Climate Impact Partners

August 2023.

Abstract

This report presents an analysis of tree growth metrics, specifically focusing on age and height, using data from various botanical names and growth rates.

A significant enhancement in this iteration is the use of the Nevergrad library for optimisation. Analyses conducted with Nevergrad have shown marked improvements in performance, yielding more accurate results compared to other methods.

Boxplots are employed to display the distribution of tree age and height for each botanical name, with the potential for differentiation based on growth rates. Additionally, an interactive species panel is integrated into the report, facilitating detailed exploration of data for specific species.

All results, visual representations, and interactive modules are systematically saved in timestamped directories for traceability and future analyses. By leveraging the capabilities of Nevergrad, this report offers a comprehensive and precise understanding of tree growth metrics.

Charles Shaw (charles@fixedpoint.io)

Contents

0 Documentation Control **4**

1 Introduction **5**

2 ARTEMIS: An Overview of Capabilities and Performance **5**

2.1 NGOpt: Adaptive Algorithm Selection in ARTEMIS 6

2.2 Interpreting the 3D Surface Plots for Optimization Landscape 7

3 Data Handling and Pre-processing **8**

4 Modelling and Growth Functions **8**

5 Optimization Techniques **9**

6 Interactive Features and User Experience **9**

0 Documentation Control

Version	Date	Reason for issue	Issued By
1	26-Jul-23	Initial version	Charles Shaw
2	28-Jul-23	added Section 2	Charles Shaw
3	12-Aug-23	Substantive rewrite of all sections	Charles Shaw

1 Introduction

ARTEMIS (Advanced Regression and Tree Estimation Model for Integrated Silviology) is a Python package developed for precise forestry statistics. Created by FixedPoint IO Ltd for Climate Impact Partners, ARTEMIS integrates the Chapman-Richards growth function, renowned for its accuracy in predicting tree growth. Additionally, the package employs Nevergrad, an advanced optimization library, to process forestry data that can often be affected by noise and outliers.

A notable feature of ARTEMIS is its ability to use prior information, or "priors," during the optimization process. This capability ensures that the optimization aligns with existing scientific knowledge, especially in the context of noisy forestry data. By enabling users to input priors based on empirical evidence, ARTEMIS produces results that combine domain expertise with robust statistical optimization.

This document provides a comprehensive overview of ARTEMIS, detailing its features, methodologies, and applications in the field of forestry analytics.

2 ARTEMIS: An Overview of Capabilities and Performance

ARTEMIS, an acronym for *Advanced Regression and Tree Estimation Model for Integrated Silviology*, is a Python package specifically tailored for forestry statistics. Its primary goal is to deliver precise estimations of tree and forest growth variables. Given the inherent complexities and variabilities present in forestry data, sophisticated techniques are indispensable. ARTEMIS addresses this need by leveraging some of the most advanced algorithms in the domain.

A hallmark of ARTEMIS is its adoption of the **Chapman-Richards growth function**. This empirically grounded function has consistently demonstrated efficacy in predicting tree growth across varied contexts. The integration of this function ensures that ARTEMIS's estimations are anchored in established methodologies.

However, the true differentiator for ARTEMIS is its use of state-of-the-art optimization techniques. Specifically, ARTEMIS employs:

- **Differential Evolution (DE)**: An evolutionary algorithm renowned for its prowess in global optimization. DE operates by generating mutant vectors and recombination, ensuring a comprehensive search for the optimal solution. This approach is especially adept at managing multi-modal functions frequently observed in forestry datasets.
- **Genetic Algorithms (GAs)**: Search heuristics inspired by the natural selection process, GAs emulate evolutionary processes to identify approximate solutions to optimization and search challenges. Through mechanisms like selection, crossover, and mutation, GAs ensure ARTEMIS's capability to navigate expansive solution spaces in forestry datasets, converging to optimal or near-optimal solutions.

The synergy of the Chapman-Richards growth function with advanced optimization strategies bestows ARTEMIS with exceptional performance. This ensures that ARTEMIS not only yields accurate estimations but maintains this precision consistently, mitigating the risk of anomalous results. For stakeholders, this equates to dependable insights and forecasts, vital for informed decision-making in silviology.

2.1 NGOpt: Adaptive Algorithm Selection in ARTEMIS

A pivotal component enhancing the capabilities of ARTEMIS is its use of **NGOpt**, a meta-optimizer within the *Nevergrad* optimization library. 'NGOpt' stands out for its adaptive algorithm selection approach, enabling ARTEMIS to determine the most suitable optimization strategy for the given task.

