Species Distribution Models: Models and Application for Sustainable Species Conservation Day 2

Koissi Savi (Ph.D.)

Founder and CEO of HealthDataInsights

April 27, 2024

1 Understanding Climate Change Data: IPCC Reports, Climate Models, and Projections

2 Notion of Equilibrium and Habitat Saturation

3 Scales of climate models

4 Timescales of future projections

5 climate change scenario IPCC, 2021

6 Downscaling Climate Data for SDMs: Bias Correction and Spatial Interpolation

Climate change models

- **Global climate models (GCMs)** are the most widely used method to understand what the climate may be like in the future as a result of emissions of greenhouse gases (global warming).

Climate change models

- **Global climate models (GCMs)** are the most widely used method to understand what the climate may be like in the future as a result of emissions of greenhouse gases (global warming).
- Because they work at a global scale, the resolution of GCM results is typically quite coarse. Each grid cell is roughly 200 Œ 200 km.

Climate change models

- **Global climate models (GCMs)** are the most widely used method to understand what the climate may be like in the future as a result of emissions of greenhouse gases (global warming).
- Because they work at a global scale, the resolution of GCM results is typically quite coarse. Each grid cell is roughly 200 Œ 200 km.
- **Regional climate models (RCMs)** are applied to smaller spatial areas to produce results with greater local detail. However, RCMs still rely on GCMs for input data and therefore are not necessarily more reliable or more accurate.

Climate change models

- **Global climate models (GCMs)** are the most widely used method to understand what the climate may be like in the future as a result of emissions of greenhouse gases (global warming).
- Because they work at a global scale, the resolution of GCM results is typically quite coarse. Each grid cell is roughly 200 Œ 200 km.
- **Regional climate models (RCMs)** are applied to smaller spatial areas to produce results with greater local detail. However, RCMs still rely on GCMs for input data and therefore are not necessarily more reliable or more accurate.
- Climate models produce numerical values for key climate parameters, such as temperature, humidity and wind speed, for specific points and at different levels on the Earths surface and in the atmosphere and oceans.

Climate change models

- **Global climate models (GCMs)** are the most widely used method to understand what the climate may be like in the future as a result of emissions of greenhouse gases (global warming).
- Because they work at a global scale, the resolution of GCM results is typically quite coarse. Each grid cell is roughly 200 Œ 200 km.
- **Regional climate models (RCMs)** are applied to smaller spatial areas to produce results with greater local detail. However, RCMs still rely on GCMs for input data and therefore are not necessarily more reliable or more accurate.
- Climate models produce numerical values for key climate parameters, such as temperature, humidity and wind speed, for specific points and at different levels on the Earths surface and in the atmosphere and oceans.
- One major concern is that the spatial scale of climate data used in most SDMs may be inadequate due to the coarse resolution of global climate models (GCM) or downscaled derivatives of those models (Seo et al., 2009)

Climate change models

- **Global climate models (GCMs)** are the most widely used method to understand what the climate may be like in the future as a result of emissions of greenhouse gases (global warming).
- Because they work at a global scale, the resolution of GCM results is typically quite coarse. Each grid cell is roughly 200 Œ 200 km.
- **Regional climate models (RCMs)** are applied to smaller spatial areas to produce results with greater local detail. However, RCMs still rely on GCMs for input data and therefore are not necessarily more reliable or more accurate.
- Climate models produce numerical values for key climate parameters, such as temperature, humidity and wind speed, for specific points and at different levels on the Earths surface and in the atmosphere and oceans.
- One major concern is that the spatial scale of climate data used in most SDMs may be inadequate due to the coarse resolution of global climate models (GCM) or downscaled derivatives of those models (Seo et al., 2009)
- These local topoclimatic habitats may need to be accounted for when evaluating climate-change impacts on biodiversity as they may constitute refugia for local populations and increase habitat connectivity for dispersal and migration (Jackson and Overpeck, 2000; Ackerly et al., 2010; Austin and Van Niel, 2011).

