



VIDEREUDVIKLING AF MODEL FOR METALLER I VANDLØBSVAND

MetalStat

Videnskabelig rapport fra DCE – Nationalt Center for Miljø og Energi

nr. 522

2023



AARHUS
UNIVERSITET
DCE – NATIONALT CENTER FOR MILJØ OG ENERGI

Videreudvikling af model for metaller i vandløbsvand

MetalStat

Videnskabelig rapport fra DCE – Nationalt Center for Miljø og Energi

nr. 522

202

Peter B. Sørensen¹
Christian F. Damgaard¹
Poul L. Bjerg²
Hans E. Andersen¹
Peter E. Holm³
Jesper L. Bak¹
Dorte Rasmussen⁴
Ane Kjeldgaard¹

¹Institut for Ecoscience, Aarhus Universitet

²Institut for Miljø- og Ressourceteknologi, Danmarks Tekniske Universitet

³Institut for Plante- og Miljøvidenskab, Københavns Universitet

⁴Dansk Hydraulisk Institut



AARHU
UNIVERSITET

DCE – NATIONALT CENTER FOR MILJØ OG ENERGI

Datablad

Serietitel og nummer:	Videnskabelig rapport fra DCE - Nationalt Center for Miljø og Energi nr. 522
Kategori:	Rådgivningsrapporter
Titel:	Videreudvikling af model for metaller i vandløbsvand
Undertitel:	MetalStat
Forfatter(e):	Peter B. Sørensen ¹ , Christian F. Damgaard ¹ , Poul L. Bjerg ² , Hans E. Andersen ¹ , Peter E. Holm ³ , Jesper L. Bak ¹ , Dorte Rasmussen ⁴ , Ane Kjeldgaard ¹ ,
Institution(er):	¹ Institut for Ecoscience, Aarhus Universitet, ² Institut for Miljø- og Ressourceteknologi, Danmarks Tekniske Universitet, ³ Institut for Plante- og Miljøvidenskab, Københavns Universitet, ⁴ Dansk Hydraulisk Institut
Udgiver:	Aarhus Universitet, DCE – Nationalt Center for Miljø og Energi ©
URL:	http://dce.au.dk
Udgivelsesår:	Marts 2023
Redaktion afsluttet:	Februar 2023
Faglig kommentering:	Martin Mørk Larsen
Kvalitetssikring, DCE:	Iben Boutrup Kongsfelt
Sproglig kvalitetssikring:	Else Vihlborg Staalsen
Ekstern kommentering:	Der er ikke kommenteret på rapporten.
Finansiel støtte:	Miljøstyrelsen har ydet finansiel støtte.
Bedes citeret:	Sørensen, P.B., Damgaard, C.F, Bjerg, P.L., Andersen, H.E., Holm, P.E., Bak, J.L., Rasmussen, D., Kjeldgaard, A., 2023. Videreudvikling af model for metaller i vandløbsvand, MetalStat. Aarhus Universitet, DCE – Nationalt Center for Miljø og Energi, 107 s. - Videnskabelig rapport nr. 522 http://dce2.au.dk/pub/SR522.pdf
	Gengivelse tilladt med tydelig kildeangivelse
Sammenfatning:	Denne rapport indgår i et større arbejde med udvikling af statistiske modeller til estimering af koncentrationer af miljøfarlige forurenende stoffer i overfladevand for Miljøstyrelsen ved at videreudvikle på en pilotudgave af en statistisk model (MetalStat) for forekomsten af 5 metaller som opløst koncentration i vandløbsvand. Rapporten tager således udgangspunkt i afrapportering af MetalStats principper givet i Sørensen et al., 2022. Modellen er sat op, så metalkoncentrationen i vandløbet regnes for at være bestemt af det ID15 opland, som målestationen er placeret i, hvor et ID15 opland har en varierende udstrækning på typisk 10-15 km ² . Opgaven for 2022 er at få pilotudgaven af MetalStat valideret og modnet, så den kan indgå i Miljøstyrelsens forvaltning.
Emneord:	metaller, vandløb, koncentrationsniveau, model, MetalStat
Foto forside:	Peter Borgen Sørensen
ISBN:	978-87-7156-735-9
ISSN (elektronisk):	2244-9981
Sideantal:	107
Internetversion:	Rapporten er tilgængelig i elektronisk format (pdf) som http://dce2.au.dk/pub/SR522.pdf

Indhold

Forord	5
Sammenfatning	6
Summary	9
1 Indledning	12
2 Nye målinger for perioden 2020-2022	14
2.1 Direkte sammenligning af målinger mellem perioderne 2011-2019 og 2020-2022	15
3 Modificering af nærværende version af MetalStat i forhold til 2021	18
3.1 Regneteknisk optimering	19
3.2 Massebalancekorrektion	20
4 Validering	28
4.1 På baggrund af målinger fra 2011-2019	29
4.2 På baggrund af målinger fra 2020-2022	31
5 Videreudvikling af MetalStat	40
5.1 Forbedring af beregningseffektivitet og numerisk præcision	40
Forbedring af X matricen	41
5.2 Forbedret inddragelse af grundvand	42
5.3 Inddragelse af udsivning fra forurenede grunde	44
6 Konklusion og implementering af MetalStat	46
6.1 Konklusion	46
7 Referencer	49
Bilag 1	50

Forord

Denne rapport beskriver resultatet af et projekt, hvis formål var at udvikle en pilot-model til prædiktion af forekomst af metaller i vandløb. Projektet er udført for Miljøstyrelsen med henblik på at kunne understøtte vandplanlægningen med modellering af koncentrationen af miljøfarlige forurenende stoffer i overfladevand bl.a. i forhold til, om der er risiko for forekomst af koncentrationer over miljøkvalitetskravene.

Projektgruppen består af: Peter Borgen Sørensen og Christian Frølund Damgaard har primært haft ansvar for udførelse af validering samt den matematiske konstruktion af modellen. Hans Estrup Andersen har primært bidraget med datagrundlag og miljøkemisk ekspertise. Poul L. Berg, Peter Engelund Holm og Jesper Leth Bak har primært bidraget med miljøkemisk forståelse til modellen samt forståelse af mulige kilder. Modelleringsprincipperne har været fremlagt overfor og diskuteret med Dorte Rasmussen. Ane Kjeldgaard har lavet rapportens GIS beregninger og figurer. Projektet har været fulgt af en gruppe med repræsentanter fra Miljøstyrelsen primært bestående af Rasmus Dyrmosé Nørregaard, Maj-Britt Bjergager, samt Kasper Lauge Thomassen, og fra DCE har projektet primært været varetaget af Iben Boutrup Kongsfelt. Projektet har været drøftet i denne gruppe. Projektet har desuden været fulgt af en styregruppe med repræsentanter for Miljøstyrelsen og miljøministeriets departement. Styregruppen er i løbet af projektperioden blevet orienteret om projektets fremdrift. Miljøstyrelsen har haft et udkast af rapporten til kommentering.

Projektet er udført som et udviklingsprojekt og den primære projektperiode var fra juni 2022 til december 2022.

Sammenfatning

Denne rapport indgår i et større arbejde under Miljøstyrelsen med udvikling af modeller til estimering af koncentrationer af miljøfarlige forurenende stoffer i overfladevand. Udgangspunktet i denne rapport er en pilot version af en statistisk model (MetalStat) for landsdækkende forekomst af metaller i vandløbsvand. Rapporten tager således udgangspunkt i afrapportering af MetalStats principper givet i Sørensen et al., 2022. Modellen er sat op, så metalkoncentrationen i vandløbet regnes for at være bestemt af det ID15 opland, som målestationen er placeret i, hvor et ID15 opland har en varierende udstrækning på typisk 10-15 km². Opgaven for 2022 er at få pilotudgaven af MetalStat valideret og modnet, så den kan indgå i Miljøstyrelsens forvaltning. Af regnetekniske grunde var det kun muligt i 2021 at afslutte beregninger for de tre metaller bly, cadmium og nikkel, selvom der også er omfattende data på kobber og zink, Sørensen et al., 2022. Dette skyldtes, at der er blevet justeret på modellens struktur helt hen til udgangen af 2021, og der stilles større krav til beregningstid med fem metaller end med tre. I denne rapport er MetalStat derfor udvidet med kobber og zink, så den behandler fem metaller i en multivariat statistik model, der udnytter korrelationen med metallerne i bestræbelsen på at levere en optimal anvendelse af monitoringsdata.

For de oplande, hvor metalkoncentrationen er blevet målt, anvendes de estimerede latente variable, som modellerer middelværdien af koncentrationen på en logskala. Denne middelværdi på log skala kan omregnes til mediankoncentrationen, hvilket betyder at hovedresultatet fra MetalStat er prædiktioner af mediankoncentrationer. For de oplande, der ikke har målinger laver MetalStat prædiktioner af mediankoncentrationen, med tilhørende usikkerhedsinterval, og er således en model, der dækker hele landet.

De estimerede latente variabler er tilpasset til en multinormal målemodel og renset for effekten af år, men effekten af år kan medregnes hvis ønsket. Prædiktionerne bliver beregnet som sandsynlighedsfordelinger, og disse kan summeres for de enkelte oplande ved f.eks. fordelings median eller andre fraktiler over et år. Usikkerheden af prædiktionen kan f.eks. gives ved et 95% usikkerheds interval (der er 95 % sandsynlighed, for at den rigtige værdi ligger inden for intervallet). Det er også muligt at angive den prædikterede sandsynlighed for, at der i oplandet er koncentrationer over miljøkvalitetskrav. Størrelsen af modellens usikkerhedsintervaller afhænger af hvor omfattende et datagrundlag, der indgår i modellen. Det er således forventeligt at MetalStat vil få betydeligt mindre usikkerhed når alle data fra perioden 2020-2022, der nu ikke er med i modellen, bliver brugt i modellen. Det vil betyde at de prædikterede øvre 95 % usikkerhedsintervaller vil blive reduceret. Det har dog ikke været muligt indenfor tidsrammen for denne rapport at få opdateret MetalStat med alle målinger fra perioden 2011-2022.

MetalStat kan dokumentere betydningen af forskellige kilder i forhold til den samlede andel af metalkoncentrationer i vandløb, som kan forklares med oplandets attributter. MetalStat vil også kunne understøtte en identifikation af med hvilken andel naturgivne kilder bidrager til metalkoncentrationen.

Hvis et ID15 opland, der ikke omfatter nogle målestationer og derfor fremstår som umålt, udveksler vand med et andet opland, i hvilken der er målinger, vil disse målinger i det målte opland kunne bidrage med viden om

koncentrationsniveauet i det umålte opland, da noget af vandet i det målte opland også optræder i det umålte opland. For at tage fuldt hensyn til denne rummelige sammenhæng, er der udviklet en massebalancekorrektions for at forbedre modellens prædiktioner, så målinger i målte ID15 oplande kan bidrage med viden i andre oplande uden målinger, i det omfang udveksling af vand berettiger til det.

I MetalStat indgår en X matrice, der samler oplysning fra GIS om hvert opland, (se Sørensen et al., 2022). Som det også er beskrevet i Sørensen et al., 2022, så vil det derfor være muligt løbende at evaluere og forbedre X matricen, i takt med at flere målinger og nye datakilder bliver tilgængelige. En regressionskoefficient R^2 på logtransformerede koncentrationer er beregnet ud fra, hvor godt X matricen sammen med sæsonvariationen kan prædiktere den latente koncentration i de målte oplande. For fire ud af fem metaller var intervallet 0,19-0,3, mens Nikkel faldt ud med en lav værdi på 0,06. Så der er klart plads til forbedring af især X matricen. Det er projektgruppens vurdering, at det første skridt mod en forbedring er at inddrage grundvandsbidraget mere detaljeret. En lille værdi for R^2 betyder, at MetalStat beskriver det meste af variationen i koncentrationsværdierne, dels som tilfældige effekter mellem oplande og år og dels som residualer, der varierer tilfældigt fra den ene måling til den næste.

Nye målinger fra perioden 2020-2022 er anvendt til at evaluere MetalStat for de fem metaller. Der har været et ret intensivt måleprogram i denne periode, hvor der er indsamlet ca. dobbelt så mange prøver i perioden 2020-2022 som for hele perioden 2011-2019, fra ca. 400 oplande. En deskriptiv statistisk analyse af de målinger fra perioden 2020-2022, der var målt over detektionsgrænsen, indikerer generelt et lavere koncentrationsniveau sammenlignet med målingerne for perioden 2011-2019, især for zink. MetalStat bliver dog ikke udfordret af sådanne årlige forskydninger i koncentrationsniveau i det omfang disse ikke er større end den forskel mellem år, der optræder i perioden 2011-2019. Den efterfølgende analyse af de enkelte metaller viser at den årlige forskydning udfordrer MetalStat for zink, men ikke for de andre fire metaller.

MetalStat er brugt til at beregne realistiske intervaller for mediankoncentrationen for de nye målte oplande, så de kan sammenlignes med de nye målinger fra perioden 2020-2022. Disse intervaller afspejler niveau og usikkerhed for mediankoncentrationen i det pågældende opland og kun i relativt få tilfælde, udfordres de beregnede intervaller af de nye målinger. En nærmere granskning af de enkelte større afvigelser for især bly viser, at oplandsstørrelsen på 10-15 km² i nogle få tilfælde godt kan dække over en heterogenitet, som ikke synes helt at være repræsentativt dækket ind af de valgte placeringer af NO-VANA stationer. For zink giver MetalStat generelt en overprædiktion af usikkerhedsintervallet, som konsekvens af det voldsomme fald, der generelt har været i koncentrationsniveauet mellem periode 2011-2019 og periode 2020-2022. Når MetalStat opdateres med målingerne fra perioden 2020-2022, vil det derfor være en fordel at estimere variablerne for zink fra år 2020 som en ny latent variabel, så der er to versioner af latente variabler for zink i modellen, et før 2020 og et andet fra 2020.

Der er blevet inkorporeret nogle regnetekniske forbedringer i den nærværende version af MetalStat i forhold til afrapporteringen i Sørensen et al., 2022. Det er især brug af en oplandslokal likelihood funktion for de latente koncentrationer, der har mindsket beregningstiden. Det betyder, at MetalStat i en beregningsgang, der tester værdien for en latent koncentration, nu kun skal

regne specifikt på det ID15 opland, som denne latente koncentration estimeres i, uden at genberegne alle de andre oplande. I bestræbelserne på at optimere X matricen afprøves såkaldte "horseshoe priors", der finder vigtigste betydende faktorer med mindre beregningstid, men denne afprøvning er ikke tilendebragt med udgivelsen af denne rapport.

Under udarbejdelse af denne rapport blev det klart, at en driftsversion af MetalStat ikke blev færdiggjort til forvaltningen i 2022, men at denne aktivitet vil blive søgt udført i 2023. Der er dog foretaget en planlægning af kodning af driftsversionen for at fremskynder arbejdet. Under denne planlægning udarbejdedes en klar plan for at parallelisere MetalStat til at accelerere beregningshastigheden over flere CPU-kerner. Så der foregår løbende en udvikling, der minimerer beregningstiden, dels ved at regne smartere og derfor mindske behovet for beregninger, dels ved at øge den computerkraft som kan udnyttes til at regne hurtigere.

Den væsentligste forestående forbedring af X matricen vurderes at være en bedre beskrivelse af grundvandsbidraget, for bedre at beskrive det geogene bidrag, der ikke skyldes menneskelig aktivitet. En sådan forbedring bør laves som led i udvikling af en driftsversion af MetalStat. Det anbefales, at der tages udgangspunkt i eksisterende undersøgelser, der relaterer metaller i grundvand til geologi, pH og redoxforhold. X matricen kan udvides med disse attributter, og MetalStat vil kunne finde den reelle betydning af disse på vandløbskoncentrationerne. Da relationen mellem disse attributter og metalkoncentrationen i grundvand er stærkt ikke-lineær, er det bedst at bruge en model, der anvender afskæringsværdier i stedet for at forudsætte en kontinuert funktion mellem pH og redox på den ene side og metalkoncentrationen på den anden side. Derfor anbefales det, at der foretages en indledende analyse af sammenhænge mellem metalkoncentration i grundvand, pH, Redox og geologi, som skal danne input til beskrivelsen med X matricen i MetalStat. Første skridt skal være at sikre adgang til relevante GIS lag i samarbejde med relevante institutioner.

Forurenede grunde kunne også inddrages i X matricen, men det vurderes ikke at være så betydende som grundvandsbidraget for de fem metaller i MetalStat, hvorfor denne aktivitet skal prioriteres efter en inddragelse af grundvand. Med disse forbehold gives der i denne rapport et forslag til en model for forurenede grunde, som kunne bygges ind i MetalStat med udgangspunkt i en eksisterende screeningsmodel for forurenede grundenes potentielle påvirkning af vandløb.

Summary

This report is part of a major work under the Danish Environmental Protection Agency to develop models for estimating concentrations of environmentally hazardous pollutants in surface water. The starting point in this report is a pilot version of a statistical model for the Danish Environmental Protection Agency (MetalStat) for the nationwide occurrence of dissolved metals in running waters. The report is, thus, based on reporting of MetalStat's principles given by Sørensen et al., 2022. Nationwide, the model distributes the metal concentration in the running waters into the ID15 catchments, in which the measuring stations are located, where an ID15 catchment has a varying area of typically 10-15 km². The task for 2022 is to validate and improve the pilot version of MetalStat to be included in the Environmental Administration's management. For calculation reasons, in 2021 it was only possible to complete calculations for the three metals lead, cadmium and nickel, although there are also extensive data on copper and zinc. This was due to the fact that the structure of the model has been adjusted right up to the end of 2021, and there are greater requirements for calculation time with five metals than with three. In this report, MetalStat has therefore expanded with copper and zinc, so that it deals with five metals in a multivariate statistical model that utilizes the correlation with the metals in the effort to provide an optimal use of monitoring data.

For the catchment areas where the metal concentration has been measured, the mean of the concentration on a log scale is estimated as a latent variable. As the mean value on the log scale corresponds to the median concentration, the main outcome from MetalStat is predictions of median concentrations. For those catchment areas that do not have measurements, MetalStat makes predictions of the median concentration, with an associated uncertainty interval, and is, thus, a model that covers the entire country.

The estimated latent variables are adapted to a multinormal model, in which the effect of years is quantified. All estimations and predictions are calculated as probability distributions for every ID15 catchment. So, the uncertainty of the prediction can be given at a 95% uncertainty interval (there is a 95% probability that the correct value is within the range). It is also possible to predict the probability that there are concentrations above environmental quality standards in the ID15 catchment. The size of the model's uncertainty depends on the amount of monitoring data applied in the model. Thus, it is expected that MetalStat will predict significantly less uncertainty when all data from the period 2020-2022, which in this report are not included in the model, are applied in the next version of the model. This will mean that the predicted upper 95% uncertainty intervals will be reduced. However, it has not been possible within the time frame for this report to update MetalStat with all measurements from the period 2011-2022.

MetalStat can document the importance of different metal sources in relation to the total share of metal concentrations in watercourses, which can be explained by the attributes of each ID15 catchment. MetalStat will also be able to support identification of the proportion of natural mineralogical sources that contribute to the metal concentration.

If an ID15 catchment area does not include any measuring stations, and therefore appears as unmeasured, exchanges water with another ID15 catchment area, in which there are measurements, then these measurements in the measured ID15 catchment will contribute knowledge about the concentration level in the unmeasured ID15 catchment, as some of the water in the measured ID15 catchment also occurs in the unmeasured ID15 catchment. This is utilized in MetalStat using a mass balance correction to improve the certainty of the model's predictions, so that measurements in measured ID15 catchments can contribute knowledge in other ID15 catchments without measurements.

MetalStat includes an X matrix that collects information from GIS about each catchment area (see Sørensen et al., 2022). As also described in Sørensen et al., 2022, it will therefore be useful to continuously evaluate and improve the X matrix as more measurements and new GIS data sources become available. A regression coefficient R^2 on log-transformed concentrations is calculated based on how well the X matrix, together with seasonal variation, can estimate the latent concentration in the measured catchments. For four out of five metals, the range was 0.19-0.3, while Nickel fell out with a low value of 0.06. Thus, there is clearly room for improvement of the X matrix in particular. It is the project group's assessment that the first step towards an improvement is to include the groundwater contribution in more detail. A small value for R^2 means that MetalStat describes most of the variation in concentration values on a logarithmic scale, partly as random effects between catchments and years, and partly as residuals that vary randomly from one measurement to the next.

New measurements from the period 2020-2022 have been used to evaluate MetalStat for the five metals. There was an intensive measurement program during this period, with approximately twice as many samples collected in the period 2020-2022 as in the entire period 2011-2019. These measurements cover approximately 400 ID15 catchments. A descriptive statistical analysis of the measurements from the period 2020-2022 that were measured above the detection limit generally indicates a lower concentration level compared to the measurements for the period 2011-2019, especially for zinc. MetalStat is not necessarily challenged by annual shifts in concentration levels, if these are not much greater than those occurring in the period 2011-2019. However, the subsequent analysis of the individual metals shows that the annual displacement challenges MetalStat for zinc, but not for the other four metals.

From MetalStat, a probability density function was generated to map the realistic intervals of the median concentration for each ID15 catchment that was measured during the period 2020-2022. Only in very few cases, the new measured concentrations were considerably outside the predicted interval of MetalStat. A closer examination of the individual major deviations for especially lead shows that for a few cases, the ID15 catchment area size of 10-15 km² may not be fully representatively covered by the selected locations of NOVANA stations. For zinc, MetalStat generally provides an overprediction of the uncertainty interval because of the sharp decrease that has generally occurred in the concentration levels between period 2011-2019 and the period 2020-2022. When MetalStat is updated with the measurements from the period 2020-2022, it will therefore be advantageous to estimate the variables for zinc from the year 2020 as a new latent variable, so that there are two versions of latent variables for zinc in the model, one before 2020 and another from 2020.

Additional calculations of technical improvements have been made in the present version of MetalStat in relation to the reporting in Sørensen et al., 2022. In particular, the use of a local likelihood function in each catchment for the latent concentrations has reduced the calculation time. This means that a calculation process that tests the value of a latent concentration now only must calculate specifically on the ID15 catchment area in which this latent concentration is estimated, without recalculating all the other catchment areas. In the effort to optimize the X matrix, so-called “horseshoe priors” are being tested, which will find most important significant factors with less calculation time, but this test has not been completed with the release of this report.

During the preparation of this report, it became clear that an operating version of MetalStat was not finalized for administration in 2022, but that this activity will be sought to be carried out in 2023. However, a planning of coding of the operating version has been made to speed up the work. During this planning, a clear plan was drawn up to parallelize MetalStat to accelerate the computational speed using multiple CPU cores. Thus, there is an ongoing development that minimizes the calculation time, partly by calculating smarter, and therefore reducing the need for calculations, and partly by increasing the available computing power.

The most obvious improvement of the X matrix to be considered as a next step is to make a better description of the groundwater contribution to better describe the geogenic contribution that is not due to human activity. Such an improvement should be made as part of the development of an operating version of MetalStat. It is recommended that existing studies that relate metals in groundwater to geology, pH and redox conditions are taken as a starting point to extend the X matrix. Since the relationship between these attributes and the metal concentration in groundwater is highly non-linear, it is best to use a model that uses cut-off values rather than assuming a continuous function between pH and redox on one side and the metal concentration on the other. Therefore, it is recommended that an initial analysis discloses correlations between metal concentration in groundwater, pH, Redox and geology, which will constitute input to the description with the X matrix in MetalStat. The first step will be to ensure access to relevant GIS layers in cooperation with relevant institutions.

Contaminated sites could also be included in the X matrix, but this is not considered to be as important as the groundwater contribution for the five metals in MetalStat, which is why this activity must be prioritized after a groundwater inclusion. With these reservations, this report proposes a model for contaminated sites that could be built into MetalStat based on an existing screening model for the potential impact of contaminated sites on water flows.

1 Indledning

Denne rapport indgår i et større arbejde med udvikling af statistiske modeller til estimering af koncentrationer af miljøfarlige forurenende stoffer (MFS) i overfladevand for Miljøstyrelsen ved at videreudvikle en pilotudgave af en statistisk model (MetalStat) for forekomsten af metaller i vandløbsvand. Rapporten tager således udgangspunkt i afrapportering af MetalStats principper givet i [Sørensen et al., 2022](#). Modellen er sat op, så metalkoncentrationen i vandløbet regnes for at være bestemt af det ID15 opland, som målestationen er placeret i, hvor et ID15 opland har en varierende udstrækning på typisk 10-15 km². Opgaven for 2022 er at få pilotudgaven af MetalStat valideret og modnet, så den kan indgå i Miljøstyrelsens forvaltning, med følgende fokuspunkter:

- Udvidelse med to metaller (kobber og zink) ud over de tre metaller (bly, cadmium, nikkel), som indgik i afrapportering for 2021 (Sørensen et al., 2022). Af regnetekniske grunde var det kun muligt i 2021 at afslutte beregninger for de tre metaller bly, cadmium og nikkel, selvom der også er omfattende data på kobber og zink. Dette skyldtes, at der er blevet justeret på modellens struktur helt hen til udgangen af 2021, og der stilles større krav til regnetid med fem metaller end med tre. Det er ikke optimalt at beregne for et metal ad gangen med modellen, da man så ikke udnytter den multivariate facilitet i modellen, der nyttiggør korrelationen mellem metallerne til at kvalificere prædiktionerne.
- Sammenligning af modellens prædiktioner med nye målinger for perioden 2020-2022. Siden opsætningen af modellen begyndte, er der kommet nye målinger i NOVANA programmet, hvorfor det nu er muligt både at validere prædiktioner for ID15 oplande målt i perioden 2020-2022, fra hvilke der ikke indgår målinger i modellen, og estimerede latente koncentrationer for oplande, fra hvilke der både indgår målinger i modellen, og hvor der er nye målinger fra perioden 2020-2022.
- En model bliver aldrig fuldkommen, så det vil altid være en fordel at undersøge, om der kan findes forbedringer, så en model afspejler den bedste udnyttelse af tilgængelig viden og med troværdig bestemmelse af den tilknyttede usikkerhed. I denne rapport er MetalStat videreudviklet i forhold til:
 - Optimering af pilotmodellens beregningsteknik, så modellen hurtigere estimerer parametrene betydning som forklarende variable.
 - Fastlæggelse af strategi for en bedre beskrivelse af metalkilder fra jordforurening. I den nuværende opsætning af MetalStat tages der ikke hensyn til en evt. forekomst af en forurenende grund i et opland. Der foreslås en mulig metode, der kan rette op på denne mangel.
 - Fastlæggelse af strategi for en bedre beskrivelse af metalkilder fra geogene kilder. I den nuværende version af MetalStat regnes en grundvandsindstrømning som mulig kilde til metaller dog uden at tage hensyn til hvorhenne i landet den foregår. I rapporten gives anvisning på, hvordan denne beskrivelse kan forbedres, så grundvandbidraget kan differentieres efter geologien og andre forhold i den aktuelle grundvandsforekomst.
 - Forbedring af oplandsbeskrivelsen ved at inddrage massebalance mellem ID15 oplande, der er oplande på typisk 1500 ha (Troldborg et al, 2015). Vandløbsstrækningen gennem et sådant

opland er typisk på et begrænset antal km, så det tager kun nogle få timer for vandet at passere et ID15 opland. Viden om målinger i et ID 15 opland kan derfor komme andre ID15 oplande til gode, i det omfang, de udveksle vand med hinanden.

2 Nye målinger for perioden 2020-2022

Den nuværende version af MetalStat er udviklet med brug af målinger fra perioden 2011 til 2019, der er målt med varierende intensitet. For at undersøge modellens prædiktioner er output fra modellen sammenlignet med målinger fra den efterfølgende periode 2020 frem til den 13/9-2022. En samlet oversigt over målinger fra hhv. 2011-2019 og 2020-2022 er vist i Tabel 2.1, og det fremgår tydeligt, at der har været en intens aktivitet med nye målinger i perioden 2020-2022. Desuden ses det, at den procentvise andel af prøver, der er målt under detektionsgrænsen er vokset, hvilket indikerer et faldende koncentrationniveau. Der er dog sket mindre forskydninger i detektionsgrænsen for nogle metaller, men dette kan ikke forklare hele forskellen.

Tabel 2.1. Oversigt over prøveantal fordelt på metaller og de to perioder hhv. 2011-2019 og 2020-2022, hvor 2022 ikke er helt fuldkommen, da analysen i dette projekt skulle afsluttes før udgangen af 2022.

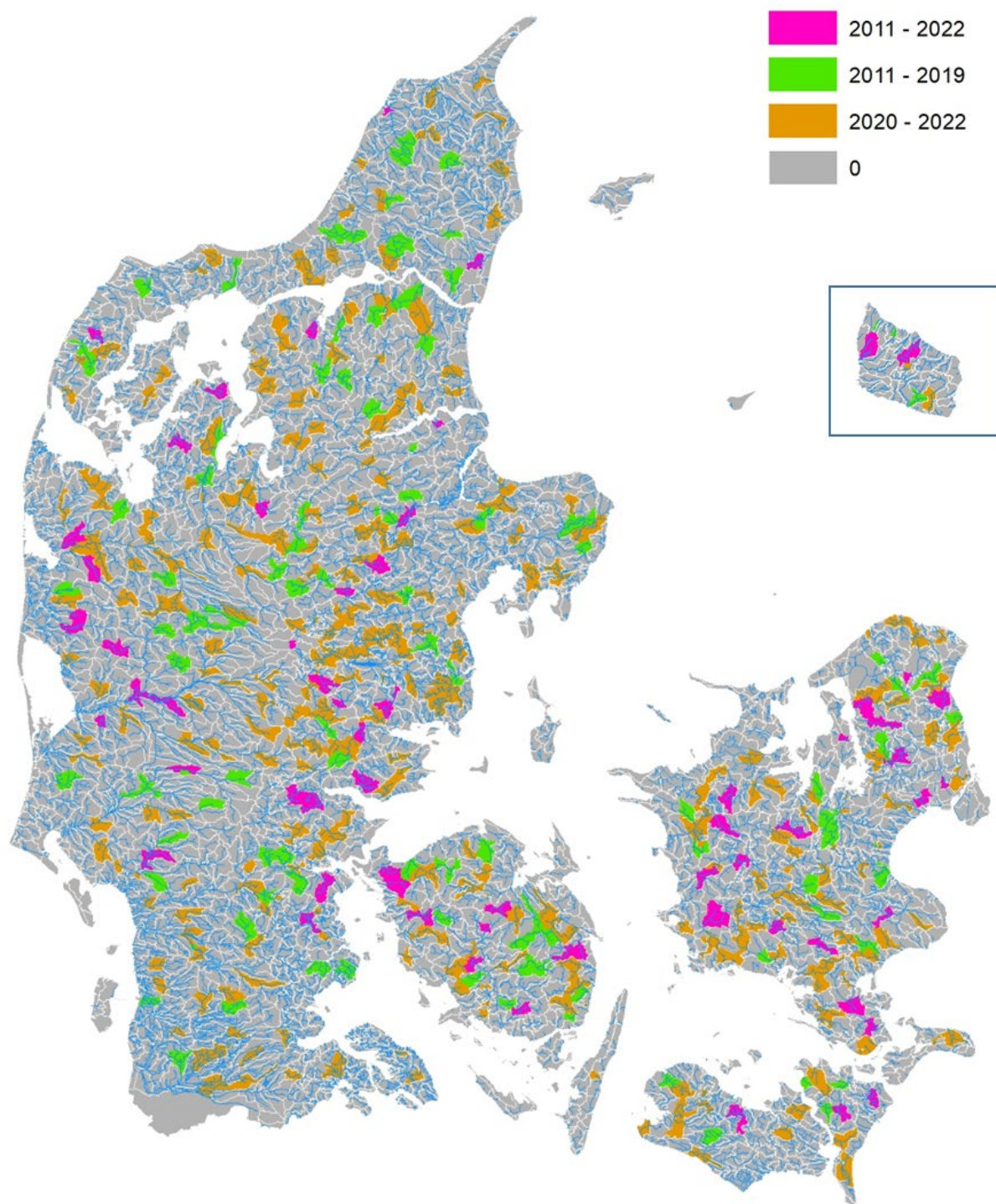
Metal	Periode 2011-2019			Periode 2020-2022 (13/9-22)		
	I alt antal	Under det.gr. Antal	%	I alt antal	Under det.gr. Antal	%
Bly	2249	589	26	4364	1920	44
Cadmium	2296	219	10	4630	734	16
Kobber	1309	3	0	3516	27	1
Nikkel	2247	5	0	4362	49	1
Zink	1127	7	1	3447	462	13

Målingerne i Tabel 2.1 er fordelt ud på ID15 oplande i Tabel 2.2 og inddelt i tre grupper: (1) Dem der har målinger fra både 2011-2019 og 2020-2022; (2) Dem med målinger alene fra 2011-2019; (3) Dem med målinger alene fra 2020-2022.

Tabel 2.2. Antal målte ID15 oplande fordelt på de perioder, i hvilke der er målt, hvor de oplande med perioden 2011-2022 både har været målt som input til MetalStat for perioden 2011-2019 og efterfølgende også er blevet målt i perioden 2020-2022.

Periode	Antal ID 15 oplande
2011-2022	66
2011-2019	107
2020-2022	343
I alt:	516

Den geografiske fordeling af de målte oplande fremgår af Figur 2.1, inddelt i tre grupper, alt efter om de indgår i hele perioden (2011-2022) eller kun i henholdsvis 2011-2019 eller 2020-2022. For de oplande, der indgår i hele perioden 2011-2022, vil de målinger, som er indsamlet i perioden 2011-2019, være med i kalibreringen af MetalStat, mens de prøver som er indsamlet senere, ikke vil være det.

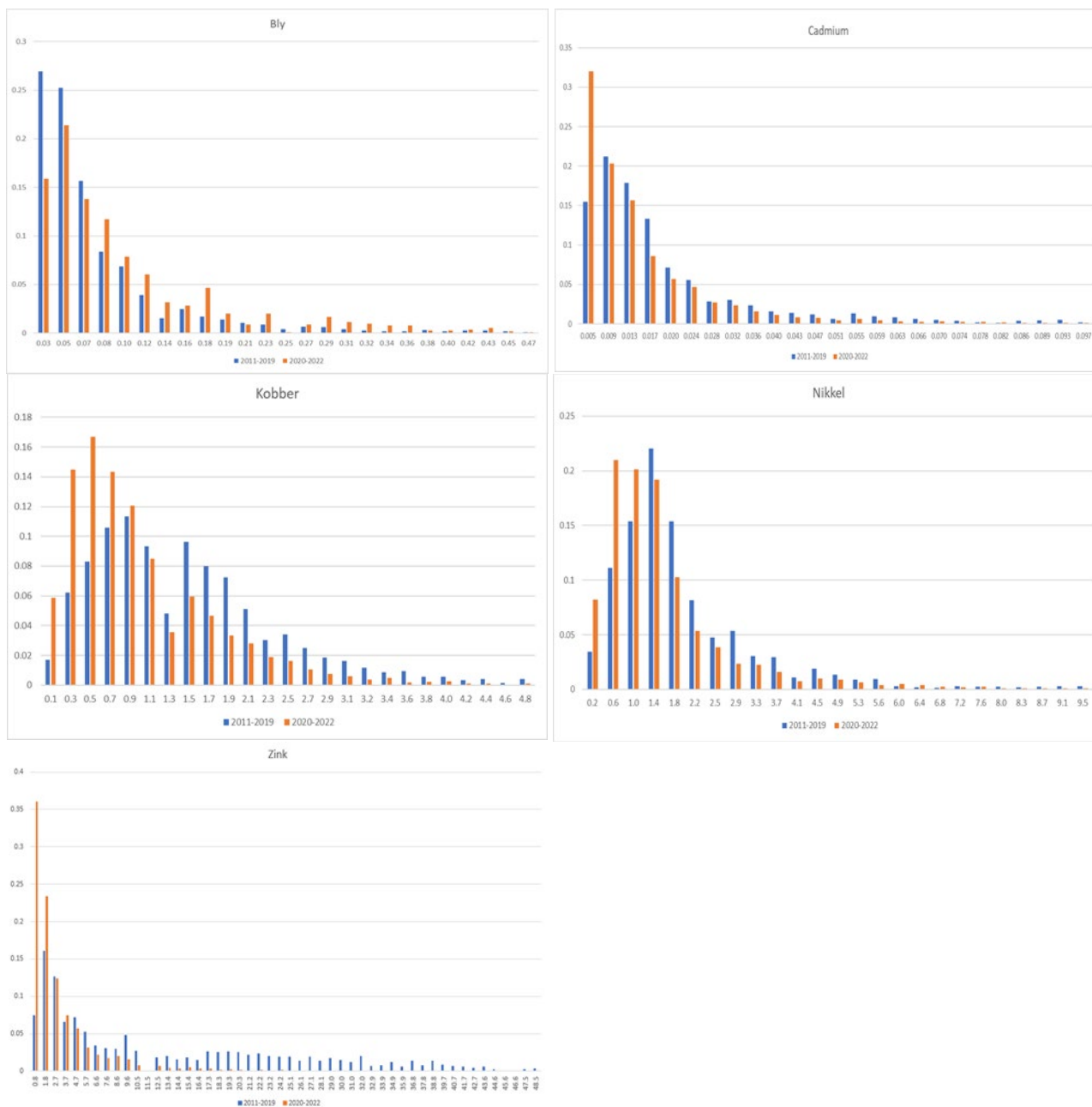


Figur 2.1. Målte oplande, hvor farvekode viser, hvilken periode prøverne er indsamlet i.

2.1 Direkte sammenligning af målinger mellem perioderne 2011-2019 og 2020-2022

Som optakt til at MetalStat bliver valideret med brug af måling fra perioden 2020-2022 (frem til 13/9-22), vil alle målinger over detektionsgrænsen indgå i en deskriptiv sammenligning af målte værdier for hhv. perioden 2011-2019 og 2020-2022. Den viste deskriptiv sammenligning tager i modsætning til MetalStat ikke hensyn til målinger under detektionsgrænsen, samt forskelle mellem oplande, år og sæson. Med det forbehold vil større afvigelser kunne indikere en uoverensstemmelse, som potentielt kan give et aftryk, når MetalStat sammenlignes med målingerne. På Figur 2.2 ses histogrammer, der viser

fordelingen af målte metalkoncentrationer. Kun de værdier, der er målt over detektionsgrænsen, er inkluderet. Bortset fra bly viser histogrammerne på Figur 2.2 en overhyppighed af lave værdier for de målte koncentrationer i perioden 2020-2022 sammenlignet med perioden 2011-2019. Især for zink er dette meget tydeligt, hvor der i perioden 2011-2019 er en lang hale af høje værdier, som er helt fraværende i perioden 2020-2022. Der ses også en nedgang i niveau for kobber, som er knap så voldsom som for zink, men zink og kobber er de to metaller, der synes at have det største fald i koncentration mellem de to perioder. Denne nedgang i de målte zink- og kobberkoncentrationer falder sammen med en ændret regulering af disse to metaller i landbruget, hvilket kunne betyde, at den påvirkning fra landbruget på zink- og kobberkoncentrationen, som er kalibreret ind i MetalStat i perioden 2011-2019, før regulering af zink og kobber trådte i kraft, ikke er optimal for prædiktering af disse to metaller i perioden 2020-2022. En anden forklaring kunne være, at dataindsamlingen for perioden 2020-2022 havde mere fokus på oplande i byområder, sammenlignet med perioden 2011-2019, og i så fald vil MetalStat potentielt kunne tage højde for forskellen gennem X matricen.



Figur 2.2. Histogram over målte koncentrationer, hvor x akser er koncentration, $\mu\text{g/l}$ og y akser er andel af målinger i intervallet fra hhv. perioden 2011-2019 og 2020-2022. Kun målinger over detektionsgrænsen er taget med.

3 Modificering af nærværende version af MetalStat i forhold til 2021

Siden MetalStat blev beskrevet i Sørensen et al., 2022, er der foretaget nogle forbedringer, som er implementeret i de beregninger, der er grundlaget for denne rapport. Desuden indførelse af en model, der forbedrer overensstemmelserne i massebalancerne mellem de oplande, der udveksler vand. Som optakt til en nærmere beskrivelse af modellen, er det nyttigt at definere den forskel mellem estimat og prædiktion, som er anvendt i rapporten og beskrevet i følgende afsnit.