The forestry domain presents a myriad of optimization challenges, ranging from multi-modal functions to high-dimensional spaces, each requiring distinct algorithmic strategies. Instead of relying on a single optimization method, 'NGOpt' assesses the nature of the problem and dynamically selects the most effective algorithm from its extensive repertoire, which includes but is not limited to Differential Evolution, Genetic Algorithms, and other evolutionary strategies.

This adaptive approach ensures that ARTEMIS remains versatile, effectively tackling a wide array of optimization problems inherent to forestry data. By leveraging 'NGOpt', ARTEMIS optimizes its performance across tasks, providing more accurate and consistent results. The integration of this meta-optimizer exemplifies ARTEMIS's commitment to harnessing state-of-the-art techniques, ensuring that the tool remains at the forefront of forestry analytics.

In the landscape of automatic optimization, challenges emerge from the diverse requirements of real-world problems. These requirements range from the intricacies of problem models to the computational resources at hand. Addressing these challenges requires algorithm selection wizards—tools that are versatile, robust, and adept at selecting the most effective algorithm for a specific problem instance.

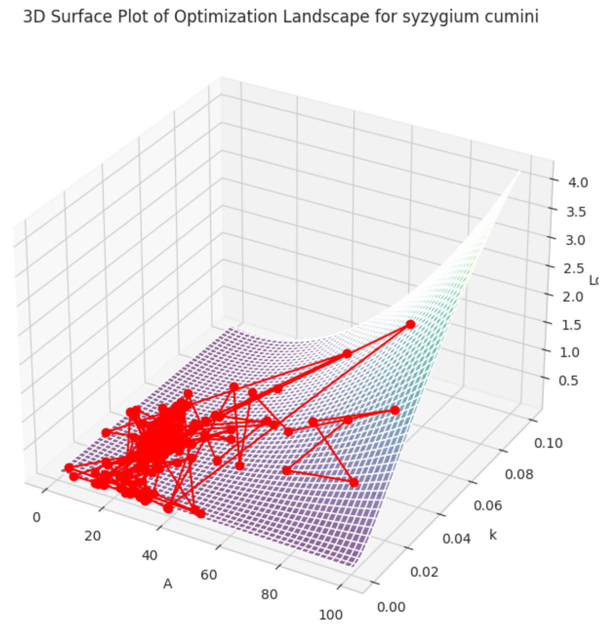
The creation of a competitive algorithm selection wizard is a daunting task. It necessitates not only defining the rules for algorithm selection but also configuring the parameters of the selectable algorithms—a challenge in its own right. While automated wizards have been crafted for specific domains like SAT problems, many algorithm selection tools, especially those designed for broader applications, are hand-crafted.

Among these, **NGOpt** stands out as a paragon of hand-crafted excellence. Integrated within the *Nevergrad* optimization platform, NGOpt is a product of meticulous research and iterative refinement. Its design was informed by a thorough evaluation of the performance of numerous optimizers across diverse benchmark suites. Based on these insights, hand-crafted rules were devised to strategically select the best optimizer tailored to specific problem features.

The evolution of NGOpt involved iterative enhancements. Over time, certain rules were refined, and specific optimizers were replaced to enhance performance. This rigorous and iterative process birthed a sophisticated algorithm selection wizard. The resultant NGOpt not only embodies the complexity of its design process but also showcases superior performance. When pitted against renowned standalone optimizers across various benchmark suites, NGOpt consistently emerges as a top performer.

For ARTEMIS, the integration of NGOpt means access to a tool that is not just versatile but also empirically validated. By leveraging NGOpt's capabilities, ARTEMIS ensures that its optimization processes are always aligned with the best available strategies, guaranteeing optimal outcomes across diverse forestry challenges.

Figure 1: 3D surface plots for optimization landscape:



2.2 Interpreting the 3D Surface Plots for Optimization Landscape

1. **Landscape Contours**:
   - The plot represents a three-dimensional space where two axes correspond to the model parameters (A and k in this case) and the third axis (vertical) represents the loss (error or objective function value).
   - The 'height' or 'elevation' at any point on this plot represents the error/loss associated with a specific combination of parameters A and k.
   - Low areas or valleys represent parameter combinations with low error, whereas peaks or mountains represent parameter combinations with high error.

