**2022 AI Challenge for Biodiversity**

**기후변화 민감 종 선별 모델 구현 및 통계적 검정**

**- 나비를 기준으로 -**

Chart, sunburst chart

Description automatically generated

**2022. 10. 31.**

**데생 팀**

**목차**

**Ⅰ. 서론 3**

**Ⅱ. 재료 및 방법 4**

1. 데이터 4

2. 모델 4

3. 통계적 추론 / 검정 5

**Ⅲ. 결과 및 고찰 7**

**Ⅳ. 인용 문헌 8**

**V. 소스 코드 9**

**VI. 활용한 외부 데이터 19**

**VII. 부록 20**

**1. 서론**

산체스-바요 교수는 “지난 4억 년 간 곤충은 다른 동물의 없어서는 안 될 중요한 먹이가 되고, 끊임없이 양산되는 썩은 쓰레기 등을 분해해 식물에게 필요한 양분으로 변화시키며, 식물의 꽃을 피워 열매를 맺게 하는 등 생태계의 핵심적인 역할을 해왔다” 고 말했다. 그러나 “세계적으로 곤충의 생물 다양성이 위협받고 있으며, 이 상황이 그대로 이어질 경우 수십 년 안에 살아있는 곤충 10마리 중 4마리가 사라질 것” 을 경고하였다.(이강봉, 2019. 02. 12.) 따라서 최근 곤충의 감소원인으로 주목받는 기후변화와 관련한 종 다양성 연구가 시급하다. 하지만 한국에서의 이와 같은 연구는 미진한 데다 연구를 하는 데 필수적인 장기데이터도 특정 종을 제외하고는 거의 없는 실정이다. 따라서 본 연구는 새롭게 축적되는 데이터로도 의미 있는 곤충 종 다양성 변화를 도출하고자 하였다.

나비는 조사자료가 많이 축적되어 있어 기후변화로 인한 군집 변화 연구가 많이 진행되었다. 또한 나비는 기후변화에 민감한 종으로 연구에 활용되고 있다. (권태성 외, 2017) 따라서 나비를 기후변화와 연관성이 높은 신뢰할 수 있는 지표로 판단하여 연구의 기준으로 삼았다. 나비와 같이 기후변화에 민감한 곤충들을 선별하는 모델을 만들어 기후변화에 따른 종 다양성 감소를 경고하고자 하였다.

**2. 재료 및 방법**

**2-1 데이터**

데이터는 국립생태원에서 수행한 전국자연환경조사 곤충분야에서 조사된 정보를 활용하였다. 곤충의 위치정보와 한글 명, 학명, 조사날짜 등이 있으며 2014년부터 2018년까지의 데이터가 있었다. EXAMIN\_BEG(조사 날짜)를 활용하여 해당 개체가 발견된 계절을 구했다. 이때 12월, 1월, 2월은 겨울로 3월, 4월, 5월은 봄으로 6월, 7월, 8월은 여름으로 9월, 10월, 11월은 가을로 설정하였다. 데이터가 적은 겨울과 2015년 데이터는 삭제하였고 곤충들 중 모든 년도, 계절에 존재하는 곤충들만 (총 5466종 중 309종) 필터링하였다. 필터링 된 곤충들의 발견 년도, 계절별로 그룹화하여 종의 위도, 경도 평균을 계산하여 데이터 전처리를 진행하였다.

**2-2 모델**

기후변화 지표종인 나비만 추출하여 Group2를 생성한다. 나비를 제외한 종들을 One Class SVM을 사용해서 나비와 비슷한 그룹(Group 1)과 비슷하지 않은 그룹(Group 2)으로 나누었다.

사용한 One Class Support Vector Machine 이란, Unsupervised learning 모델로, 정상 데이터를 학습시켜, 정상 데이터의 범주에서 벗어나는 이상치를 탐지할 때 사용하는 알고리즘이다. 본 연구에서는 나비를 정상 데이터로 학습 시킨 후, 나비를 제외한 나머지 데이터들을 집어 넣어 정상(나비와 같은 특징을 가지는 종)과, 비정상(나비와 다른 특징을 가지는 종)으로 분류했다.

