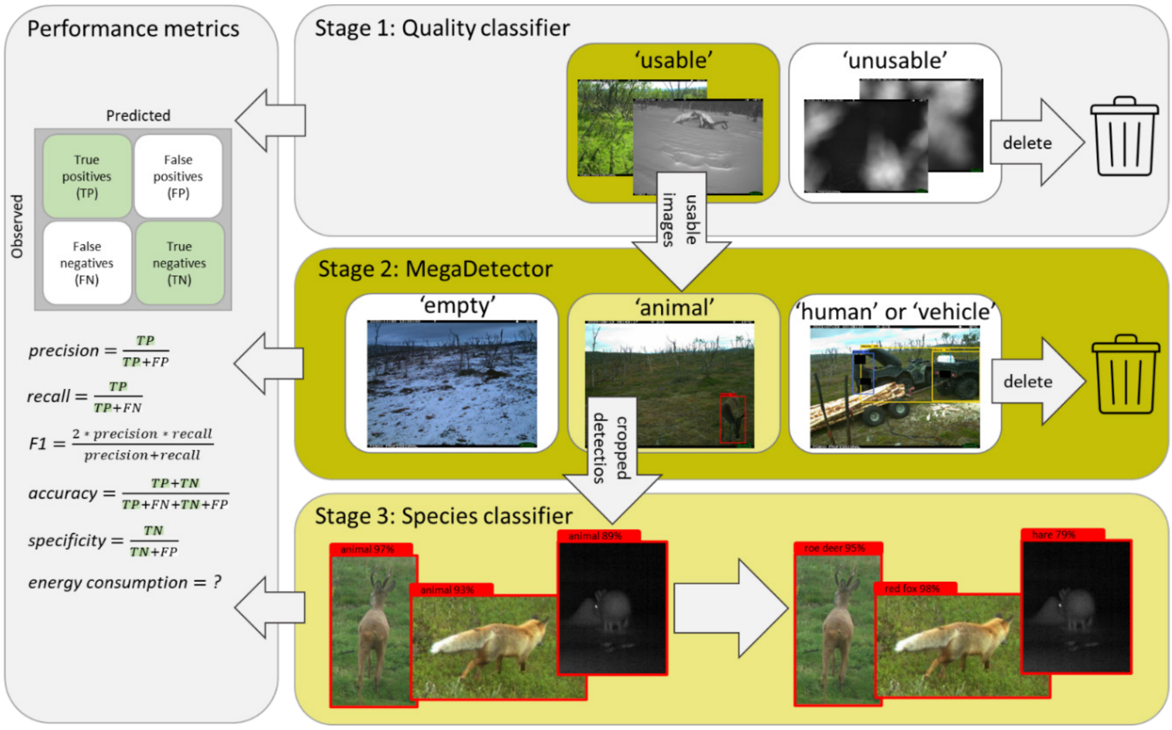

Figure 1. *Graphical summary of the work-flow through stages 1-3. All performance metrics listed except energy consumption are well established in the ML/deep learning literature.*

Stage 3. Species-level annotation and identification: *MegaDetector* detects animals on pictures and does not classify into taxa. In most applications however, the taxon of the animal is needed (e.g. indexes of species richness, species abundance or occurrence). We will develop the pipeline by using the animal detections of *MegaDetector* for training a custom species classifier (Fig. 1) for one of the cases in Table 1.

Stage 4. Documentation and reproducibility: Scripts and a pipeline documentation will be stored in a GitHub repository so that it is possible to reproduce the entire pipeline. Moreover, the goal of this project is to provide tools accessible to everyone at NINA to process project camera trap pictures and as such, user-friendly scripts will be provided. In addition to storing the code on GitHub we will document the findings of this project, including the performance statistics, in a scientific publication. A scientific paper, however, is outside the scope of a 1-year pilot and will be achieved using FU-contributions from the authors.

Organization

This project is initiated by ecologists with strong interests [4, 22-25] in image-based research and monitoring (JUJ, JS, SCD, AL) and a data scientist (BC) with the required expertise [26] in deep learning and associated tools. JUJ, JS, SCD and AL currently have responsibility for image-based monitoring of terrestrial herbivores and carnivores as well as seabirds which supply three of the test cases for this pilot project. In addition, the core group includes an early career researcher (SG) engaged in camera-based monitoring in NINA as part of his PhD-project (supervised by AL). This work supplies the fourth test case for the project. We will engage several other NINA researchers as discussion partners and reviewers of our activities; AT will be engaged in discussions related to energy consumptions budgets for ML workflows as a representative for *Et Grønnere NINA*. SB and JO will be engaged as representatives *ScandCam* the largest camera-based monitoring program currently run within NINA. Discussions with FF will ensure that the current project accomodate previous technical developments made to process camera data from *ScandCam*. LR and NEE bring in similar expertise as SB and JO, but from the perspective of another of NINA's large camera-based monitoring initiatives; the Arctic fox monitoring program. The pipeline will be presented in a NINA seminar towards the end of 2023, encouraging the wider NINA community to challenge and provide feedback on the developed pipeline with regards to their own camera-based monitoring data.