Forskel mellem "estimat" og "prædiktion"

De to ord "estimat" og "prædiktion" kan umiddelbart virke som synonymmer, men for at opnå en mere præcis forståelse af modellen, indføres følgende skelnen mellem disse to begreber.

De oplande, der har målinger fra perioden 2011-2019 vil indgå i MetalStats likelihood-funktion, hvor der *estimeres* en latent koncentration for hver af disse oplande, og en usikkerhed på de enkelte målinger. Den latente koncentration svarer til logaritmen til mediankoncentrationen, i de oplande der har målinger. Derudover *estimeres* de parametre, som indgår i MetalStat, også som f.eks. koefficienterne til X matricen.

Estimeringerne fra likelihood-funktionen bruges til at *prædiktere* mediankoncentrationen i alle oplande, inklusive dem, hvor der ikke er målinger fra perioden 2011-2019. Der er også muligt for alle oplande at *prædiktere* en forventet målt koncentration.

I de oplande, hvorfra der er målinger fra perioden 2011-2019, er der således både en *estimeret* og *prædikteret* mediankoncentration, hvor den estimerede mediankoncentration typisk vil være mindre usikker, sammenlignet med den prædikterede, fordi estimatet er tæt knyttet til de målinger, der er foretaget i det specifikke opland.

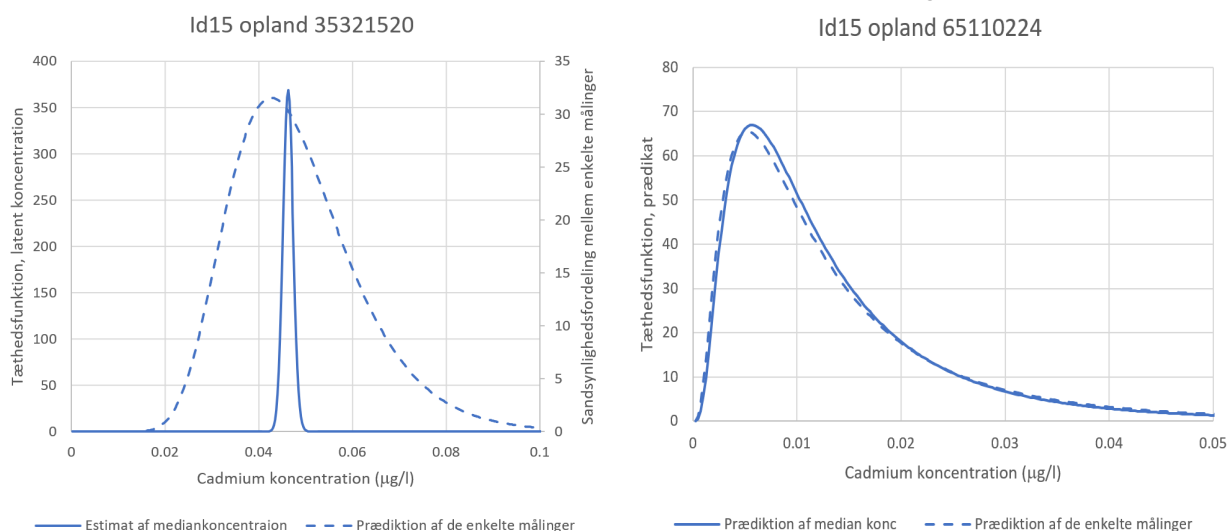
Opdeling af usikkerhed

Det er vigtigt at skelne mellem usikkerhed på mediankoncentrationen og usikkerheden mellem enkelte målte koncentrationer, da MetalStat opdeler kilder til usikkerhed i to typer:

- Måleusikkerhed, der beskriver hvor godt, en enkeltmåling repræsenterer koncentrationsniveauet i ID15 oplandet (mediankoncentration)
- Usikkerhed på prædiktion af mediankoncentrationen.

Forskellen mellem disse to kilder til usikkerhed er fundamental, da måleusikkerheden er en samlet usikkerhed, der indeholder prøvetagningens stikprøveusikkerhed i forhold til at beskrive ID15 oplandet, og usikkerheden ved selve analysen. Forskellen mellem måleusikkerhed og usikkerhed på mediankoncentration illustreres på Figur 3.1. Til venstre på Figur 3.1 (ID15 35321520), ses to fordelinger: (1) Tæthedsfunktion af mediankoncentration; (2) Tæthedsfunktion for de enkelte målte koncentrationer. Da ID15 35321520 er et af de oplande, der er målt på i perioden 2011-2019, så er mediankoncentration

bestemt med lille usikkerhed ud fra estimatet, sammenlignet med den usikkerhed der er afspejlet i de enkelte målinger. Det ses på Figur 3.1, ved at tæthedsfunktionen for mediankoncentrationen er meget mere snæver end tæthedsfunktionen for de enkelte målinger. Der er altså så mange målinger med i MetalStat fra dette målte opland, at usikkerheden på estimatet af mediankoncentrationen er relativt begrænset. Den stiplede sandsynlighedsfordeling for det målte oplande beskriver de værdier, der kan forventes at blive målt i oplandet, og er således et resultat af både usikkerheden på mediankoncentrationen og måleusikkerheden. Da den stiplede tæthedsfunktion er meget ”breddere” end den fuldt optrukne, viser det, at måleusikkerheden dominerer over usikkerheden på mediankoncentrationen. Ekstramålinger indsamlet i ID15 35321520 vil derfor ikke bidrage meget til at mindske usikkerhed ved estimering af mediankoncentrationen. Sandsynlighedsfordelingen til højre på Figur 3.1 viser et andet opland (ID15 65110224), hvor der ikke er nogle målinger fra perioden 2011-2019, og hvor MetalStat derfor prædikerer koncentrationen alene ved hjælp af information fra de andre målte oplande. Pga. manglende målinger for dette opland ses en usikkerhed på mediankoncentrationen, som er langt større end for ID15 35321520, og forskellen mellem måleusikkerhed og usikkerhed på den prædikerede mediankoncentration er stort set elimineret på figuren. I dette tilfælde vil ekstra målinger i oplandet have stor betydning for at mindske usikkerheden på den prædikerede mediankoncentration.



Figur 3.1. Tæthedsfunktion for hhv. mediankoncentration og for de enkelte målinger. ID15 opland 35321520 er målt, hvilket understøtter estimatet af den latente koncentration, mens det ikke er tilfældet for ID15 opland 65110224.

3.1 Regneteknisk optimering

I forhold til den tidligere rapporterede model er den tilfældige effekt af år flyttet til målemodellen, så de målte metalkoncentrationer, $\log(C_{m,i,y,s})$ for metal m i et opland i , i år y og sæson (årstid) s nu bliver modelleret ved hjælp af en Bayesiansk hierarkisk model, hvor målemodellen er

$$\log(C_{m,i,y,s}) \sim MN(lc_{L,i,s} + \sigma_y \gamma_y, \Sigma_m) \quad (1)$$

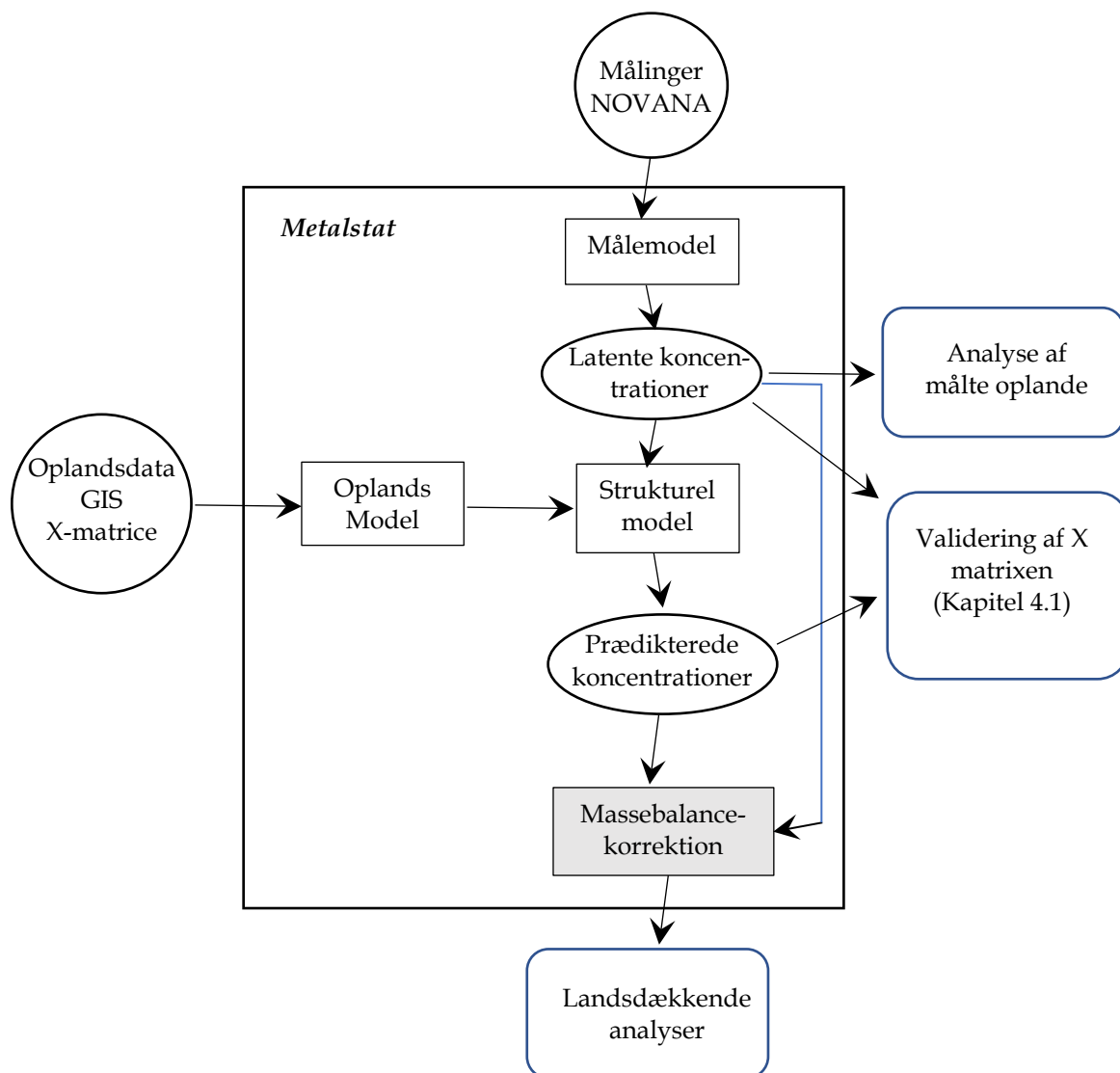
hvor $lc_{L,i,s}$ er latente variabel, som modellerer middelværdier af $\log(C_{m,i,y,s})$ uden effekten af årsvariation ($\sigma_y \gamma_y$) på en logskala, og Σ_m er den estimerede kovariansmatrice for de forskellige metaller af dimension $m \times m$. Effekten af år udgår af strukturmodellen.

Modellen er tilpasset de fem metaller, og udvælgelsen af data er ændret, så kun prøver, hvor mindst tre af de fem metaller er målt, er anvendt til modellen. I alt blev der anvendt 2.236 prøver, fra 168 oplande, hvoraf de fleste indeholder målinger for 5 metaller, og der var i alt 665 latente variable (oplande * sæson) i det samlede datasæt.

Der er opstillet nye log-likelihood funktioner for målemodellen, således at de latente variabler for hvert opland tilpasses uafhængigt. Dette har givet en betydelig hastighedsforbedring (ca. 50 gange ved tilpasning af latente variabler).

3.2 Massebalancekorrektion

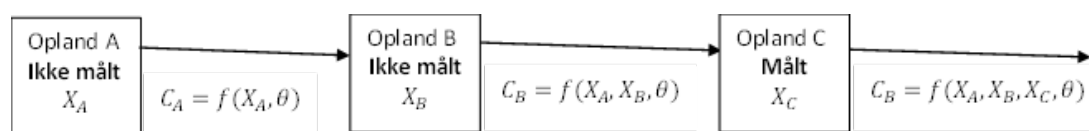
Den version af MetalStat som blev afleveret i 2022 (Sørensen et al., 2022), bliver i dette afsnit udbygget med en korrektion, der sikrer en bedre rummelig beskrivelse. Dette illustreres nedenfor i Figur 3.1., der tager udgangspunkt i Figur 4.2. i Sørensen et al., 2022, hvor der er tilføjet en massebalance korrektion (grå boks i Figur 3.2) inden resultatet fra modellen vil indgå i landsdækkende analyser af koncentrationsniveauer.



Figur 3.2. Samlet modelstruktur, hvor målinger fra NOVANA programmet indgår dels i en målemodel, der kan analysere koncentrationsniveauerne i de målte oplande, dels sammen med oplandsmodellen i en strukturel model og dels i en massebalance-korrektion til prædiktion af koncentrationer for hele landet

Når MetalStat prædikerer et koncentrationsniveau i et opland, anvendes dels X matrixen og dels de statistiske parametre, såsom bl.a. de tilfældige effekter for oplande og år. Dette princip er vist på Figur 3.3, hvor i alt tre ID15 oplande udveksler vand, med opland A som det øverste opland, opland B som det næste og opland C som det nederste nedstrøms opland, og hvor C_A , C_B , C_C er de prædikterede værdier af den latente metalkoncentration for hhv. opland A, B og C. Prædiktionen afhænger dels af X matrixen (X) og dels af de statistiske parametre (θ). Mens X matrixen afhænger af forholdene i oplandet (landbrug, spildevandtilledning, etc.), så er de statistiske parametre generiske for hele landet. Betydningen af X matrixen illustreres på Figur 3.3, hvor udløbskoncentrationen fra opland A (C_A) er en funktion, dels af X matrixen i opland A (X_A) og dels af de statistiske parametre (θ). Udløbskoncentrationen fra opland B er derimod både en funktion af forholdene i opland A (X_A) og opland B (X_B) samt de statistiske parametre (θ). Efter samme princip er udløbskoncentrationen fra opland C både en funktion af forholdene i opland A (X_A), opland B (X_B) og opland C (X_C) samt de statistiske parametre (θ). I dette illustrative eksempel er der målinger i opland C i perioden 2011-2019, hvilket

styrker den statistiske model i dens estimering af den latente koncentration C_B .



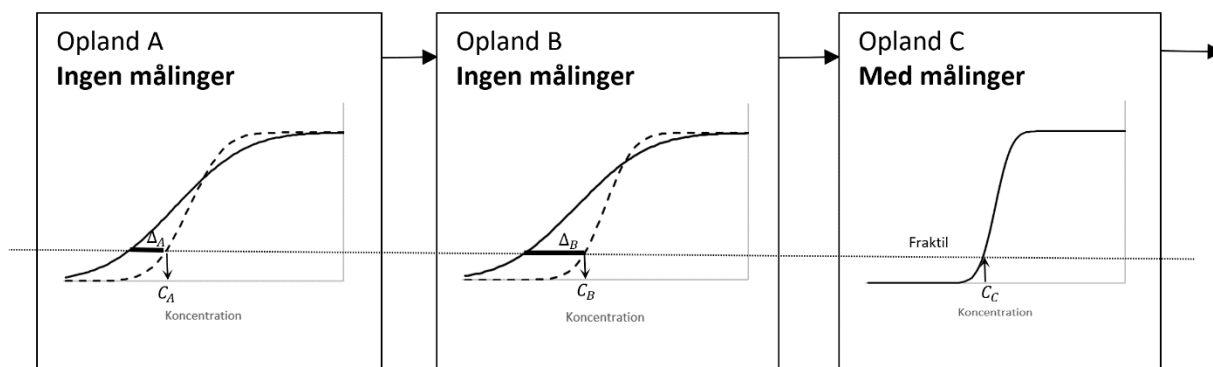
Figur 3.3. Princip for hvordan X matricen giver en sammenhæng i estimerer nedstrøms eller mellem oplande, der udveksler vand. Opland A er øverst opstrøms, hvorfra vandet løber til opland B og videre til opland C.

Da vand fra opland A løber til opland B, vil X matricens oplysning for opland A også indgå i beregningen af koncentrationsniveauet i opland B, hvilket automatisk vil give en vis korrelation mellem de to oplandes prædikterede koncentration. På samme måde vil der være en korrelation fra oplandene A og B til C. I det hypotetiske tilfælde hvor oplandsmodellen var helt ideel og derfor uden usikkerhed, ville X matricen således sikre en realistisk korrelation i koncentrationsniveauet mellem oplandene og sikre, at måleresultaterne i opland C beskrives med modellen uden afvigelser.

Reelt set udgør oplandsmodellen dog ikke en fuldkommen beskrivelse af de forventede koncentrationer. En del af forskellen mellem oplandene skal beskrives med den statistiske model som tilfældige effekter mellem stationerne, og i denne del af MetalStat tages der ikke hensyn til oplande, der udveksler vand med hinanden. Det betyder, at den viden man har i opland C ikke øger præcisionen af prædiktionen for opland A og B, selvom noget af det vand, der er målt på i opland C kommer fra oplandene A og B. Så en prædiktion baseret på estimerede parametre vil ikke fuldt ud tage hensyn til den forventede rumlige korrelation, der vil være mellem oplande, hvor vandet strømmer fra det ene opland til det næste. Modelteknisk svarer det til, at massebalancen ikke er helt afstemt mellem oplande, der udveksler vand. En løsning kunne være at lade koncentrationen i et opland være en funktion af alle oplande som løber ind i oplandet, men det vil udvide antallet af latente variable i log-likelihood funktioner dramatisk, hvorfor der vælges en tilnærmet løsning. Derfor er der i nedenstående afsnit udviklet en massebalancejustering, der tager hensyn til den korrelation, der vil være mellem nabo-oplande pga. vandudveksling.

Princip

I hvert opland prædikteres koncentrationerne ved en sandsynlighedsfordeling, så opgaven er at justere alle fordelingerne, så massebalancen overholdes mellem oplandene, uden at der justeres så kraftigt, at det medfører en underestimering af usikkerheden for det enkelte opland. Generelt er det bedre at overestimere usikkerheden frem for at underestimere den, da sidstnævnte tilfælde let giver falske konklusioner. For at sikre mod en overjustering vælges det derfor at massebalancejustere fraktilværdierne for alle oplande, som illustreret på Figur 3.4, koncentrationsniveauerne i de umålte oplande (A og B) bliver justeret med værdierne hhv. Δ_A og Δ_B i hhv. opland A og B, så massebalancen er afstemt med det målte koncentrationsniveau i opland C. Denne procedure beregnes for et tilstrækkeligt antal fraktiler for, at en ny sandsynlighedsfordeling for koncentrationer kan etableres i hhv. opland A og B.



Figur 3.4. Justering af oplande uden målinger (A og B) med samme fraktil som opland (C), hvor der er målt.

Numerisk metode

Der er udtrukket 26 prædiktioner med brug af Marco Chain Monte Carlo fra MetalStat for hvert opland, og disse 26 er rangordnet i forhold til hinanden for hvert opland. Massebalancejusteringen er forgået mellem koncentrationsniveauer for samme rang mellem oplandene. Således forudsættes det at den prædiktion for opland A, der har rang 5 blandt de 26 prædiktioner for A, at tilhøre den samme fraktil, som prædiktionen for opland B, der også har rang 5. I et gennemløb justeres alle 26 prædiktioner for alle oplande en enkelt gang. Og når justeringerne er tilendebragt, kan der konstrueres en ny koncentrationsfordeling baseret på de 26 justerede koncentrationsniveauer.

Følgende oplande vil ikke blive justeret: (1) Oplande, som er målt i perioden 2011-2019, (2) Oplande, der ikke modtager vand fra andre oplande, og som strømmer ind i et opland sammen med andre indstrømmende oplande; (3) Oplande, der ikke modtager vand fra andre oplande, og som har udløb ved kysten.

Følgende massebalance gælder for et vilkårligt opland A:

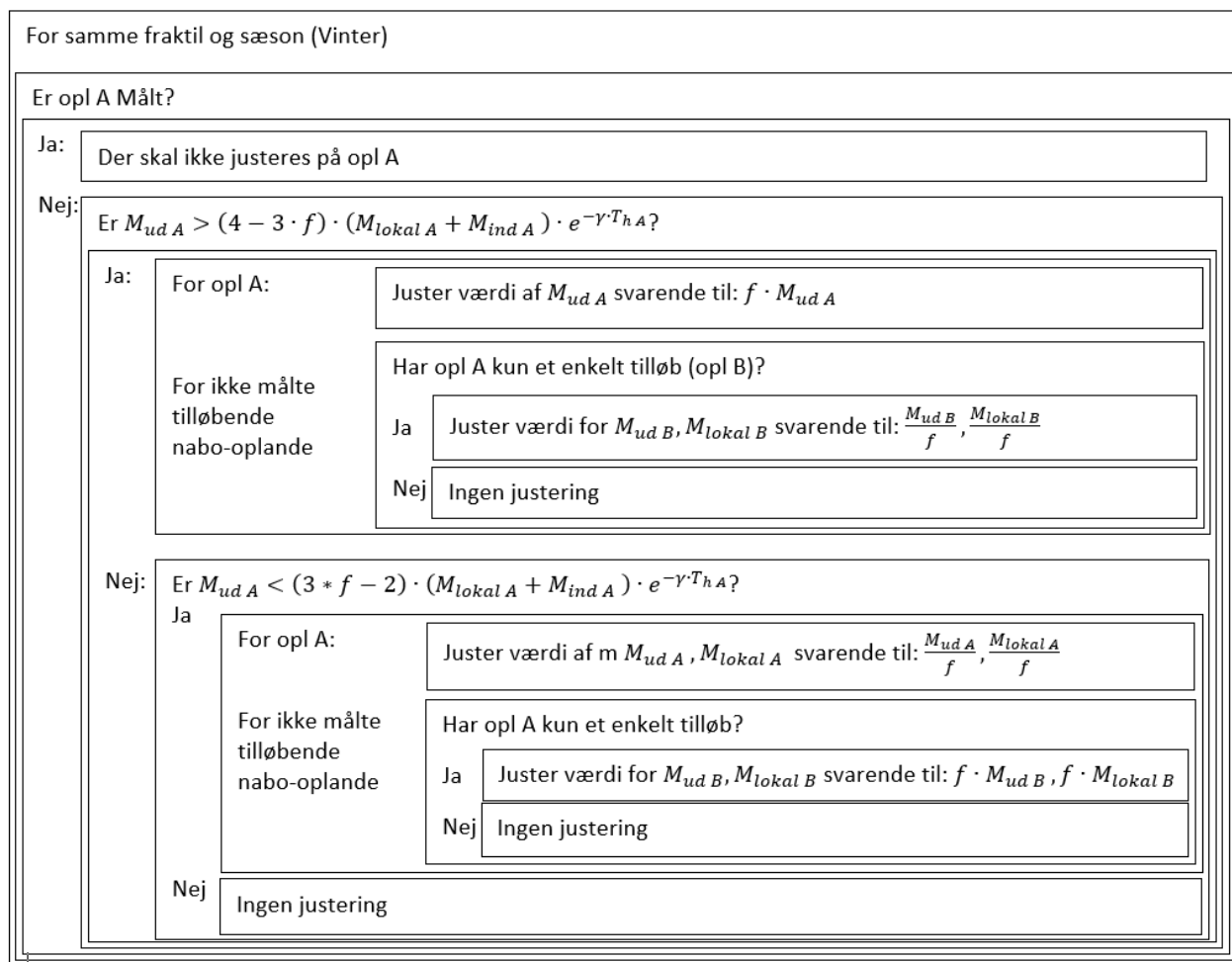
$$M_{ud\ A} = M_{ind\ A} + M_{lokal\ A} - M_{ret\ A} \quad (2)$$

Hvor $M_{ind\ A}$ er samlet massetilførsel til opland A fra opstrøms oplande, $M_{lokal\ A}$ er massebidrag fra lokale kilder i opland A, $M_{ret\ A}$ er massen der tilbageholdes i opland A (retention), $M_{ud\ A}$ er massen der løber ud af opland A. $M_{ind\ A}$ beregnes som sum af M_{ud} for opstrøms oplande. $M_{lokal\ A}$ beregnes som startværdi med brug af X matricen, som den masse af metal, der kommer fra kilder i opland A. $M_{ret\ A}$ bliver beregnet ud fra retentionen i oplandet (Sørensen et al., 2022).

I hvert gennemløb af algoritmen, som er illustreret i Figur 3.5 kan oplande, der udveksler vand blive justeret med faktoren f , hvor $f < 1$ i det omfang der er diskrepans i massestrømmen hhv. ind og ud af oplandene. For beregningerne vælges $f = 0,95$, hvilket betyder, at der trinvist justeres med 5 % af en masse hver gang, der identificeres en diskrepans mellem massestrømmene, jvf kriterierne i Figur 3.5. For at finde den bedst afstemte massebalance mellem alle oplande gentages algoritmen med mange iterationer for alle oplande og for hver fraktilværdi. Når masserne er blevet justeret, kan der beregnes nye, justerede koncentrationer baseret på de justerede fraktilværdier.

Hvis massebalancen (Ligning 2) viser, at der er en diskrepans mellem ind og udløb, som ligger uden for en tolerance fra 85 % til 115 % vil der blive udført en justering efter algoritmen som vist på Figur 3.5. Disse grænser er sat for at afspejle en rimelig overholdelse af massebalancen mellem oplande. Et mere snævert kriterium (f.eks. en tolerance fra 95% til 105%) vil ikke påvirke resultatet mærkbart.

Der justeres for samme sæson, hvor der her vælges vinter, der i modellen er valgt som referencepunkt (intercept), og justerede værdier for de andre sæsoner findes efterfølgende ved at lægge sæsoneffekten til de justerede værdier for vinter. Her udnyttes det, at modellen bruger logtransformerede koncentrationer, hvilket betyder, at alle disse effekter er relative. Hvis der f.eks. er en effekt af sæsonen "sommer" på -0,5, så betyder det, at sommerkoncentrationerne estimeres til at være en faktor $e^{-0,5} = 0,61$ mindre end vinterkoncentrationerne. Det betyder, at de justerede sommerkoncentrationer er 61 % mindre end de justerede vinterkoncentrationer, og derfor kan de justerede sommerkoncentrationer let beregnes ud fra de justerede vinterkoncentrationer.

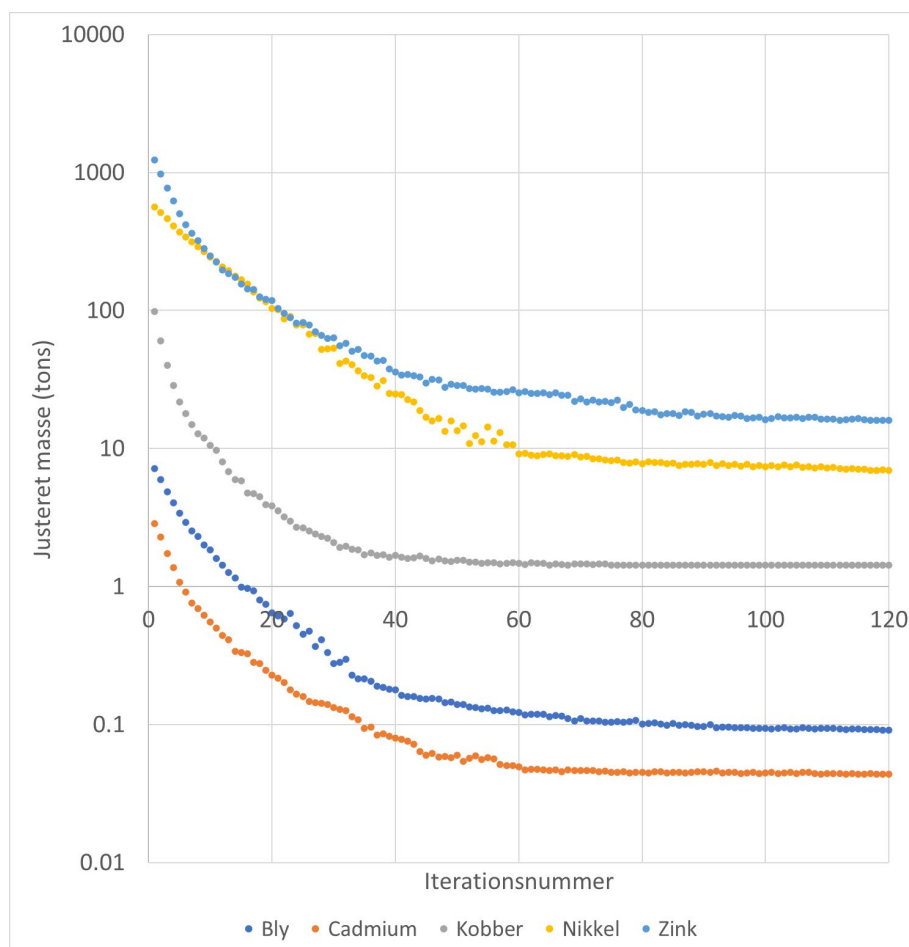


Figur 3.5. Algoritme for massebalancejustering mellem oplandet A og opstrøms nabo oplande. Denne algoritme skal gennemløbes mange gange for alle oplande for hver fraktilværdi.

På Figur 3.6 ses den samlede justerede masse for hvert metal, summeret over alle oplande og afbildet med iterations nummer angivet på x akse. De største justeringer ses for zink, der er justeret med lidt over 1000 tons ved første gennemløb, hvorefter justeringen falder, efterhånden som massebalancerne udlignes mod hinanden til et stabilt niveau på 16 tons pr iteration. Justeringerne konvergerer ikke helt mod 0, da der er oplande, som ikke justeres i

algoritmen, og som derfor ikke kan afstemmes. Dette giver svingende ”op og ned” justeringer, der ikke kan udjævnes yderligere med algoritmen. For alle fem metaller ses en justering der falder drastisk gennem de første ca. 50 iterationer for derefter at flade ud mod en konstant justering, der er ca. 100 gange mindre end startværdierne.

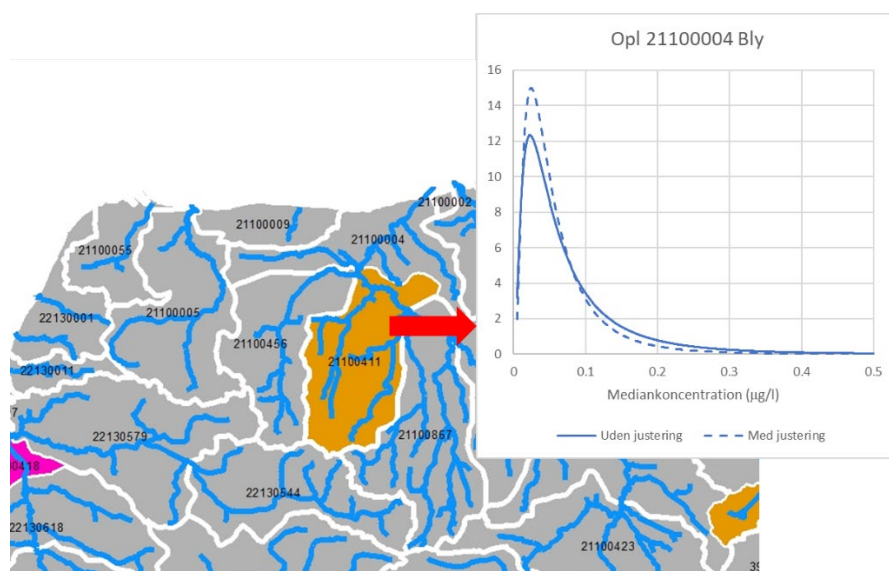
Figur 3.6. Samlet justeret masse for hver iteration, hvor justeringsalgoritmen gennemløbes.



Illustrative eksempler

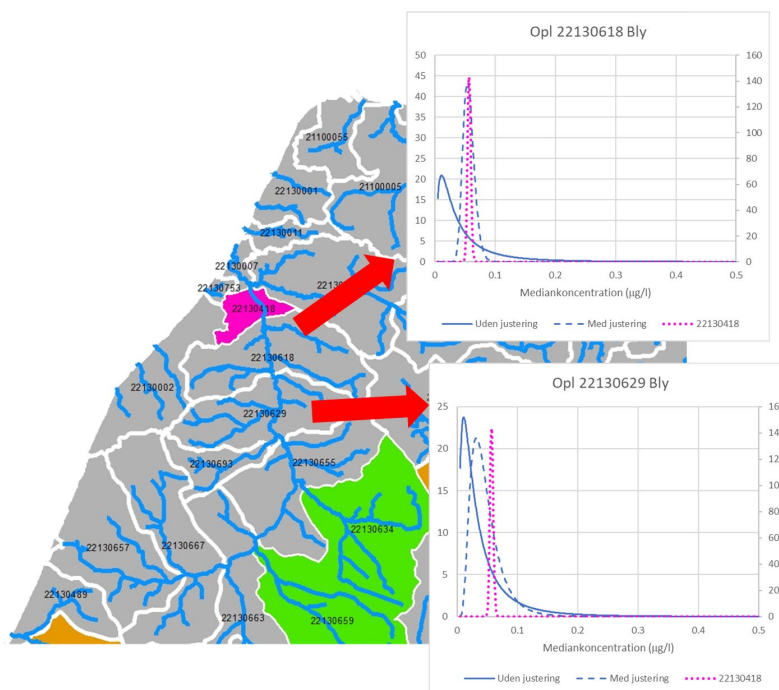
Massebalancejusteringen skaber en bedre overensstemmelse for massetransporten mellem oplandene og vil derfor påvirke prædiktionen af koncentrationer for alle de oplande, der indgår i justeringen. På Figur 3.7 ses resultatet for bly fra Uggerby Ås nederste opland i Vendsyssel, der er et vandløbsopland uden målinger fra perioden 2011-2019. Koncentrationen er derfor prædikeret i dette opland helt uden brug af lokale målinger. I denne prædiktion indgår dels X matricen, hvor alle oplande er korrelerede, og dels tilfældige effekter for oplandene. Da de tilfældige effekter er regnet uafhængig af hinanden, betyder det at massebalancer mellem oplandene ikke er helt afstemte, hvilket igen betyder at massebalancejusteringen vil påvirke prædiktion af mediankoncentration for oplandene, hvor de enkelte oplande er regnet uafhængig af hinanden. Da de tilfældige effekter er regnet uafhængigt af hinanden, betyder det, at massebalancer mellem oplandene ikke er helt afstemte, hvilket igen betyder, at massebalancejusteringen vil påvirke prædiktion af koncentrationen.

Figur 3.7. Massebalance justering for prædiktation af mediankoncentration i nederste ID15 opland i Uggerby Å.



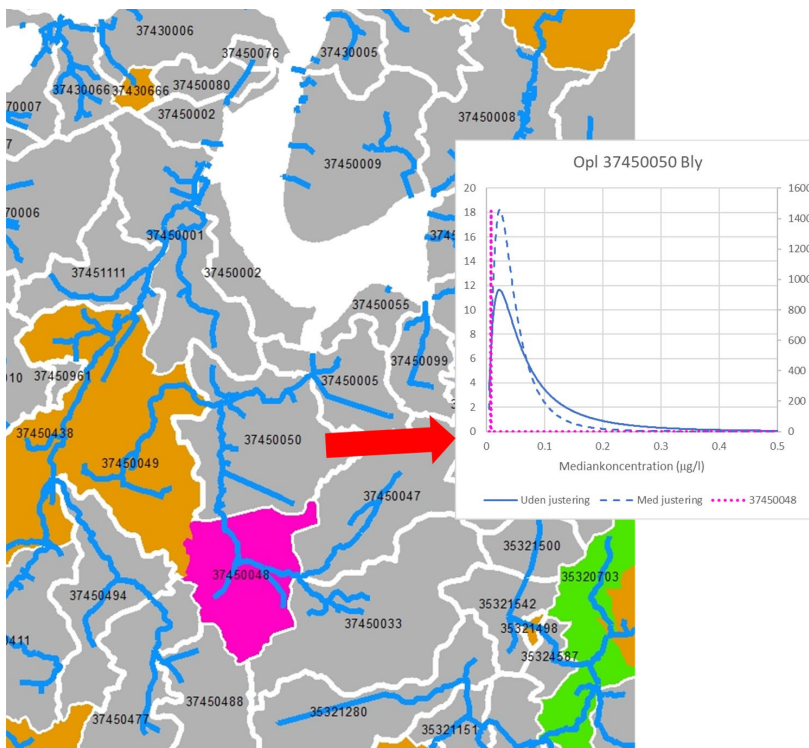
Massebalancejusteringen bliver især betydningsfuld, når der er målte ID15 oplande i et vandløbsopland, hvilket illustreres nedenfor på Figur 3.8, hvor en justering for bly af to ID15 oplande vises fra Liver Å i Vendsyssel. Der er målt i perioden 2011-2019 i opland ID15 22130418 nær udløbet af Liver Å. Det giver en prædiktation af koncentrationen med relativt lille spredning, som vist på Figur 3.8 med en lilla stiptet kurve. Opland ID15 22130618 har ingen målinger, men det ligger lige opstrøms opland ID15 22130418; her tager prædiktationen før massebalancejustering ikke hensyn til, at oplandet lige nedstrøms er målt, hvilket vises ved en fuldt optrukket kurve med stor spredning. En massebalancejustering af ID15 22130618, der tager hensyn til massestrømmen af bly ud af ID15 22130418, er vist som den stiplede blå kurve med meget mindre spredning. Da det er Liver Ås nedre løb, er der en stor gennemstrømning af vand igennem de nederste ID15 oplande, og derfor bliver massen af bly, der er målt for opland ID15 22130418, meget bestemmende for den masse bly, der kan løbe ud af det umålte opland lige opstrøms (ID15 22130618). Det betyder en lang mindre usikkerhed og dermed spredning, for den massebalancejusterede tæthedsfunktion for ID15 22130618, sammenlignet med tæthedsfunktionen før massebalancejusteringen. En sådan tæt sammenhæng ses derfor mellem den målte koncentration i ID15 22130418 og den justerede værdi af den prædikterede koncentration for opland ID15 22130618. Oplandet opstrøms opland ID15 22130618 (ID15 22130629) bliver også påvirket af målinger i ID15 22130418, fordi det påvirker oplandet ID15 22130618, som igen påvirker massebalancen for ID15 22130629. Men massebalancejusteringen bliver ikke helt så voldsom, da der er et umålt opland mellem ID15 22130629 og ID15 22130418, hvilket giver en større usikkerhed og heraf en tæthedsfunktion med større spredning end det er tilfældet for ID15 22130629. Desuden vil ID15 22130629 også i stigende grad blive massebalancejusteret efter de to målte oplande ID15 22130634 og ID15 22130659, der ligger opstrøms, og dermed blive justeret m.h.t. flere målte ID15 oplande i Live Ås opland.

Figur 3.8. Massebalancejustering af oplandene ID15 22130618 og 22130629, der er en del af oplandet i Liver Å. Den prædikterede mediankoncentration for nedstrøms opland (ID15 22130418) er vist som en lilla fordelingskurve.



Massebalancejusteringen betyder også, at oplande der er placeret nedstrøms de målte oplande, kan blive justeret, hvilket ses på Figur 3.9 for Fiskbæk Å ved Viborg. Her bliver ID15 37450050 justeret, så massebalancen passer med det målte opstrøms opland (ID15 37450048).

Figur 3.9. Massebalancejustering af ID15 37450050, der er en del af oplandet i Fiskbæk Å. Den prædikterede mediankoncentration for opstrøms opland (ID15 37450048) er vist som en lilla fordelingskurve.



4 Validering

MetalStat har tre hovedformål: (1) Alle måledata samles i en landsdækkende statistisk beskrivelse, der kan bruges til at prædiktere en statistisk tæthedsfordeling, både i de oplande, der er målt samt i de umålte. (2) Oplysninger fra oplandene samles i en X matrice med det formål at prædiktere metalkoncentrationerne i oplandene ud fra de mulige kilder i oplandet og i de opstrøms oplande, hvis disse kilder findes. (3) Årlige og årstidsafhængige forskelle i koncentrationsniveauer identificeres og bruges til prædiktioner.