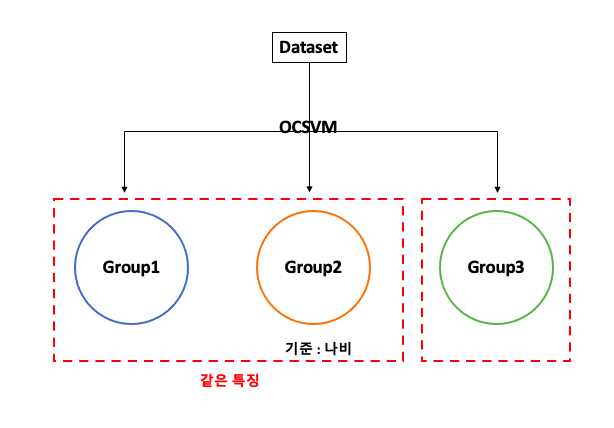


그림 1. OCSVM으로 Grouping (Group1 : 나비와 같은 특징 Group2 : 나비 Group3 : 나비와 다른 특징)

그룹을 나눌 때는 년도와 계절 간의 위도 경도 개별 변화를 기준으로 하였다. (예시 : 위도 1 : 100, 위도 2 : 113, 위도 3 : 118 ~~ 데이터가 있으면, 데이터간 차이 생성 -13, -5 ~~)

데이터가 계속 갱신됨에 따라 짧은 기간에 생성한 데이터를 활용하기 위해 비교적 짧은 기간 이동거리 차이를 사용하였다.

Graphical user interface, text

Description automatically generatedGraphical user interface, text

Description automatically generated

그림 2. OneClassSVM 예측 결과 01

Graphical user interface, application

Description automatically generated

그림 3. OneClassSVM 예측 결과 02

총 279개의 종에 대해서 예측을 했다. 예측 결과 비정상(나비와 다른 특징을 가지는 종) 238종, 정상(나비와 같은 특징을 가지는 종) 41종으로 예측 되었다.

구축이 완료된 우리의 모델에 추후 들어오는 곤충들의 위도 경도 데이터를 넣는다면 정상과 비정상으로 분류를 해주기 때문에, 활용 가능성이 높을 것으로 기대한다.

**2-3 통계적 추론/검정**

짧은 기간 동안의 차이로 나눠진 그룹이 장기적인 이동 거리 차이에서도 유의미한지 검증하기 위해 2 sample independent t-test를 진행하였다. 그룹별로 누적 위도 경도 차이를 추출하고 평균값을 사용하여 통계적 검정을 진행하였다.

첫번째 대립가설은 ‘Group1(나비와 같은 특징을 가지는 종)은 Group2(나비)와 차이가 있다’ 이고, 귀무가설은 ‘두 그룹간 차이가 없다’로 설정하였다. 유의수준(a)은 0.05로 설정하였다. 검정 결과 p-value가 0.2429가 나와 나비와 같은 특징을 가지는 종과 나비의 이동 거리 차이에는 유의미한 차이가 없었다.

두번째 대립가설은 ‘Group2(나비)은 Group3(나비와 다른 특징을 가지는 종)와 차이가 있다’ 이고, 귀무가설은 ‘두 그룹간 차이가 없다’로 설정하였다. 유의수준(a)은 0.05로 설정하였다. 검정 결과 p-value가 0.0035가 나와 나비와, 나비와 다른 특징을 가지는 종간의 이동 거리 차이에는 유의미한 차이가 있었다.

세번째 대립가설은 ‘Group1(나비와 같은 특징을 가지는 종)은 Group3(나비와 다른 특징을 가지는 종)와 차이가 있다’ 이고, 귀무가설은 ‘두 그룹간 차이가 없다’로 설정하였다. 유의수준(a)은 0.05로 설정하였다. 검정 결과 p-value가 0.0307이 나와 나비와 같은 특징을 가지는 종과 다른 특징을 가지는 종간의 이동 거리 차이에는 유의미한 차이가 있었다.