<u>Milestones and deliverables</u>

The main milestones and deliverables of the project are outlined below. Stage 1 and 2 will be performed for all cases. Given that this proposal is for a pilot project and therefore of short duration, Stage 3 will be performed for one or two of the most illustrative cases. We will use github for both documentation and reporting on model performance to ensure that our workflow can be adopted and developed by others within NINA and the wider scientific community.

Table 2. *Proposed quarterly timing of project Stage 1-4 in 2023 with associated deliverables.*

Stage	Deliverables	Q-1	Q-2	Q-3	Q-4
1	1. A trained classifier which separates images into two classes; 'unusable' and 'usable'. 2. Performance statistics of the classifier. 3. A set of 'usable' images for each case for use in Stage 2.	X			
2	1. Deployment of *MegaDetector* on a HPC 2. A classification of the 'usable' sample for each case into the four *MegaDetector* classes; 'empty', 'animal', 'human' and 'vehicle'. 3. Performance statistics of the classifier. 4. Cropped image sections of all animal objects detected by *MegaDetector* for use in Stage 3.	X	X		
3	1. A trained classifier which classifies cropped image sections of animal objects to species for selected cases. 2. Performance statistics of the classifier.			X	
4	1. Written documentation on project repository on NINA's github 2. 1st draft of a scientific paper*				X

* not funded by the SATS contribution.

Budget

Researchers each contribute between 2 and 5 days of FU time for image sample preparation (JS, JUJ, SCD) and project management (JUJ). The funds will be allocated approximately as follows between stages in the project: Project management: 15.000 (100% JUJ); Image sample preparation and manual annotation: 150.000 (25% JS, 25% JUJ, 25% SCD, 25% AL); Stage 1 (Quality classifier): 50.000 (100% BC); State 2 (*MegaDetector*): 200.000 (60% BC, 13% JS, 13% JUJ, 13% SCD); Stage 3 (Species classifier): 150.000 (70% BC, 30% to contact person(s) for selected case(s)); Stage 4 (Documentation): 35.000.

Table 3 *Project budget.*

	2023	2024	SUM
Timekostnader NINA, nfr-satser	600.000	0	600.000
Reiseutgifter	0	0	0
Innkjøp, utstyr	0	0	0

Andre driftskostnader	0	0	0
Utgifter totalt	600.000	0	600.000
Bidrag fra FU	100.000	0	100.000
Bidrag fra andre kilder	0	0	0
Sum egenfinansiering	100.000	0	100.000
Søknadssum sats	500.000	0	500.000