Det er en fordel at have så meget målt variation som muligt beskrevet i punkterne 2 og 3, da det vil minimere prædiktioners usikkerheden. I det følgende bliver MetalStats evne til at prædiktere koncentrationer jf. punkterne 2 og 3 testet med målinger fra perioden 2011-2019, ved at udnytte at MetalStat både har et estimat og en prædiktion for hvert af de målte oplande. Denne metode illustreres i Figur 3.2, hvor de estimerede latente variable sammenholdes med prædiktioner af for de målte oplande. Derefter i kapitel 4.2. testes modellens evne til at identificere landsdækkende statistiske tæthedsfordelinger af koncentrationer i ID15 oplande (punkt 1) generelt med brug af målinger fra perioden 2020-2022, som ikke indgik i kalibreringen af MetalStat. Nogle af målingerne fra perioden 2020-2022 er foretaget på prøver fra ID15 oplande, hvorfra der også var indsamlet prøver i perioden 2011-2019, mens andre prøver er indsamlet i ID15 oplande fra hvilke der ikke er indsamlet prøver i 2011-2019.

Konkret vil MetalStat således blive valideret på to niveauer: (1) X matrixens evne til at forklare forskelle i koncentrationsniveauer testes ved at finde den forklarede varians, som er udtrykt ved den log-lineære regressionskoefficient (R^2). (2) Strukturmodellens evne til at udtrykke koncentrationsniveauer som tæthedsfordelinger testes ved visuelt at sammenligne estimerede tæthedsfordelinger med nye målinger, som ikke har indgået i kalibreringen. Der er ikke nogen specifikke krav til at R^2 værdien skal være over en bestemt værdi før modellen er tilstrækkelig sikker. Det er dog en fordel at R^2 værdien er så høj som muligt, og en lav værdi tæt på nul indikerer, at X matrixen bør granskes for forbedringer. Strukturmodellens evne til at udtrykke tæthedsfunktionerne skal være udtalt, forstået på den måde at midtpunktet af det interval som de nye målte koncentrationer skaber, kun sjældent bør ligge udenfor hvad tæthedsfunktionerne prædikterer som værende realistisk.

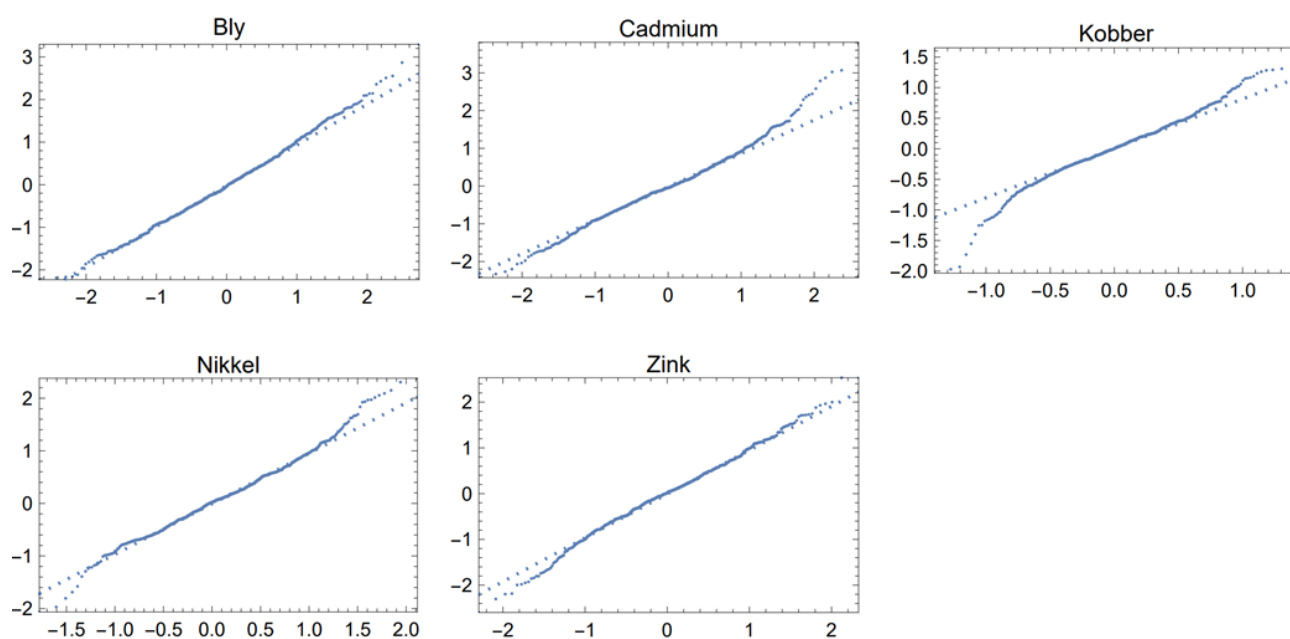
Da MetalStat anvender logaritmen til koncentrationsniveauerne (Sørensen et al, 2022) vil den forventede værdi for et opland være identisk med logaritmen til mediankoncentrationen, og det er derfor naturligt at validere MetalStat mod målte mediankoncentrationer. I de tilfælde hvor Miljøstyrelsen vælger at bruge mediankoncentrationer, vil det således svare til resultaterne fra MetalStat. Det er dog også muligt at beregne aritmetrisk gennemsnitskoncentration, hvis det viser sig ønskeligt, så MetalStat binder ikke Miljøstyrelsen til at bruge mediankoncentrationer som tilstandsvurdering.

4.1 På baggrund af målinger fra 2011-2019

Residualplots

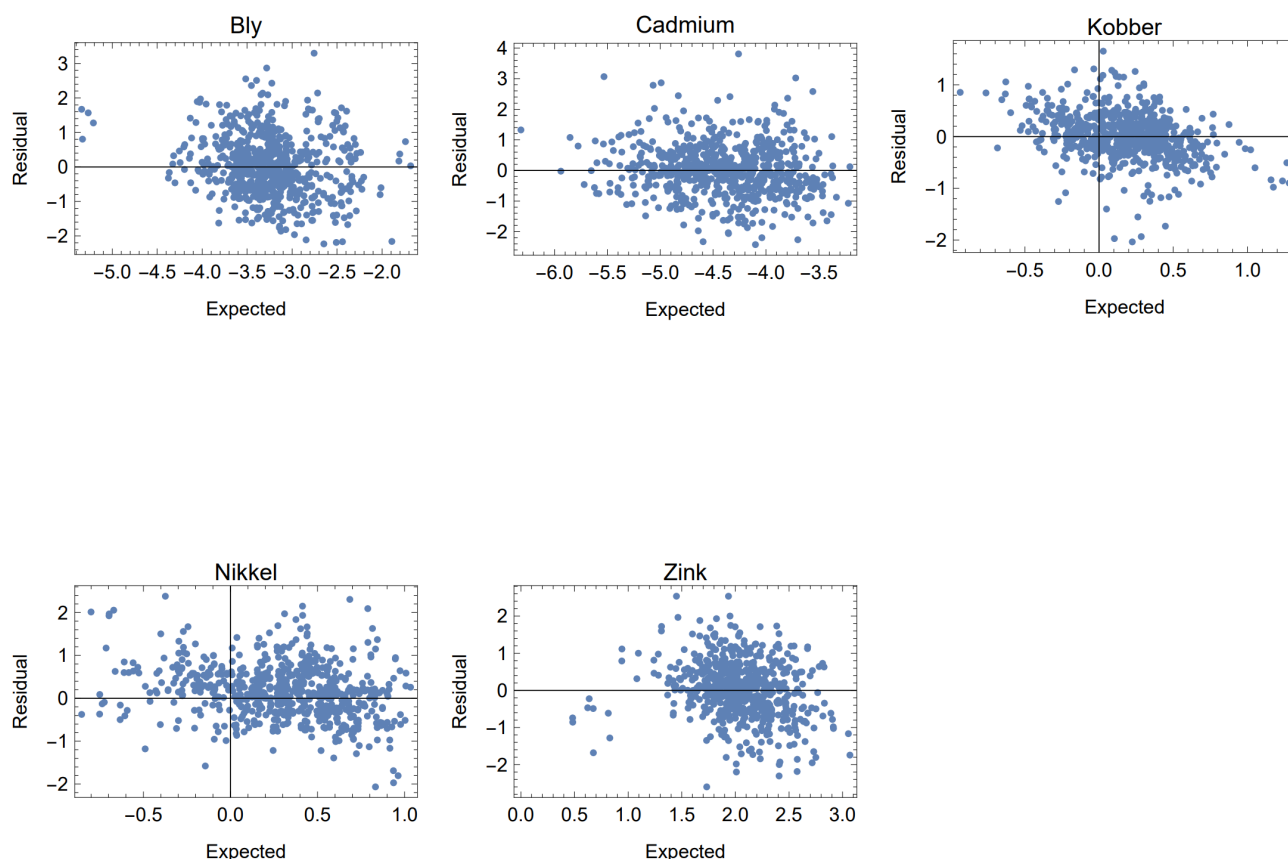
I de oplande, hvor der er målte værdier fra perioden 2011-2019, giver MetalStat både en prædiktion og et estimat af logaritmen til mediankoncentrationen. Prædiktionen foretages uden hensyntagen til at der er målt netop i dette opland. Estimatet derimod (af den latente variable i målemodellen) tager hensyn til, at der er målt i oplandet, og er derfor det mindst usikre, se Figur 3.2. Ved at sammenligne residualerne mellem disse to beregninger (latent variable fratrukket prædiktion) er det muligt at undersøge MetalStat, i forhold til om modellen kan prædiktere forskelle mellem oplande.

MetalStat forudsætter normalfordelte residualer mellem måling og prædiktion (Sørensen et al., 2022). Derfor er det oplagt først at undersøge om denne forudsætning om normalfordelte residualer holder. På Figur 4.1 vises et Q-Q plot for residualerne, der har den egenskab, at der fremkommer en lineær relation, hvis residualerne er normalfordelte. På et Q-Q plot afbildes fraktilerne for data (her residualer) på den ene akse, og tilsvarende fraktiler for en normalfordeling på den anden akse. Hvis der er en lineær relation mellem disse to måder at beregne fraktilerne på, så er data normalfordelte. Da der ses en klar lineær relation for alle fem metaller, betyder det altså, at forudsætningen om normalfordelte residualer vurderes som værende tilstrækkeligt opfyldt.



Figur 4.1. Q-Q plot af residualerne mellem prædikeret og målt værdi, for de stationer, der er målte.

Residualerne er vist på Figur 4.2 som funktion af den forventede værdi. For alle fem metaller ses en ret homogen varians af residualerne. Dog udviser kobber og zink en tendens til faldende værdi med den forventede værdi. I det omfang residualerne har en tendens til at stige eller falde i værdi, peger det på, at oplandsmodellen mangler en faktor, som har betydning for metalkoncentrationen. Specifikt for zink er der sket store ændringer i koncentrationsniveau i løbet af måleprogrammet som ikke kan beskrives som en tilfældig effekt mellem årene. Det ses også af Figur 2.2, at der er sket store forskydninger i koncentrationsniveau for zink mellem perioden 2011-2019 og 2020-2022, så MetalStat bør modificeres i forhold til zink med to værdier en historisk værdi og en nyere værdi, for at dække disse forskydninger.



Figur 4.2. Residual plots af middelværdien af den forventede koncentration per målt opland givet sæson og opland og middelværdien af de tilhørende estimerede latente variable (begge på en log skala).

Forklaret varians

Oplandsmodellen er den del af MetalStat, der tilstræber at forklare den latente koncentration ud fra forholdene i oplandet (X matricen), samt landsdækkende sæsonforskelle. Det er derfor ønskeligt at få mest mulig variation beskrevet med oplandsmodellen, da det vil styrke MetalStats evne til at forudsige koncentrationerne under hensyntagen til lokale forhold i oplandet. I den nedenstående analyse indgår log værdier af koncentrationsniveauerne, så den kan tolkes som en analyse for relative bidrag. Det betyder, at oplandsmodellen påvirker R^2 , uafhængigt af, om den latente koncentration er lille eller stor. Som illustration: Hvis koncentrationen i et opland estimeres af oplandsmodellen til $1 \mu\text{g/l}$, hvor den latente koncentration i dette opland estimeres til $2 \mu\text{g/l}$, så er oplandsmodellen 50 % for lille, denne forskel påvirker R^2 med samme styrke, som hvis oplandsmodellen for et andet opland er estimeret til $2 \mu\text{g/l}$, og hvor den latente variabel i dette andet opland estimeres til $4 \mu\text{g/l}$, da denne afvigelse også er 50 %.

Andelen af variansen i metalkoncentrationerne, som forklares af oplandsmodellen på en logskala, beregnes ved

$$R^2 = \frac{var_{fit}}{var_{fit} + var_{random} + var_{res}}$$

hvor var_{fit} er variansen af de forventede koncentrationer i de målte oplande ved brug af oplandsmodellen, var_{random} er middelværdien af den estimerede tilfældige variation mellem oplande, og var_{res} er middelværdien af den estimerede residuelle variation (Gelman et al. 2019). Resultatet ses i Tabel 4.1.

Tabel 4.1. Andelen af variansen (R^2) i metalkoncentrationerne (log værdier) som blev forklaret af oplandsmodellen på en logskala.

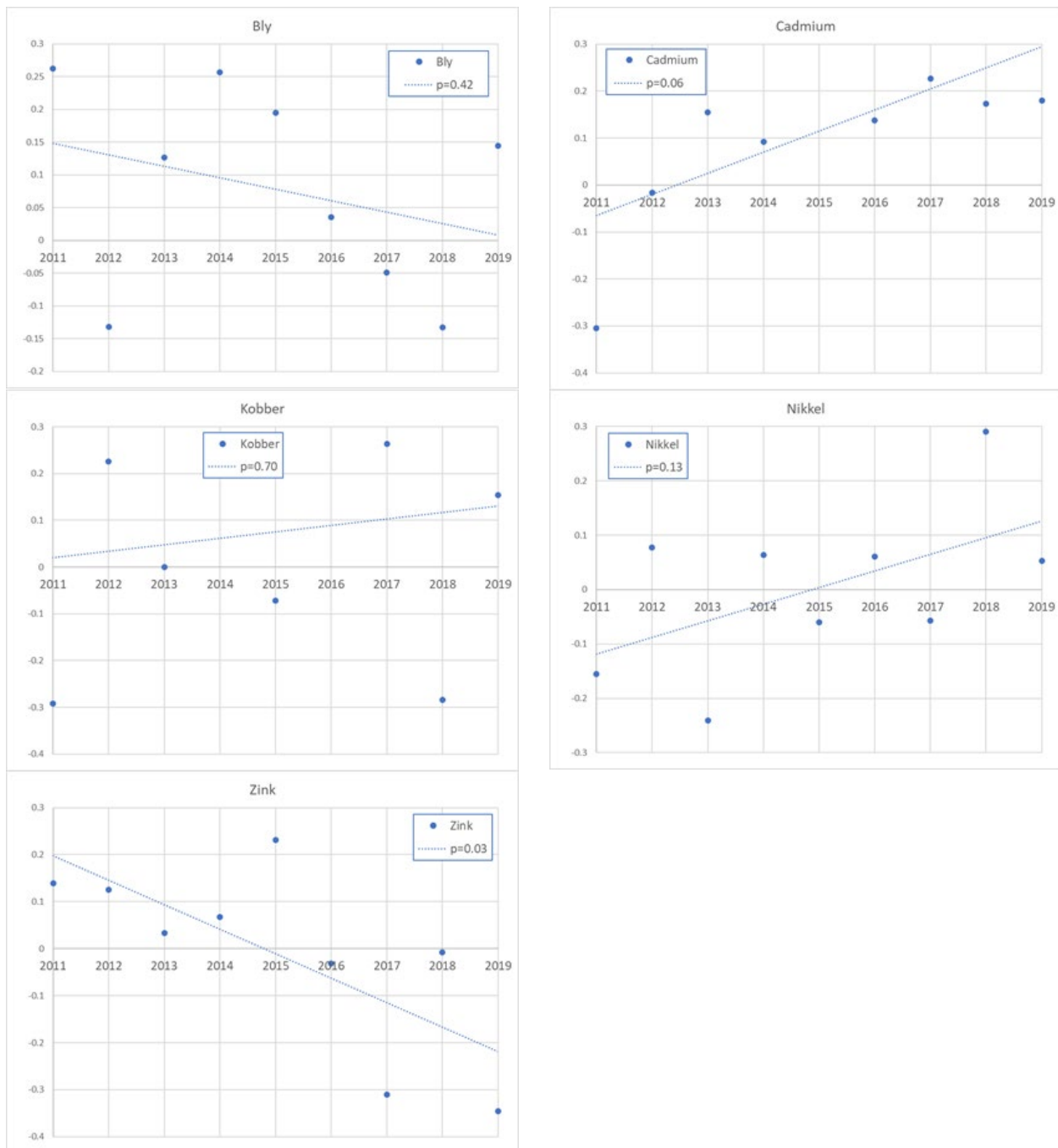
Bly	Cadmium	Kobber	Nikkel	Zink
0.19	0.22	0.30	0.06	0.19

Andelen af variansen, som forklares af oplandsmodellen, er relativt beskeden. Den højeste værdi er 30 % for kobber, og den mindste værdi er 6 % for nikkel. Det vurderes, at den mest sandsynlige grund til at R^2 er relativt lille, er, at der mangler en bedre beskrivelse af grundvandsbidraget. Senere i denne rapport fremsættes forslag til at forbedre især beskrivelsen af grundvandsbidraget i MetalStat, som en del af den driftsversion, der udvikles i 2023.

4.2 På baggrund af målinger fra 2020-2022

I MetalStat modelleres årlige landsdækkende variationer som tilfældige effekter, der forudsættes normalfordelt for log-koncentrationen. Det betyder, at hvis koncentrationsniveauet et år er reduceret/øget med en vis procentdel i forhold til de andre år, så beskrives det i modellen som en tilfældig effekt på en additiv log-skala. F.eks. hvis et metal har en tilfældig effekt på 0,2 et bestemt år, så betyder det, at koncentrationsniveauet i alle ID15 oplande har en tendens til at være $\exp(0,2)=1,22$ gange større dette specifikke år sammenlignet med medianåret for samme opland. Så hvis et ID15 opland har en median årlig forventning på 2 $\mu\text{g/l}$, så er estimeret for et år med random-effekt på 0,2: $2 \cdot 1,22 = 2,44$ $\mu\text{g/l}$, mens et andet opland der har en årlig median på 0,5 $\mu\text{g/l}$, estimeres til at have en koncentration på $0,5 \cdot 1,22 = 0,61$ $\mu\text{g/l}$. De tilfældige effekter for hvert metal er vist for hvert år på Figur 2.5, hvor der ses en typisk variation i intervallet fra -0,4 til 0,3, svarende til, at koncentrationsniveauet gennem perioden mellem årene varierer mellem 0,67-1,35 gange årlig mediankoncentration. Hvis de nye målinger for perioden 2020-2022 generelt ligger inden for intervallet beskrevet med tilfældige effekter, så vil disse forskelle kunne prædikteres af MetalStat, men hvis der sker store forskydninger, så de nye målinger falder helt ud af intervallet for de tilfældige effekter for året, så vil MetalStat ikke kunne prædiktere denne effekt.

De tilfældige effekter pr år i perioden 2011-2019 er brugbare til at estimere en landsdækkende tidstrend. I Figur 4.3, ses det netop at zink udviser et signifikant fald i koncentrationsniveauet med årene. De målte zinkkoncentrationer for perioden 2020-2022 synes dog at være på et helt andet og lavere niveau end det var tilfældet for perioden 2011-2019, og det kan derfor forventes, at MetalStat generelt vil overestimere de målte zink koncentrationer, hvilket diskuteres yderligere i senere afsnit.



Figur 4.3. Tilfældige effekter for år, vist for hvert metal, hvor p værdien er signifikansniveauet for den estimerede fortegn af linjes hældningen (stigende eller faldende). Det ses, at zink isoleret set har et signifikant ($p=0,03$) fald.

Estimeret fordeling og målte koncentrationsniveauer for oplande målt i perioden 2020-2022

Grafisk sammenligning af målinger af MetalStat

For hvert metal er der ca. 400 ID15 oplande med måledata fra perioden 2020-2022, og for de fem metaller giver det i alt ca. 2000 grafer, hvor MetalStat beregner en sandsynlighedsfordeling for mediankoncentrationen, der kan sammenlignes med målinger, se bilag 1. Før sammenligningen med målinger er der anvendt en massebalancejustering, som forklaret i kapitel 3. I bilag 1 vises oplande med målte koncentrationer for de fem metaller i perioden 2020-2022 samt de beregnede tæthedsfordelinger af mediankoncentrationer ved brug af målinger fra perioden 2011-2019. Hver graf er angivet med "metal:oplandsnummer" som titel. Farven af titlen følger farvekoden fra Figur 2.3, hvor et

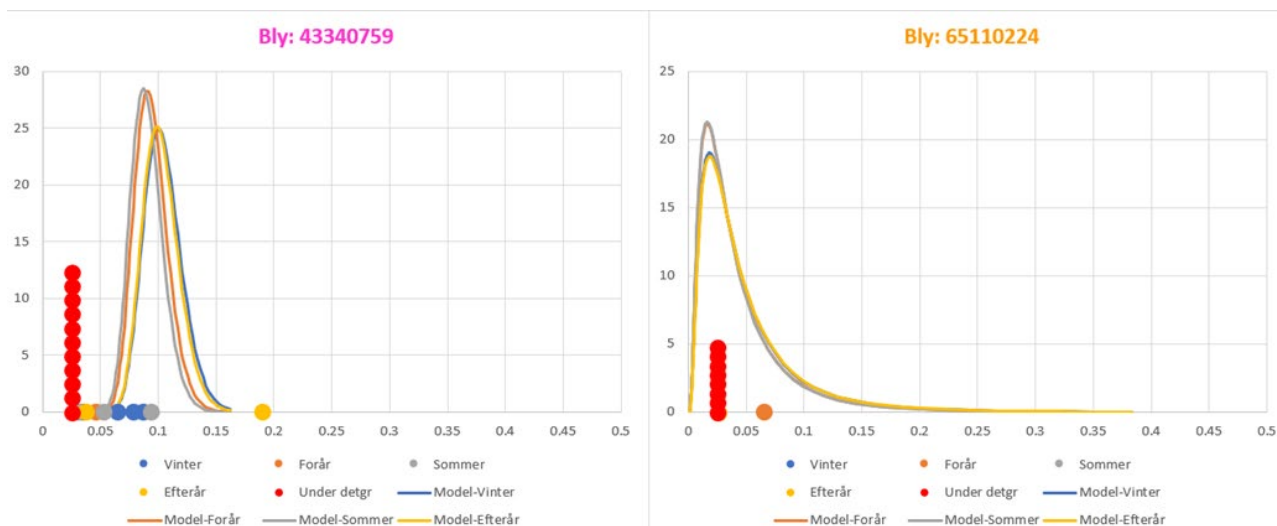
opland, der er målt i hele perioden 2011-2022, har farven lilla, mens oplande, der kun er målt i perioden 2020-2022, har farven gul. For de oplande hvor der er målt i perioden 2011-2019, svarer den beregnede tæthedsfordeling til den estimerede latente koncentration tilføjet tilfældige effekter for variablene år, og sæson. For de øvrige oplande er tæthedsfunktionen beregnet ved prædiktion af ikke målte oplande.

De målte værdier er afsat på x-aksen med den værdi, hvor der er brugt en farvekode for at adskille målinger fra forskellige sæsoner. Målinger under detektionsgrænsen er markeret med rødt og placeret svarende til detektionsgrænsen på x-aksen, og hvis der er flere målte værdier under detektionsgrænsen, er de stablet ovenpå hinanden, så hver måling kan ses. Der er beregnet en tæthedsfordeling for hver sæson, og hver af dem er vist som en kurve med samme farvekode som den sæson, der er brugt til visning af målingerne på x-aksen. Tæthedsfunktionerne viser således mediankoncentrationen, svarende til den fuldt optrukne kurve på Figur 4.4, hvor spredningen på kurven afspejler usikkerheden i beregning af mediankoncentrationen. Hvis der er overensstemmelse mellem MetalStat og de viste målinger på x-aksen, så vil målingerne fordele sig under kurverne med størst tæthed, der hvor kurverne er højest (størst sandsynlighedstæthed). Da kurven viser en beregnet tæthedsfordeling af mediankoncentrationen, er det især målepunkternes intervalmidte (median), som skal sammenlignes med tæthedsfunktionernes dækningsgrad. Generelt er der ikke mange af sådanne afvigelse. De afvigelser, der er fundet omhandler primært bly og vil blive diskuteret nedenfor i et specifikt afsnit for bly. Desuden er der en generel tendens til at målte zinkkoncentrationer er i den lave ende af de beregnede intervaller, eller i nogle få tilfælde ligger helt uden for kurven og under det beregnede interval, hvilket diskuteres nedenfor i afsnittet om zink.

Der ses en generel tendens til at de oplande, der ikke har målinger for perioden 2011-2019 har større usikkerhed forbundet ved prædiktion af mediankoncentrationen, hvilket giver mere højreskæve tæthedsfunktioner. Dette er en naturlig konsekvens af, at der er mere viden med i MetalStat til bestemmelse af mediankoncentrationen for de oplande, hvorfra der er målinger for perioden 2011-2019.

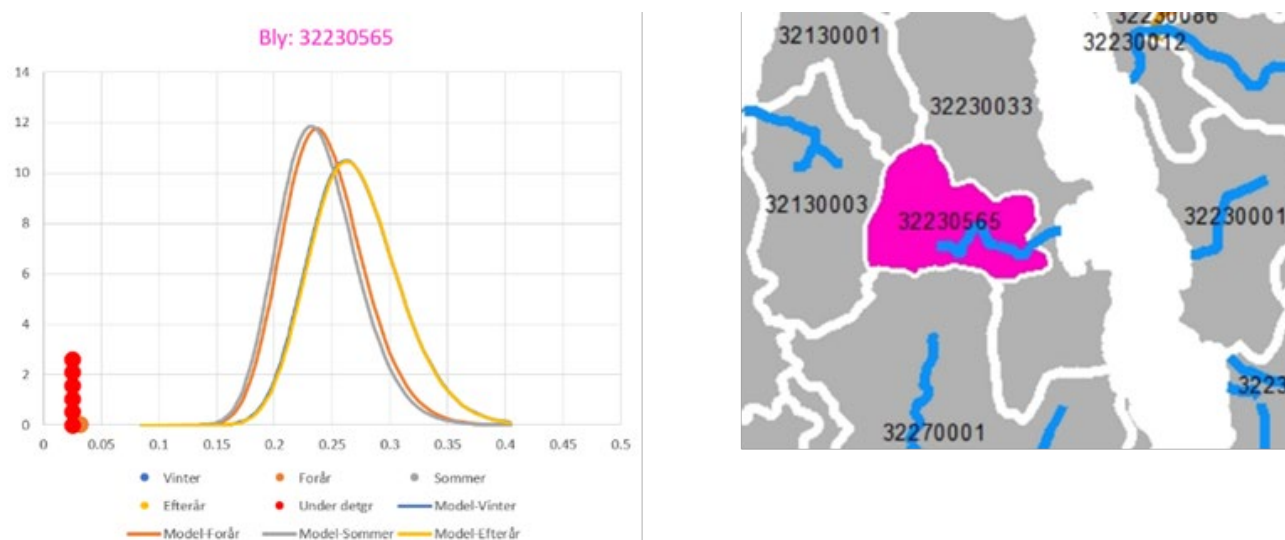
Bly

Generelt ses det, at bly har en god sammenhæng mellem de estimerede/prædikterede tæthedsfunktioner og de målte værdier. Sæsonforskellen er ikke særlig udtalt, da tæthedsfunktionerne for de forskellige sæsoner overlapper hinanden. På Figur 4.4 ses to tæthedsfunktioner for et opland, der hhv. er målt i perioden 2011-2019 (venstre) og ikke målt (højre). Kurven til venstre er for et opland, som har målinger i perioden 2011-2019 som derfor forventeligt fremstår med mindre usikkerhedsinterval sammenlignet med kurven til højre, hvor oplandet ikke er målt i perioden 2011-2019. Den venstre kurve på Figur 4.4 viser, en tendens for ID15 43340759, at den prædikterede mediankoncentration ligger i et interval over de målte værdier, men da der er et tydeligt overlap mellem målte værdier og prædikteret interval, vurderes denne uoverensstemmelse som værende af mindre betydende karakter.



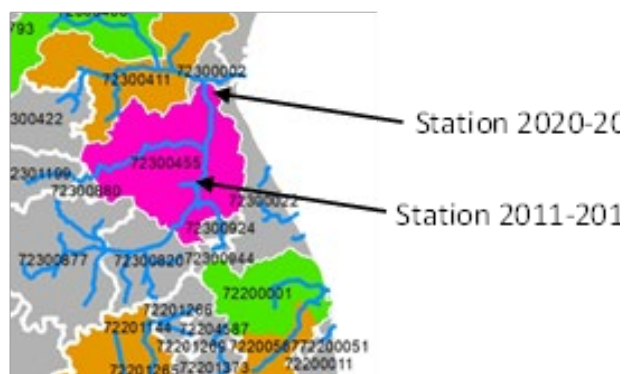
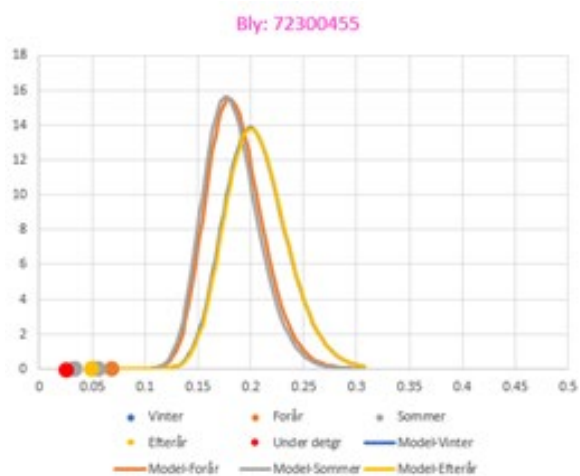
Figur 4.4. Tæthedsfunktion for bly i et opland, der er målt i perioden 2011-2019 (ID15 43340759), og et opland, der ikke målt (ID15 65110224). Enhed på x-aksen er opløst vandkoncentration i $\mu\text{g/l}$.

Den mest tydelige afvigelse blandt de 400 grafer for bly, ses ved fire grafer, hvor tæthedsfunktionerne i alle fire tilfælde ligger betydeligt over de målte koncentrationsværdier, samtidig med at det drejer sig om oplande, der også har målinger fra perioden 2011-2019, og hvor MetalStat derfor bør have relativt sikre estimater. Disse afvigelser vil typisk skyldes særlige lokale forhold i de pågældende oplande. På Figur 4.5 vises bækken Mademose Å syd for Tørslev med udløb i Roskilde Fjord, der løber i åbent land. Det er ikke umiddelbart muligt at adressere årsagen til, at netop denne bæk har en afvigelse i koncentrationsniveau mellem de to måleperioder.



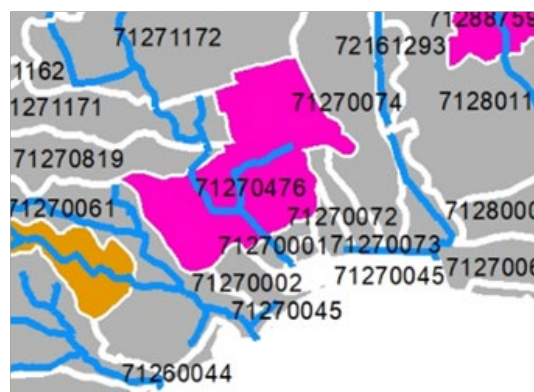
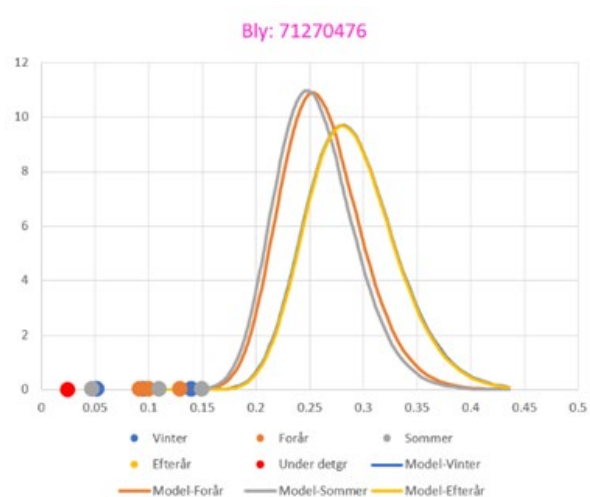
Figur 4.5. Graf for oplandet ID15 32230565 (Mademose Å), der er en lille bæk syd for Tørslev med udløb i Roskilde Fjord, hvor MetalStat præsenterer tæthedsfunktioner, som viser et koncentrationsniveau, der er noget større end det målte. Enhed på x-aksen er opløst vandkoncentration i $\mu\text{g/l}$.

Et andet af de fire afvigende oplande (ID15 72300455) vises på Figur 4.6. Det er et opland, der gennemstrømmes af Usserød Å, der løber til Nivå, som igen løber ud i Øresund nord for København. NOVANA Målestationen er rykket nedstrøms i Usserød Å mellem de to måleperioder, fra at være mellem Stampedammen og Frabrikdammen i perioden 2011-19 til at være ca. 5 km længere nedstrøms ved Nivå Mølle i perioden 2020-22. Især i byområder kan der være store skift langs et vandløb inden for samme ID15 opland, så oplandsstørrelsen kan blive udfordret for nogle ID15 oplande i byområder.



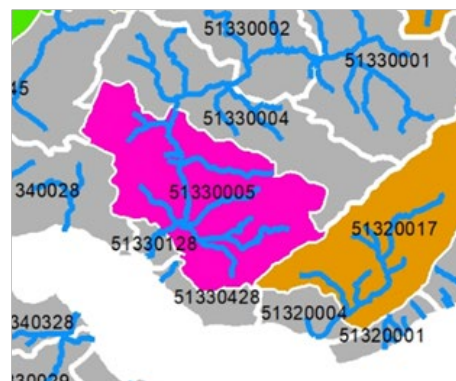
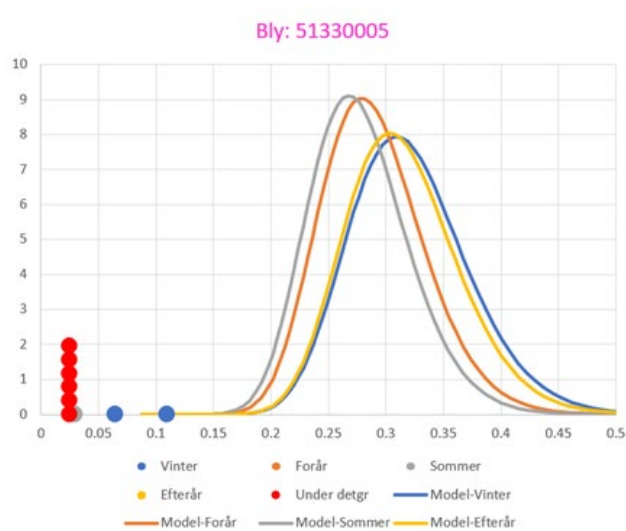
Figur 4.6. Graf for oplandet ID15 372300455 (Usseørd Å), der er udløb til Nivå, hvor MetalStat beregner tæthedsfunktioner, som viser et koncentrationsniveau der er noget større end det målte. Enhed på x akse er opløst vandkoncentration i $\mu\text{g/l}$.

Det tredje tilfælde hvor MetalStat over prædikerer mediankoncentrationen er i ID15 71270476, der er en mindre å i det sydlige København (Store Vejle Å), som løber ud i Brøndby Strand, se Figur 4.7. På samme måde, som for ID15 372300455 (Figur 4.6), så er ID15 71270476 også karakteriseret ved at bestå af et heterogent byområde.



Figur 4.7. Graf for oplandet ID15 71270486 (Store Vejle Å), med udløb ved Brøndby Strand, hvor MetalStat beregner tæthedsfunktioner, som viser et koncentrationsniveau, der er noget større end det målte. Enhed på x akse er opløst vandkoncentration i $\mu\text{g/l}$.

Det sidste af de fire udvalgte oplande for bly er ID15 51330005 (Rohden Å), der løber ud i Vejle Fjord, se Figur 4.8. Denne å er middelstor og løber primært gennem åbent land, så afvigelsen mellem MetalStat og målingerne må skyldes en lokal faktor, der ikke er umiddelbart identificerbar, og derfor kunne lægge op til en nærmere granskning af oplandet.



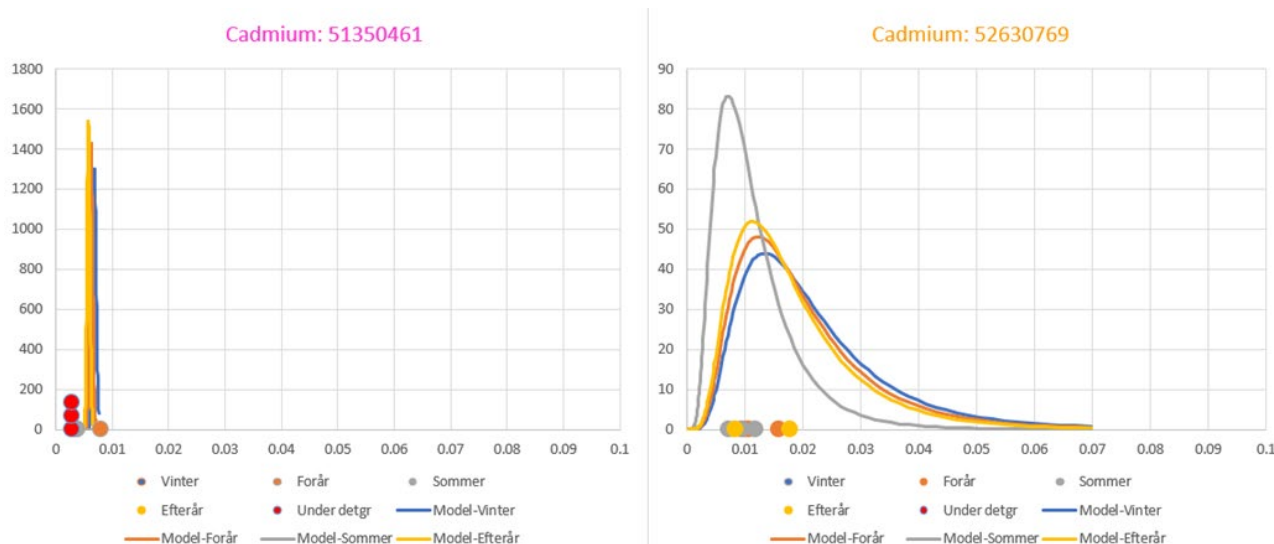
Figur 4.8. Sammenligning for opland ID15 51330005 (Rodan Å) med udløb til Vejle Fjord, hvor MetalStat beregner tæthedsfunktioner, som viser et koncentrationsniveau der er noget større end det målte. Enhed på x akser er opløst vandkoncentration i $\mu\text{g/l}$.

En dansk å vil typisk have en strømningshastighed mellem 0,2-0,4 m/s (Ovesen et al., 2000, Bilag 2), der dog kan variere meget. Ud fra en overordnet størrelsesorden vil vandet i et vandløb derfor flytte sig med ca. 1 km per time. ID15 oplandene har typisk en størrelse på 10-15 km^2 , hvilket betyder, at et vandløb gennemstrømmer et opland over en strækning på relativt få km, så hvis der ikke er opstemninger og søer i oplandet, vil den hydraulisk opholdstid i et opland typisk være nogle få timer. Set i det perspektiv er ID15 oplandene en relativt finkornet opdeling af landet, men det er klart, at et opland i nogle tilfælde kan være heterogen i et kritisk omfang, som det synes at være tilfældet i de viste oplande for bly.