Climate change models

- **Global climate models (GCMs)** are the most widely used method to understand what the climate may be like in the future as a result of emissions of greenhouse gases (global warming).
- Because they work at a global scale, the resolution of GCM results is typically quite coarse. Each grid cell is roughly 200 Œ 200 km.
- **Regional climate models (RCMs)** are applied to smaller spatial areas to produce results with greater local detail. However, RCMs still rely on GCMs for input data and therefore are not necessarily more reliable or more accurate.
- Climate models produce numerical values for key climate parameters, such as temperature, humidity and wind speed, for specific points and at different levels on the Earths surface and in the atmosphere and oceans.
- One major concern is that the spatial scale of climate data used in most SDMs may be inadequate due to the coarse resolution of global climate models (GCM) or downscaled derivatives of those models (Seo et al., 2009)
- These local topoclimatic habitats may need to be accounted for when evaluating climate-change impacts on biodiversity as they may constitute refugia for local populations and increase habitat connectivity for dispersal and migration (Jackson and Overpeck, 2000; Ackerly et al., 2010; Austin and Van Niel, 2011).

Notion of Equilibrium and Habitat Saturation

- Using current environmental correlates of a species' distribution to project its future occurrence also assumes that the current distribution is in equilibriumsuitable habitat is fully occupied or saturated.

Notion of Equilibrium and Habitat Saturation

- Using current environmental correlates of a species' distribution to project its future occurrence also assumes that the current distribution is in equilibriumsuitable habitat is fully occupied or saturated.
- Suitable places may be unoccupied, however, if recent disturbances have eradicated a species from an area [as visualized in metapopulation theory], if a species is expanding into areas that have only recently become available, or if regional population density is inadequate to support colonization of suitable areas.

Notion of Equilibrium and Habitat Saturation

- Using current environmental correlates of a species' distribution to project its future occurrence also assumes that the current distribution is in equilibriumsuitable habitat is fully occupied or saturated.
- Suitable places may be unoccupied, however, if recent disturbances have eradicated a species from an area [as visualized in metapopulation theory], if a species is expanding into areas that have only recently become available, or if regional population density is inadequate to support colonization of suitable areas.
- On the other hand, time lags associated with longevity of individuals established under previous conditions ["legacy" effects in plant distributions] or with breeding-area philopatry in birds may result in the occurrence of individuals in areas that no longer fall within their environmental niche space.

Notion of Equilibrium and Habitat Saturation

- Using current environmental correlates of a species' distribution to project its future occurrence also assumes that the current distribution is in equilibriumsuitable habitat is fully occupied or saturated.
- Suitable places may be unoccupied, however, if recent disturbances have eradicated a species from an area [as visualized in metapopulation theory], if a species is expanding into areas that have only recently become available, or if regional population density is inadequate to support colonization of suitable areas.
- On the other hand, time lags associated with longevity of individuals established under previous conditions ["legacy" effects in plant distributions] or with breeding-area philopatry in birds may result in the occurrence of individuals in areas that no longer fall within their environmental niche space.
- Thus, ecological niche models may be prone to both omission errors (leaving out of the niche space information from places that could be occupied) and commission errors (including in the niche space places that cannot sustain the species).

Scales of climate models

- Climate model portals allow the climate science research community to compare model results, assess errors and identify improvements as part of ongoing model development (such as CMIP5 for the global level and CORDEX3 for the regional level).

Scales of climate models

- Climate model portals allow the climate science research community to compare model results, assess errors and identify improvements as part of ongoing model development (such as CMIP5 for the global level and CORDEX3 for the regional level).
- To produce projections, the data files need to be processed and critical decisions have to be made about which aspects of the data are most relevant for the audience. These decisions relate to the:

Scales of climate models

- Climate model portals allow the climate science research community to compare model results, assess errors and identify improvements as part of ongoing model development (such as CMIP5 for the global level and CORDEX3 for the regional level).
- To produce projections, the data files need to be processed and critical decisions have to be made about which aspects of the data are most relevant for the audience. These decisions relate to the:
  1. Future time period of interest (e.g. 2016-2035, or 2046-2065, or 2081-2100)

Scales of climate models

- Climate model portals allow the climate science research community to compare model results, assess errors and identify improvements as part of ongoing model development (such as CMIP5 for the global level and CORDEX3 for the regional level).
- To produce projections, the data files need to be processed and critical decisions have to be made about which aspects of the data are most relevant for the audience. These decisions relate to the:
  1. Future time period of interest (e.g. 2016-2035, or 2046-2065, or 2081-2100)
  2. spatial scale (e.g. regional, national or subnational level)