One Class SVM 모델을 활용한 곤충 종별 분류된 결과가, 통계적으로 유의미한 결과임을 알 수 있었다.

**3. 결과 및 고찰**

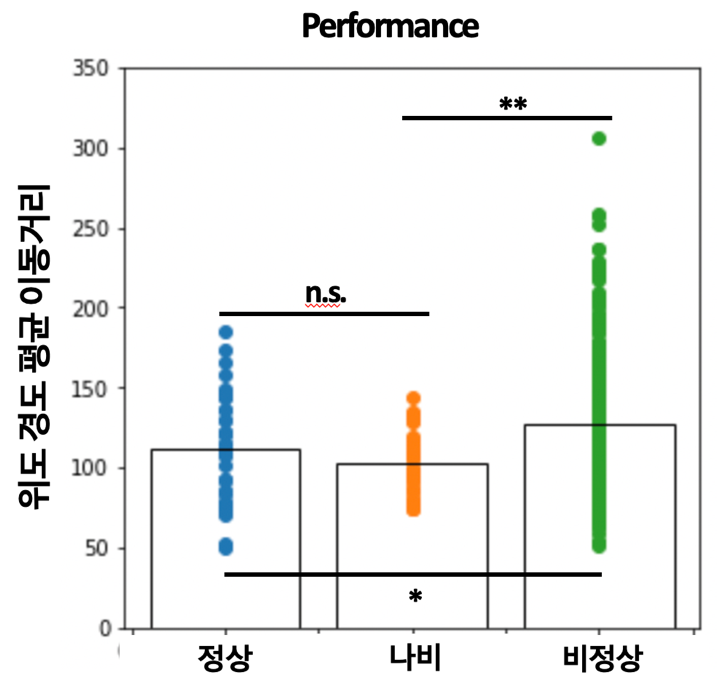


그림 4. 그룹별 이동거리 시각화(파랑: Group1(나비와 같은 특징) 주황: Group2(나비) 초록: Group3(나비와 다른 특징)

위 그림은 그룹별 이동거리 분포도를 나타낸 것이다. Group1, 2가 Group3보다 이동거리가 작은 것을 확인했다. 일부의 이주성 종들은 매년 기후변화에 빠르게 반응할 수 있지만 대부분의 정주성 종은 그 같이 신속한 대응을 할 수 없다. (박해철, 2005) 따라서 Group1, 2는 Group3보다 기후변화에 더 영향을 받을 것이라고 추정한다.

나비를 제외한 종은 총 279종이다. 그 중 Group1으로 분류된 종은 41종이고 Group3으로 분류된 종은 238종이다. Group1로 분류된 종이 전체 종에서 14.695%이다. 스웨덴 스톡홀름 대학·핀란드 오울루 대학 등의 연구진은 따뜻하고 긴 동절기가 나비에게 부정적 결과를 준다는 사실을 밝혔다. (최우리, 2022. 04. 06) 또한 미국 대학 공동연구팀은 나비의 감소 원인이 기후변화에 따른 가을철 온난화 때문이라는 사실을 밝혀냈다.(이근영, 2021. 03. 05) Group1을 나비와 같이 개체 수 감소 위험 종으로 분류했을 때, 기후변화가 지속됨에 따라 곤충다양성의 감소는 심각할 것으로 예상된다.

**4. 인용문헌**

권태성 외 (2017), 한국나비 분포변화 분석, 국립산림과학원

박해철(2005), 기후변화가 곤충에 미치는 영향에 관한 국내외 연구 분석, 자연보존 130 :38-47

이근영(2021. 03. 05), 가을이 따뜻해지자 나비가 줄었다, 한겨례, <https://www.hani.co.kr/arti/science/science_general/985512.html>

최우리(2022. 04. 06), 나비가 사라진다… 2022년 봄, 자연의 경고 들리나요, 한겨례, <https://www.hani.co.kr/arti/society/environment/1037703.html>

**5. 소스코드**

Text

Description automatically generated

그림 5. 데이터 리딩 및 위도 경도 처리

Fiona library를 사용해 shp 형식의 파일을 읽어와 한국 좌표계 기준인 epsg:4326으로 변환해서 위도와 경도를 추출해 낸다.