Cadmium

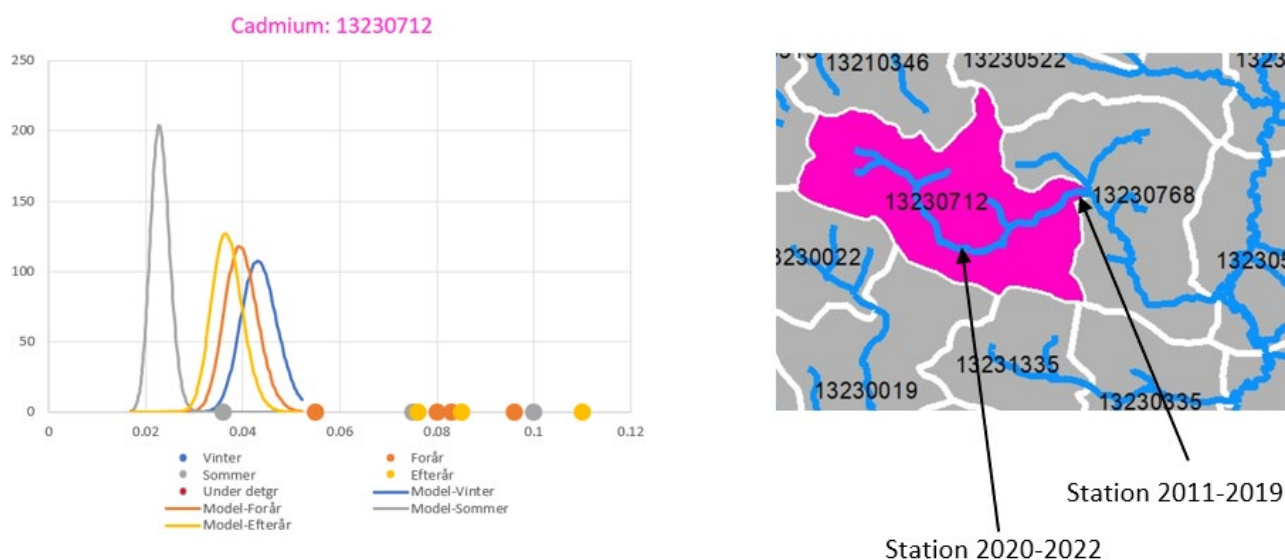
For cadmium beregner MetalStat koncentrationsniveauet om sommeren til at være lavere end resten af året med vinterkoncentrationen som den højeste. Generelt ses en god sammenhæng mellem tæthedsfunktionerne og de målte værdier. Der synes dog at være en tendens til, at variationsbredden beskrevet af tæthedsfunktionerne er større end variationen i de målte værdier.

På Figur 4.9 ses to kurver for cadmium og for et opland, der hhv. er målt i perioden 2011-2019 (venstre) og ikke målt (højre). Kurven til venstre fremstår på samme måde som for bly med et mindre usikkerhedsinterval for mediankoncentrationen sammenlignet med kurven til højre, hvor oplandet ikke er målt i perioden 2011-2019. Der ses en tendens til, at tæthedsfunktionen for de ikke-målte oplande er mere højreskæv end for målte oplande.



Figur 4.9. Grafer for bly og to oplande der hhv. er målt i perioden 2011-2019 (ID15 51350461) og ikke målt (ID15 52630769). Enhed på x-aksen er opløst vandkoncentration i µg/l.

En granskning af graferne i Bilag 1 kan identificere en enkelt graf, som viser en stor uoverensstemmelse mellem beregnede og målte koncentrationer fra perioden 2020-2022, for oplandet ID15 13230712 (Herborg Bæk), der ligger i Skjern Ås opland som tilløb til Vorgod Å via Egeris Mølleå, se Figur 4.10. I dette opland er der målt nogle af de højeste cadmiumkoncentrationer i perioden 2020-2022 på landsplan, og de er også noget over prædiktionen fra MetalStat for oplandet. NOVANA målestationen er flyttet opstrøms fra perioden 2011-2019 til perioden 2020-2022, hvilket kan være årsagen til det ændrede koncentrationsniveau. Når en station flyttes opstrøms, kan der være en tendens til mere ekstreme lave/høje målinger, alene fordi enkelte kilder får større betydning pga. lavere vandføring og dermed mindre opblandning.

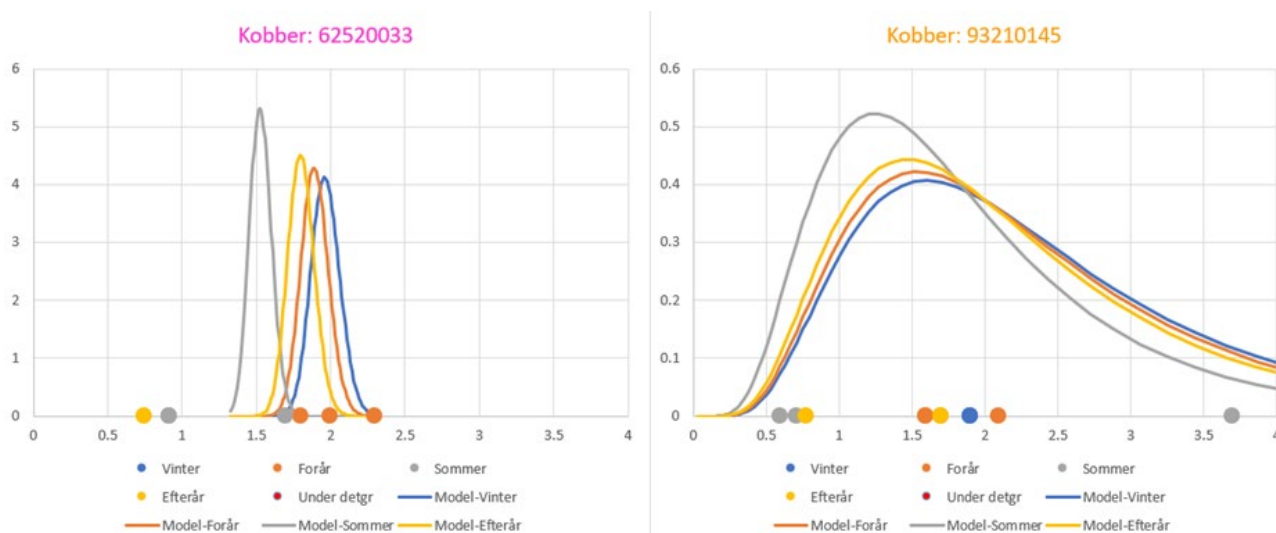


Figur 4.10. Graf for oplandet ID15 13230712 (Herborg Bæk), der ligger i Skjern Ås opland, som tilløb til Vorgod Å, via Egeris Mølleå. Til højre vises at NOVANA stationen er blevet flyttet opstrøms fra periode 2011-2019 til periode 2020-2022. Enhed på x-aksen er opløst vandkoncentration i µg/l.

Kobber

Generelt ses en god sammenhæng mellem tæthedsfunktionernes intervaller og de målte værdier.

På Figur 4.11 ses to kurver for kobber for et opland, der hhv. er målt i perioden 2011-2019 (venstre) og ikke målt (højre). Kurven til venstre er for et opland, som også har målinger i perioden 2011-2019. På samme måde som for bly og cadmium, så fremstår det målte opland med et mindre usikkerhedsinterval for den beregnede mediankoncentrationen sammenlignet med tæthedsfunktionen til højre, hvor oplandet ikke er målt i perioden 2011-2019, med en tendens til at tæthedsfunktionen for de ikke-målte oplande er mere højreskævet end for de målte oplande (Figur 4.11). Ligesom for cadmium, synes kobber at have et lavere koncentrationsniveau om sommeren sammenlignet med resten af året, især om vinteren.

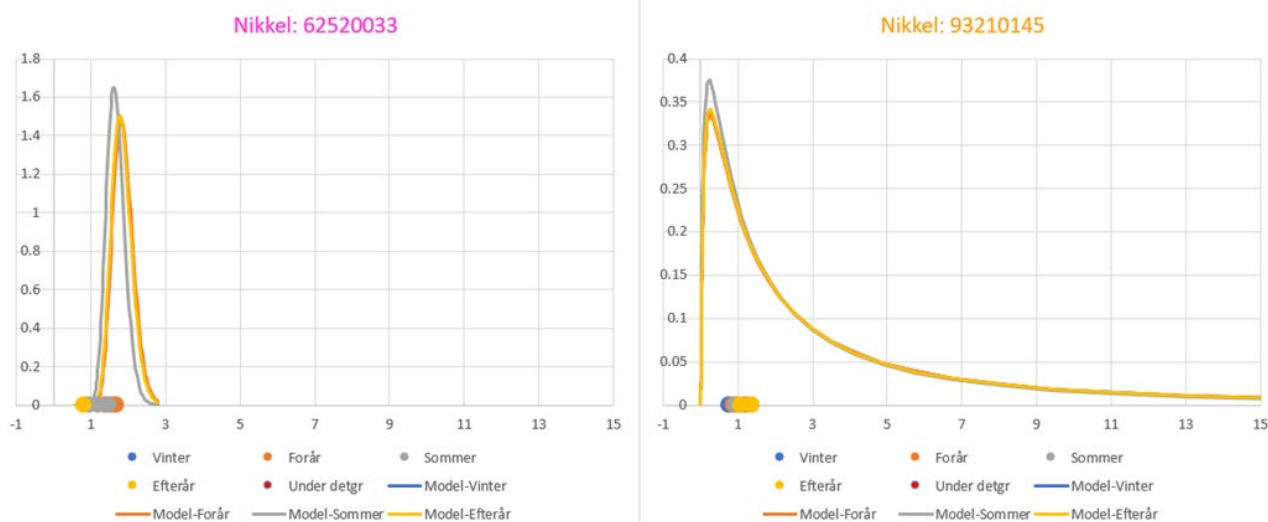


Figur 4.11. Grafer for kobber og to oplande, der hhv. er målt i perioden 2011-2019 (ID15 62520033) og ikke målt (ID15 93210145). Enhed på x-aksen er opløst vandkoncentration i $\mu\text{g/l}$.

Nikkel

Generelt ses der en god sammenhæng mellem tæthedsfunktionerne og de målte værdier for nikkel. Dog er der en tendens til, at oplande, der ikke er målt i perioden 2011-2019, har en stor variationsbredde. Men der er dog typisk et godt sammenfald mellem de målte værdier og de forventede værdier jf. tæthedsfunktionerne, hvilket kan være en konsekvens af den lave R^2 værdi i Tabel 2.1, som viser, at modellen har svært ved at adskille oplandene fra hinanden og derfor giver generisk ens og brede tæthedsfordelinger for alle oplande med udbredt effekt fra tilfældige effekter mellem oplandene.

Sæsonforskellen er ikke særlig udbredt, da kurverne for hver sin sæson overlapper hinanden. På Figur 4.12 ses to kurver for et opland, der hhv. er målt i perioden 2011-2019 (venstre) og ikke målt (højre). Kurven til venstre er for et opland, som har målinger i perioden 2011-2019 og derfor fremstår med et mindre usikkerhedsinterval, sammenlignet med kurven til højre, hvor oplandet ikke er målt i perioden 2011-2019. For nikkel ses også en udbredt højreskævet fordelingskurve for de ikke-målte oplande (Figur 4.12). Denne forskel mellem de målte oplande og de ikke-målte oplande er særlig tydelig for nikkel, hvilket skyldes, at de tilfældige effekter mellem oplande er relativt stor.

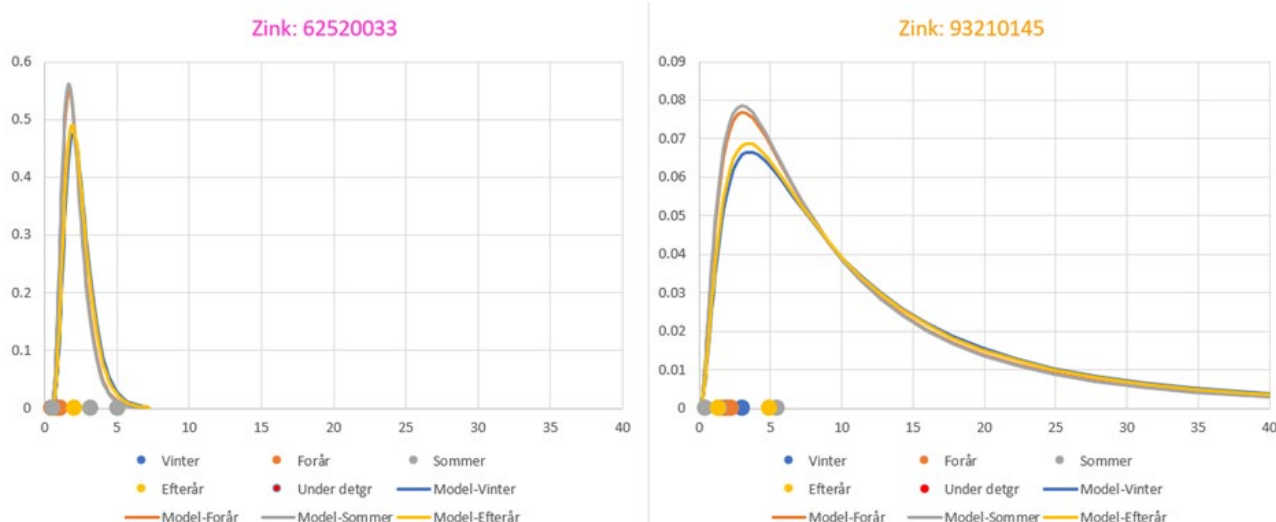


Figur 4.12. Kurver for nikkel og to oplande, der hhv. er målt i perioden 2011-2019 (62520033) og ikke-målt (93210145). Enhed på x-aksen er opløst vandkoncentration i $\mu\text{g/l}$.

Zink

Den tendens, som er vist på Figur 2.2, hvor zinkkoncentrationerne er faldet fra perioden 2011-2019 til perioden 2020-2022, ses tydelig, når MetalStat sammenlignes med målinger fra de enkelte stationer. Generelt viser figurerne i Bilag 1 for zink en overestimering af de målte koncentrationer, når tæthedsfunktionerne sammenlignes med målte værdier.

Sæsonforskellen er ikke særlig udbredt, da kurverne for hver sin sæson overlapper hinanden. På Figur 4.13 ses to tæthedsfunktioner for et opland, der hhv. er målt i perioden 2011-2019 (venstre) og ikke målt (højre). Tæthedsfunktionen til venstre er for et opland, som har målinger i perioden 2011-2019 og derfor fremstår med mindre usikkerhedsinterval sammenlignet med kurven til højre, hvor oplandet ikke er målt i perioden 2011-2019.



Figur 4.13. Kurver for zink og to oplande, der hhv. er målt i perioden 2011-2019 (62520033) og ikke målt (93210145). Enhed på x-aksen er opløst vandkoncentration i $\mu\text{g/l}$.

5 Videreudvikling af MetalStat

I arbejdet med løbende at minimere usikkerheden på MetalStats prædiktioner beskrives de næste tiltag, som vurderes at være de mest oplagte at implementere, nedenfor. Disse forbedringer giver dels en forbedret beregningsmæssig performance med hurtigere beregninger, der mere specifikt kan adskille betydende faktorer fra ikke betydende faktorer. Og dels beskrives de mest oplagte forbedringer af X matricen i bestræbelse på bedre at kunne forudsige metalkoncentrationerne baseret på viden om oplandet.

5.1 Forbedring af beregningseffektivitet og numerisk præcision

Forbedring af beregningseffektivitet og numerisk præcision kan forgå på to fronter-. Den ene tilgang er at optimere koden af modellen, så det kræver mindre beregningsarbejde at komme frem til resultatet. Den anden tilgang er kodning af modellen på en måde, der bedre udnytter tilgængelig CPU og evt. GPU-kapacitet for derved at øge den regnekraft modellen har til rådighed.

Minimering af beregningstid

MetalStat er bygget op om en såkaldt Metropolis–Hastings algoritme (Sørensen et. Al., 2022), hvor modellen gennem iterationer afsøger troværdige værdier for alle modellens variable og kortlægger deres usikkerhed. Disse iterationer er regnekrævende approksimation til en "sand" men ukendt løsning. Metropolis–Hastings algoritmen vil således altid konvergere mod den sande løsning, og det er løbende muligt at tjekke, hvor tæt iterationerne er kommet til den sande løsning, men nogle typer af sådanne algoritmer er hurtigere til at konvergere mod den sande løsning end andre. Formålet med en forbedret algoritme er altså ikke, at MetalStat skal være mere korrekt i dens estimer, men at modellen skal kunne opnå gode approksimationer til den sande løsning med mindre beregningstid, og derfor blive mere fleksibel og brugbar. Dette skal også ses i lyset af, at modellen derved forberedes til andre miljøfarlige forurenende stoffer, hvor hver stofgruppe vil stille krav til ekstra beregningsarbejde.

Ved den bayesianske metode, som MetalStat bygger på, kan de simple uniforme prior's for koefficienterne i X matricen erstattes af såkaldte "horseshoe priors" (Bhadra et al. 2017). Navnet horseshoe dækker over en billedmetafor med udgangspunkt i en hestesko, der vender buen nedad, hvor en koefficient som udgangspunkt forudsættes at have en betydning, der enten er stor eller lille og sjældnere er i midten. Med dette udgangspunkt vil Metropolis–Hastings algoritmen, ved brug af mindre beregningstid, kunne adskille de betydende faktorer fra dem med mindre betydning. Brugen af horseshoe priors vil således accelerere iterationerne til hurtigere at identificere og separere de faktorer i X matricen, der har hhv. ingen og stor betydning. Denne information vil så kunne udnyttes til at optimere X matricen. En version af MetalStat med horseshoe priors er under testning, men resultatet når ikke at komme med i denne afrapportering.

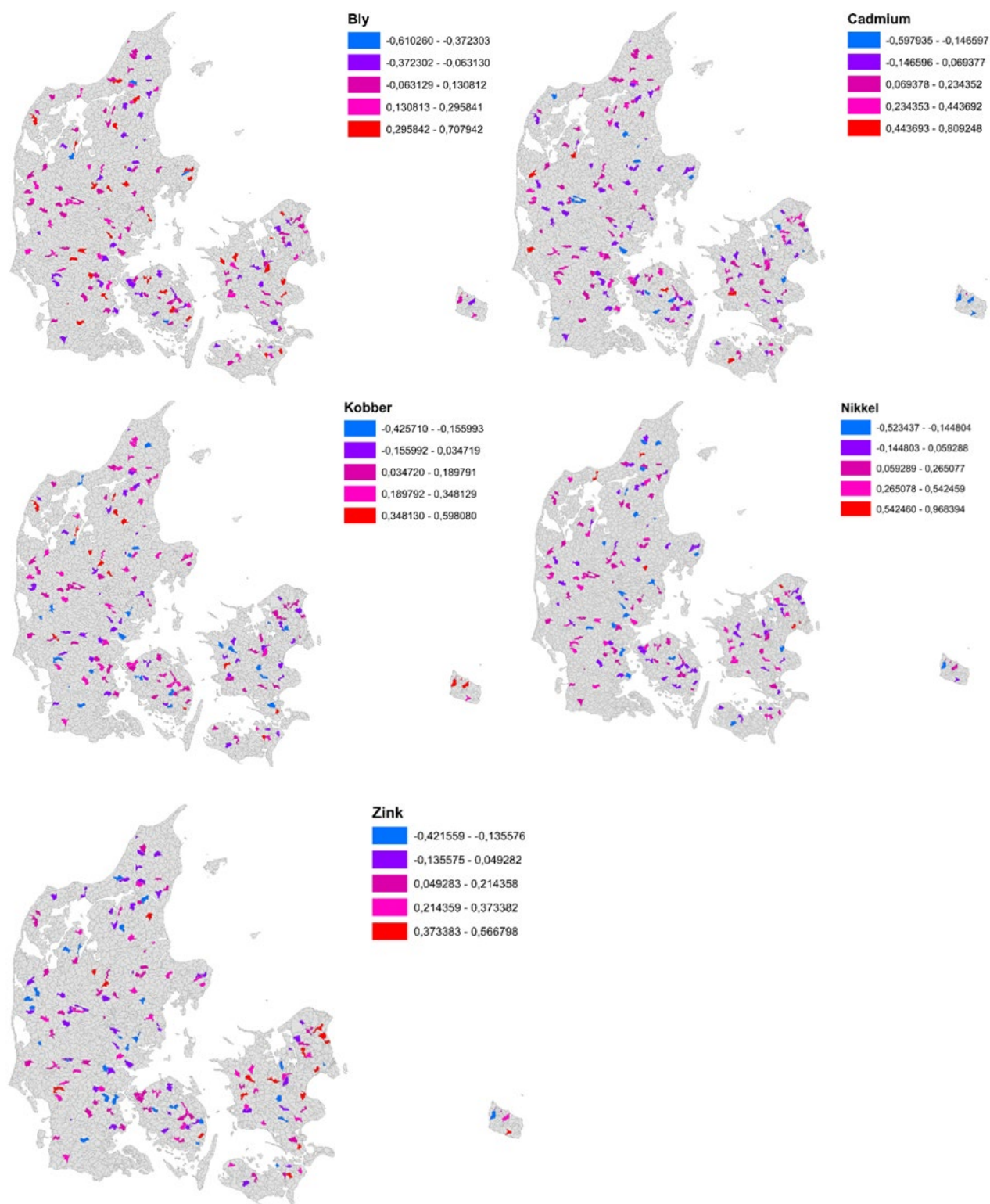
Øgning af regnekraft

I nærværende version af MetalStat bestemmes troværdig værdisætning af 800 variabler, og når modellen i 2023 bliver opdateret med de nye målinger fra 2020-2022, vil antallet af variabler øges betydeligt med de nye latente koncentrationer fra de nye ca 400 oplande, der er målt ekstra i perioden 2020-2022. I en iteration testes hver variabel for en ny potentiel værdi, hvor alle de andre variablers værdi fastholdes. Dvs. at variabler kan beregnes parallelt i flere CPU-kerner inden for den samme iteration (parallelisering af koden) og principielt kan hver variabel blive beregnet i hver sin kerne, hvorved den beregningstid som en enkelt kerne kan levere ved at skulle testet alle 800 variabel kan erstattes af beregningstiden for test af en enkelt variabel i hver sin kerne. Det er dog ikke realistisk at beregne alle variabler på hver sin kerne, da det stiller alt for store krav til kodning og hardware, men det er realistisk at bruge omkring 50 kerner i stedet for kun en enkelt, som bruges i den nærværende version af MetalStat.

Forbedring af X matricen

Af Tabel 2.3 fremgår det, at R^2 værdien er relativt beskeden især for nikkel, hvilket påpeger, at der generelt er plads til forbedring af X matricen for metallerne generelt. Der har derfor været en faglig diskussion i arbejdsgruppen bag rapporten, hvad der vil være det mest oplagte skridt i bestræbelserne på at forbedre X matricens forklaringskraft. En oplagt forbedring af X matricen vurderes til at være relateret til en bedre beskrivelse af grundvandsbidraget fra dybere, typisk mineralogiske kilder til metallerne. Dernæst kan de forurenede grunde også tænkes at have en betydning, hvilket dog ikke vurderes til at være lige så betydningsfuldt som grundvandsbidraget, for metaller. Det anbefales derfor, at en forbedret modellering af grundvandsbidraget prioriteres over en inddragelse af de forurenede grunde.

De tilfældige effekter på oplandsniveau beskriver de uforklarlige effekter, hvor et opland har en latent koncentration, der ikke fuldt ud kan forklares med X matricen, som tilfældige effekter mellem år, eller som sæsoneffekter. En uforklarlig effekt kan enten afsløre, om et opland har et uforklarligt lavt eller højt koncentrationsniveau sammenlignet med andre oplande. Derfor afslører en kortlægning af tilfældige effekter, de målte oplande hvor X matricen ikke slår til ved prædiktion af latent koncentration baseret på målinger fra perioden 2011-2019. Det geografiske mønster af tilfældige effekter er et godt input til diskussionen om X matricen. På Figur 5.1 ses den geografiske fordeling af de tilfældige effekter for hvert metal. For bly er der en tendens til, at der er en relativt beskeden uforklarlig variation mellem de oplande, der dækker de store Vestjyske åer (Storåen, Skjern Å og Varde Å), da de tilfældige effekter er heterogent fordelt med både lave og høje værdier rundt om i resten af landet. Cadmium har en jævn heterogen fordeling af de tilfældige effekter i hele landet, dog med en tendens til, at Bornholmske oplande har uforklarligt lave koncentrationer. Kobber har ligeledes jævnt fordelte tilfældige effekter, med en tendens til, at der er uforklarligt høje koncentrationer på Bornholm. Nikkel udviser ikke nogen klart geografisk mønster. Zink udviser som det eneste metal en tendens til at være uforklarligt høj omkring København, hvilket kunne skyldes kilder så som zinktagrender ikke er tilstrækkeligt godt inddraget i X matricen, men for zinks vedkommende er der, som vist på Figur 2.2 sket så store forskydninger i zinkkoncentrationerne i perioden 2020-2022, så den geografiske fordeling af de tilfældige effekter i nærværende analyse, der er baseret alene på målinger fra perioden 2011-2019, skal tages med forbehold, især for oplande der er placeret i landbrugsområder.



Figur 5.1. Tilfældige effekter fordelt på de oplande, hvorfra der er målinger i perioden 2011-2019. De tilfældige effekter er enhedsløse, da de beskriver relative forskelle som beskrevet i afsnit 4.2.

5.2 Forbedret inddragelse af grundvand

I den nuværende version af MetalStat er grundvand inddraget alene i form af mængden af grundvand, der som middel strømmer til vandløbet hvert ID15 opland per år. Så i det omfang, der er geografiske forskelle på grundvandskoncentrationen vil modellen i dens nuværende version ikke afspejle dette.

Derfor er det oplagt at udbygge X matricen til bedre at kunne inddrage forskellige grundvandskoncentrationer. Denne diskussion blev allerede åbnet i afrapportering af MetalStat, Sørensen et al., 2022, men der blev ikke lavet yderligere tiltag gennem 2022.

Sammenhængen mellem geokemiske parametre, geokemisk oprindelse og forekomsten af metaller i grundvandet er undersøgt af Mortensen et al., 2021. En hovedkonklusion fra denne undersøgelse er, at de fem metaller afhænger primært af pH og redoxforhold, som vist i Tabel 5.1.

Tabel 5.1. Fordeling af hhv. pH og Redox som betydende for metallerne naturlige baggrundsværdier, efter Mortensen et al., 2021, (Tabel 7.4).

Metal	pH	Redox
Bly		X
Cadmium	X	X
Kobber	X	X
Nikkel	X	X
Zink	X	X

Redox og pH kan kombineres med geologien i grundvandsmagasinet, der leverer vand til vandløbet i det pågældende ID15 opland som vist i Tabel 5.2 (udpluk af Tabel 7.4. i Mortensen, et al., 2021).

Tabel 5.2. Forslag til klassificering for at adskille høje og lave grundvandskoncentrationer (efter Mortensen et al., 2021, Tabel 7.4).

Parameter	Klassificeringer
Geologi i grundvandsmagasin	Prækvartær sand (ps), kvartær sand (ks), kalkbjergarter (kalk), blandede bjergarter på Bornholm (uu)
pH	≤ 6 og >6
Nitrat (redox)	≤ 2 mg/l og >2 mg/l

Med udgangspunkt i Tabel 5.2, ses det, at der er fire klassificeringer for geologi og to for hhv. pH og nitrat. Da metalkoncentrationen kan tænkes at afhænge af interaktionen mellem disse kategorier, kan de ikke behandles hver for sig som kategoriske variabler, men bør indgå i indbyrdes kombination som interaktioner, hvor pH f.eks. har forskellig betydning, alt efter hvilken geologi der er til stede og afhængig af redox niveauet. Det giver med udgangspunkt i Tabel 5.2 i alt $4 \times 2 \times 2 = 16$ forskellige kombinationer, af hhv. geologi, pH og redox. Hver af disse med en potentiel forskellig betydning for metalkoncentrationen. Det vil være en klar fordel for parameterbestemmelsen i MetalStat at reducere disse 16 kombinationer til de 10 vigtigste, samt fastlægge mere specifikke redoxgrænser. Dette kan gøres med udgangspunkt i en korrelationsanalyse med brug af Artificial Intelligens som f.eks. en Random Forest analyse mellem boredata fra Jupiterdatabasen og klassificeringen i Tabel 5.2. En sådan analyse vil kunne bestemme betydningen af hver kombination af klassificeringen i Tabel 5.2 set i forhold til at bestemme metalkoncentrationer i boreanalyser og bruges som model, der understøtter X matricen i MetalStat. Denne analyse vil også kunne bruges til specificere redoxgrænser og estimere usikkerhed ved tærskelværdierne (i Tabel 5.2 er tærskelværdien forudsat fast på hhv. pH 6 og 2 mg/l nitrat).

Næste skridt for at forbedre inddragelsen af grundvandsbidrag vil være at opstille relevante data inklusive GIS lag i samarbejde med relevante institutioner.

5.3 Inddragelse af udsivning fra forurenede grunde

Jordforureningens påvirkning af overfladevand er kvantificeret til brug i en screeningsmodel i MST, 2014a. Her er der estimeret en forureningsflux for forskellige modelstoffer. Som modelstof for metaller anvendes arsen, og fluxen beregnes med brug af følgende ligning (MST, 2014b, Lign 3.1):

$$J_{\text{worst case}} = C_{\text{worst case}} \cdot A_{\text{kilde}}(\text{branche}) \cdot N_{\text{kommune}} \quad (3)$$

Hvor:

$C_{\text{worst case}}$ er koncentration i vandfasen (betegnes som $C_{\text{worst case}}$ i MST, 2014b)

Brancegruppe 1: Diameter = 3 m $\rightarrow A_{\text{kilde}} = 7.1 \text{ m}^2$

Brancegruppe 2: Diameter = 10 m $\rightarrow A_{\text{kilde}} = 79 \text{ m}^2$

Brancegruppe 3: Diameter = 30 m $\rightarrow A_{\text{kilde}} = 710 \text{ m}^2$

Brancegruppe 4: Diameter = 100 m $\rightarrow A_{\text{kilde}} = 7900 \text{ m}^2$

N_{kommune} er nettonedbør i den pågældende kommune.

Der er stor forskel på de forskellige branchers reelle betydning som kilde til metaller, så det er nødvendigt at kategorisere dem efter deres potentielle betydning. Det kan eventuelt foregå ved en eksperworkshop i dialog med Regionerne og jordforureningseksperter bliver bedt om at rangordne brancher i forhold til deres potentielle risiko for at afgive metaller til jorden. En sådan rangordning kan evt. foretages for hvert metal for at kunne differentiere mellem dem for hhv. V1 og V2 vidensniveau. En rang kan fordele brancher i fire kategorier: (1) Høj risiko; (2) Mellem Risiko; (3) Lav Risiko; (4) Ingen risiko.

Hvis Metalstat skal udvides til at inddrage forurenede grunde i ID15 oplandene, kan et eksponentielt fald i koncentration med afstanden f.eks. antages og transporttiden fra kilde til vandløb og dermed indgå i stedet for worst case koncentrationen vist i ligning 3., da formålet med MetalStat er at estimere sandsynlighedsfordelingen for koncentrationen. Ved samtidigt at have inddelet brancher i risikoklasser for hvert metal, kan følgende model anvendes i MetalStat:

$$J = X_{\text{ned}} \cdot (\beta_1 \sum_1^N [A1_n e^{-\theta \cdot X_{\text{ned}} \cdot l1_n}] + \beta_2 \sum_1^M [A2_m e^{-\theta \cdot X_{\text{ned}} \cdot l2_m}] + \beta_3 \sum_1^Z [A3_z e^{-\theta \cdot X_{\text{ned}} \cdot l3_z}]) \quad (4)$$

Hvor:

J : er mass transport over en specificeret tidsperiode (et år) til vandløbet i et ID15 opland for et miljøfarligt forurenende stof, X_{ned} er nettonedbør i oplandet, som i øvrigt er med i nærværende version af MetalStat, β_1 er ukendt effektiv betydning for massetransport til vandløb fra højrisikokilder, der forlader den forurenede jord, $A1_n$ er arealet den n 'th højrisikokilde i ID15 oplandet, $l1_n$ er korteste afstand fra den n 'th kilden til vandløbet, der bestemmes som fast værdi, med en GIS model, θ er en ukendt koefficient i udtrykket for det eksponentielle fald i vandkoncentrationen ved transporten fra kilde mod vandløbet, β_2 er ukendt effektiv betydning for massetransport til vandløb fra mellemrisikokilder, $A2_m$ er arealet den m 'th mellemrisikokilde i ID15 oplandet, $l2_m$ er korteste afstand fra kilden til vandløbet for den m 'th mellemrisikokilde, der bestemmes som fast værdi, med en GIS model, β_3 er ukendt

effektiv betydning for massetransport til vandløb fra lavrisikokilder, $A3_z$ er arealet den z 'th lavrisiko kilde i ID15 oplandet, $l3_z$ er korteste afstand fra kilden til vandløbet, der bestemmes som fast værdi, med en GIS model. For hvert ID15 opland vil der således være opgivet N høj-risikogrunde, M mellem risikogrunde og Z lavrisikogrunde, hvor hver af disse grunde er angivet med en nærmeste afstand til vandløb. For hvert metal vil der således være behov for at estimere fire koefficienter: Tre værdier (β_1 , β_2 , β_3) der beskriver massefrigivelsen per areal fra hhv. høj-, mellem- og lavrisikobrancher, samt en transportkoefficient (θ), der beskriver nettoresultatet af retention og mobilitet ved transport fra den forurenede grund til vandløbet for hvert metal. Transportkoefficienten (θ) kan evt. gøres afhængig af den dominerende jordtype i ID15 oplandet, så den hurtigere transport gennem f.eks. sandjord sammenlignet med lerjord kan komme med som en faktor i modellen. Det kunne gøres ved at lade θ være en kategorisk variabel, der afhænger af jordtypen. Arealer fordelt på branche og afstand til nærmeste vandløb i hvert ID15 opland skal fremskaffes primært ved at analysere GIS lag. Der skal etableres et nyt datasæt, som er opdateret for fx 2022. Dette skal ske i et tæt samarbejde med Miljøstyrelsen og danske regioner. Der bør skelnes mellem V1 og V2 kortlagte lokaliteter

6 Konklusion og implementering af MetalStat

6.1 Konklusion

Som beskrevet i Sørensen et al., 2022, stiller MetalStat store krav til beregningsarbejdet, og der er derfor udført nogle regnetekniske forbedringer i den nærværende version af MetalStat i forhold til afrapporteringen i Sørensen et al., 2022. Det er især brug af en lokal likelihood funktion for de latente koncentrationer, der har mindsket beregningstiden. Det betyder, at MetalStat i en beregningsgang, der tester værdien for en latent koncentration, nu kun skal regne specifikt på det ID15 opland, som denne latente koncentration estimeres i, uden at genberegne alle de andre oplande.

Beregningsarbejdet for MetalStat kan overordnet inddeles i to trin, hvor første trin er en estimering af variabler ud fra målinger og andet trin er at levere prædiktioner med brug af de estimerede variabler. I det første trin indgår kun de ID15 oplande, hvor der er målt i NOVANA programmet og de enkelte oplande regnes for at være uafhængige af hinanden, hvilket er en gyldig forudsætning under første trin. Til gengæld er denne forudsætning ikke optimal i andet trin, når MetalStat skal prædiktere koncentrationsniveauer for alle ID15 oplande, inklusiv dem uden målinger. Hvis et ID15 opland uden målinger udveksler vand med et andet opland med målinger, vil disse målinger i det målte opland, kunne bidrage med viden om koncentrationsniveauet i det umålte opland, da noget af vandet i det målte opland også optræder i det umålte opland. Derfor bliver der i den version af MetalStat som valideres mod nye målinger i denne rapport, lavet en massebalancejustering for at forbedre modellens prædiktioner, så viden om de målte ID15 oplande får indflydelse på prædiktioner for oplande uden målinger, i det omfang udveksling af vand med målte oplande berettiger til det.

I MetalStat indgår en X matrice, der samler oplysning fra GIS om hvert opland, (se Sørensen et al., 2022). Som det også er beskrevet i Sørensen et al., 2022, så vil det være nyttigt løbende at evaluere og forbedre X matricen, i takt med at flere målinger og nye datakilder bliver tilgængelige. En regressionskoefficient R^2 på logtransformerede koncentrationer er beregnet ud fra, hvor godt X matricen sammen med sæsonvariationen kan prædiktere den latente koncentration i de målte oplande. Fire ud af fem metaller var i intervallet 0,19-0,3, mens Nikkel faldt ud med en lav værdi på 0,06. Så der er klart plads til forbedring af især X matricen og især for Nikkel. Det er projektgruppens vurdering, at det første skridt mod en forbedring er at inddrage grundvandsbidraget mere detaljeret. En lille værdi for R^2 betyder, at MetalStat beskriver det meste af variationen i koncentrationsværdier dels som tilfældige effekter mellem oplande og år og dels som residualer, der varierer tilfældigt fra den ene måling til den næste.

Nye målinger fra perioden 2020-2022 er anvendt til at evaluere MetalStat for de fem metaller. Der har været et ret intensivt måleprogram i denne periode, hvor der er indsamlet ca. dobbelt så mange prøver i perioden 2020-2022 som for hele perioden 2011-2019. En deskriptiv statistisk analyse af de målinger fra perioden 2020-2022, der var målt over detektionsgrænsen, indikere et generelt lidt lavere niveau for bly, og et lidt højere niveau for cadmium, kobber og nikkel, mens niveauet for zink var markant lavere i perioden 2020-2022, sammenlignet med perioden 2011-2019. Hvis denne forskydning mellem de to

perioder er betydelig større end forskydninger mellem de enkelt år i perioden 2011-2019 kan det udfordre MetalStat. Den efterfølgende analyse af de enkelte metaller viser at denne forskydning udfordrer MetalStat for zink, men ikke for de andre fire metaller.

For ca. 400 oplande og for hvert metal er tæthedsfunktion, der viser mediankoncentrationens sandsynlighedsfordeling sammenlignet med den aktuelle måling fra perioden 2020-2022. Nogle af disse oplande var der også målt på i perioden 2011-2019, så her er den estimerede latente koncentration brugt som udtryk for mediankoncentrationen, mens mediankoncentrationen er prædikeret for de resterende 400 nye målte oplande. Denne tæthedsfunktion afspejler niveau og usikkerhed for mediankoncentrationen i det pågældende opland. Kun i relativt få tilfælde, ligger de nye målte koncentrationer uden for MetalStats prædikerede intervaller, og disse er domineret af bly. En nærmere granskning af de enkelte større afvigelser for især bly viser, at oplandsstørrelsen på 10-15 km² i nogle tilfælde godt kan dække over en heterogenitet, som ikke synes helt at være repræsentativt dækket ind af de valgte placeringer af NOVANA stationer. For zink giver MetalStat generelt en overprædiktation af usikkerhedsintervallet, som konsekvens af det voldsomme fald, der har været i koncentrationsniveauet mellem periode 2011-2019 og periode 2020-2022. Når MetalStat opdateres med målingerne fra perioden 2020-2022 vil det derfor være en fordel at estimere variablerne for zink fra år 2020 som en ny latent variabel, så der er to versioner af latente variabler for zink i modellen, et før 2020 og et andet fra 2020.

I bestræbelserne på at optimere X matricen afprøves såkaldte "horseshoe priors", der finder de vigtigste betydende faktorer med mindre beregningstid, men denne afprøvning er ikke tilendebragt med udgivelsen af denne rapport. Alt tyder dog på, at denne tilgang vil vise sig fordelagtig og derfor bør indgå i en kommende driftsversion af MetalStat. Dette skal også ses i lyset af, at modellen derved forberedes til andre miljøfarlige forurenende stoffer, hvor hver stofgruppe vil stille krav til beregningstid. Under udarbejdelse af denne rapport blev det klart, at der ikke blev færdiggjort en driftsversion af MetalStat til forvaltningen i 2022, men at denne aktivitet vil blive søgt udført i 2023. Der er dog foretaget en planlægning af kodning af driftsversionen. Under denne planlægning udarbejdedes en klar plan for at parallelisere MetalStat til at accelerere beregningshastigheden over flere CPU-kerner. Så der foregår løbende en udvikling, der minimerer beregningstiden, dels ved at regne smartere og derfor mindske behovet for beregninger, dels ved at øge den computerkraft som kan udnyttes til at regne hurtigere.