Scales of climate models

- Climate model portals allow the climate science research community to compare model results, assess errors and identify improvements as part of ongoing model development (such as CMIP5 for the global level and CORDEX3 for the regional level).
- To produce projections, the data files need to be processed and critical decisions have to be made about which aspects of the data are most relevant for the audience. These decisions relate to the:
  1. Future time period of interest (e.g. 2016-2035, or 2046-2065, or 2081-2100)
  2. spatial scale (e.g. regional, national or subnational level)
  3. choice of climate parameters (most commonly temperature and rainfall)

Scales of climate models

- Climate model portals allow the climate science research community to compare model results, assess errors and identify improvements as part of ongoing model development (such as CMIP5 for the global level and CORDEX3 for the regional level).
- To produce projections, the data files need to be processed and critical decisions have to be made about which aspects of the data are most relevant for the audience. These decisions relate to the:
  1. Future time period of interest (e.g. 2016-2035, or 2046-2065, or 2081-2100)
  2. spatial scale (e.g. regional, national or subnational level)
  3. choice of climate parameters (most commonly temperature and rainfall)
  4. number of climate models to be used (it is good practice to use multiple models)

Scales of climate models

- Climate model portals allow the climate science research community to compare model results, assess errors and identify improvements as part of ongoing model development (such as CMIP5 for the global level and CORDEX3 for the regional level).
- To produce projections, the data files need to be processed and critical decisions have to be made about which aspects of the data are most relevant for the audience. These decisions relate to the:
  1. Future time period of interest (e.g. 2016-2035, or 2046-2065, or 2081-2100)
  2. spatial scale (e.g. regional, national or subnational level)
  3. choice of climate parameters (most commonly temperature and rainfall)
  4. number of climate models to be used (it is good practice to use multiple models)
  5. greenhouse gas emission scenario (from best-case to worst-case)

Scales of climate models

- Climate model portals allow the climate science research community to compare model results, assess errors and identify improvements as part of ongoing model development (such as CMIP5 for the global level and CORDEX3 for the regional level).
- To produce projections, the data files need to be processed and critical decisions have to be made about which aspects of the data are most relevant for the audience. These decisions relate to the:
  1. Future time period of interest (e.g. 2016-2035, or 2046-2065, or 2081-2100)
  2. spatial scale (e.g. regional, national or subnational level)
  3. choice of climate parameters (most commonly temperature and rainfall)
  4. number of climate models to be used (it is good practice to use multiple models)
  5. greenhouse gas emission scenario (from best-case to worst-case)
  6. format and level of technical detail of the information (e.g. maps, tables, time series).

Timescales of future projections

- Climate models can project out to 2100, but the time periods used in the CMIP5 models are generally 2016-2035 (for near term); 2046-2065 (for medium term); and 2081-2100 (for long term).

Timescales of future projections

- Climate models can project out to 2100, but the time periods used in the CMIP5 models are generally 2016-2035 (for near term); 2046-2065 (for medium term); and 2081-2100 (for long term).
- GCM results are available on daily time steps (data are generated in daily increments) and can be aggregated into monthly, seasonal and annual values.

Five future scenarios

- The U.N. climate panel report released on August 9th, 2021, about the physical science of climate change uses five possible scenarios for the future.

Five future scenarios

- The U.N. climate panel report released on August 9th, 2021, about the physical science of climate change uses five possible scenarios for the future.
- The scenarios are the result of complex calculations that depend on how quickly humans curb greenhouse gas emissions.

Five future scenarios

- The U.N. climate panel report released on August 9th, 2021, about the physical science of climate change uses five possible scenarios for the future.
- The scenarios are the result of complex calculations that depend on how quickly humans curb greenhouse gas emissions.
- But the calculations are also meant to capture socioeconomic changes in areas such as population, urban density, education, land use and wealth.