A screenshot of a computer

Description automatically generated with medium confidence

그림 6. 데이터 프레임 출력결과

불러온 데이터의 출력 결과이다. 우리는 위도 경도 정보인 geometry, 종 정보인 SPCS\_LCNM, 조사 날짜인 EXAMIN\_YEA, EXAMIN\_BEG를 활용하기로 했다.

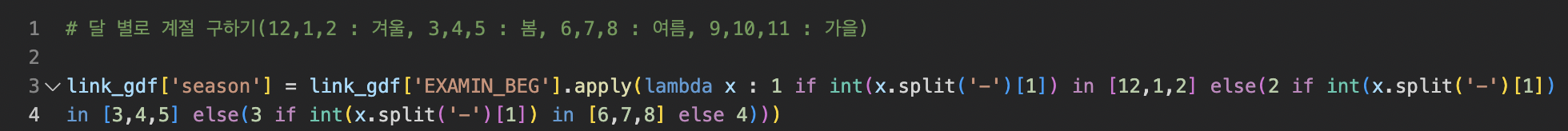


그림 7. 조사 날짜를 활용한 계절 추출

조사 날짜의 달 을 활용해 계절을 추출해냈다. 자세한 설명은 2-1 데이터 부분에 서술해 놓았다.

Text

Description automatically generated

그림 8. 위도 경도 추출

병합되어 있던 위도 경도를 각각의 컬럼으로 나누어 주었다.

Text

Description automatically generated

그림 9. 겨울 계절, 2015년도 데이터, 국명미정 데이터 삭제

상대적으로 데이터가 적은 겨울 계절, 2015년도 데이터와 종이 불분명한 국명미정 데이터를 삭제처리 해주었다.

Text

Description automatically generated

**그림 10. 모든 년도, 계절에 존재하는 곤충 필터링**

년도와 계절 별 위도 경도 차이를 계산하기 위해, 모든 년도 및 계절에 존재하는 곤충들만 필터링 한다.

Text

Description automatically generated

**그림 11. 종별 년도, 시즌 그룹화 및 위도 경도 평균화 작업**

종별로 년도와 시즌으로 그룹화를 진행해서 년도와 시즌 별로 한 종당 여러 개의 위도 경도 데이터가 존재하는 경우 평균화를 진행해, 대체적으로 해당 종이 발견되는 위도 경도 지역을 찾아냈다.

Text

Description automatically generated

**그림 12. 필터링 및 살아남은 종 저장**

최종적으로 필터링이 끝나고 살아남은 종들에 대한 정보를 저장한다. 총 309종이 필터링 되었다.

Text

Description automatically generated

**그림 13. 종별 위도 경도 merge**

종별로 년도와 시즌별로 분리되어 있던 위도 경도 데이터를 묶어주었다.

Graphical user interface, text, application

Description automatically generated

**그림 14. 종별 위도 경도 merge 결과**

종별로 년도와 시즌 별 merge 결과이다. 총 24개의 columns이 존재하는데 총 4개년도(2014, 2016, 2017, 2018)와 총 3개 계절(봄, 여름, 가을)에 대한 위도 경도 데이터 이다.

Text

Description automatically generated

**그림 15. 위도 경도 간 차이 계산**

종별로 년도와 계절 변화에 따른 위도 경도 차이를 계산하기 위한 구간이다. 위도 12개와 경도 12개를 분리해서 각 step별(년도, 계절 진행 순서)로 차이를 계산했다. 이를 통해 곤충이 계절 진행에 따라 얼만큼 움직였는지를 파악할 수 있었다.

Graphical user interface, text, application

Description automatically generated

**그림 16. 위도 경도 간 차이 계산 결과물 데이터프레임 화**

앞서 계산한 위도 경도 차이에 대한 수치들을 데이터프레임화 시켜주었다.