Den væsentligste forestående forbedring af X matricen vurderes at være en bedre beskrivelse af grundvandsbidraget, hvorfor en sådan forbedring bør laves som led i udvikling af en driftsversion af MetalStat. Det anbefales, at der tages udgangspunkt i eksisterende undersøgelser, der relaterer metaller i grundvand til geologi, pH og redoxforhold. X matricen kan udvides med disse attributter, og MetalStat vil kunne finde den reelle betydning af disse vandløbskoncentrationer. Da relationen mellem disse attributter og metal-koncentrationen i grundvand er stærkt ikke-lineær, er det bedst at bruge en model, der anvender afskæringsværdier i stedet for at forudsætte en kontinuert funktion mellem pH og redox på den ene side og metalkoncentrationen på den anden side. Derfor anbefales det, at der foretages en indledende analyse af sammenhænge mellem metalkoncentration i grundvand, pH, redox og geologi, som skal danne input til beskrivelsen med X matricen i MetalStat. Første

skridt skal være at sikre adgang til relevante GIS lag i samarbejde med relevante institutioner.

Forurenede grunde kunne også inddrages i X matricen, men det vurderes ikke at være så betydende som grundvandsbidraget for de fem metaller i MetalStat, hvorfor denne aktivitet skal prioriteres efter en inddragelse af grundvand. Med disse forbehold gives der i denne rapport et forslag til en model for forurenede grunde, som kunne bygges ind i MetalStat med udgangspunkt i en eksisterende screeningsmodel for forurenede grundenes potentielle worst case påvirkning af vandløb, med et opdateret datagrundlag.

Hovedformålet med MetalStat er at beregne metalkoncentrationer for alle oplande i Danmark. MetalStat arbejder med mediankoncentrationer, men disse kan efterfølgende omregnes til gennemsnitskoncentrationer, i det omfang det er ønskeligt i sagsbehandlingen. For de oplande, hvor metalkoncentrationen er blevet målt, udføres beregninger ved hjælp af estimerede latente variabler, som modellerer middelværdien af koncentrationen på en logskala. De estimerede latente variabler er tilpasset en multinormal målemodel og renset for effekten af år og sæson, men effekten af både år og sæson kan medregnes hvis ønsket. For de oplande, hvor metalkoncentrationen ikke er blevet målt, laves prædiktionen ved hjælp af den estimerede oplandsmodel, den estimerede tilfældige effekt af forskellige oplande samt en strukturel proces-kovariansmatrice. Effekten af både år og sæson kan også her medregnes hvis ønsket. Prædiktionerne bliver beregnet som sandsynlighedsfordelinger, og disse kan summeres for de enkelte oplande ved f.eks. fordelings middelværdi, median eller andre fraktiler over et år. Usikkerheden af prædiktionen kan f.eks. gives ved et 95% usikkerhedsinterval (der er 95 % sandsynlighed, for at den rigtige værdi ligger inden for intervallet). Det er også muligt at angive den prædikterede sandsynlighed for, at oplandet overskrider et fastsat miljøkvalitetskriterie i den sæson, hvor det biotiske respons er mest følsomt. Det må forventes, at nye data, som inkluderes i MetalStat, vil føre til en præcisering af prædiktionerne; så 95 % usikkerhedsintervallerne formindskes, hvilke betyder at de øvre usikkerhedsintervaller typiske mindskes.

Den estimerede oplandsmodel kan derudover anvendes til at dokumentere betydningen af forskellige kilder i forhold til den samlede andel af metalkoncentrationer i vandløb, som kan forklares med oplandets attributter. For eksempel kan man beregne, at forureningskilde X udgør 10 % af den samlede tilførsel i et givet geografisk område. Det vil også være muligt at identificere andelen af naturgivne kilder i forhold til en mediankoncentration i et opland.

7 Referencer

Bhadra A, Datta J, Polson NG, Willard BT. 2017. Lasso meets horseshoe: A survey. arXiv preprint arXiv:170610179.

Mette Hilleke Mortensen, Vibeke Ernsten, Denitza Voutchkova og Lærke Thorling, 2021, Danmarks og Grønlands Geologiske Undersøgelse. Udvikling af metode til vurdering af grundvandsforekomsters kemiske tilstand for udvalgte uorganiske sporstoffer og salte, Dokumentationsrapport, rapport 2021/19.

Gelman A, Goodrich B, Gabry J, Vehtari A. 2019. R-squared for bayesian regression models. The American Statistician. 73(3):307-309.

MST, 2014a, Jordforureningers påvirkning af overfladevand, delprojekt 6. Systematisering af data og udvælgelse af overfladevandstruende jordforureninger. Miljøprojekt nr. 1573, 2014.

MST 2014b, Jordforureningers påvirkning af overfladevand, delprojekt 3. Relationer mellem stoffer, koncentrationer og fluxe. Miljøprojekt nr. 1574, 2014.

Ovesen, N.B., Iversen, H.L., Larsen, S.E., Müller-Wohlfeil, D.-L., Svendsen, L. M., 2000, Afstrømningsforhold i danske vandløb, Faglig rapport fra DMU, nr. 340.

Sørensen P.B., m.fl., 2022: Peter Borgen Sørensen, Christian Frølund Damgaard, Poul Løgstrup Bjerg, Hans Estrup Andersen, Peter Engelund Holm, Henrik Tornberg, Jesper Leth Bak, Goswin Johann Heckrath, Ane Kjeldgaard, Patrik Fauser. 2022. Udvikling af model til understøttelse af vandplanlægningen i fastlæggelsen af koncentrationsniveauer af metaller i danske vandløb. Aarhus Universitet, DCE – Nationalt Center for Miljø og Energi, 70 s. - Videnskabelig rapport nr. 484.

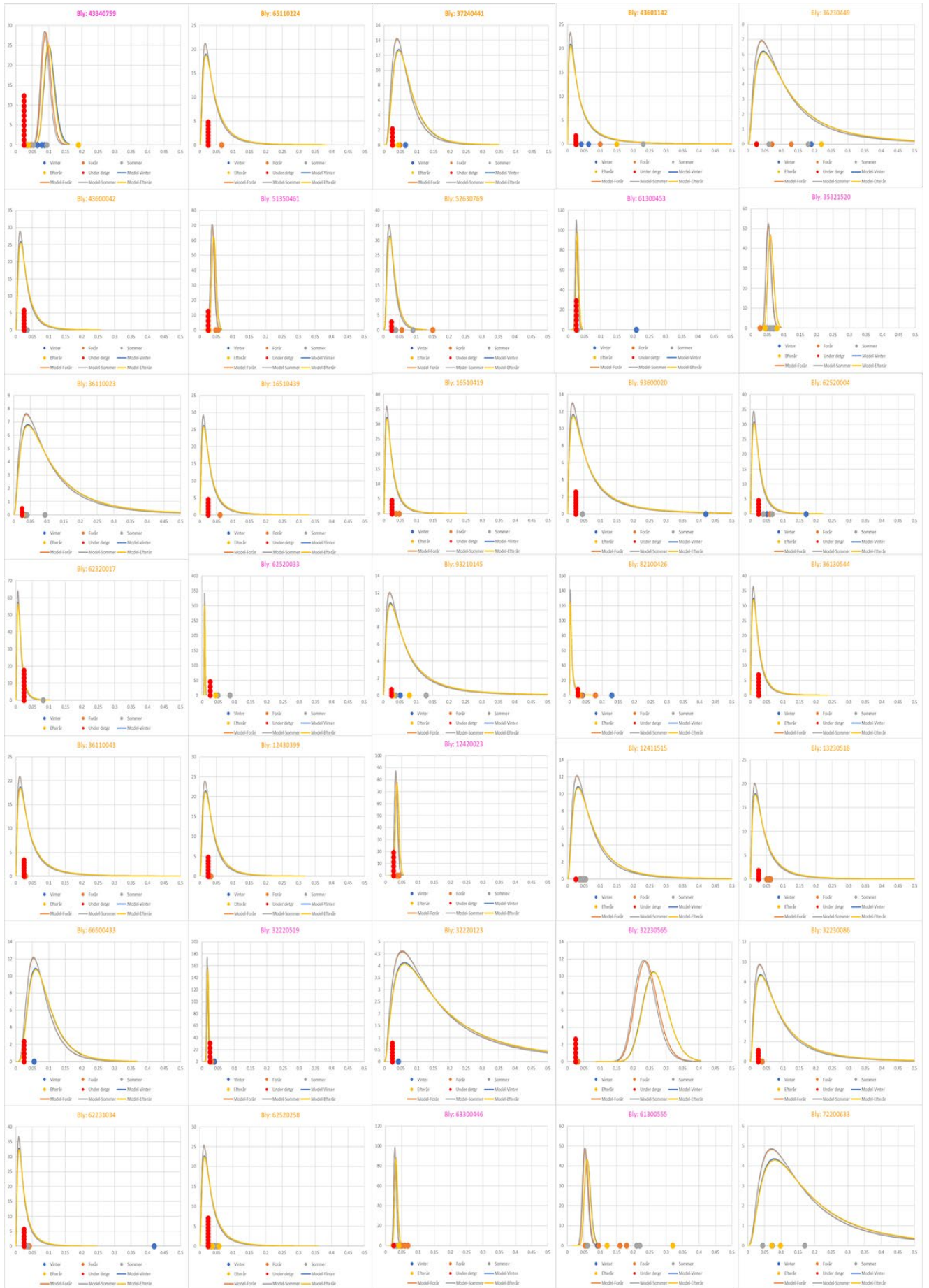
<https://dce2.au.dk/pub/SR484.pdf>

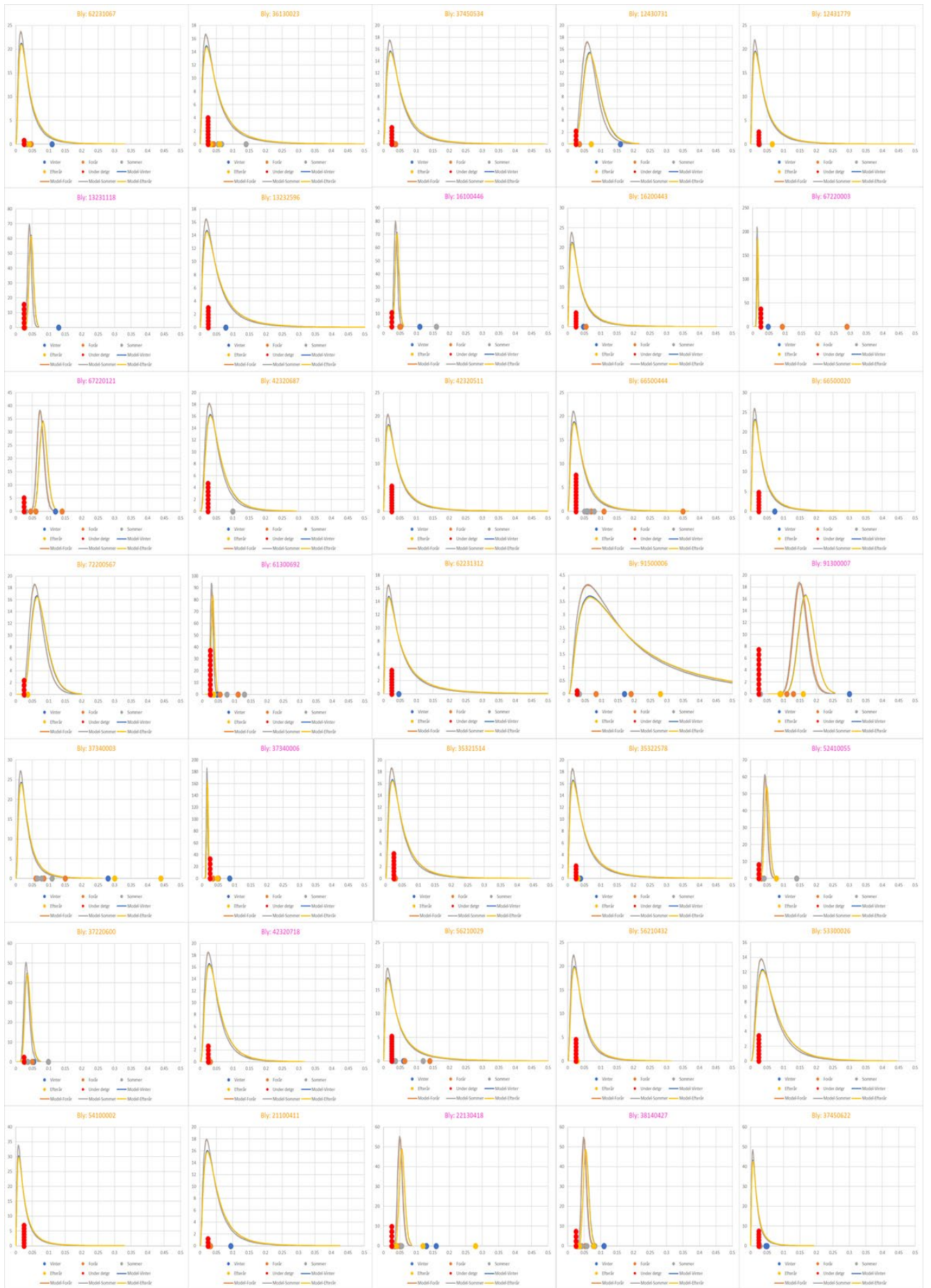
Troldborg, L., Tornbjerg, H., Blicher-Mathiesen, G., Kronvang, B., Thodsen, H., Ernsten, V., 2015, National kvælstofmodel. Oplandsmodel til belastning og virkemidler. Metode rapport. De Nationale Geologiske Undersøgelser for Danmark og Grønland, 2015.

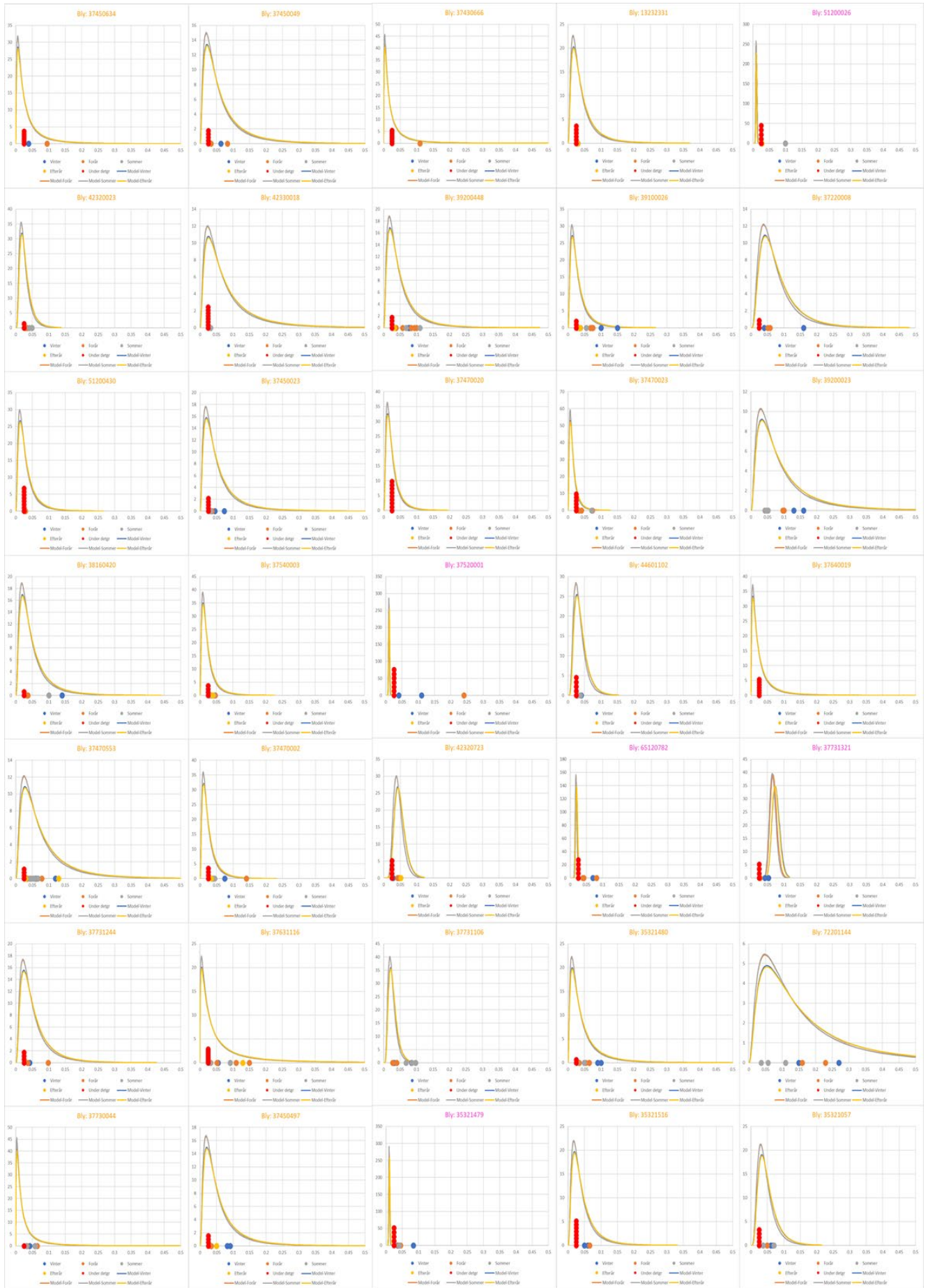
Bilag 1

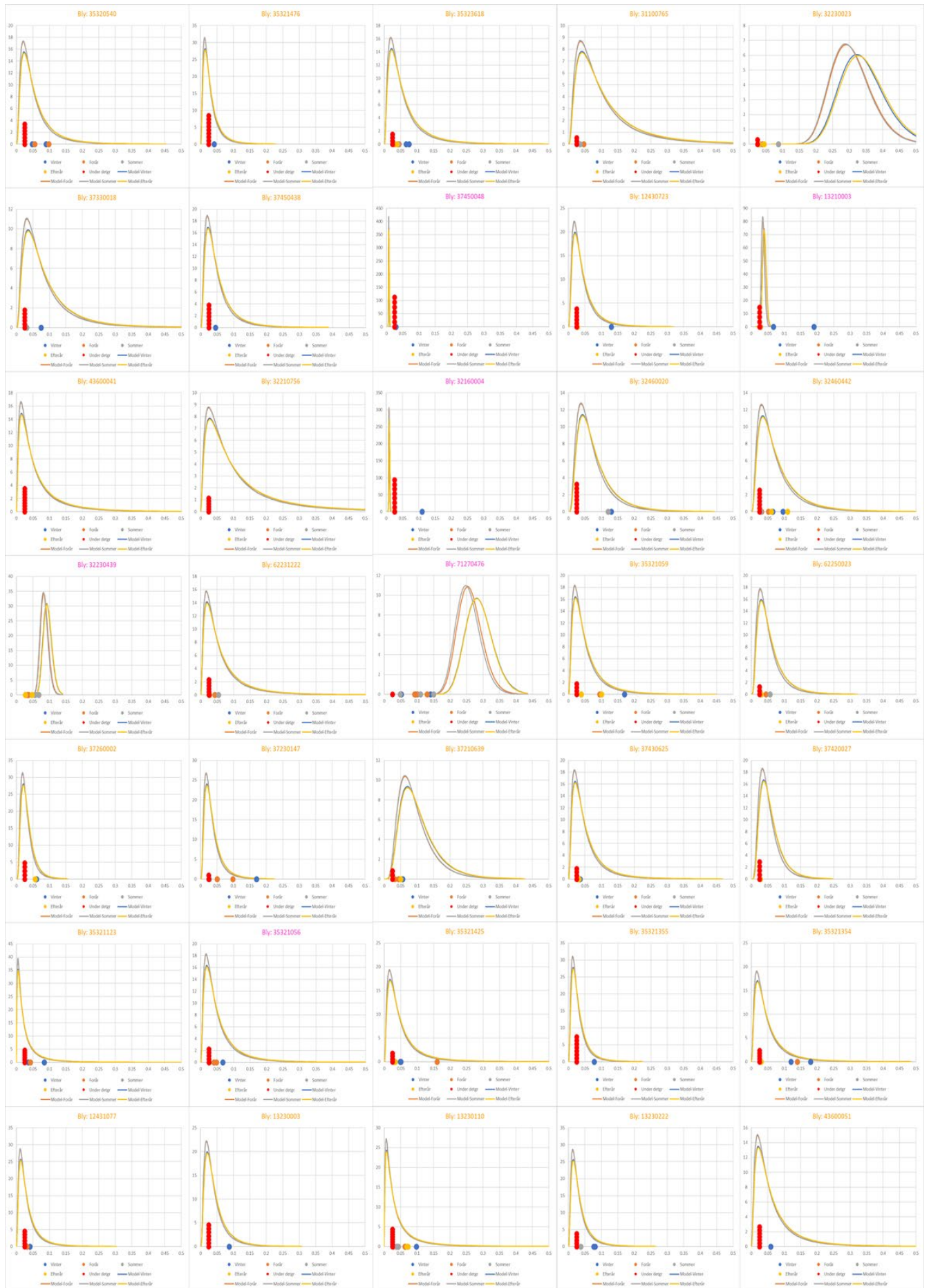
I de følgende figurer vises de målte koncentrationer for de fem metaller i perioden 2020-2022, samt de prædikterede fordelinger af den latente variabel fundet ved MetalStat ved brug af målinger fra perioden 2011-2019 og inklusiv massebalancejustering. Hver kurve er angivet med "metal:oplandsnummer" som titel. Kurven er altså estimeret fordeling af mediankoncentrationer, mens punkterne på akse er aktuelle målinger. Farver af titlen følger farvekode på Figur 4.8, hvor lilla betyder, at der er målinger for dette opland i MetalStat; gul betyder, at der ikke er brugt målinger for dette opland fra perioden 2011-2019 i MetalStat.

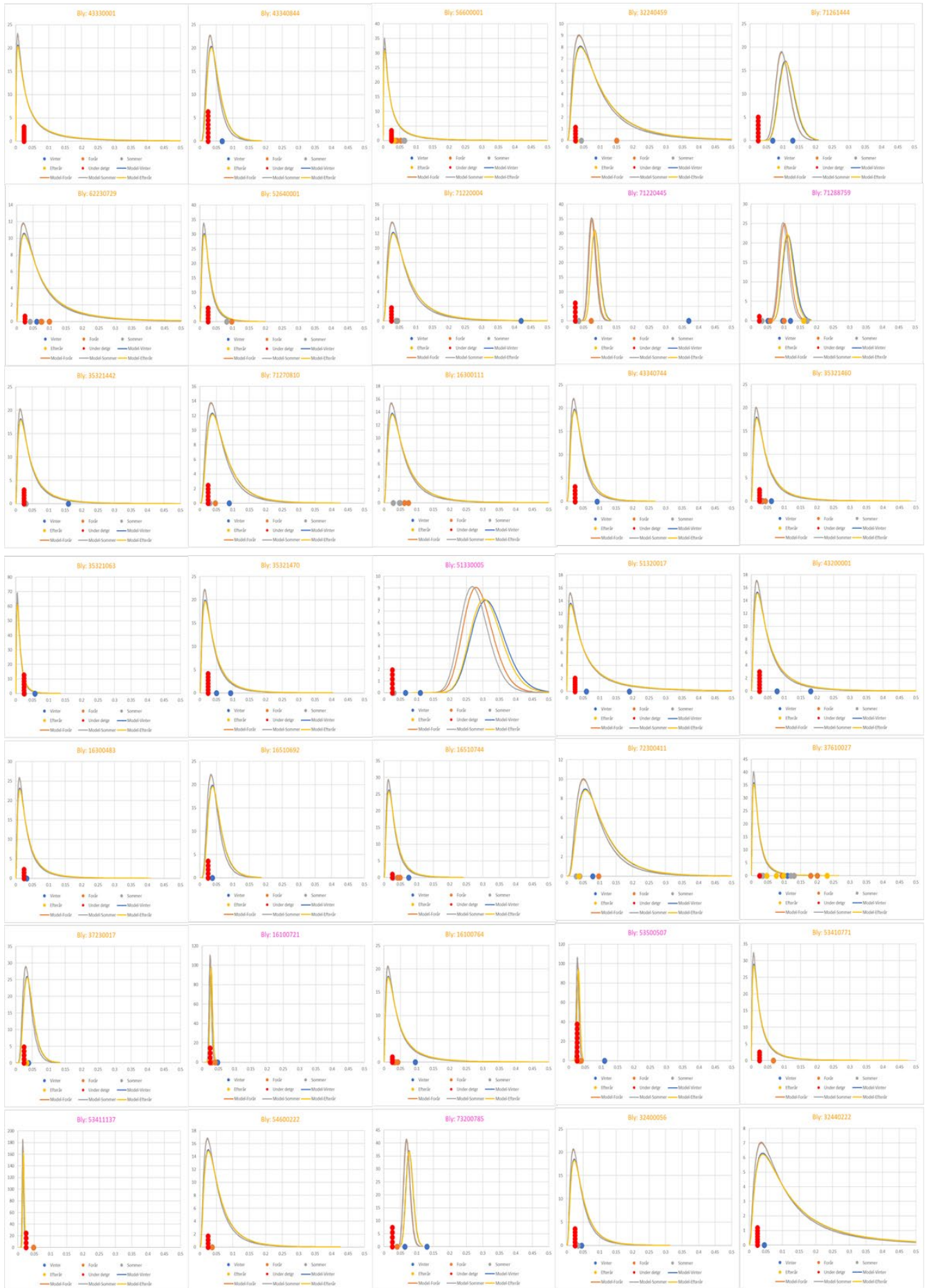
Røde punkter, der er stablet, hvis der er flere, angiver hver en måling under detektionsgrænsen. Der er anvendt farvekode for hver sæson, men samme kode på hhv. målinger og fordelinger.