Five future scenarios

- The U.N. climate panel report released on August 9th, 2021, about the physical science of climate change uses five possible scenarios for the future.
- The scenarios are the result of complex calculations that depend on how quickly humans curb greenhouse gas emissions.
- But the calculations are also meant to capture socioeconomic changes in areas such as population, urban density, education, land use and wealth.
- Each scenario is labeled to identify both the emissions level and the so-called Shared Socioeconomic Pathway, or SSP, used in those calculations.

Five future scenarios

- The U.N. climate panel report released on August 9th, 2021, about the physical science of climate change uses five possible scenarios for the future.
- The scenarios are the result of complex calculations that depend on how quickly humans curb greenhouse gas emissions.
- But the calculations are also meant to capture socioeconomic changes in areas such as population, urban density, education, land use and wealth.
- Each scenario is labeled to identify both the emissions level and the so-called Shared Socioeconomic Pathway, or SSP, used in those calculations.

Scenario 1 Most optimistic: 1.5C by 2050 SSP1-1.9:

- The IPCCs most optimistic scenario, this describes a world where global CO_2 emissions are cut to net zero around 2050.

Scenario 1 Most optimistic: 1.5C by 2050 SSP1-1.9:

- The IPCCs most optimistic scenario, this describes a world where global CO_2 emissions are cut to net zero around 2050.
- Societies switch to more sustainable practices, with focus shifting from economic growth to overall well-being.

Scenario 1 Most optimistic: 1.5C by 2050 SSP1-1.9:

- The IPCCs most optimistic scenario, this describes a world where global CO2 emissions are cut to net zero around 2050.
- Societies switch to more sustainable practices, with focus shifting from economic growth to overall well-being.
- Investments in education and health go up. Inequality falls.

Scenario 1 Most optimistic: 1.5C by 2050 SSP1-1.9:

- The IPCCs most optimistic scenario, this describes a world where global CO2 emissions are cut to net zero around 2050.
- Societies switch to more sustainable practices, with focus shifting from economic growth to overall well-being.
- Investments in education and health go up. Inequality falls.
- Extreme weather is more common, but the world has dodged the worst impacts of climate change.

Scenario 1 Most optimistic: 1.5C by 2050 SSP1-1.9:

- The IPCCs most optimistic scenario, this describes a world where global CO2 emissions are cut to net zero around 2050.
- Societies switch to more sustainable practices, with focus shifting from economic growth to overall well-being.
- Investments in education and health go up. Inequality falls.
- Extreme weather is more common, but the world has dodged the worst impacts of climate change.

Scenario 2 Next Best: 1.8C by 2100 SSP1-2.6:

- In the next-best scenario, global CO_2 emissions are cut severely, but not as fast, reaching net-zero after 2050.

Scenario 2 Next Best: 1.8C by 2100 SSP1-2.6:

- In the next-best scenario, global CO_2 emissions are cut severely, but not as fast, reaching net-zero after 2050.
- It imagines the same socioeconomic shifts towards sustainability as SSP1-1.9.

Scenario 2 Next Best: 1.8C by 2100 SSP1-2.6:

- In the next-best scenario, global CO2 emissions are cut severely, but not as fast, reaching net-zero after 2050.
- It imagines the same socioeconomic shifts towards sustainability as SSP1-1.9. But temperatures stabilize around 1.8C higher by the end of the century.

Scenario 2 Next Best: 1.8C by 2100 SSP1-2.6:

- In the next-best scenario, global CO_2 emissions are cut severely, but not as fast, reaching net-zero after 2050.
- It imagines the same socioeconomic shifts towards sustainability as SSP1-1.9. But temperatures stabilize around 1.8C higher by the end of the century.

Scenario 3 Middle of the road: 2.7C by 2100 SSP2-4.5:

- This is a middle of the road scenario. CO2 emissions hover around current levels before starting to fall mid-century, but do not reach net-zero by 2100.

Scenario 3 Middle of the road: 2.7C by 2100 SSP2-4.5:

- This is a middle of the road scenario. CO2 emissions hover around current levels before starting to fall mid-century, but do not reach net-zero by 2100.
- Socioeconomic factors follow their historic trends, with no notable shifts.

Scenario 3 Middle of the road: 2.7C by 2100 SSP2-4.5:

- This is a middle of the road scenario. CO_2 emissions hover around current levels before starting to fall mid-century, but do not reach net-zero by 2100.
- Socioeconomic factors follow their historic trends, with no notable shifts.
- Progress toward sustainability is slow, with development and income growing unevenly.