2. **Optimizer Path**:
   - The red line and markers on the plot represent the path taken by the optimization algorithm.
   - Starting from an initial guess, the optimizer tries different combinations of parameters in its quest to find the lowest point on the landscape (the best parameters).
   - By following the path, you can see how the optimizer 'searched' through the parameter space.

3. **Global vs Local Minima**:
   - A critical feature to look out for in such plots is the presence of multiple low areas (valleys). This indicates that there may be multiple sets of parameters that result in a similar low error.

- The true goal of the optimizer is to find the global minimum (the absolute lowest point across the entire landscape). However, sometimes optimizers can get stuck in local minima (a point lower than its surroundings, but not the lowest overall).
- If you see multiple valleys and the optimizer's path ends in one of them, it might be stuck in a local minimum. It's essential to know this as there might be better parameter sets the optimizer did not explore.

4. **Initial Guess and Convergence**:
   - The starting point of the optimizer path indicates the initial guess for the parameters.
   - The end of the path represents the best parameters found by the optimizer.
   - If the path shows a clear and steady descent into a valley, it indicates good convergence of the optimization algorithm. If the path seems erratic or doesn't settle into a valley, it might suggest challenges with optimization convergence.
5. **Complexity of Landscape**:
   - A smooth landscape with a clear single valley is typically easier for optimization algorithms. A rugged landscape with multiple valleys and peaks suggests a more complex optimization problem, which may require more sophisticated or randomized optimization techniques.

3 Data Handling and Pre-processing

Forestry data, inherently complex and often gathered from varied sources, requires meticulous handling and preprocessing. ARTEMIS addresses this need head-on. Its advanced data cleaning scripts, tailored specifically for forestry datasets, tackle a plethora of common issues. These scripts systematically identify and rectify missing values, spurious outliers, and inconsistent data entries. Beyond mere cleaning, ARTEMIS also offers features for data transformation, ensuring that the data conforms to the requirements of the subsequent analytical steps.

The significance of this rigorous preprocessing cannot be overstated. In the realm of forestry, even minor data inconsistencies can lead to substantial discrepancies in predictions. By ensuring that the input data is of the highest quality, ARTEMIS substantially enhances the reliability of its analytics, laying a strong foundation for all subsequent analyses.

4 Modelling and Growth Functions

Central to ARTEMIS's analytical engine is the Chapman-Richards growth function. This mathematical expression, grounded in empirical research, has consistently demonstrated its ability to capture the intricacies of tree growth across diverse contexts. The function delineates the sigmoidal growth curve commonly observed in trees, encapsulating the nuances of the rapid initial growth which gradually tapers as trees approach maturity.

While various growth functions exist in silviology, the Chapman-Richards function's empirical accuracy and theoretical soundness make it the cornerstone of ARTEMIS's modelling strategy. By anchoring its predictions in this well-established function, ARTEMIS ensures that its outputs resonate with both empirical observations and theoretical expectations.

5 Optimization Techniques

In the face of complex and often noisy forestry data, mere modelling is insufficient. The models need to be fine-tuned to the data, a task that demands sophisticated optimization techniques. ARTEMIS, understanding this imperative, integrates the Nevergrad library—a cutting-edge optimization toolbox. Within Nevergrad, ARTEMIS predominantly employs NGOpt, a versatile meta-optimizer.

NGOpt stands out for its adaptive algorithm selection. Instead of being confined to a single optimization strategy, NGOpt assesses the problem at hand and judiciously selects the most effective algorithm, be it Differential Evolution, Genetic Algorithms, or other advanced techniques. This adaptability ensures that ARTEMIS's optimization process is always attuned to the specific challenges of the dataset, yielding optimal or near-optimal solutions consistently.

6 Interactive Features and User Experience

Beyond its analytical capabilities, ARTEMIS prioritises user engagement and experience. The software's design integrates an interactive species panel, a feature that revolutionises data exploration. This panel offers real-time, species-specific data visualisation, allowing users to glean insights with unprecedented granularity. Such interactive features transform the user's engagement from passive observation to active exploration, fostering a deeper understanding of the data.

Furthermore, ARTEMIS's intuitive interface ensures that its advanced capabilities are accessible to a wide audience. The streamlined design, coupled with comprehensive documentation, ensures that users, irrespective of their technical proficiency, can navigate the software, harness its features, and extract meaningful insights with ease.