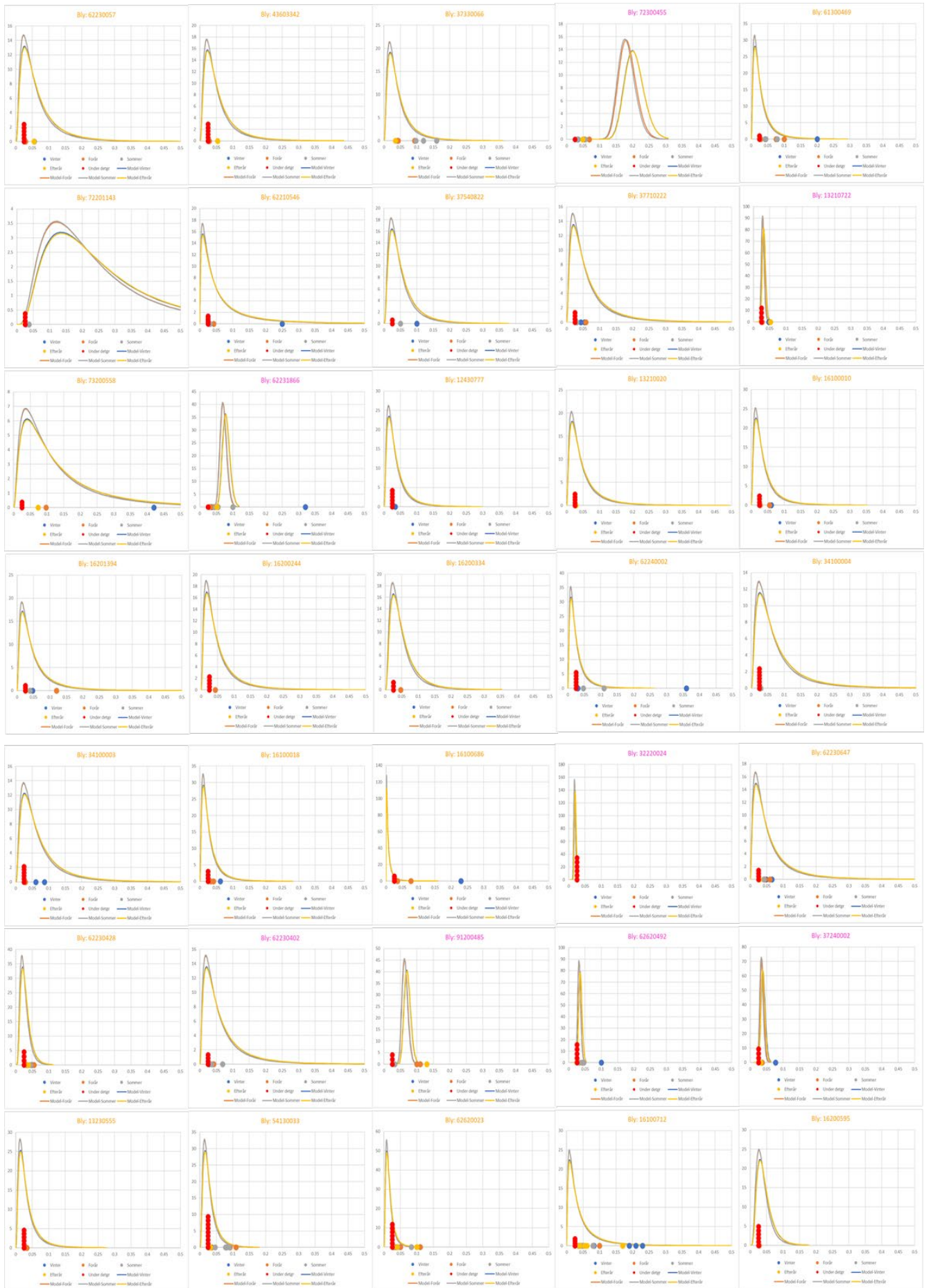


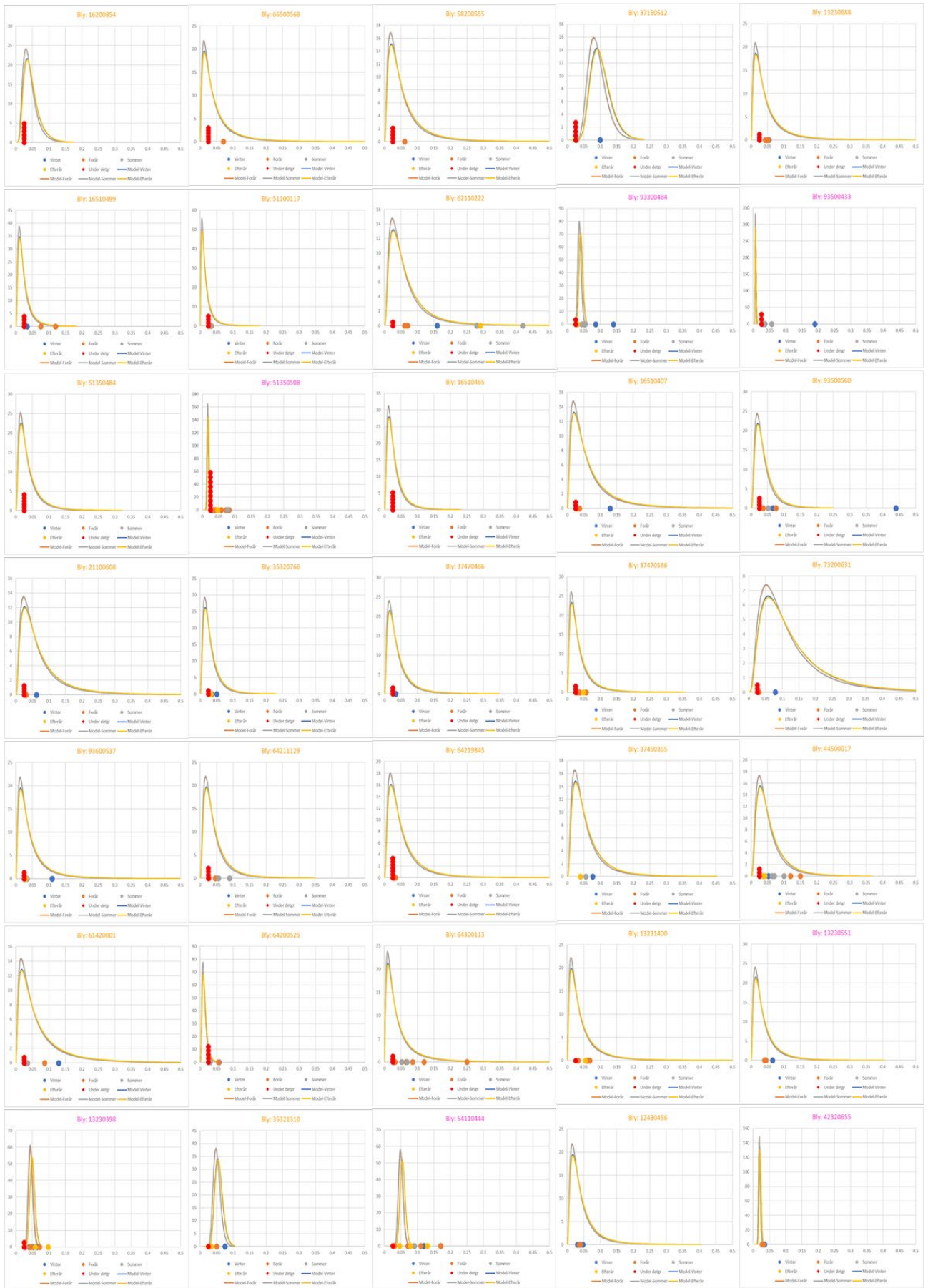


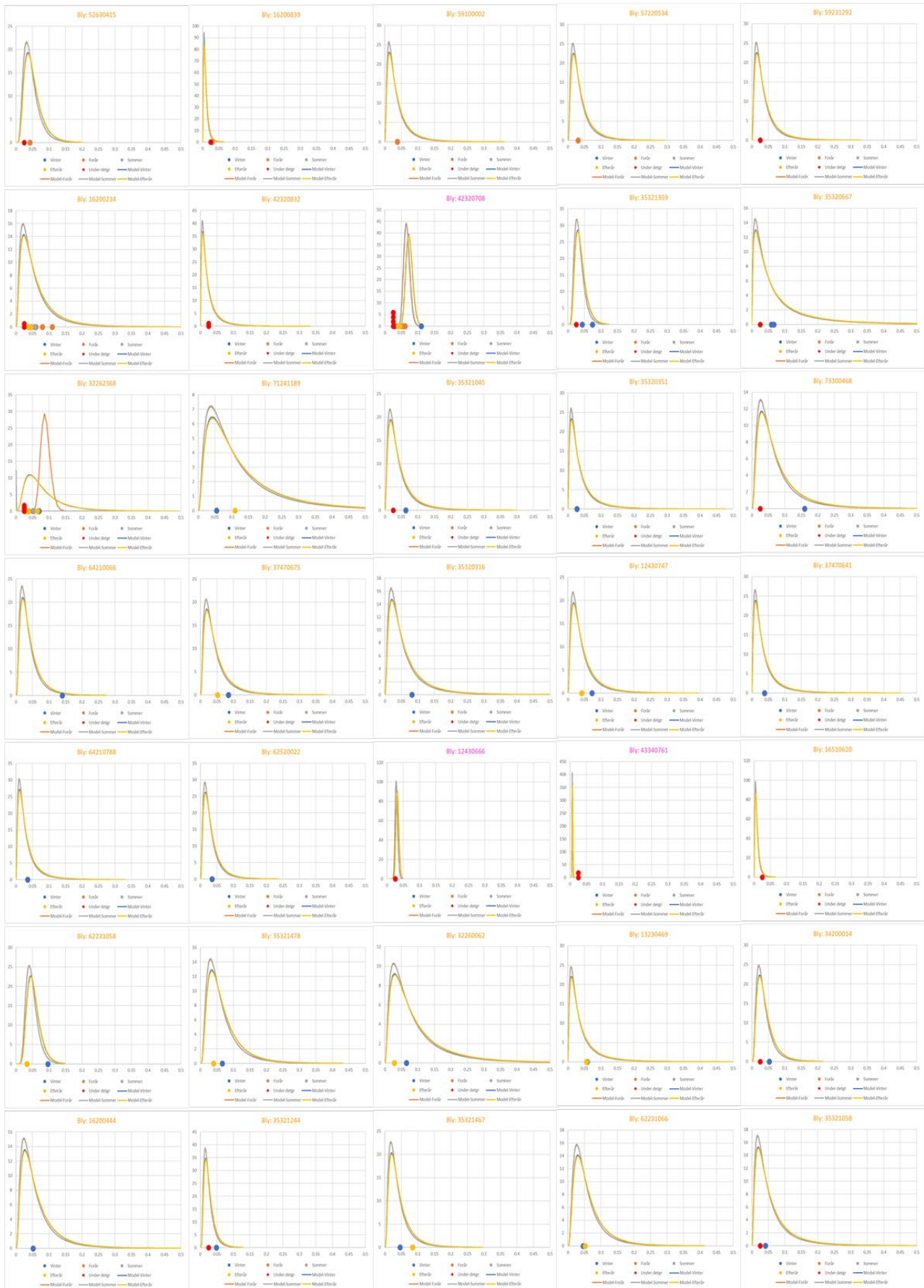


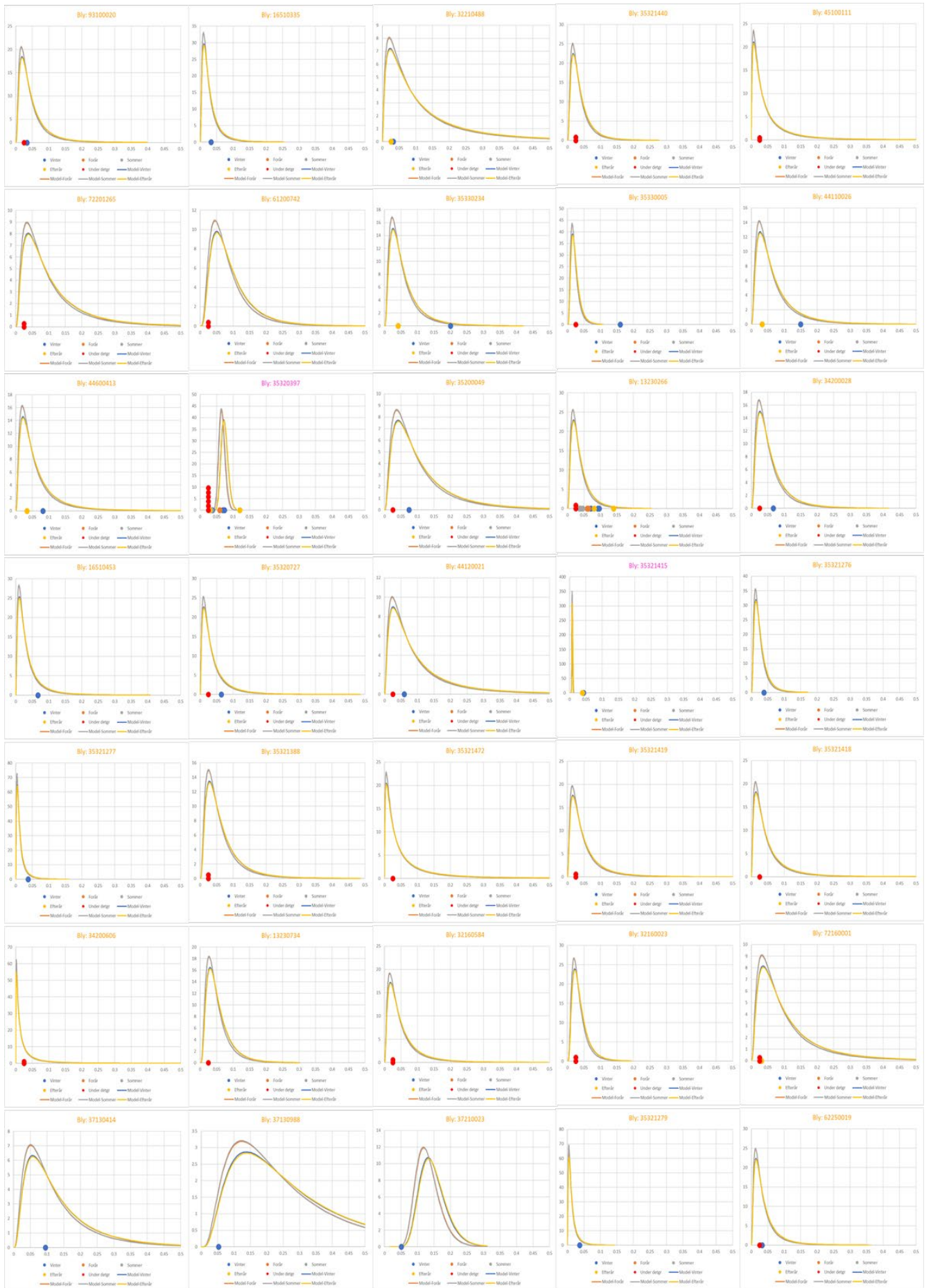


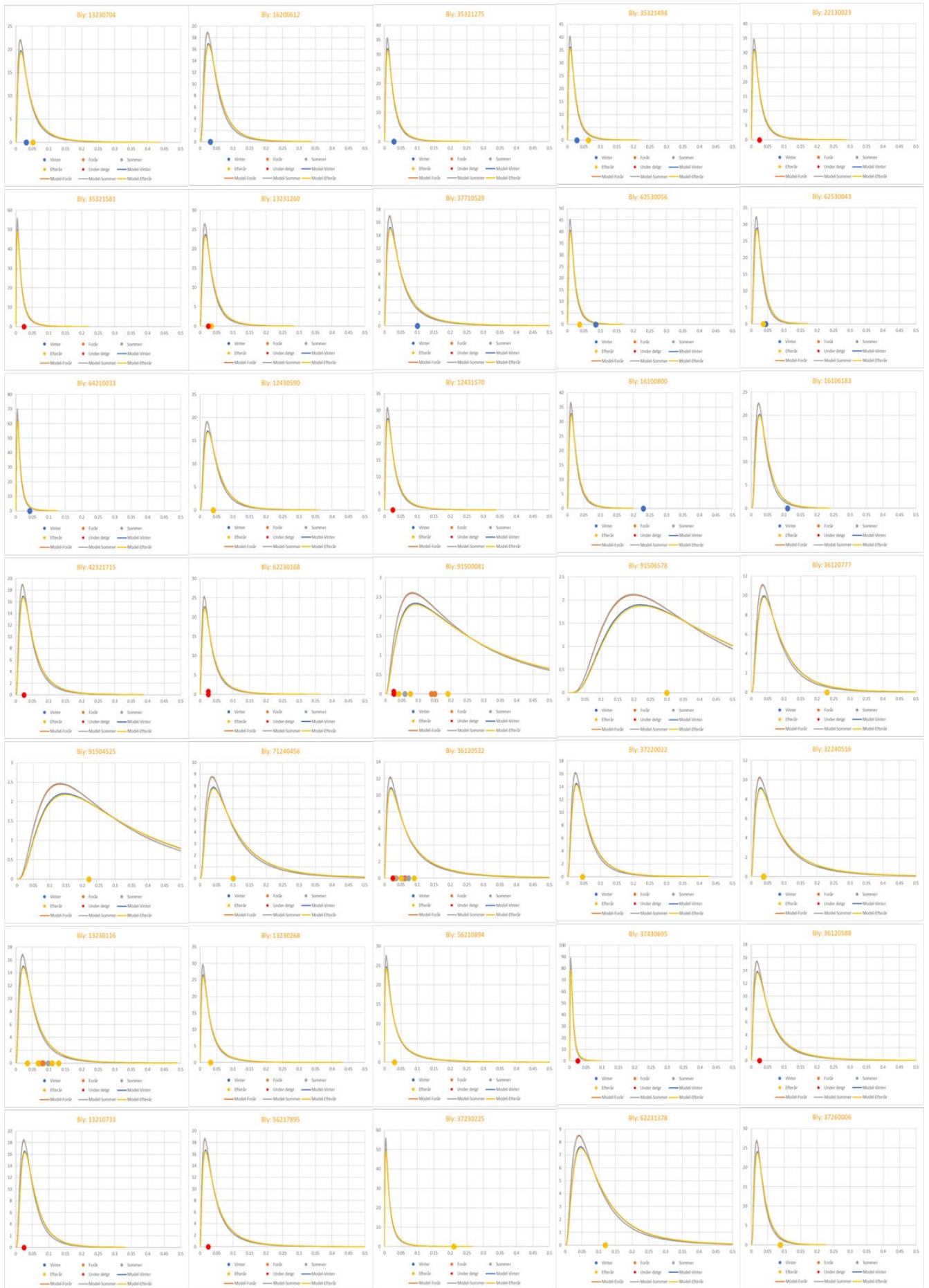


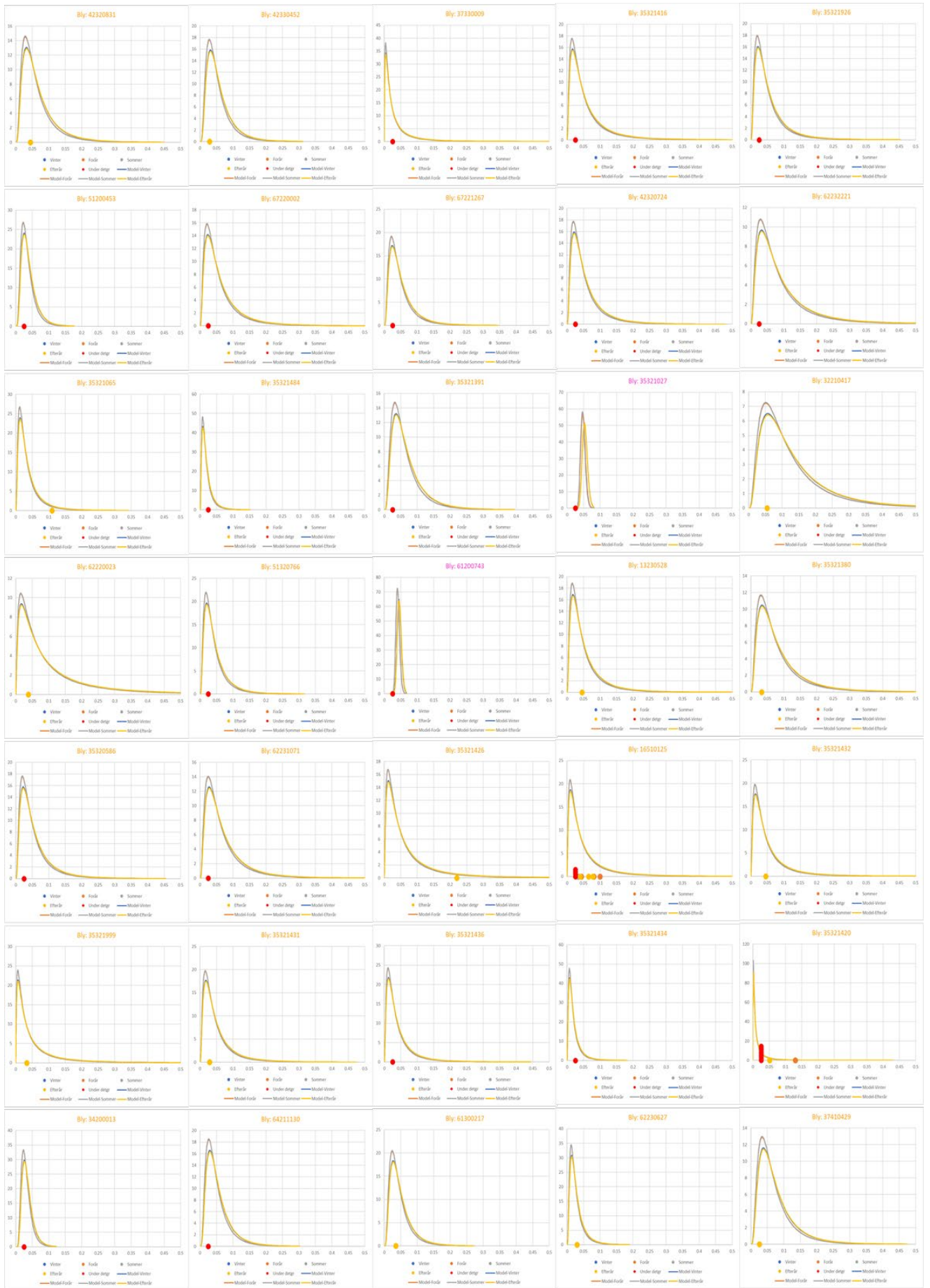


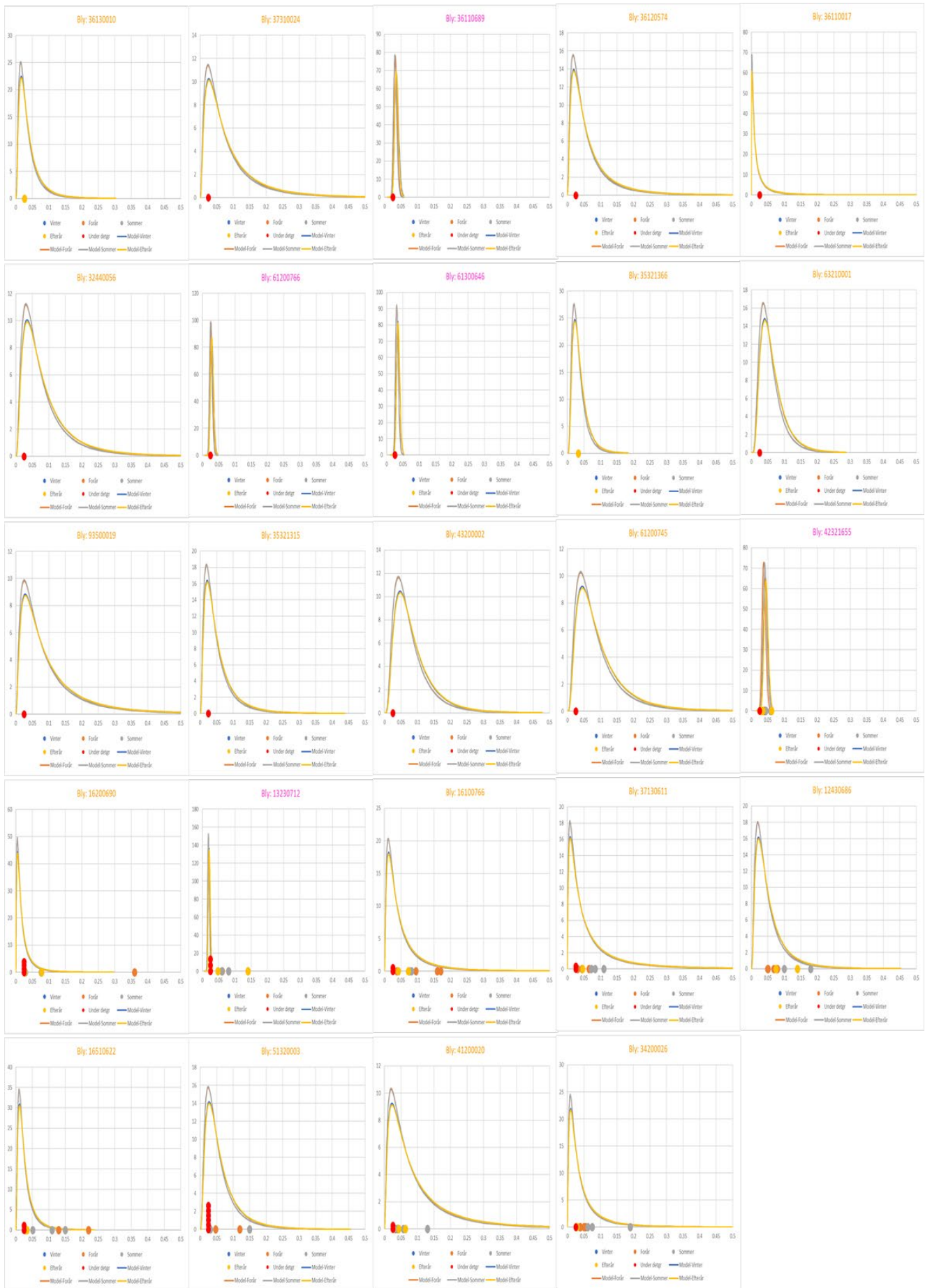


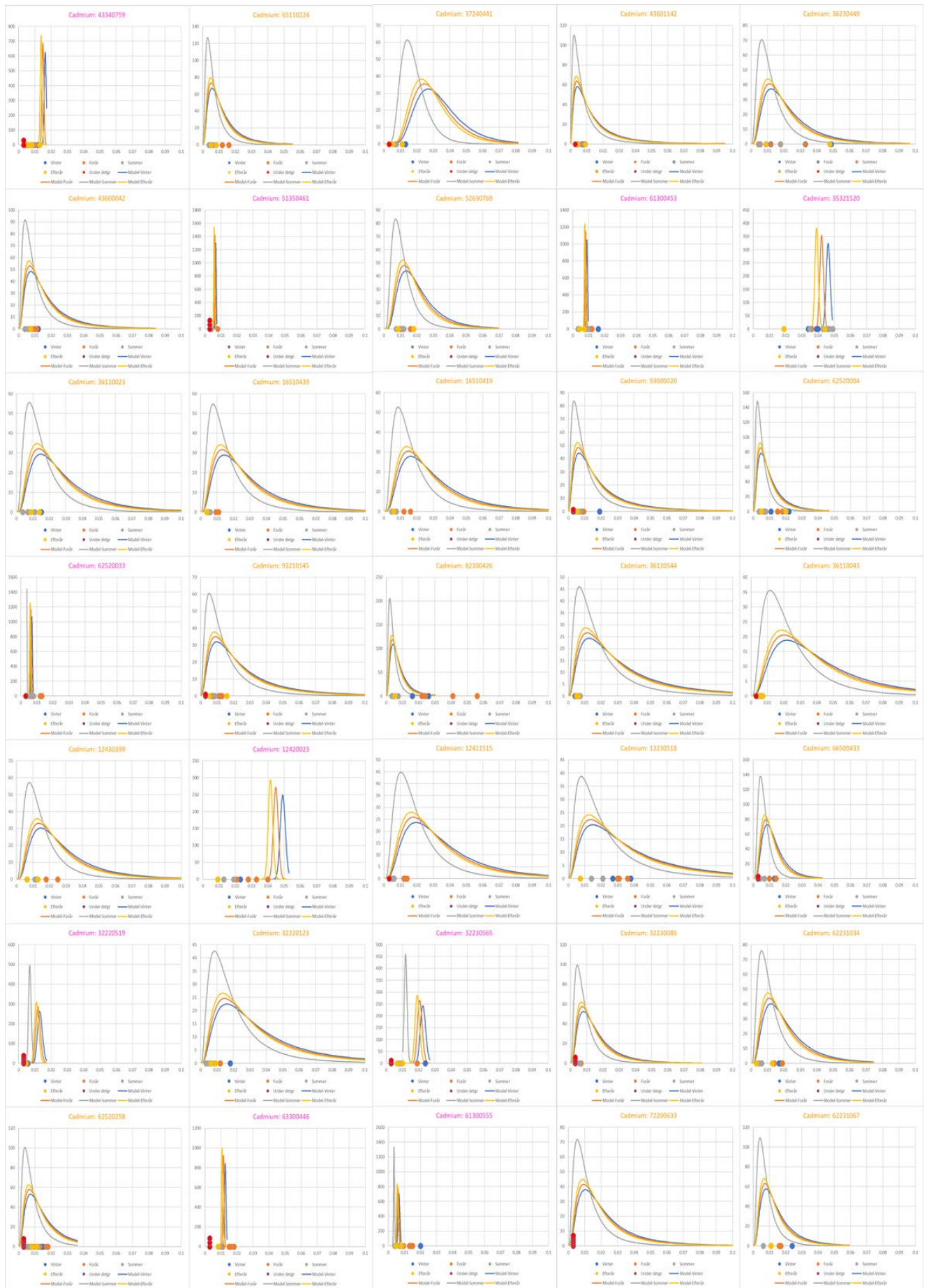


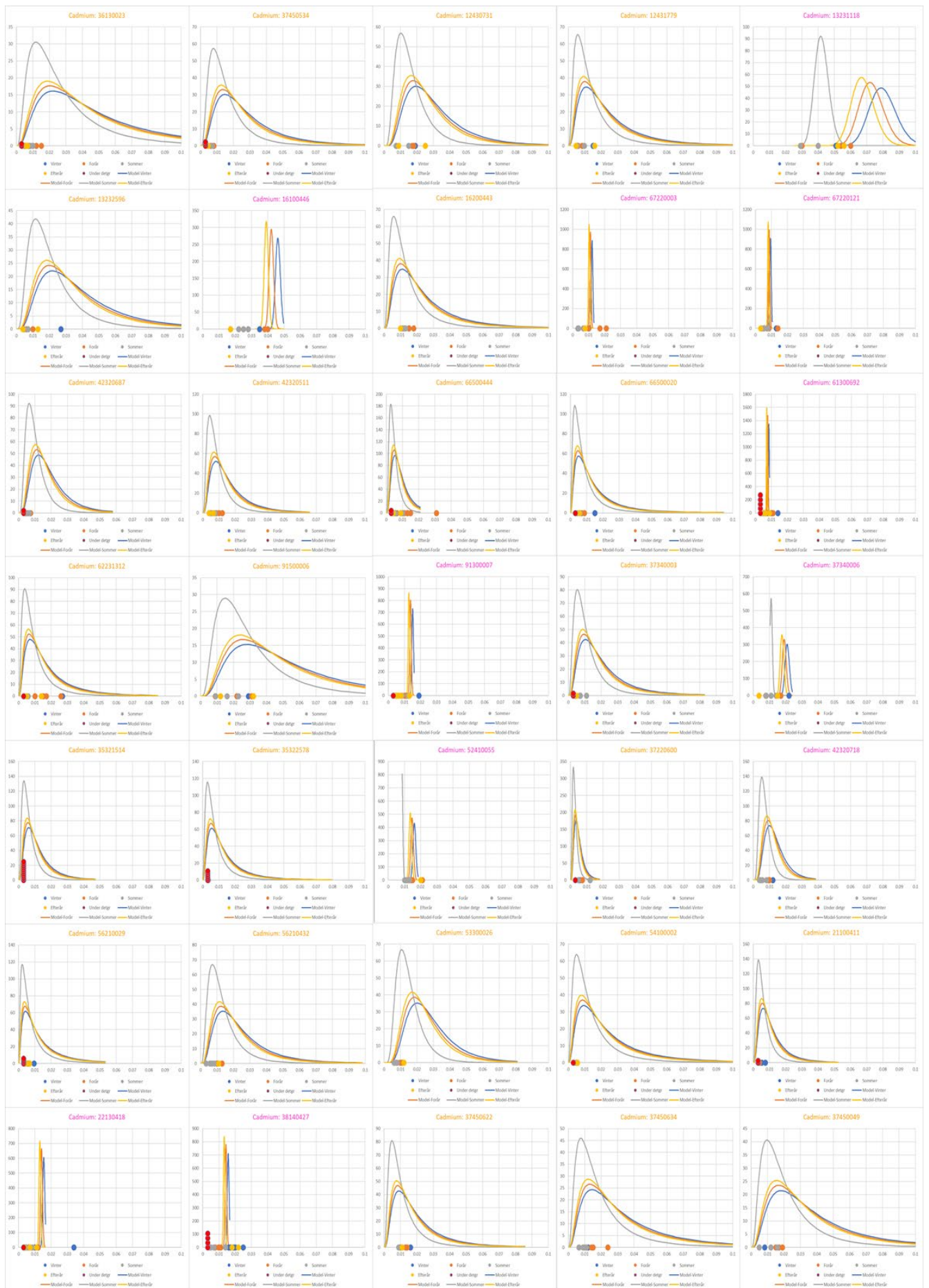


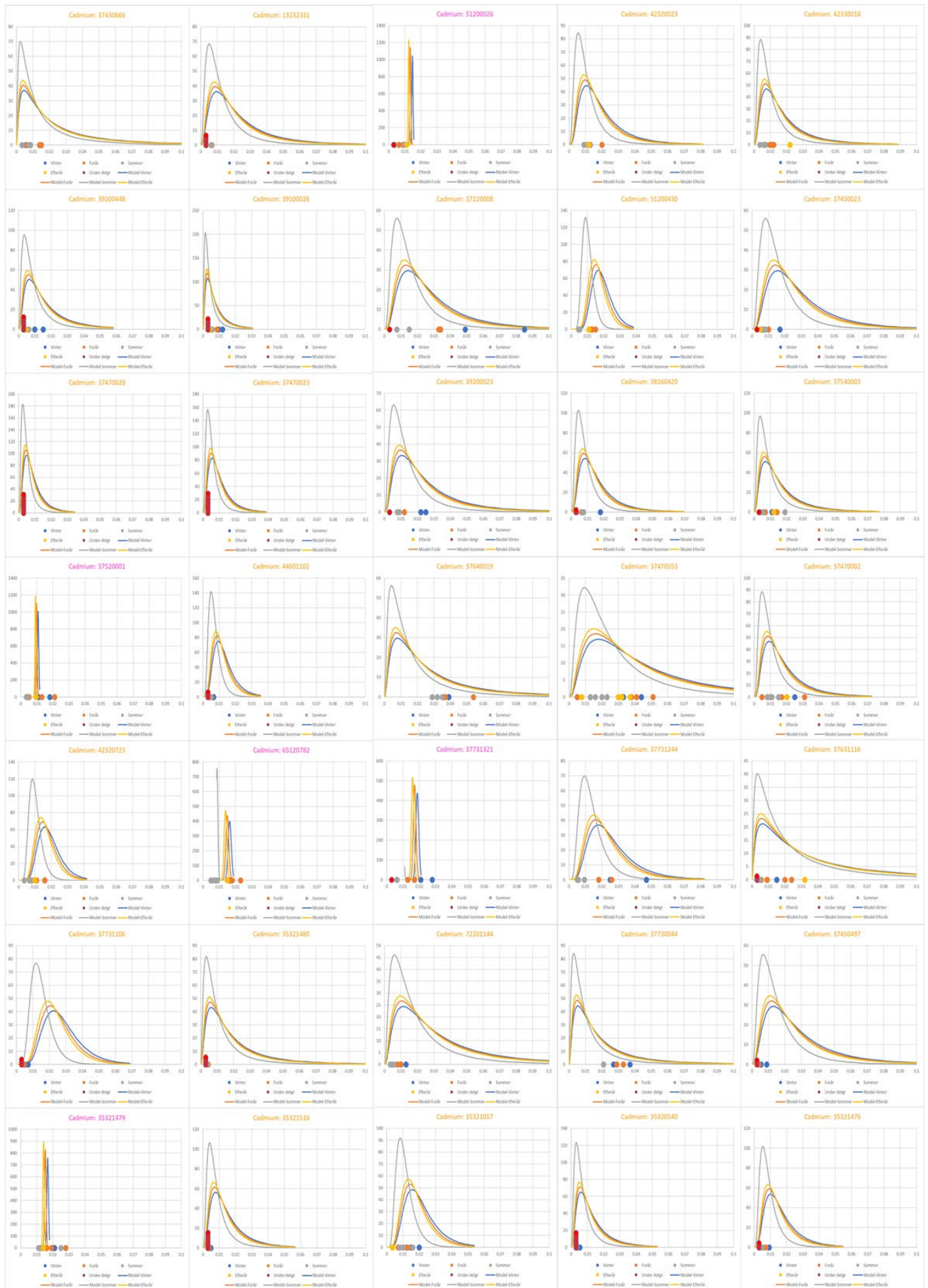


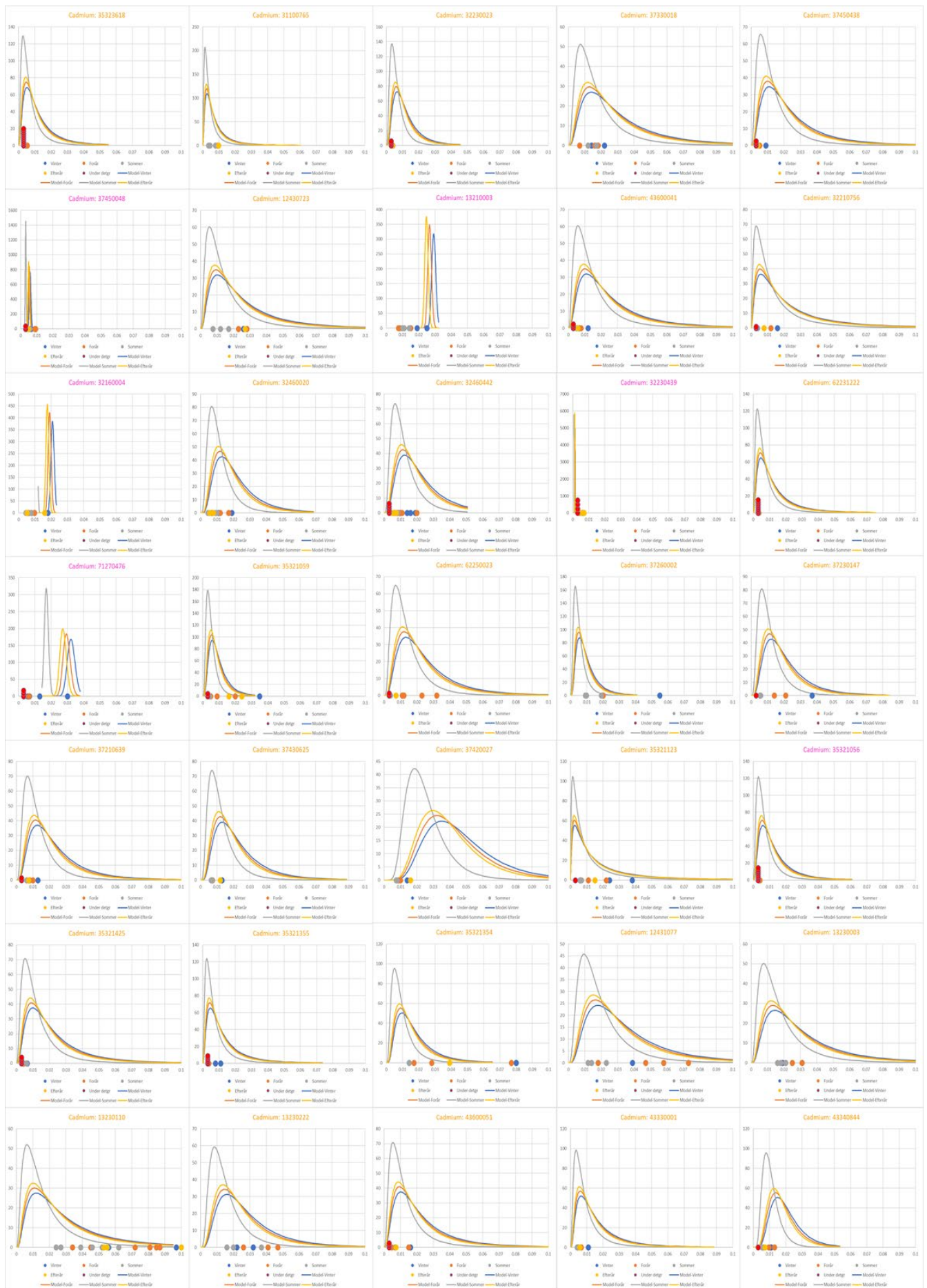


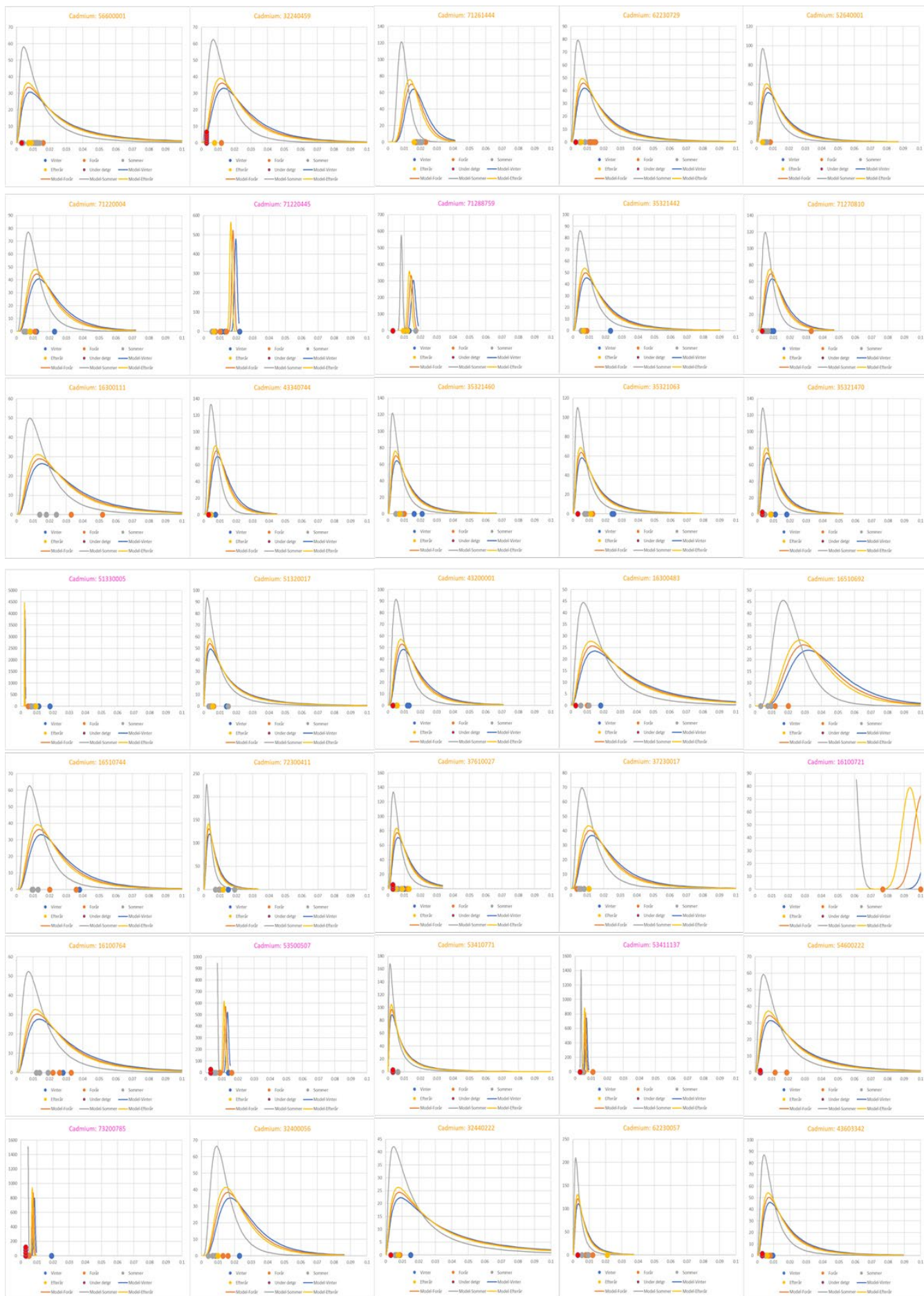