Scenario 3 Middle of the road: 2.7C by 2100 SSP2-4.5:

- This is a middle of the road scenario. CO2 emissions hover around current levels before starting to fall mid-century, but do not reach net-zero by 2100.
- Socioeconomic factors follow their historic trends, with no notable shifts.
- Progress toward sustainability is slow, with development and income growing unevenly.
- In this scenario, temperatures rise 2.7C by the end of the century.

Scenario 3 Middle of the road: 2.7C by 2100 SSP2-4.5:

- This is a middle of the road scenario. CO2 emissions hover around current levels before starting to fall mid-century, but do not reach net-zero by 2100.
- Socioeconomic factors follow their historic trends, with no notable shifts.
- Progress toward sustainability is slow, with development and income growing unevenly.
- In this scenario, temperatures rise 2.7C by the end of the century.

Scenario 4 Dangerous: 3.6C by 2100, SSP3-7.0:

- On this path, emissions and temperatures rise steadily and CO_2 emissions roughly double from current levels by 2100.

Scenario 4 Dangerous: 3.6C by 2100, SSP3-7.0:

- On this path, emissions and temperatures rise steadily and CO2 emissions roughly double from current levels by 2100.
- Countries become more competitive with one another, shifting toward national security and ensuring their own food supplies.

Scenario 4 Dangerous: 3.6C by 2100, SSP3-7.0:

- On this path, emissions and temperatures rise steadily and CO2 emissions roughly double from current levels by 2100.
- Countries become more competitive with one another, shifting toward national security and ensuring their own food supplies.
- By the end of the century, average temperatures have risen by 3.6C.

Scenario 4 Dangerous: 3.6C by 2100, SSP3-7.0:

- On this path, emissions and temperatures rise steadily and CO2 emissions roughly double from current levels by 2100.
- Countries become more competitive with one another, shifting toward national security and ensuring their own food supplies.
- By the end of the century, average temperatures have risen by 3.6C.

Scenario 5 Avoid at all costs: 4.4C by 2100, SSP5-8.5:

- This is a future to avoid at all costs.

Scenario 5 Avoid at all costs: 4.4C by 2100, SSP5-8.5:

- This is a future to avoid at all costs.
- Current CO_2 emissions levels roughly double by 2050.

Scenario 5 Avoid at all costs: 4.4C by 2100, SSP5-8.5:

- This is a future to avoid at all costs.
- Current CO2 emissions levels roughly double by 2050.
- The global economy grows quickly, but this growth is fuelled by exploiting fossil fuels and energy-intensive lifestyles.

Scenario 5 Avoid at all costs: 4.4C by 2100, SSP5-8.5:

- This is a future to avoid at all costs.
- Current CO_2 emissions levels roughly double by 2050.
- The global economy grows quickly, but this growth is fuelled by exploiting fossil fuels and energy-intensive lifestyles.
- By 2100, the average global temperature is a scorching 4.4C higher.

Scenario 5 Avoid at all costs: 4.4C by 2100, SSP5-8.5:

- This is a future to avoid at all costs.
- Current CO2 emissions levels roughly double by 2050.
- The global economy grows quickly, but this growth is fuelled by exploiting fossil fuels and energy-intensive lifestyles.
- By 2100, the average global temperature is a scorching 4.4C higher.
- Read the forecast of different models at McSweeney et al. 2015 and the definition of each acronym WCRP-CMIP CMIP6$_C Vsversion$: 6.2.58.67

Downscaling Climate Data for SDMs: Bias Correction and Spatial Interpolation

- Global Climate Models (GCMs) are the primary tools to simulate multi-decadal climate dynamics and to generate and understand global climate change projections under different future emission scenarios.

Downscaling Climate Data for SDMs: Bias Correction and Spatial Interpolation

- Global Climate Models (GCMs) are the primary tools to simulate multi-decadal climate dynamics and to generate and understand global climate change projections under different future emission scenarios.
- However, these models have a coarse spatial resolution (typically a few hundred kilometres) and suffer from substantial systematic biases when compared with observations.