A picture containing text

Description automatically generated

**그림 17. 기후변화 지표 종 나비 저장**

기후변화 지표종으로 선정된 나비 데이터에 대해 따로 추출 및 저장을 진행해주었다.

Text

Description automatically generated

**그림 18. 나비와 나머지 데이터 간의 분리 작업**

One Class SVM 모델을 활용해 학습하기 위해, 나비 데이터와 그 외 데이터 간의 분리작업을 진행해 주었다.

Text

Description automatically generated

**그림 19. OneClassSVM 학습 및 피팅 과정**

One Calss SVM 모델을 사용해 나비들의 위도 경도에 변화에 대한 특징을 학습시키고, 그 외 나머지 데이터를 넣어서 나비와 같은 위 경도 변화를 보이는 종들은 1이란 class로 다른 위 경도 변화를 보이는 종들은 -1로 Binary Classification 작업을 진행해주었다.

Text

Description automatically generated

**그림 20. 통계적 검정을 위한 전 처리 작업 01**

분류된 라벨에(1 and -1)따른 실제 위도 경도 움직임에 대한 차이가 있는지 검정하고자, haversine library를 활용해, 종별 년도 및 계절별 위 경도 거리 계산을 진행해 주었다.

Text

Description automatically generated

**그림 21. 통계적 검정을 위한 전 처리 작업 02**

**Text

Description automatically generated**

**그림 22. 통계적 검정을 위한 전 처리 작업 03**

마찬가지로 -1로 분류된 class와 나비 군집에 대해서도 동일한 작업을 진행해 주었다.

A screenshot of a computer

Description automatically generated with medium confidence

**그림 23. 각 군집 별 움직임 평균값 추출**

마찬가지로 각 군집 별 움직임(km 기준)에 대한 평균값을 추출해서 각각 데이터를 저장해주었다.

Text

Description automatically generated

**그림 24. 나비 군집과, 나비 군집과 비슷한 움직임을 가지는 군집간 t-test 진행 결과**

나비 군집과, 나비 군집과 비슷한 움직임을 가지는 군집간 t-test 진행 결과, p-value가 0.242 정도로 움직임에 대한 평균이 같다는 가설을 기각하지 못하므로, 비슷한 움직임을 가진다. 라는 결론이 나왔다.

Graphical user interface, text

Description automatically generated

**그림 25. 나비 군집과, 나비 군집과 다른 움직임을 가지는 군집간 t-test 진행 결과**

나비 군집과, 나비 군집과 다른 움직임을 가지는 군집간 t-test 진행 결과, p-value가 0.003 정도로 움직임에 대한 평균이 같다는 가설을 기각하므로, 서로 다른 움직임을 가진다. 라는 결론이 나왔다.

A screenshot of a computer

Description automatically generated with medium confidence

**그림 26. 나비 군집과 비슷한 움직임, 다른 움직임을 가지는 군집간 t-test 진행 결과**

나비 군집과 비슷한 움직임을 가지는 군집과, 나비 군집과 다른 움직임을 가지는 군집간 t-test 진행 결과, p-value가 0.03 정도로 움직임에 대한 평균이 같다는 가설을 기각하므로, 서로 다른 움직임을 가진다. 라는 결론이 나왔다.

**6. 활용한 외부 데이터**

Graphical user interface

Description automatically generated with low confidence

**그림27. 활용데이터 설명 이미지**

환경 빅데이터 플랫폼에서 제공해주는 생태 생태계 정밀조사(곤충\_점)에서 제공해주는 shp 데이터를 활용했다.

(링크 : https://www.bigdata environment.kr/user/data\_market/detail.do?id=f1c88b10-2fb5-11eb-8f72-932712f5aa3c)

해당 데이터는 Python Fiona 라이브러리를 사용해 열었으며, 내부 컬럼은 곤충 발견 위도 경도 데이터, 곤충 ID, 곤충 종, 영문명, 발견 날짜가 저장되어 있다.