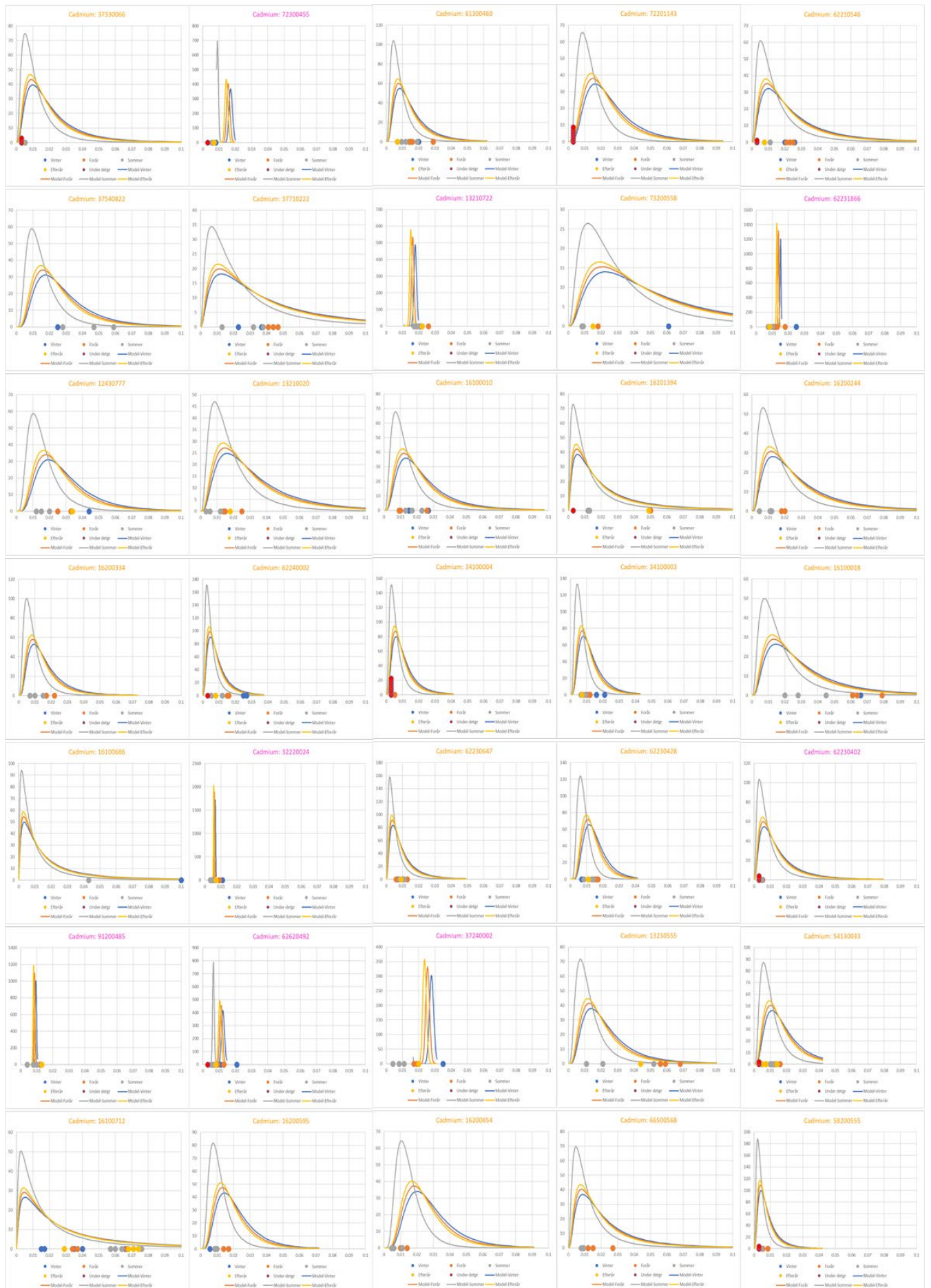


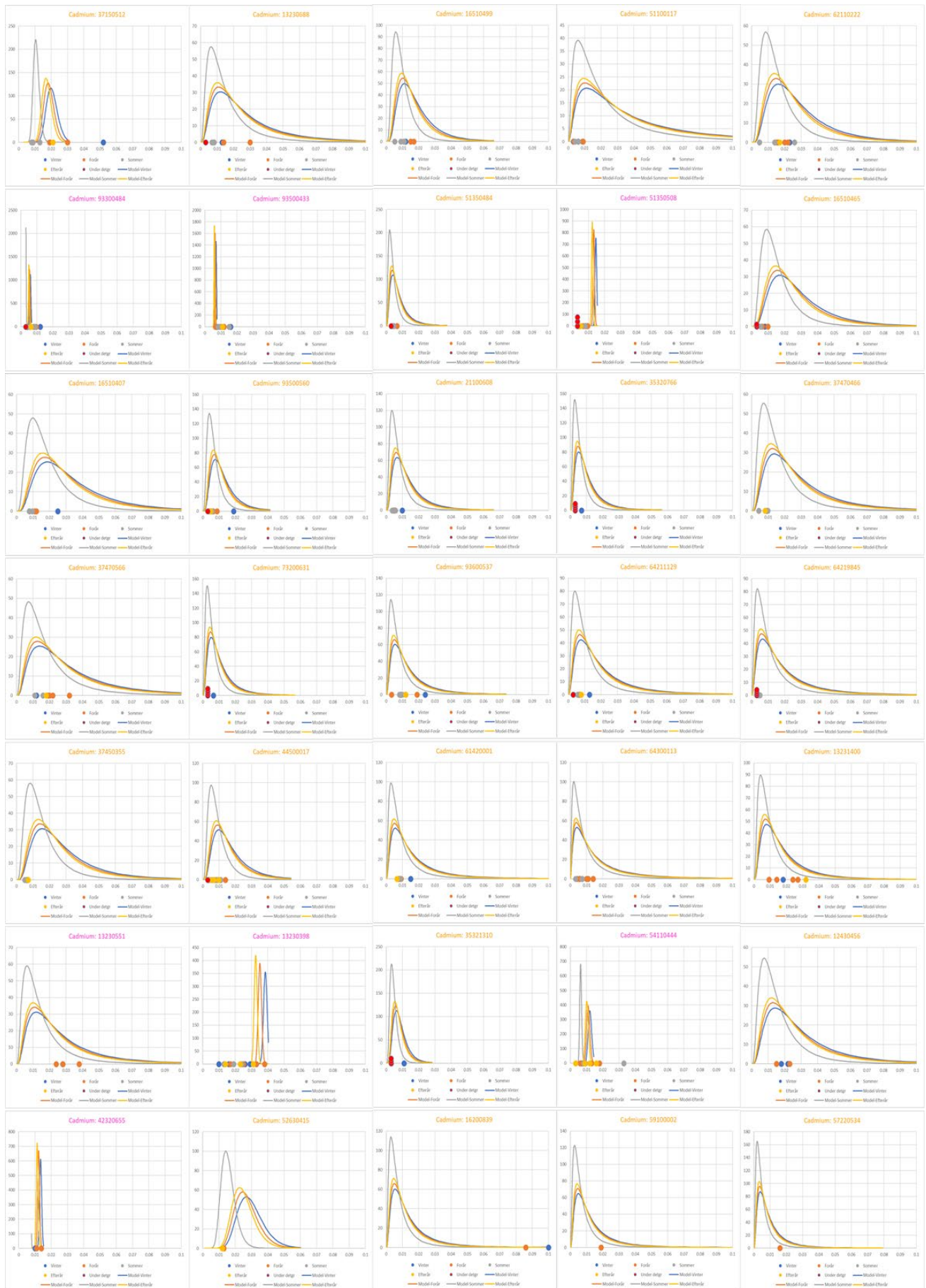


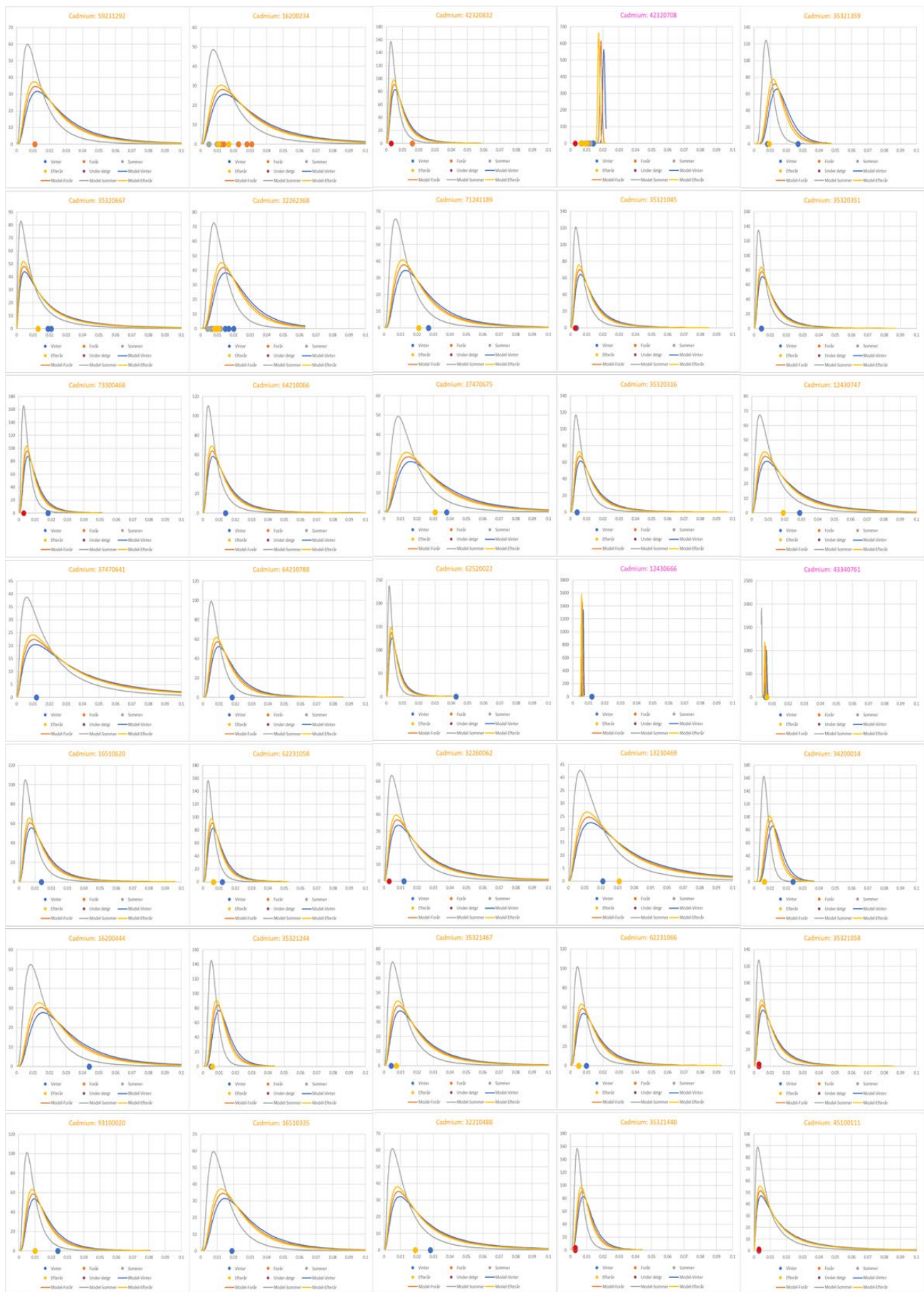


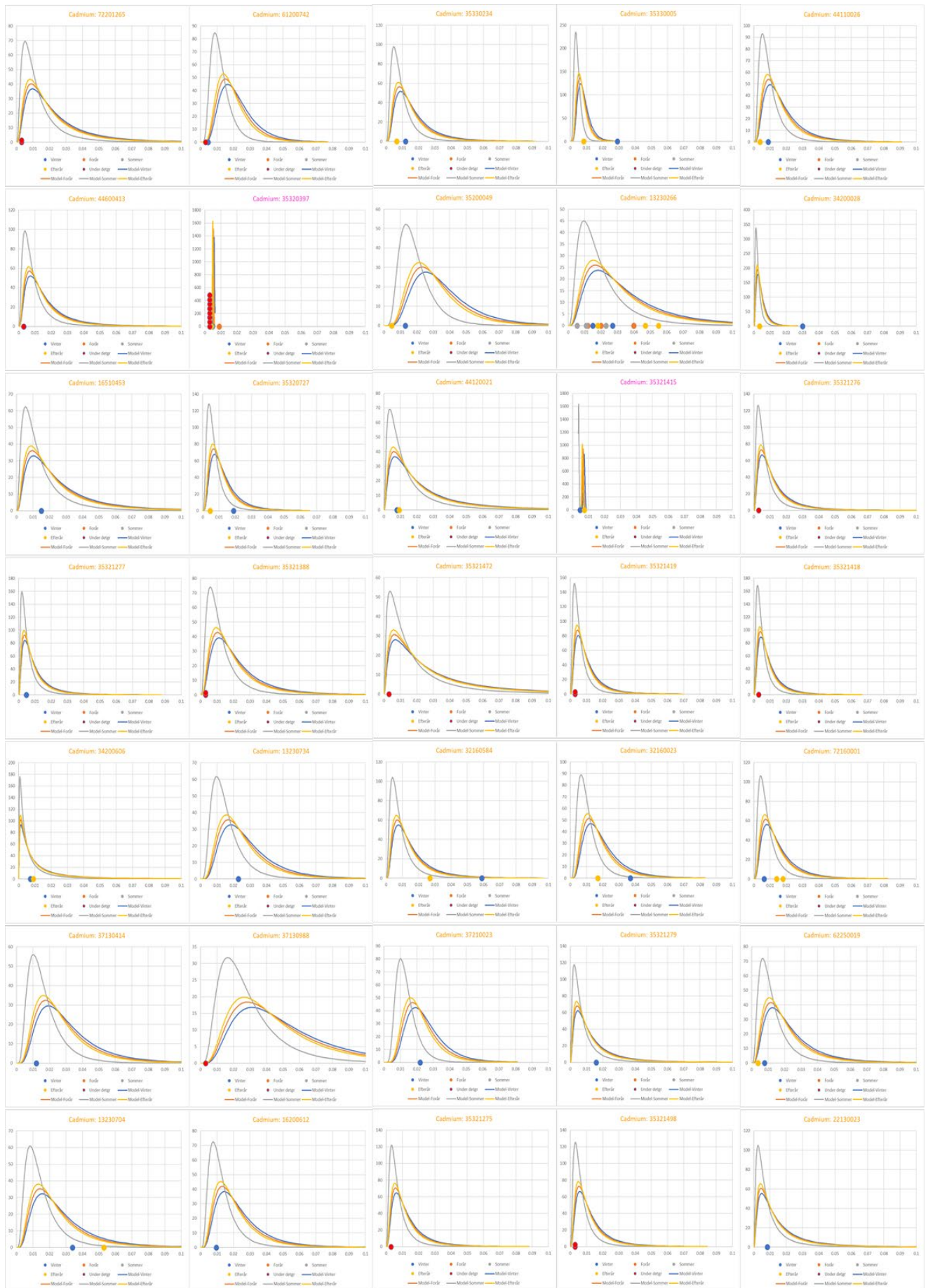


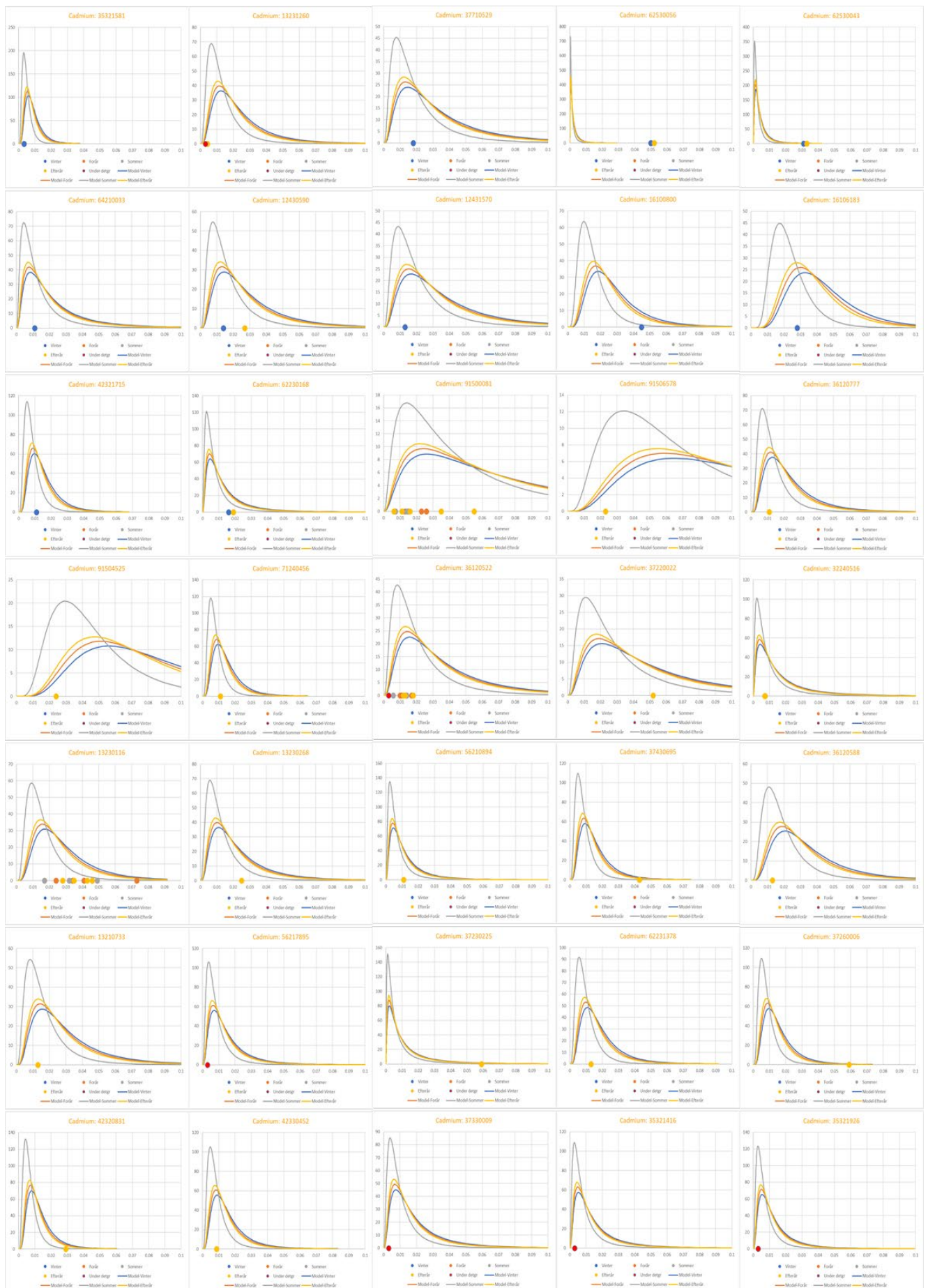


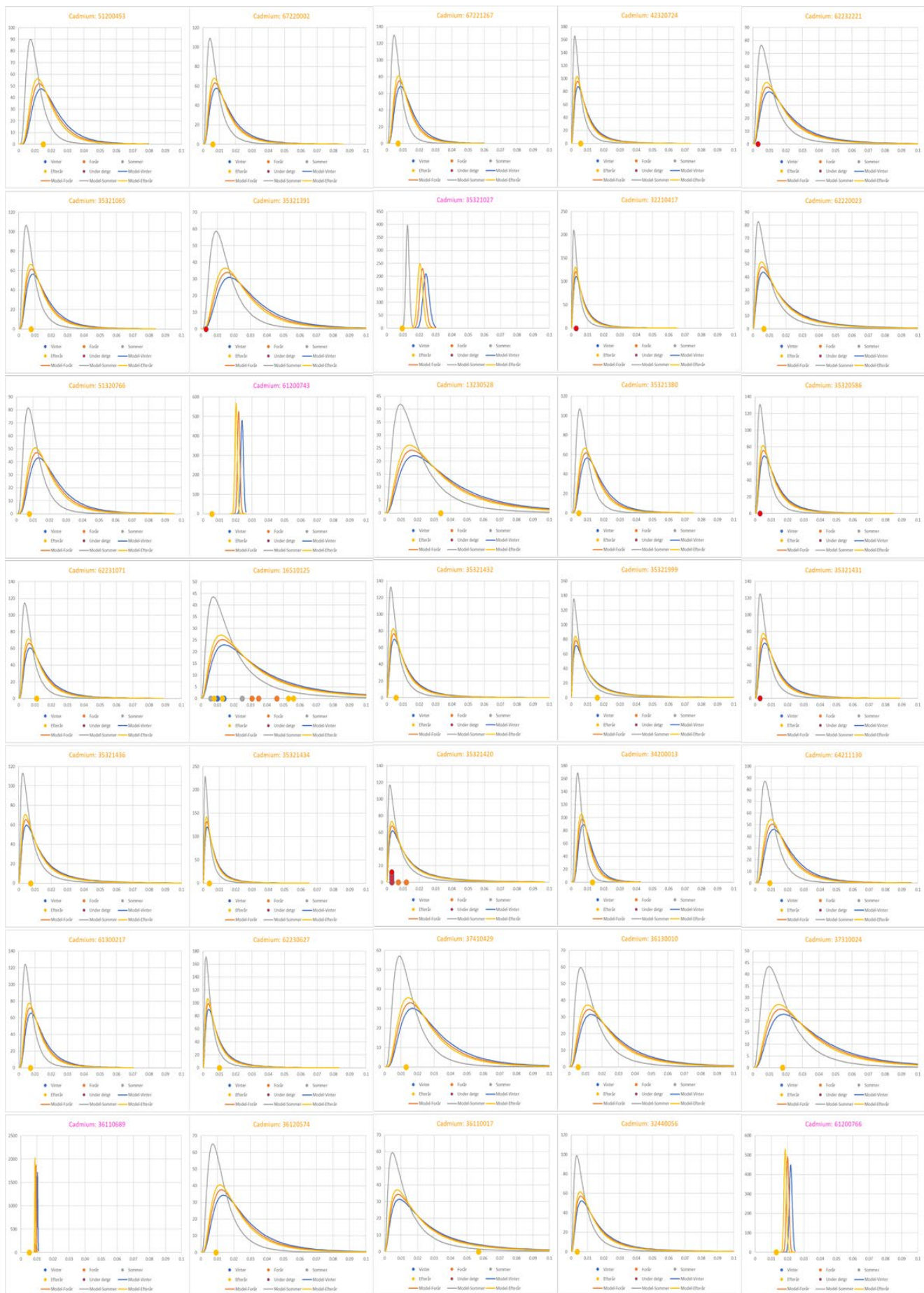


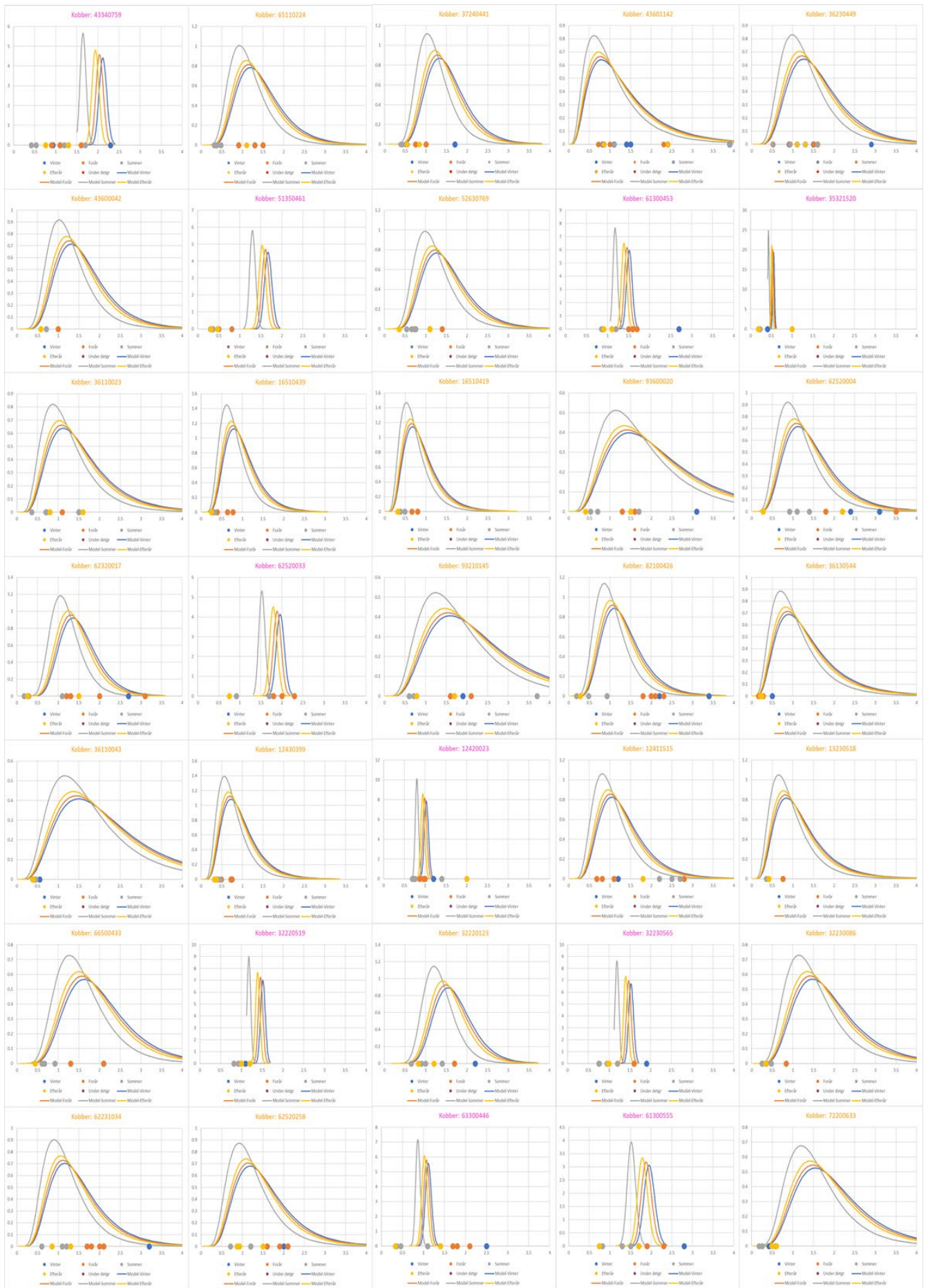


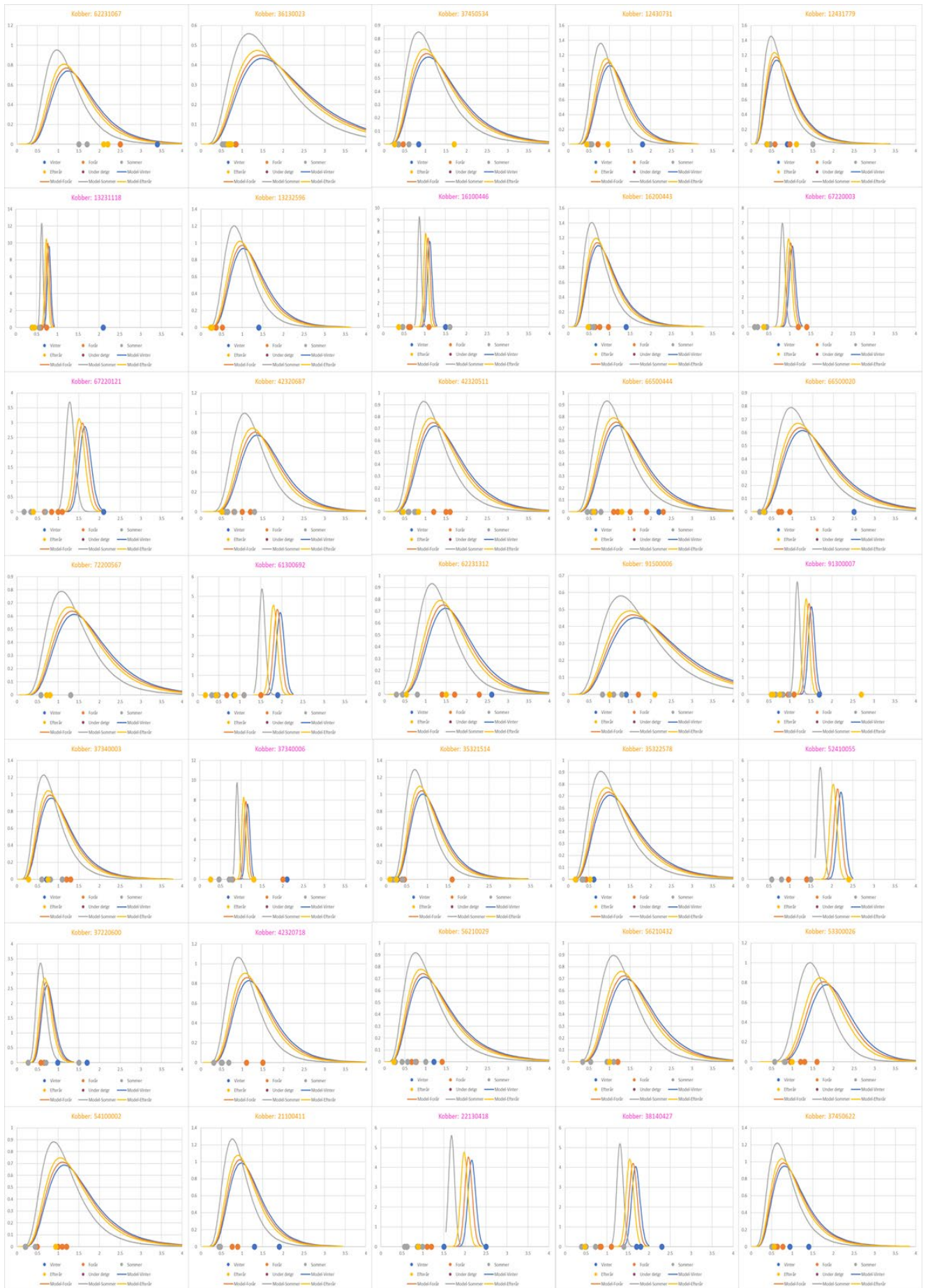


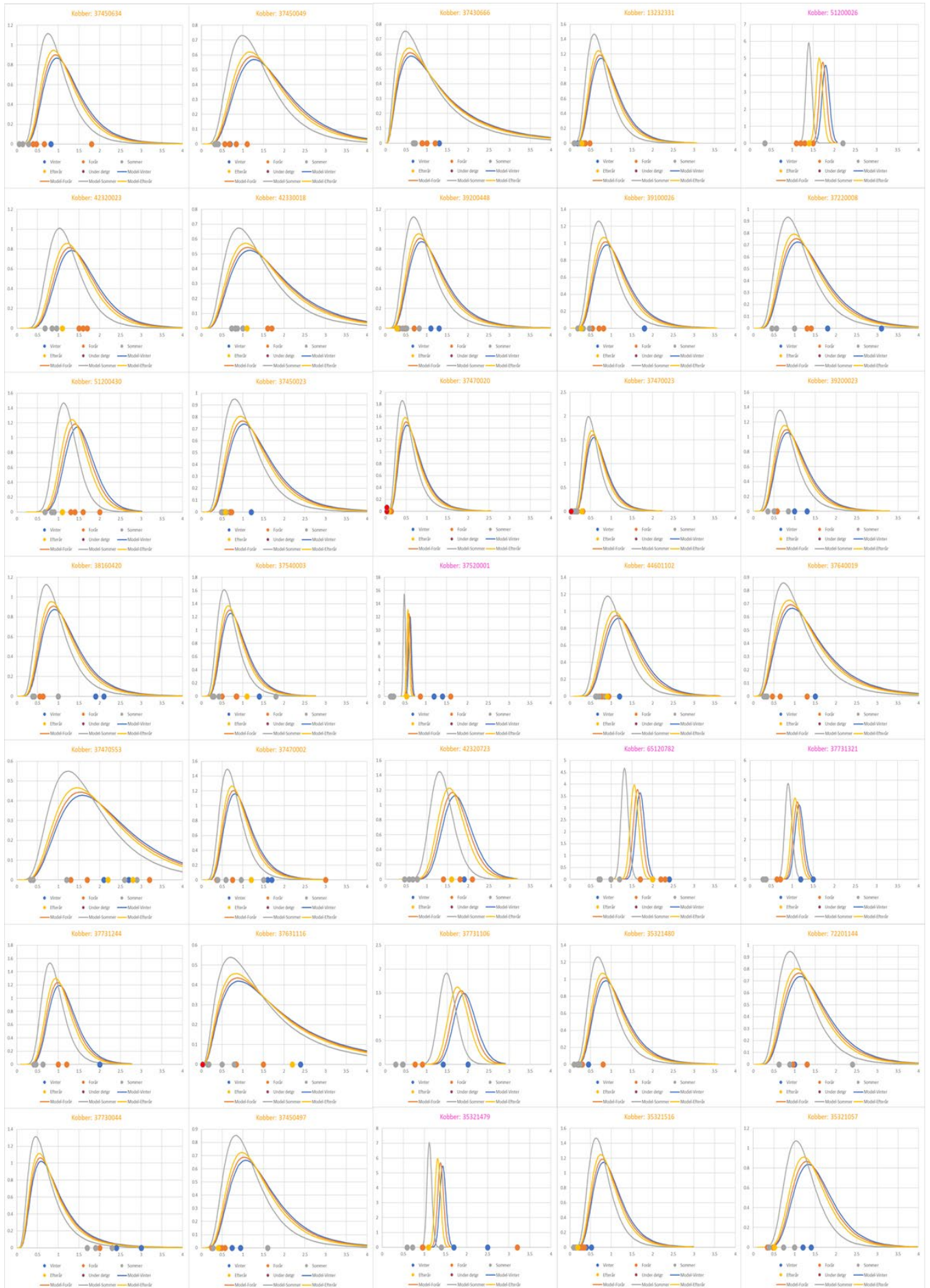


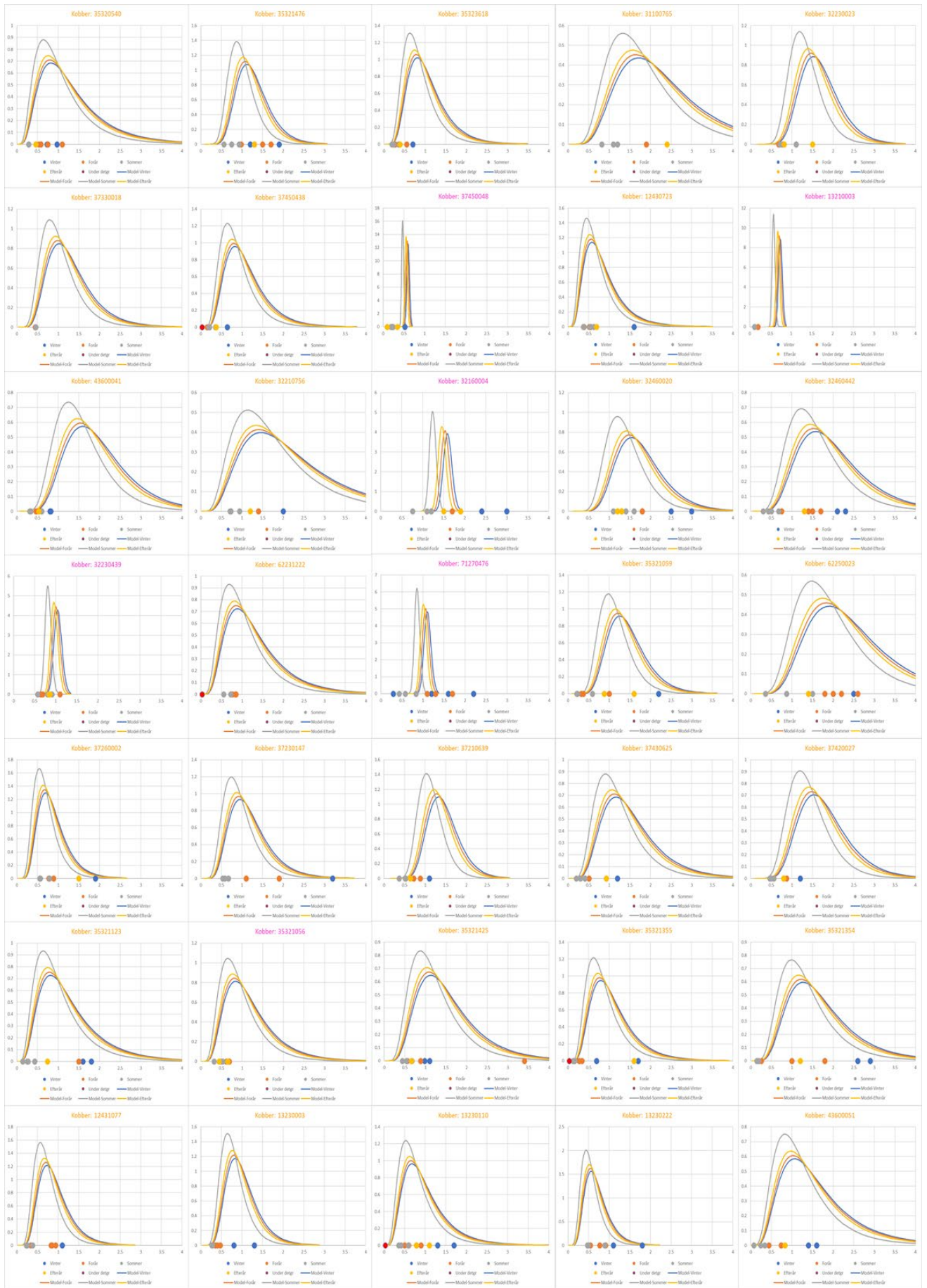


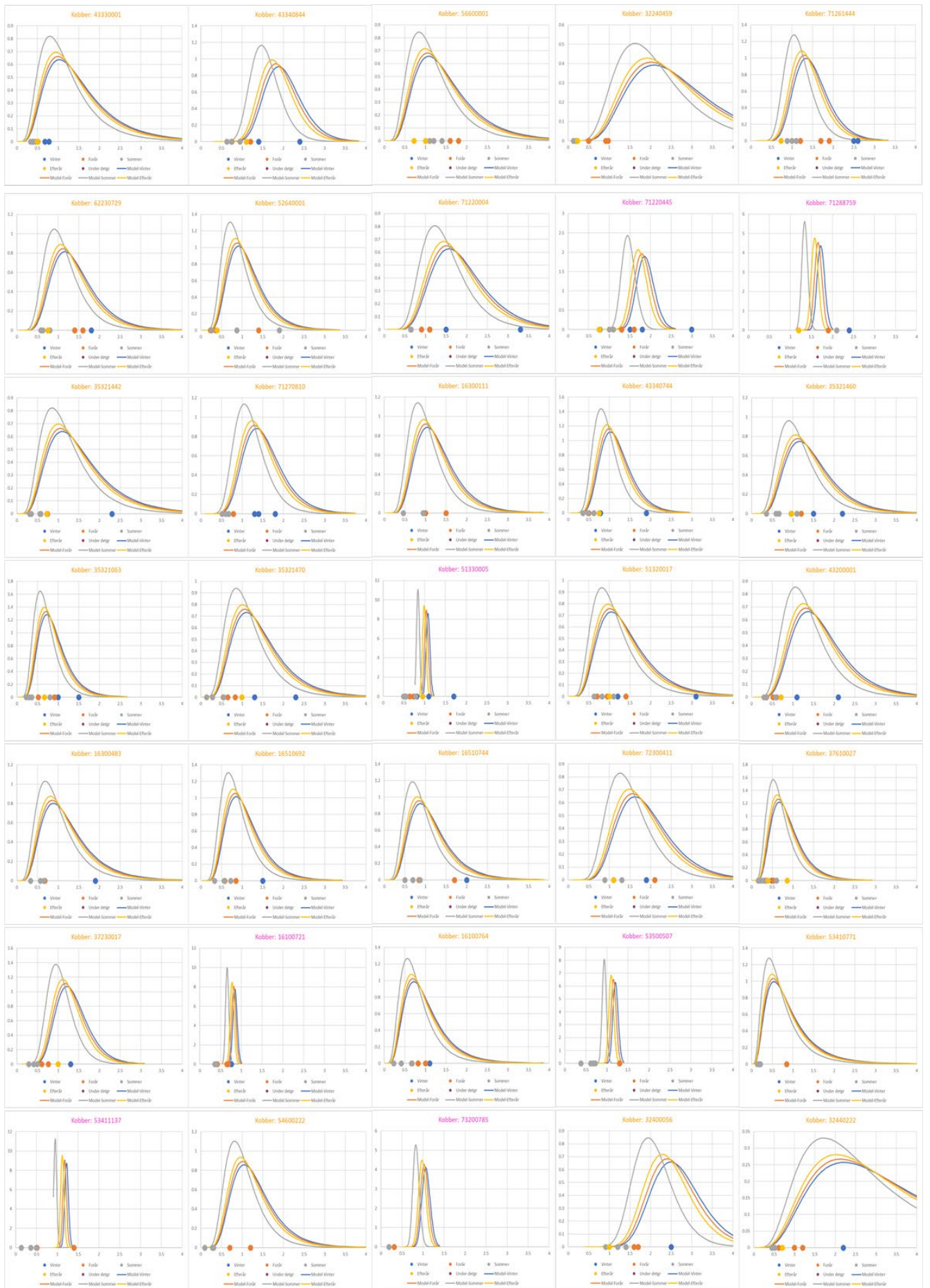




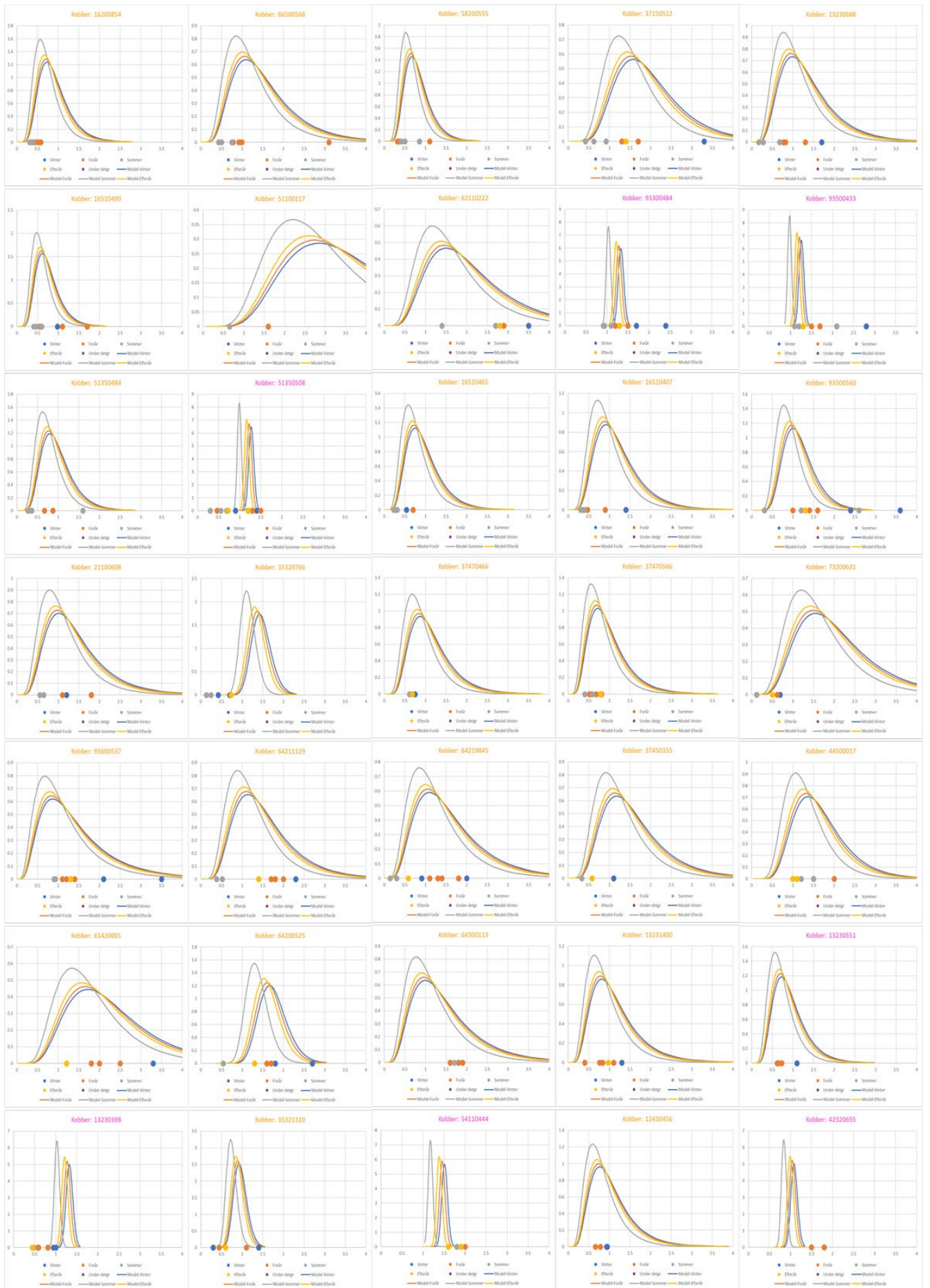