Downscaling Climate Data for SDMs: Bias Correction and Spatial Interpolation

- Global Climate Models (GCMs) are the primary tools to simulate multi-decadal climate dynamics and to generate and understand global climate change projections under different future emission scenarios.
- However, these models have a coarse spatial resolution (typically a few hundred kilometres) and suffer from substantial systematic biases when compared with observations.
- Therefore, they are unable to provide actionable information at the regional and local spatial scales required in impact and adaptation studies (e.g., Grouillet al., 2016).

Downscaling Climate Data for SDMs: Bias Correction and Spatial Interpolation

- Global Climate Models (GCMs) are the primary tools to simulate multi-decadal climate dynamics and to generate and understand global climate change projections under different future emission scenarios.
- However, these models have a coarse spatial resolution (typically a few hundred kilometres) and suffer from substantial systematic biases when compared with observations.
- Therefore, they are unable to provide actionable information at the regional and local spatial scales required in impact and adaptation studies (e.g., Grouillet al., 2016).
- Hence, higher resolution simulations and if possible bias corrected are required for the most relevant climate variables (Vaittinada Ayar et al., 2015).

Downscaling Climate Data for SDMs: Bias Correction and Spatial Interpolation

- Global Climate Models (GCMs) are the primary tools to simulate multi-decadal climate dynamics and to generate and understand global climate change projections under different future emission scenarios.
- However, these models have a coarse spatial resolution (typically a few hundred kilometres) and suffer from substantial systematic biases when compared with observations.
- Therefore, they are unable to provide actionable information at the regional and local spatial scales required in impact and adaptation studies (e.g., Grouillet al., 2016).
- Hence, higher resolution simulations and if possible bias corrected are required for the most relevant climate variables (Vaittinada Ayar et al., 2015).
- Downscaling attempts to resolve the scale discrepancy between climate change scenarios and the resolution required for impact assessment. Two main downscaling approaches have been developed since the early 1990s: Dynamical downscaling (based on Regional Climate Models, RCMs) and Statistical Downscaling Models (SDMs), which are nowadays recognized as complementary in many practical applications.

Downscaling Climate Data for SDMs: Bias Correction and Spatial Interpolation

- Global Climate Models (GCMs) are the primary tools to simulate multi-decadal climate dynamics and to generate and understand global climate change projections under different future emission scenarios.
- However, these models have a coarse spatial resolution (typically a few hundred kilometres) and suffer from substantial systematic biases when compared with observations.
- Therefore, they are unable to provide actionable information at the regional and local spatial scales required in impact and adaptation studies (e.g., Grouillet al., 2016).
- Hence, higher resolution simulations and if possible bias corrected are required for the most relevant climate variables (Vaittinada Ayar et al., 2015).
- Downscaling attempts to resolve the scale discrepancy between climate change scenarios and the resolution required for impact assessment. Two main downscaling approaches have been developed since the early 1990s: Dynamical downscaling (based on Regional Climate Models, RCMs) and Statistical Downscaling Models (SDMs), which are nowadays recognized as complementary in many practical applications.

Ensemble Model

- As ecology and conservation disciplines require a robust understanding of species distribution and habitat requirements, species distribution models (SDM) are, therefore, one of the essential techniques for spatio-temporal predictions of biodiversity at the biogeographic scale (Alvarado-Serrano and Knowles, 2014; Naimi, 2015; Srivastava et al., 2019).

Ensemble Model

- As ecology and conservation disciplines require a robust understanding of species distribution and habitat requirements, species distribution models (SDM) are, therefore, one of the essential techniques for spatio-temporal predictions of biodiversity at the biogeographic scale (Alvarado-Serrano and Knowles, 2014; Naimi, 2015; Srivastava et al., 2019).
- However, one major disadvantage of these distribution models is that there are already a large number of modeling algorithms available, and this number is expanding all the time, making it difficult to choose the optimal methodology (Elith et al., 2010; Ahmad et al., 2019b).