**7. 부록**

**- 나비 (Group2) 총 종 수 : 30마리**

거꾸로여덟팔나비, 긴꼬리제비나비, 남방노랑나비, 남방부전나비, 네발나비, 노랑나비, 대만흰나비, 먹부전나비, 물결나비, 배추흰나비, 별박이세줄나비, 부전나비, 부처나비, 뿔나비, 사향제비나비, 산제비나비, 산호랑나비, 암먹부전나비, 애기세줄나비, 애물결나비, 왕자팔랑나비, 작은멋쟁이나비, 작은주홍부전나비, 제비나비, 제이줄나비, 청띠신선나비, 큰멋쟁이나비, 큰줄흰나비, 푸른부전나비, 호랑나비

**- 나비와 비슷한 특징을 가지는 종 (Group1) 총 종 수 : 41마리**

Vespa crabro, 가시노린재, 갈색날개노린재, 검은다리실베짱이, 꼬마꽃등에, 끝검은말매미충, 끝짤룩노랑가지나방, 넓은띠담흑수염나방, 넓적배허리노린재, 노랑띠알락가지나방, 노랑테가시잎벌레, 다리무늬침노린재, 다색띠큰가지나방, 더듬이긴노린재, 덩굴꽃등에, 등검은메뚜기, 말매미충, 무당벌레, 무당알노린재, 변색장님노린재, 북쪽비단노린재, 붉은잡초노린재, 빨간긴쐐기노린재, 뿔들파리, 삿포로잡초노린재, 설상무늬장님노린재, 수수꽃다리명나방, 수중다리꽃등에, 썩덩나무노린재, 아시아실잠자리, 알락수염노린재, 알통다리꽃등에, 애긴노린재, 양봉꿀벌, 우리가시허리노린재, 좁쌀메뚜기, 칠성무당벌레, 큰허리노린재, 톱다리개미허리노린재, 호리꽃등에, 황줄점갈고리나방