References

1. Burton AC, Neilson E, Moreira D, Ladle A, Steenweg R, Fisher JT, Bayne E, Boutin S: **Wildlife camera trapping: a review and recommendations for linking surveys to ecological processes**. *Journal of Applied Ecology* 2015, **52**(3):675-685.
2. Chen C, Brodie JF, Kays R, Davies TJ, Liu RZ, Fisher JT, Ahumada J, McShea W, Sheil D, Agwanda B *et al*: **Global camera trap synthesis highlights the importance of protected areas in maintaining mammal diversity**. *Conservation Letters* 2022, **15**(2).
3. Henden JA, Stien A, Bardsen BJ, Yoccoz NG, Ims RA: **Community-wide mesocarnivore response to partial ungulate migration**. *Journal of Applied Ecology* 2014, **51**(6):1525-1533.
4. Mann HMR, Iosifidis A, Jepsen JU, Welker JM, Loonen M, Hoye TT: **Automatic flower detection and phenology monitoring using time-lapse cameras and deep learning**. *Remote Sensing in Ecology and Conservation* 2022.
5. Jachowski DS, Katzner T, Rodrigue JL, Ford WM: **Monitoring Landscape-Level Distribution and Migration Phenology of Raptors Using a Volunteer Camera-Trap Network**. *Wildlife Society Bulletin* 2015, **39**(3):553-563.
6. Rich LN, Davis CL, Farris ZJ, Miller DAW, Tucker JM, Hamel S, Farhadinia MS, Steenweg R, Di Bitetti MS, Thapa K *et al*: **Assessing global patterns in mammalian carnivore occupancy and richness by integrating local camera trap surveys**. *Global Ecology and Biogeography* 2017, **26**(8):918-929.
7. Steenweg R, Hebblewhite M, Kays R, Ahumada J, Fisher JT, Burton C, Townsend SE, Carbone C, Rowcliffe JM, Whittington J *et al*: **Scaling-up camera traps: monitoring the planet's biodiversity with networks of remote sensors**. *Frontiers in Ecology and the Environment* 2017, **15**(1):26-34.
8. Nickel BA, Suraci JP, Allen ML, Wilmers CC: **Human presence and human footprint have non-equivalent effects on wildlife spatiotemporal habitat use**. *Biological Conservation* 2020, **241**.
9. Tuia D, Kellenberger B, Beery S, Costelloe BR, Zuffi S, Risse B, Mathis A, Mathis MW, van Langevelde F, Burghardt T *et al*: **Perspectives in machine learning for wildlife conservation**. *Nature Communications* 2022, **13**(1).
10. Rigoudy N, Benyoub A, Besnard A, Birck C, Bollet Y, Bunz Y, De Backer N, Caussimont G, Delestrade A, Dispan L *et al*: **The DeepFaune initiative: a collaborative effort towards the automatic identification of the French fauna in camera-trap images**. 2022:2022.2003.2015.484324.
11. **Documentation on using HPC and containers for Machine Learning. Accessed via https://ninanor.github.io/91126800_ML_and_associated_tech/**
12. Kellenberger B, Tuia D, Morris D: **AIDE: Accelerating image-based ecological surveys with interactive machine learning**. *Methods in Ecology and Evolution* 2020, **11**(12):1716-1727.
13. Sandbrook C, Luque-Lora R, Adams WM: **Human Bycatch: Conservation Surveillance and the Social Implications of Camera Traps**. *Conservation & Society* 2018, **16**(4):493-504.
14. Fennell M, Beirne C, Burton AC: **Use of object detection in camera trap image identification: Assessing a method to rapidly and accurately classify human and animal detections for research and application in recreation ecology**. *Global Ecology and Conservation* 2022, **35**.
15. Vinuesa R, Azizpour H, Leite I, Balaam M, Dignum V, Domisch S, Fellander A, Langhans SD, Tegmark M, Nerini FF: **The role of artificial intelligence in achieving the Sustainable Development Goals**. *Nature Communications* 2020, **11**(1).
16. Skiter NN, Rogachev AF, Ketko NV, Simonov AB, Tarasova IA: **Sustainable Development of Enterprises in Conditions of Smart Ecology: Analysis of The Main Problems and Development of Ways to Solve Them, Based on Artificial Intelligence Methods and Innovative Technologies**. *Frontiers in Environmental Science* 2022, **10**.
17. Gillings S, Harris SJ: **Estimating the carbon footprint of citizen science biodiversity monitoring**. *People and Nature* 2022, **4**(4):996-1006.
18. Patterson D, Gonzalez J, Hölzle U, Le Q, Liang C, Munguia LM, Rothchild D, So DR, Texier M, Dean J: **The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink**. *Computer* 2022, **55**(7):18-28.
19. Høye TT, Arje J, Bjerge K, Hansen OLP, Iosifidis A, Leese F, Mann HMR, Meissner K, Melvad C, Raitoharju J: **Deep learning and computer vision will transform entomology**. *Proceedings of the National Academy of Sciences of the United States of America* 2021, **118**(2).
20. Kleiven E, Framstad E, Bakkestuen V, Böhner H, Cretois B, Frassinelli F, Ims R, Jepsen J, Soininen E, Eide NE: **Ny nasjonal smågnagerovervåking i fjell basert på kamerafeller. Forslag til innsamlingsdesign og dataprosessering**. *NINA Rapport nr 2170 Norsk institutt for naturforskning (NINA), Trondheim* 2022.
21. Kleiven E: **Population cycles in small rodents seen through the lens of a wildlife camera**. *PhD Thesis.* Tromsø, Norway: UiT Arctic University of Norway; 2022.
22. Agersborg JA, Luppino LT, Anfinsen SN, Jepsen JU: **Toward Targeted Change Detection with Heterogeneous Remote Sensing Images for Forest Mortality Mapping**. *Canadian Journal of Remote Sensing* 2022:1-23.
23. Stien J, Ims RA: **Absence from the nest due to human disturbance induces higher nest predation risk than natural recesses in Common Eiders Somateria mollissima**. *Ibis* 2016, **158**(2):249-260.
24. Stien J, Stien A, Tveraa T, Rod-Eriksen L, Eide NE, Killengreen ST: **Estimating abundance in unmarked populations of Golden Eagle (Aquila chrysaetos)**. *Ecological Solutions and Evidence* 2022, **3**(3).

25. Heggøy O, Christensen-Dalsgaard S, Ranke PS, Chastel O, Bech C: **GPS-loggers influence behaviour and physiology in the black-legged kittiwake Rissa tridactyla**. *Marine Ecology Progress Series* 2015, **521**:237-248.
26. Cretois B, Rosten CM, Sethi SS: **Voice activity detection in eco-acoustic data enables privacy protection and is a proxy for human disturbance**. *Methods in Ecology and Evolution* 2021.