Ensemble Model

- As ecology and conservation disciplines require a robust understanding of species distribution and habitat requirements, species distribution models (SDM) are, therefore, one of the essential techniques for spatio-temporal predictions of biodiversity at the biogeographic scale (Alvarado-Serrano and Knowles, 2014; Naimi, 2015; Srivastava et al., 2019).
- However, one major disadvantage of these distribution models is that there are already a large number of modeling algorithms available, and this number is expanding all the time, making it difficult to choose the optimal methodology (Elith et al., 2010; Ahmad et al., 2019b).
- To overcome this problem, the ensemble modeling technique implemented in the biomod2 package provides a valuable platform for determining the species' current distribution and predicting their future potential Spatio-temporal distributions under changing climatic scenarios (Gillard et al., 2017; Thuiller et al., 2019).

Ensemble Model

- As ecology and conservation disciplines require a robust understanding of species distribution and habitat requirements, species distribution models (SDM) are, therefore, one of the essential techniques for spatio-temporal predictions of biodiversity at the biogeographic scale (Alvarado-Serrano and Knowles, 2014; Naimi, 2015; Srivastava et al., 2019).
- However, one major disadvantage of these distribution models is that there are already a large number of modeling algorithms available, and this number is expanding all the time, making it difficult to choose the optimal methodology (Elith et al., 2010; Ahmad et al., 2019b).
- To overcome this problem, the ensemble modeling technique implemented in the biomod2 package provides a valuable platform for determining the species' current distribution and predicting their future potential Spatio-temporal distributions under changing climatic scenarios (Gillard et al., 2017; Thuiller et al., 2019).

Some key definitions

- **Ensemble** is a machine learning paradigm where multiple models (often called weak learners) are trained to solve the same problem and combined to get better results. The main hypothesis is that when weak models are correctly combined we can obtain more accurate and/or robust models.

Some key definitions

- **Ensemble** is a machine learning paradigm where multiple models (often called weak learners) are trained to solve the same problem and combined to get better results. The main hypothesis is that when weak models are correctly combined we can obtain more accurate and/or robust models.
- In ensemble learning theory, we call **weak learners (or base models)** models that can be used as building blocks for designing more complex models by combining several of them.

Some key definitions

- **Ensemble** is a machine learning paradigm where multiple models (often called weak learners) are trained to solve the same problem and combined to get better results. The main hypothesis is that when weak models are correctly combined we can obtain more accurate and/or robust models.
- In ensemble learning theory, we call **weak learners (or base models)** models that can be used as building blocks for designing more complex models by combining several of them.
- Most of the time, these basic models perform not so well by themselves either because they have a high bias (low degree of freedom models, for example) or because they have too much variance to be robust (high degree of freedom models, for example).

Some key definitions

- **Ensemble** is a machine learning paradigm where multiple models (often called weak learners) are trained to solve the same problem and combined to get better results. The main hypothesis is that when weak models are correctly combined we can obtain more accurate and/or robust models.
- In ensemble learning theory, we call **weak learners (or base models)** models that can be used as building blocks for designing more complex models by combining several of them.
- Most of the time, these basic models perform not so well by themselves either because they have a high bias (low degree of freedom models, for example) or because they have too much variance to be robust (high degree of freedom models, for example).
- Then, the idea of ensemble methods is to try reducing bias and/or variance of such weak learners by combining several of them together in order to create a **strong learner (or ensemble model)** that achieves better performances.

Some key definitions

- **Ensemble** is a machine learning paradigm where multiple models (often called weak learners) are trained to solve the same problem and combined to get better results. The main hypothesis is that when weak models are correctly combined we can obtain more accurate and/or robust models.
- In ensemble learning theory, we call **weak learners (or base models)** models that can be used as building blocks for designing more complex models by combining several of them.
- Most of the time, these basic models perform not so well by themselves either because they have a high bias (low degree of freedom models, for example) or because they have too much variance to be robust (high degree of freedom models, for example).
- Then, the idea of ensemble methods is to try reducing bias and/or variance of such weak learners by combining several of them together in order to create a **strong learner (or ensemble model)** that achieves better performances.

Paradox

- If we choose base models with low bias but high variance, it should be with an aggregating method that tends to reduce variance whereas if we choose base models with low variance but high bias, it should be with an aggregating method that tends to reduce bias.