**- 나비와 다른 특징을 가지는 종 (Group3) 총 종 수 : 238마리**

Vespa simillima, 가슴골좁쌀바구미, 가시점둥글노린재, 각시가지나방, 각시메뚜기, 각시얼룩가지나방, 갈고리가지나방, 갈색여치, 갈색점밤나방, 갈색줄수염나방, 거품벌레, 검은끝짤름나방, 검은두줄꼬마밤나방, 검은띠수염나방, 검은물잠자리, 검은점뾰족명나방, 검정넓적꽃등에, 검정볼기쉬파리, 검정오이잎벌레, 검정파리매, 게눈노린재, 고려꽃등에, 고마로브집게벌레, 고운날개가지나방, 고추좀잠자리, 곧은띠비단명나방, 곧은띠수염나방, 곧은줄재주나방, 곰개미, 광대노린재, 광대파리매, 광붙이꽃등에, 교차무늬주홍테불나방, 구리꼬마꽃벌, 굴뚝긴노린재, 굵은줄제비가지나방, 금록색잎벌레, 금파리, 긴날개쐐기노린재, 깜보라노린재, 꼬마긴노린재, 꼬마남생이무당벌레, 꼬마독나방, 꼬마큰날개파리, 꼬마혹수염나방, 꼭지파리, 꽃날개들명나방, 꽃등에, 꽈리허리노린재, 끝검은뿔벌레, 남생이무당벌레, 넉점박이송장벌레, 네눈가지나방, 네눈들명나방, 네점박이노린재, 노랑갈고리나방, 노랑다리들명나방, 노랑무당벌레, 노랑배불나방, 노랑쐐기나방, 노랑테불나방, 노박덩굴가지나방, 녹색박각시, 닥나무들명나방, 달무리무당벌레, 대벌레, 도토리노린재, 독나방, 동애등에, 돼지풀잎벌레, 된장잠자리, 두릅나무잎벌레, 두점배허리노린재, 두줄가지나방, 두줄점가지나방, 둥글노린재, 뒷노랑수염나방, 뒷노랑점가지나방, 뒷분홍가지나방, 등검은실잠자리, 딱부리긴노린재, 땅노린재, 땅벌, 땅별노린재, 떼허리노린재, 뚱보기생파리, 띠넓은가지나방, 루리알락꽃벌, 말굽무늬들명나방, 머루박각시, 머리폭긴노린재, 멋쟁이재주나방, 메추리노린재, 모메뚜기, 목도리불나방, 목창주둥이바구미, 목화명나방, 무늬독나방, 무늬뾰족날개나방, 무당벌레붙이, 묵은실잠자리, 물결넓적꽃등에, 뭉뚝바구미, 미디표주박긴노린재, 밀감무늬검정장님노린재, 밀잠자리, 밤색갈고리나방, 방아깨비, 방울실잠자리, 배노랑긴가슴잎벌레, 배둥글노린재, 배붉은흰불나방, 배자바구미, 배저녁나방, 배점무늬불나방, 배짧은꽃등에, 배추좀나방, 배치레잠자리, 배털가지나방, 뱀허물쌍살벌, 버드나무좀비단벌레, 벼룩잎벌레, 벼메뚜기, 별넓적꽃등에, 별노린재, 별쌍살벌, 보리장님노린재, 복숭아명나방, 북방풀노린재, 붉은등침노린재, 붉은무늬푸른자나방, 빨간촉각장님노린재, 사마귀, 산그물무늬짤름나방, 산바퀴, 세줄꼬마갈고리나방, 세줄날개가지나방, 세줄무늬수염나방, 세줄점가지나방, 솔버짐나방, 스미스개미, 시골가시허리노린재, 신부짤름나방, 실노린재, 십자무늬긴노린재, 쌍띠밤나방, 쌍복판눈수염나방, 쌍점흰가지나방, 쌍줄푸른밤나방, 썩은밤나방, 쑥잎벌레, 알꽃벼룩, 알노린재, 알락꽃등에, 알락주홍불나방, 알린콩알락파리, 암수다른장님노린재, 앞붉은명나방, 애모무늬잎말이나방, 애홍점박이무당벌레, 얼룩무늬좀비단벌레, 얼룩물결자나방, 얼룩애기들명나방, 얼룩장다리파리, 엉겅퀴밤나방, 엉겅퀴창주둥이바구미, 에사키뿔노린재, 여치, 연푸른가지나방, 오이잎벌레, 왕사마귀, 왕잠자리, 우단박각시, 은무늬밤나방, 일본잎벌레, 작은주걱참나무노린재, 장수말벌, 장수허리노린재, 재래꿀벌, 쟈바꽃등에, 적갈색긴가슴잎벌레, 점날개잎벌레, 점무늬불나방, 점박이둥글노린재, 점박이밤나방, 점분홍꼬마밤나방, 점흑다리잡초노린재, 점흰독나방, 좀말벌, 좁은가슴잎벌레, 주둥무늬차색풍뎅이, 주름개미, 줄검은들명나방, 줄고운가지나방, 줄고운노랑가지나방, 줄박각시, 줄보라집명나방, 줄우단풍뎅이, 줄점불나방, 쥐머리거품벌레, 참나무갈고리나방, 참딱부리긴노린재, 참땅벌, 청딱지개미반날개, 콩독나방, 콩잎벌레, 콩풍뎅이, 큰딱부리긴노린재, 큰무늬배짧은꽃등에, 큰뱀허물쌍살벌, 큰이십팔점박이무당벌레, 큰톱날물결자나방, 탈장님노린재, 태극나방, 털뿔가지나방, 톱날물결자나방, 투명잡초노린재, 파리매, 포도들명나방, 풀색꽃무지, 풀색노린재, 호박벌, 홍띠애기자나방, 홍맥장님노린재, 홍비단노린재, 홍색얼룩장님노린재, 홍줄불나방, 황어리호박벌, 회양목명나방, 희미무늬알노린재, 흰그물물결자나방, 흰띠거품벌레, 흰무늬긴노린재, 흰뾰족날개나방, 흰솜털검정장님노린재, 흰점박이꽃바구미, 흰점쐐기나방, 흰줄푸른자나방