Paradox

- If we choose base models with low bias but high variance, it should be with an aggregating method that tends to reduce variance whereas if we choose base models with low variance but high bias, it should be with an aggregating method that tends to reduce bias.
- We can mention three major kinds of meta-algorithms that aim at combining weak learners:
  1. **bagging**, that often considers homogeneous weak learners, learns them independently from each other in parallel and combines them following some kind of deterministic averaging process

Paradox

- If we choose base models with low bias but high variance, it should be with an aggregating method that tends to reduce variance whereas if we choose base models with low variance but high bias, it should be with an aggregating method that tends to reduce bias.

- We can mention three major kinds of meta-algorithms that aim at combining weak learners:
    1. **bagging**, that often considers homogeneous weak learners, learns them independently from each other in parallel and combines them following some kind of deterministic averaging process
    2. **boosting**, that often considers homogeneous weak learners, learns them sequentially in a very adaptative way (a base model depends on the previous ones) and combines them following a deterministic strategy

Paradox

- If we choose base models with low bias but high variance, it should be with an aggregating method that tends to reduce variance whereas if we choose base models with low variance but high bias, it should be with an aggregating method that tends to reduce bias.

- We can mention three major kinds of meta-algorithms that aim at combining weak learners:
  1. **bagging**, that often considers homogeneous weak learners, learns them independently from each other in parallel and combines them following some kind of deterministic averaging process
  2. **boosting**, that often considers homogeneous weak learners, learns them sequentially in a very adaptative way (a base model depends on the previous ones) and combines them following a deterministic strategy
  3. **stacking,** that often considers heterogeneous weak learners, learns them in parallel and combines them by training a meta-model to output a prediction based on the different weak models' predictions

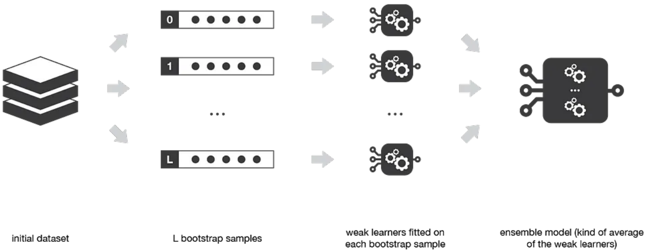
Paradox

- If we choose base models with low bias but high variance, it should be with an aggregating method that tends to reduce variance whereas if we choose base models with low variance but high bias, it should be with an aggregating method that tends to reduce bias.

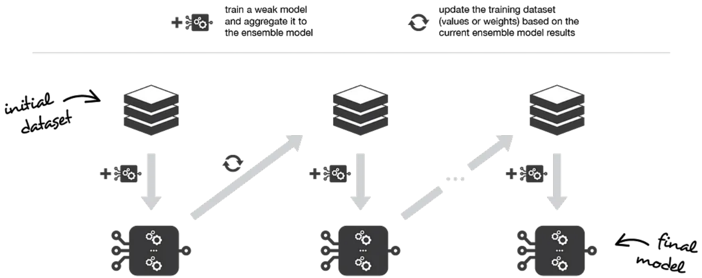
- We can mention three major kinds of meta-algorithms that aim at combining weak learners:
  1. **bagging**, that often considers homogeneous weak learners, learns them independently from each other in parallel and combines them following some kind of deterministic averaging process
  2. **boosting**, that often considers homogeneous weak learners, learns them sequentially in a very adaptative way (a base model depends on the previous ones) and combines them following a deterministic strategy
  3. **stacking,** that often considers heterogeneous weak learners, learns them in parallel and combines them by training a meta-model to output a prediction based on the different weak models' predictions

Bootstrapping

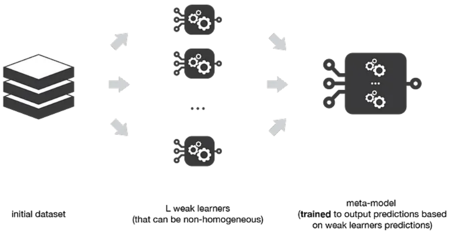
Parameter estimation while Bagging



Booting


train a weak model and aggregate it to the ensemble model
update the training dataset (values or weights) based on the current ensemble model results
initial dataset
...
final model

Stacking


initial dataset
L weak learners
(that can be non-homogeneous)
meta-model
(**trained** to output predictions based
on weak learners predictions)

Thank You