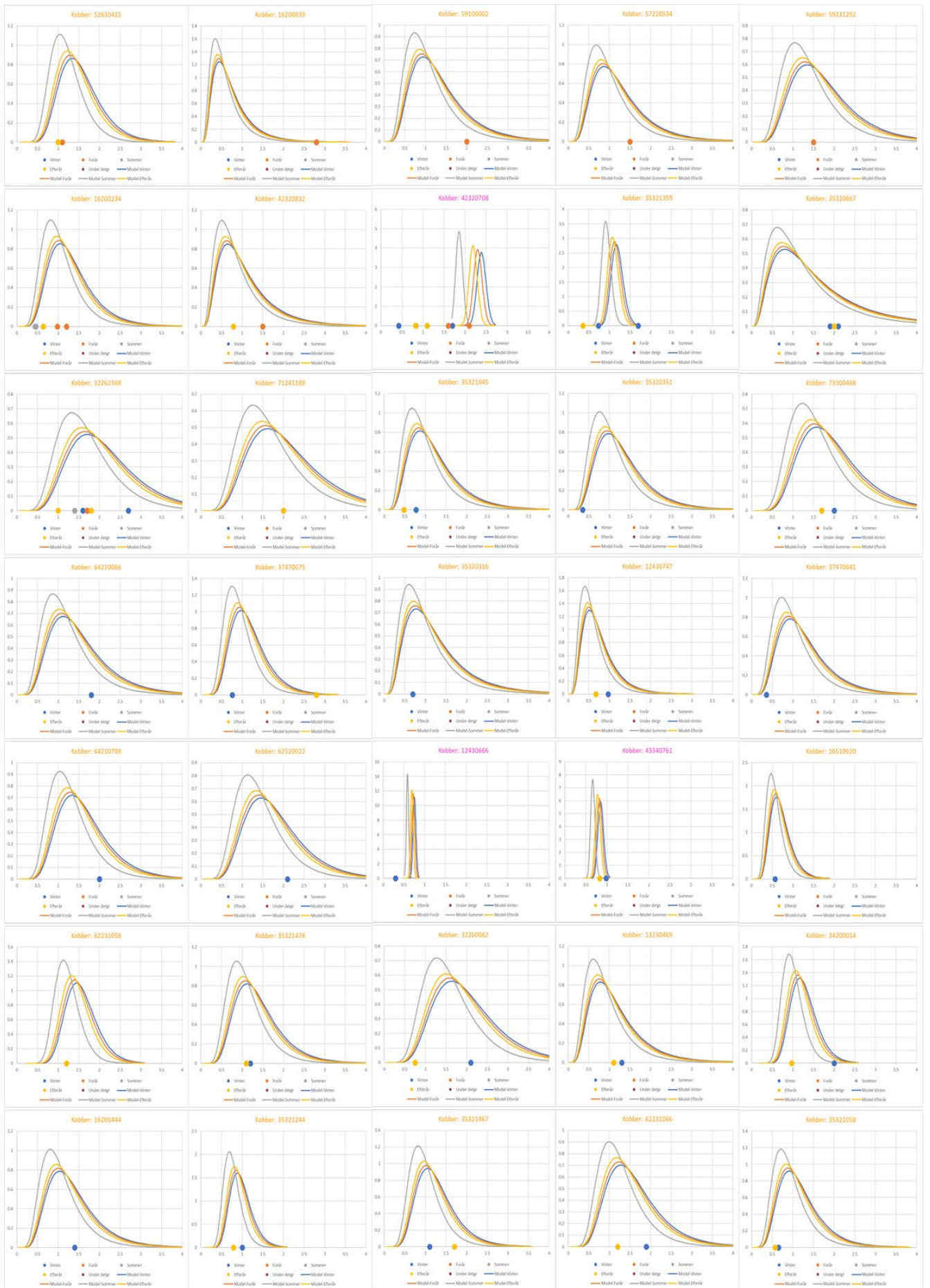


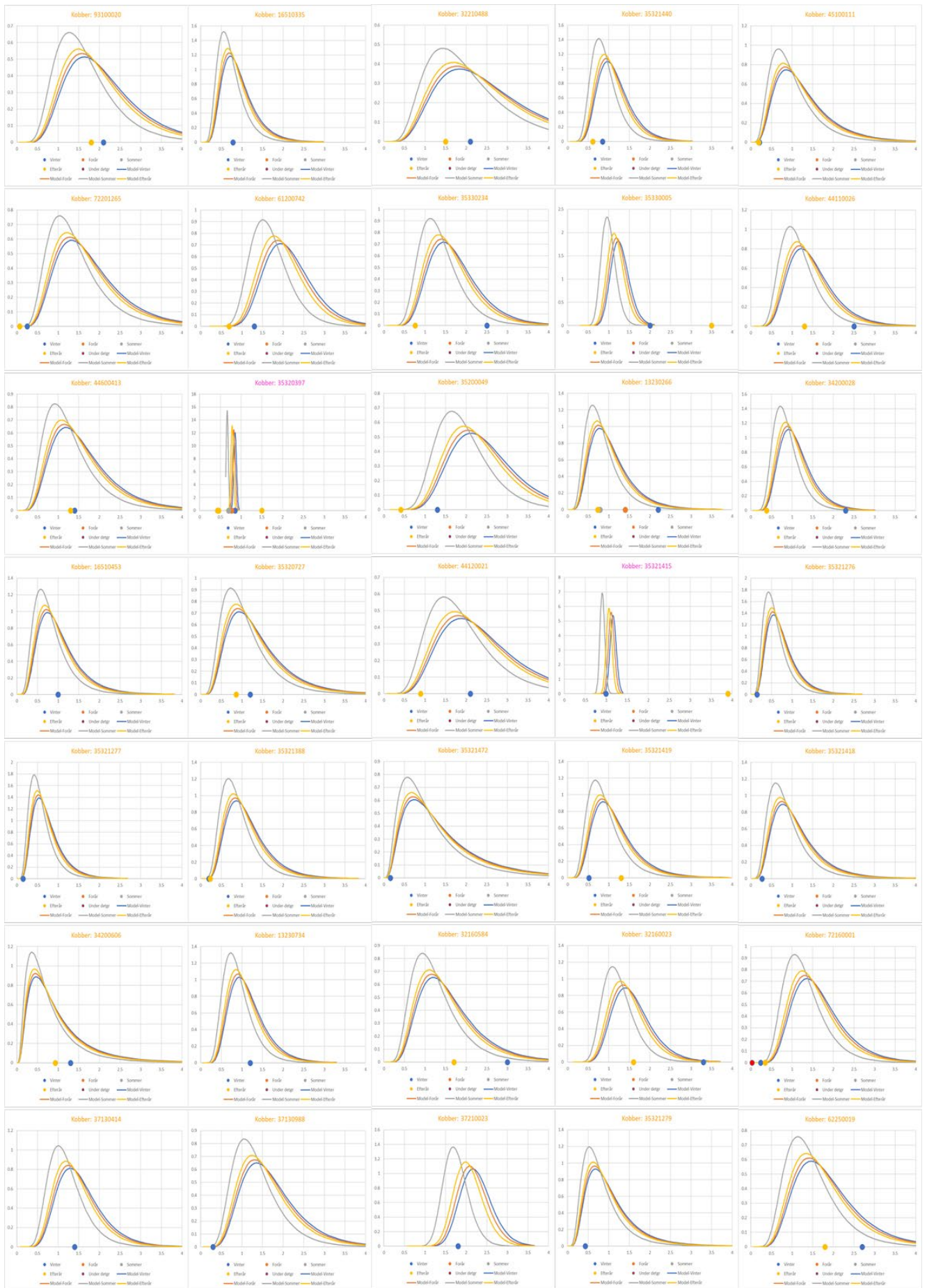


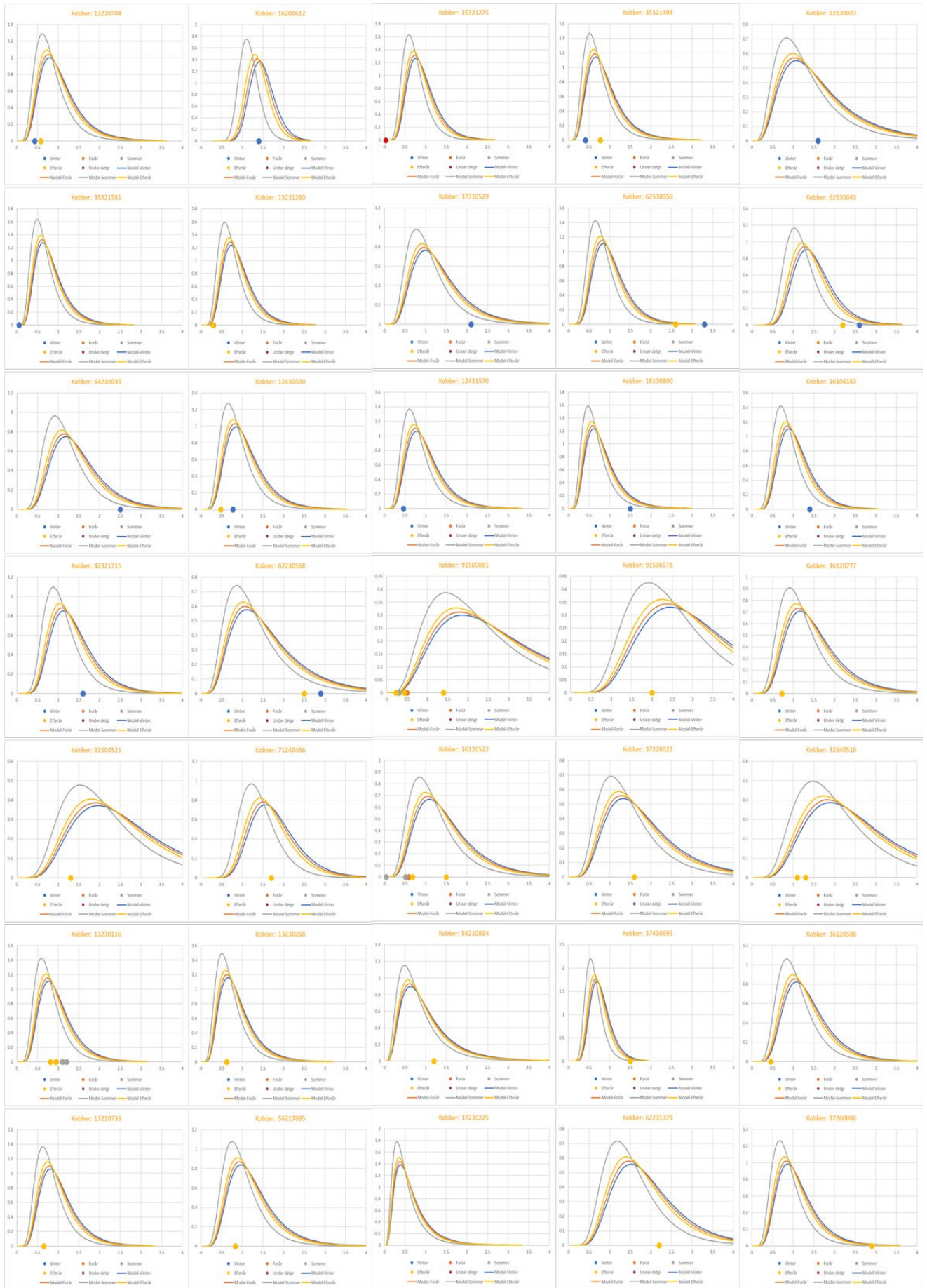


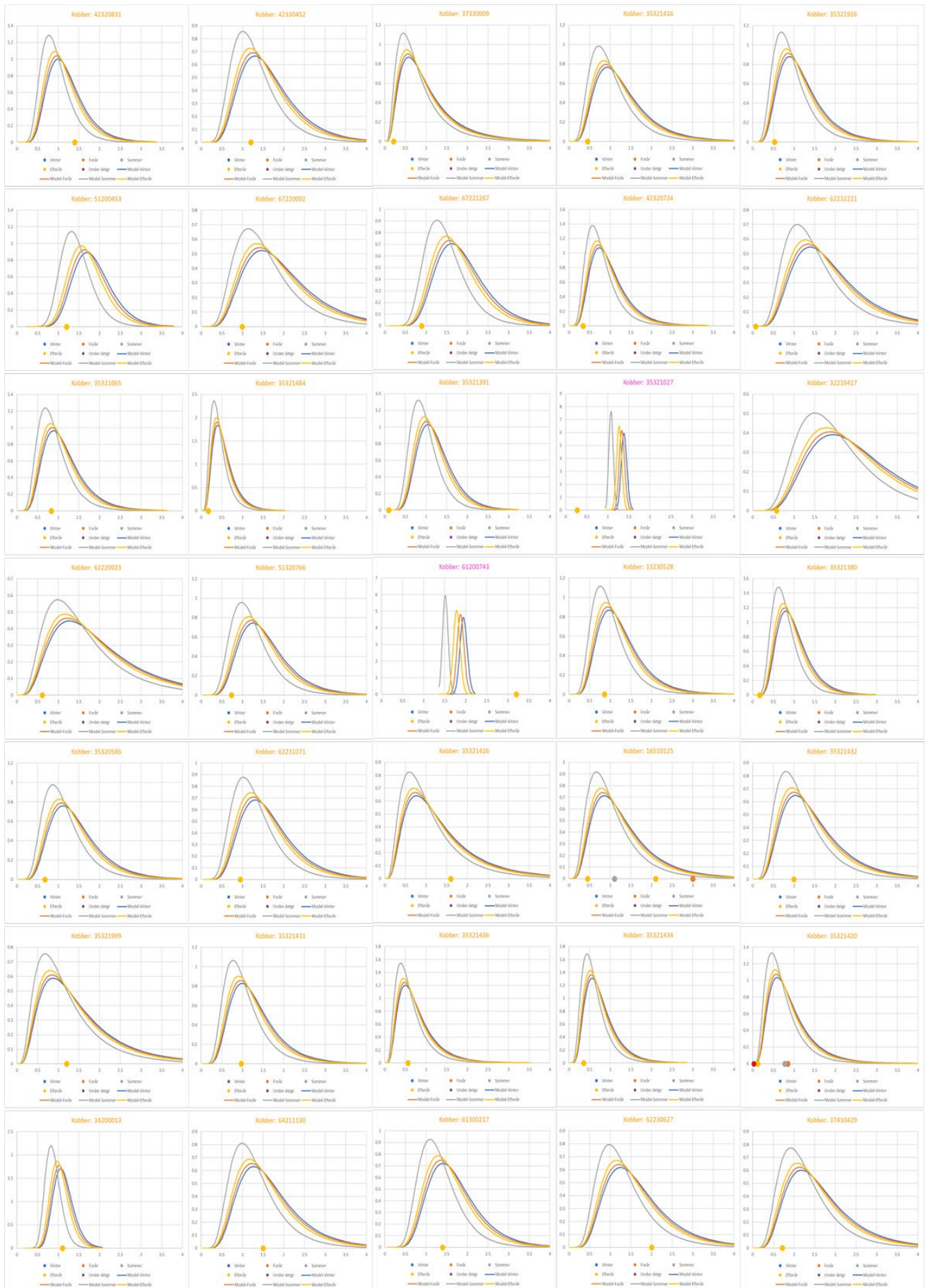


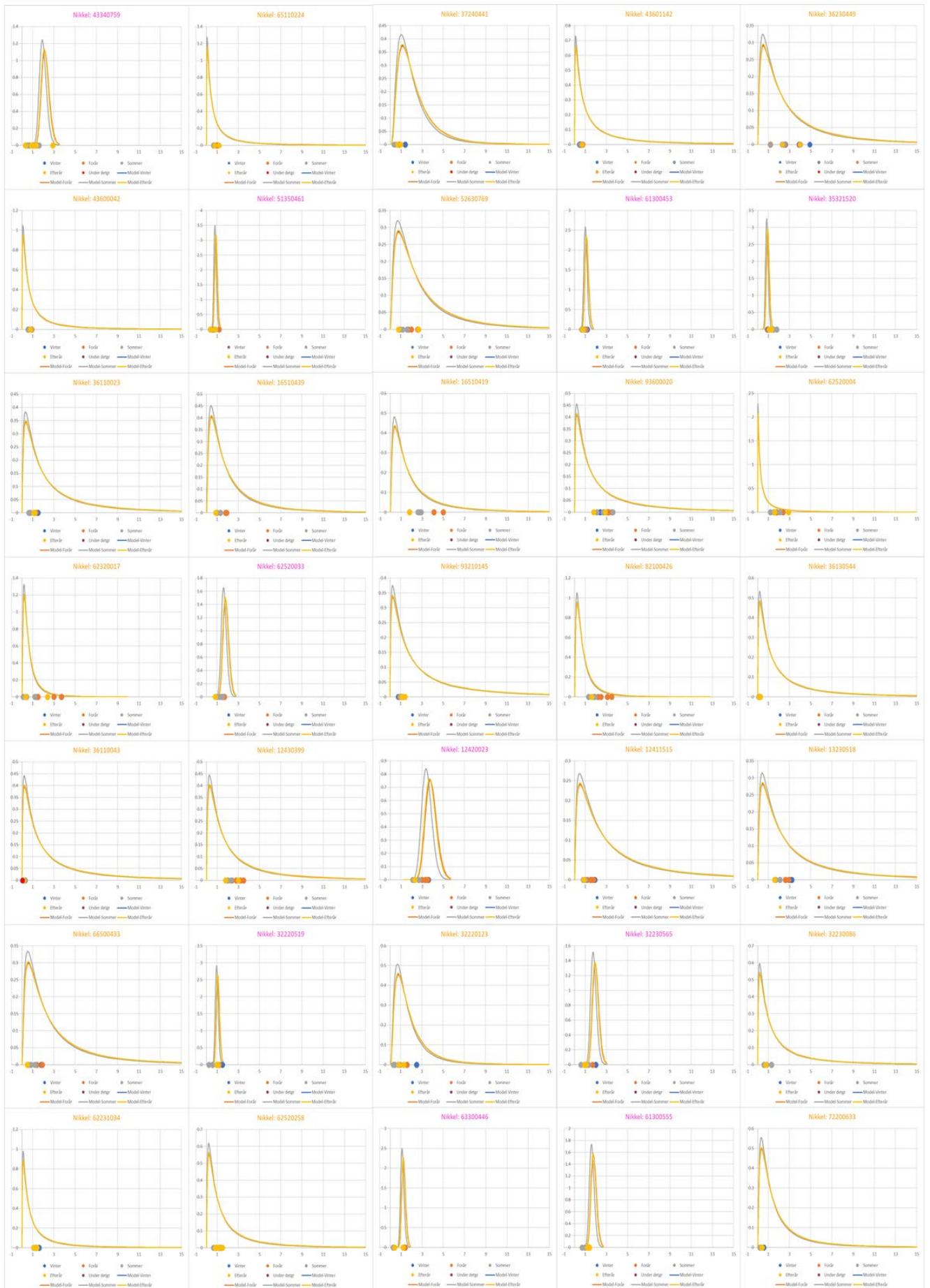


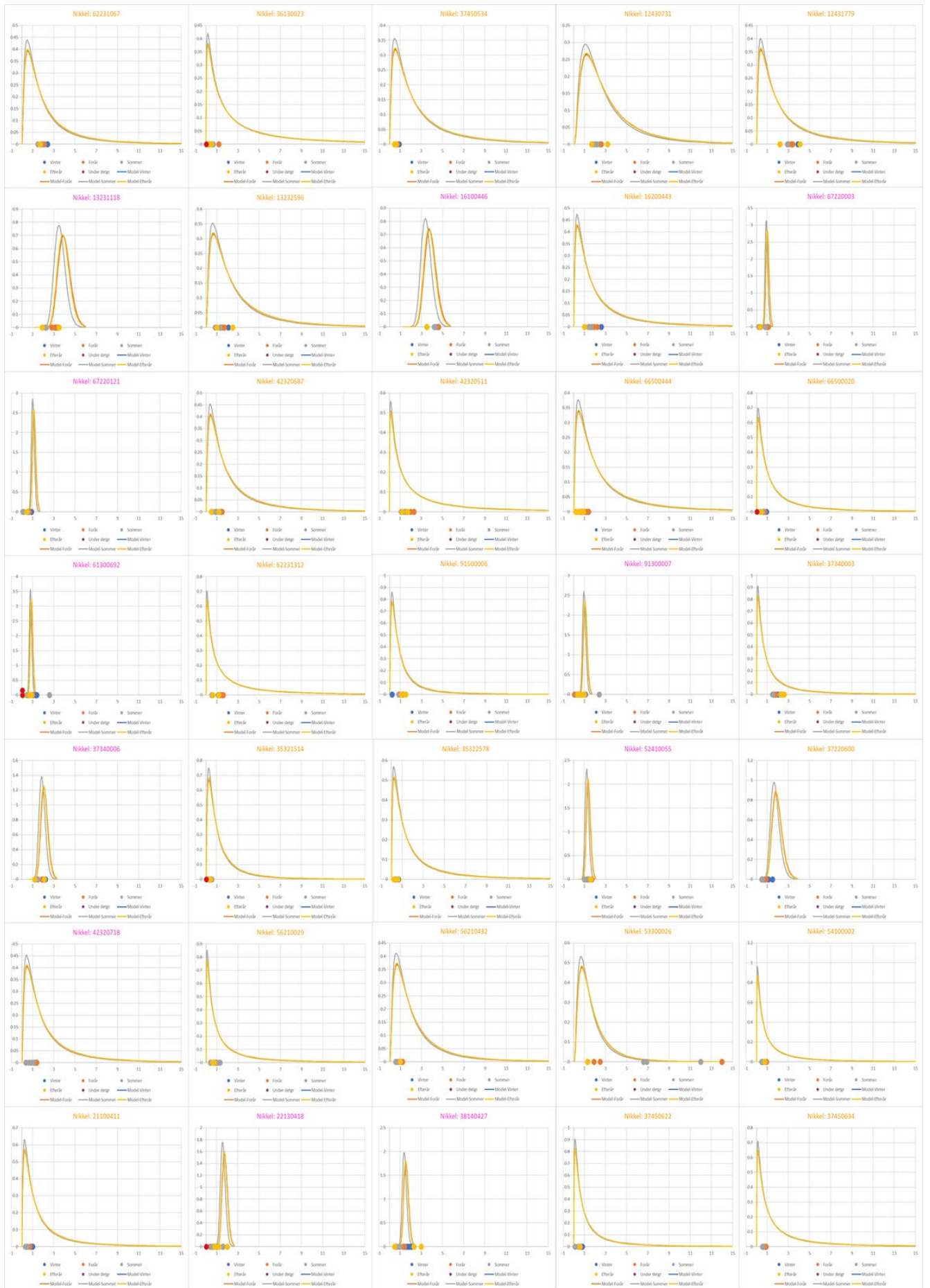


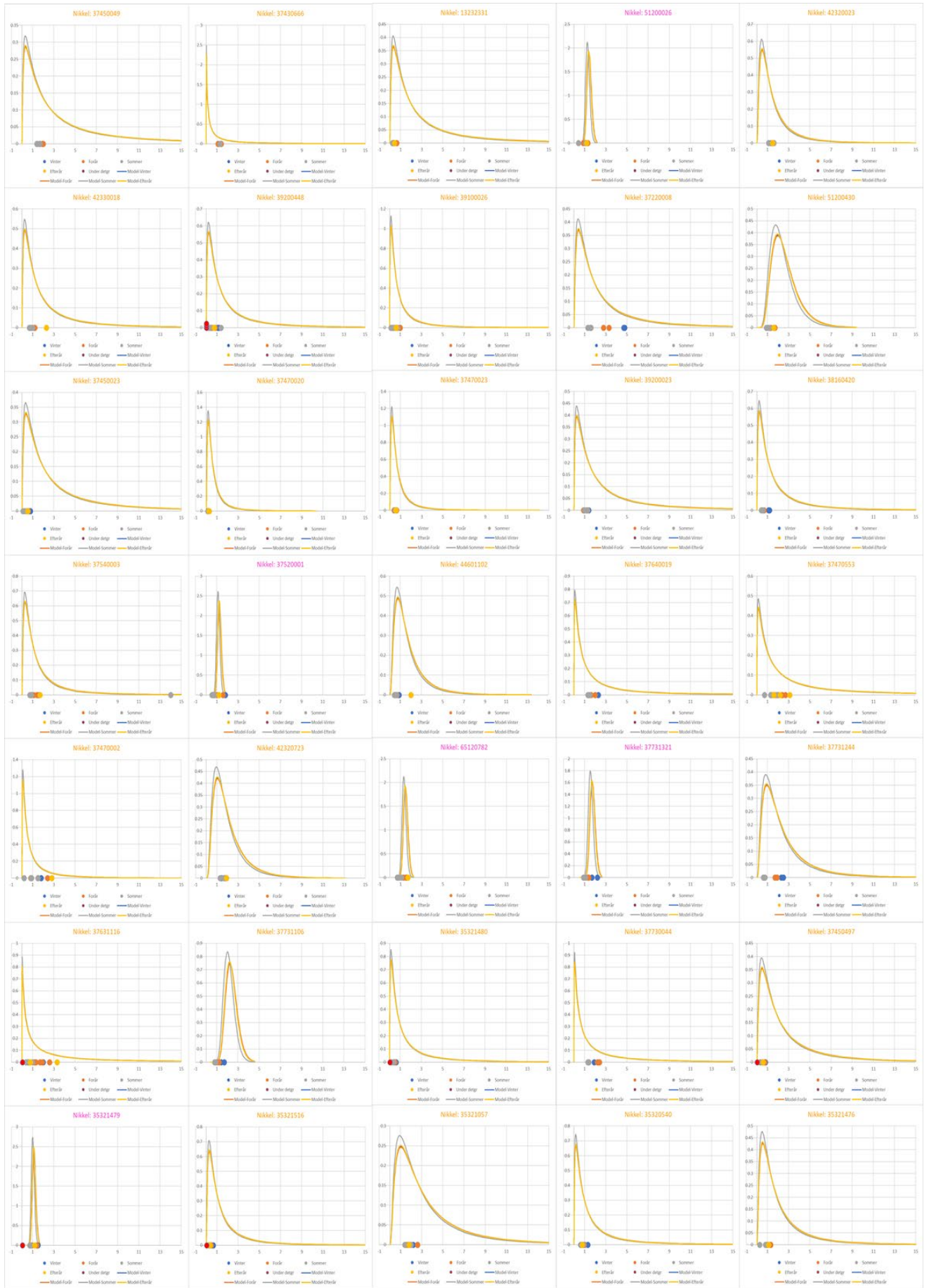


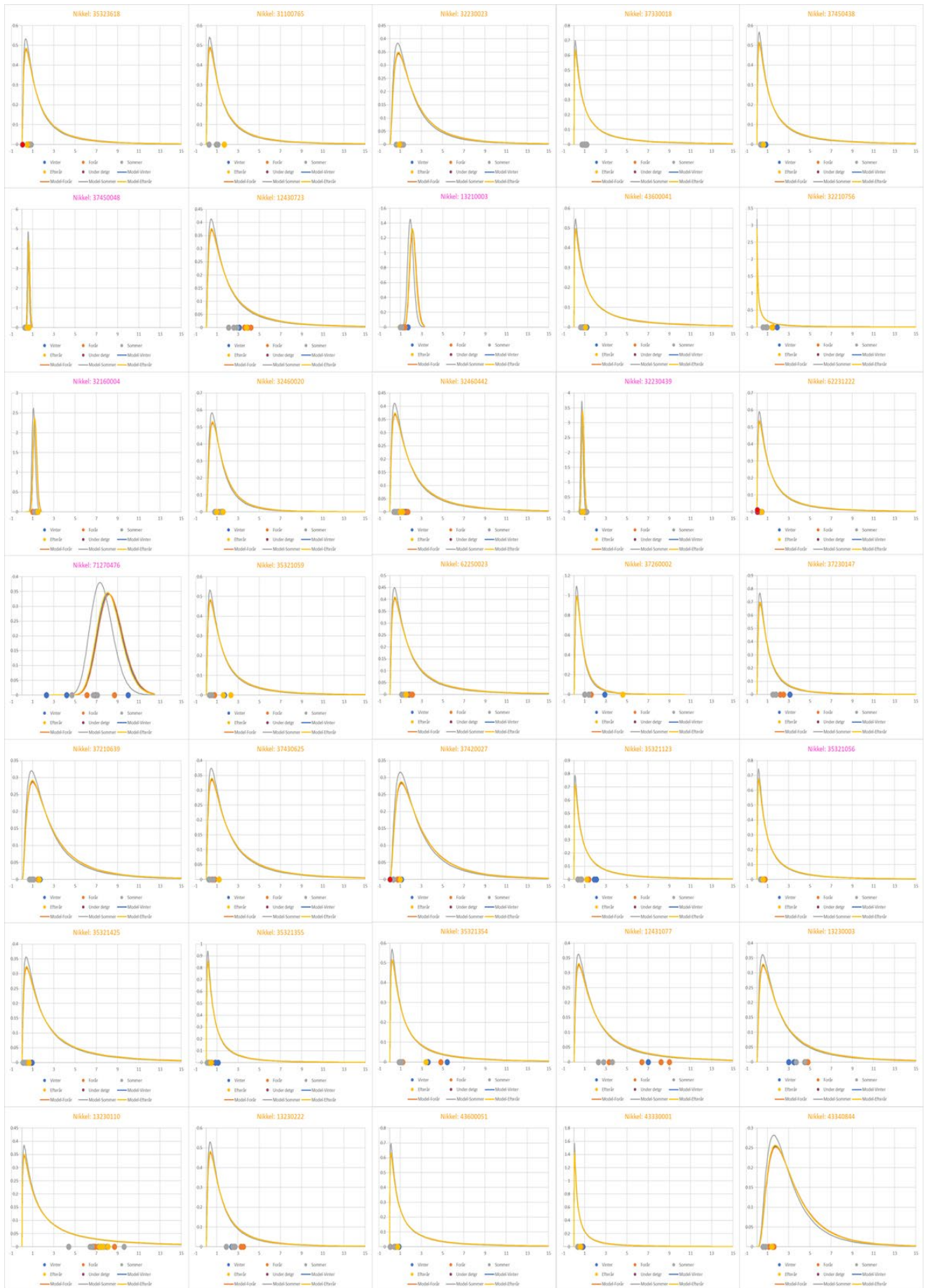


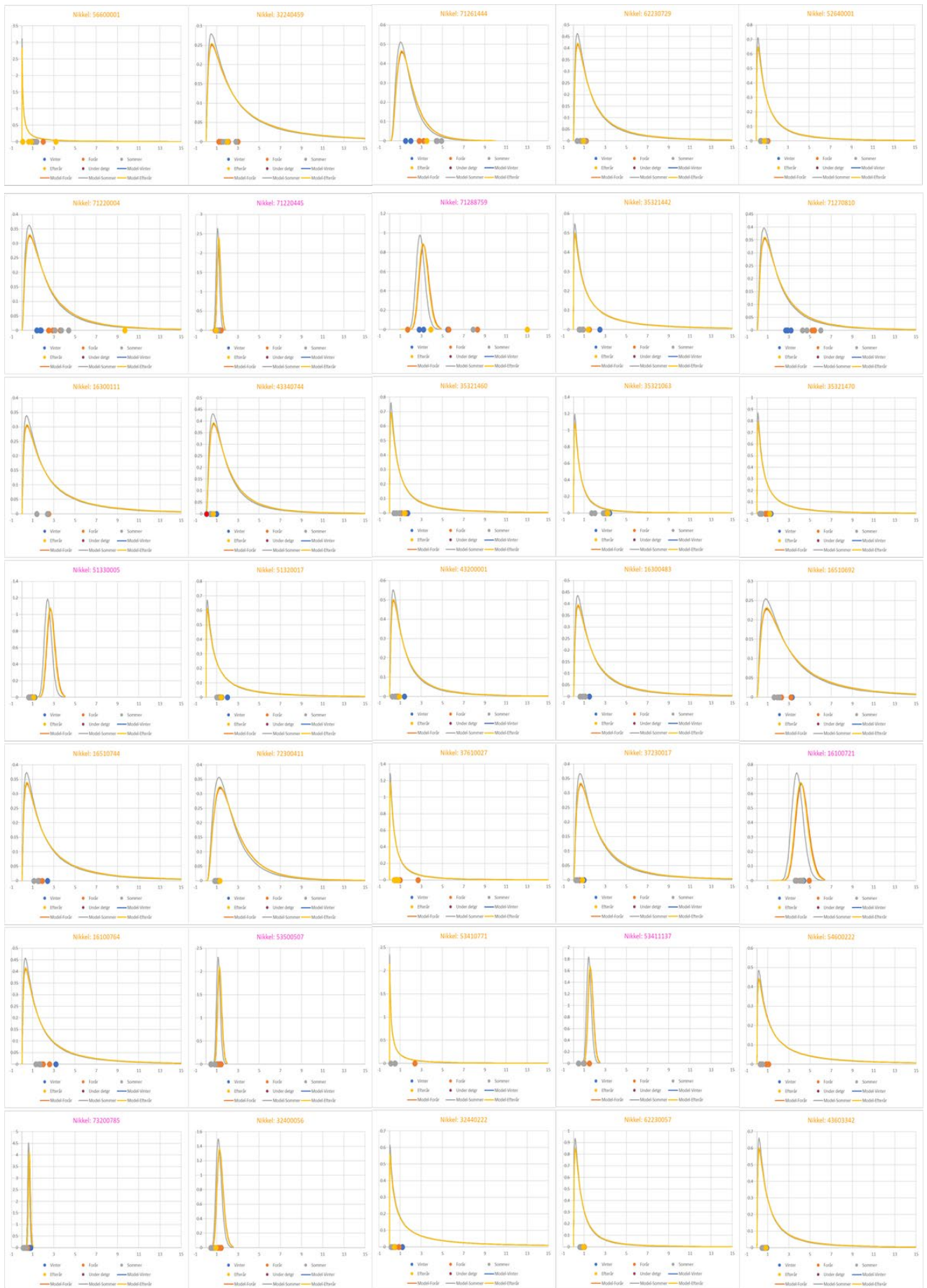


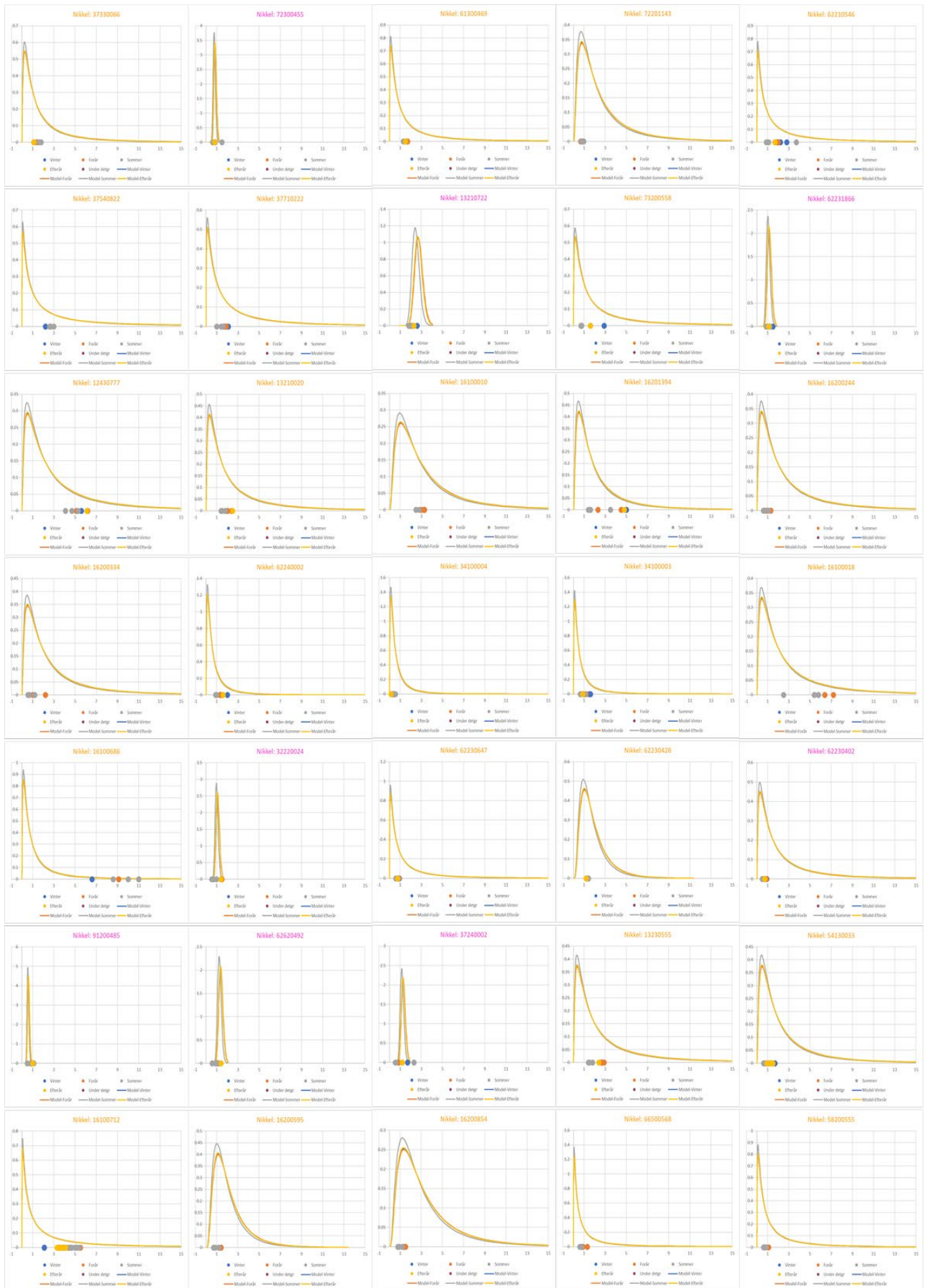


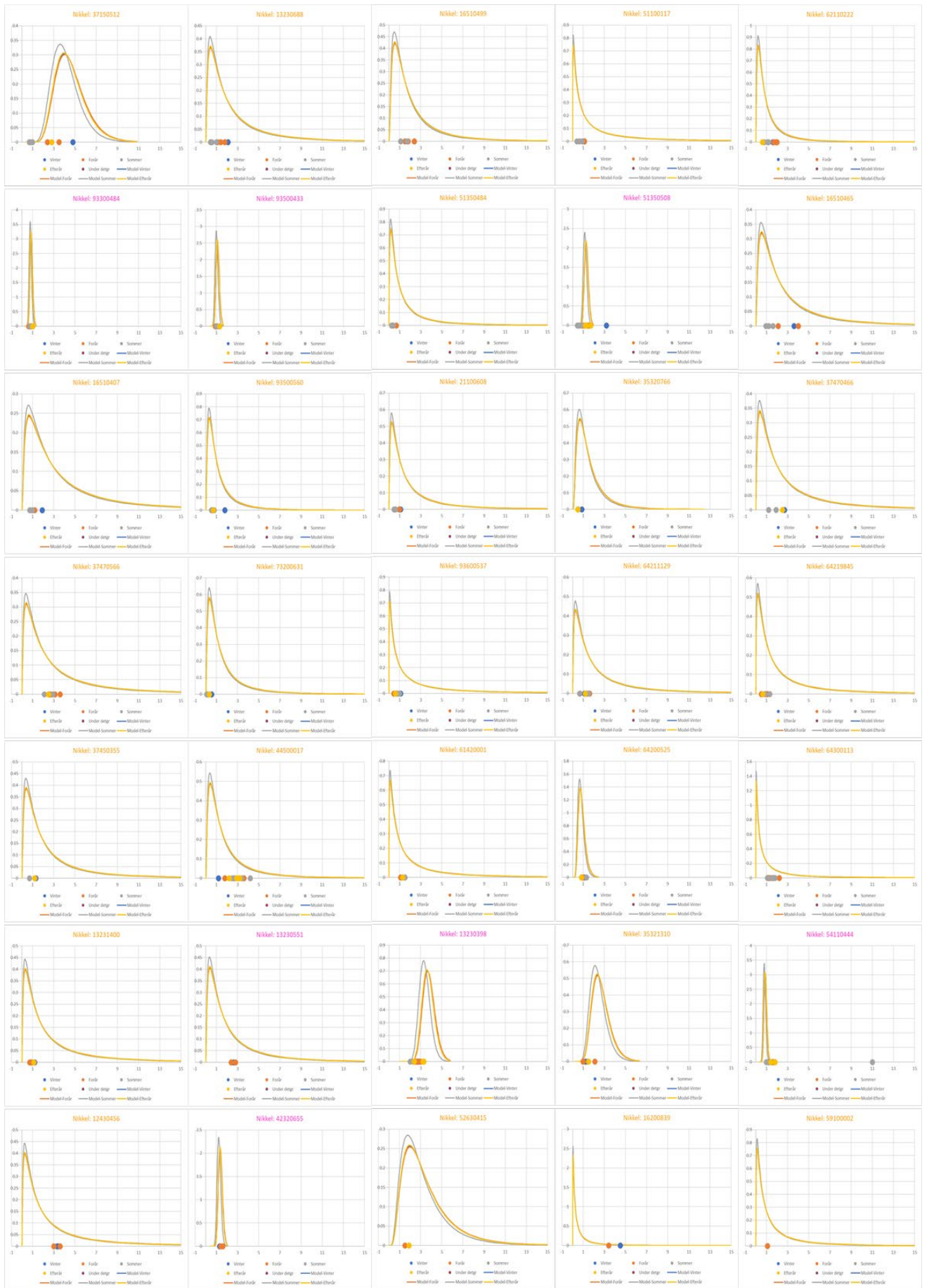


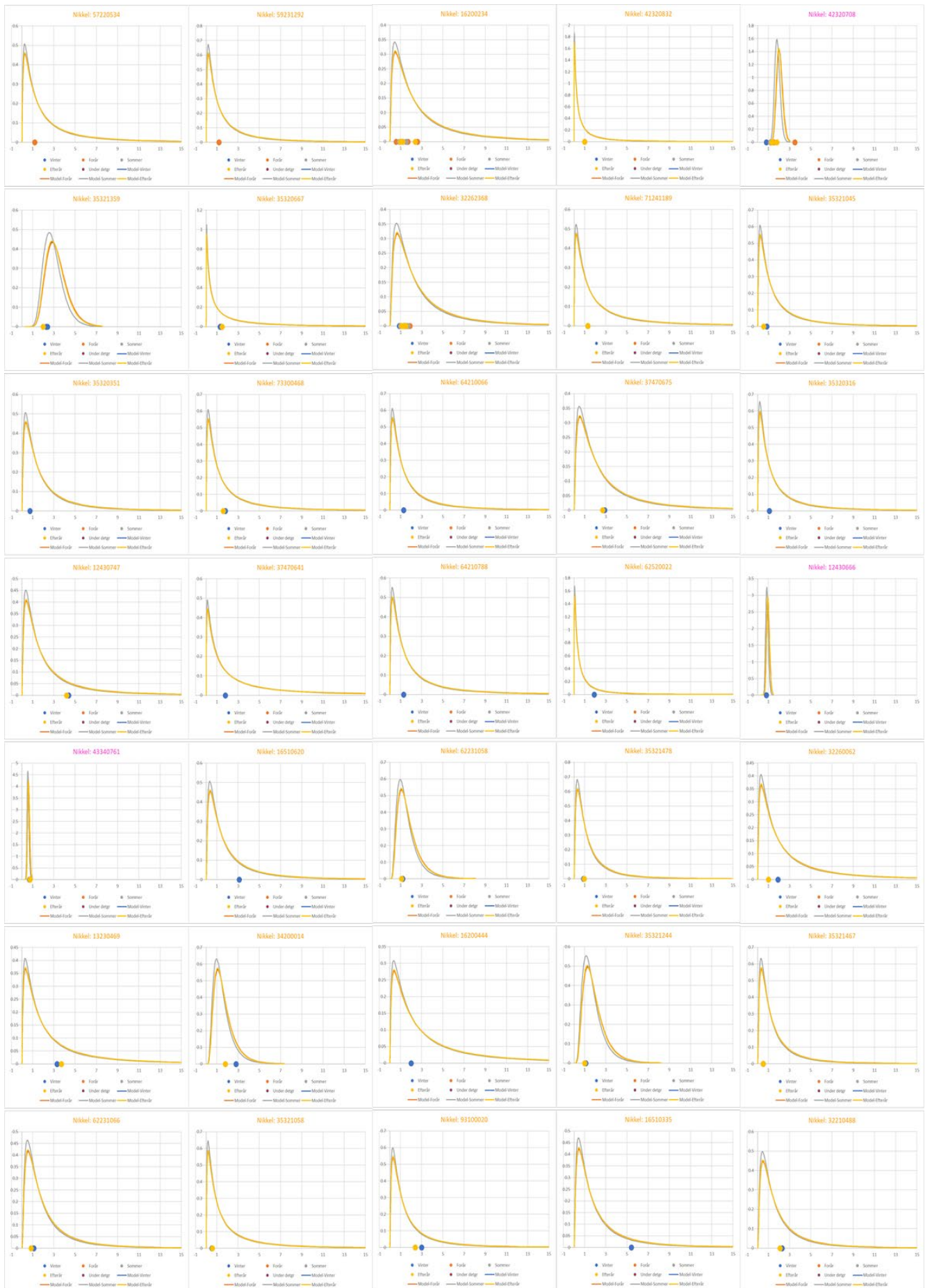


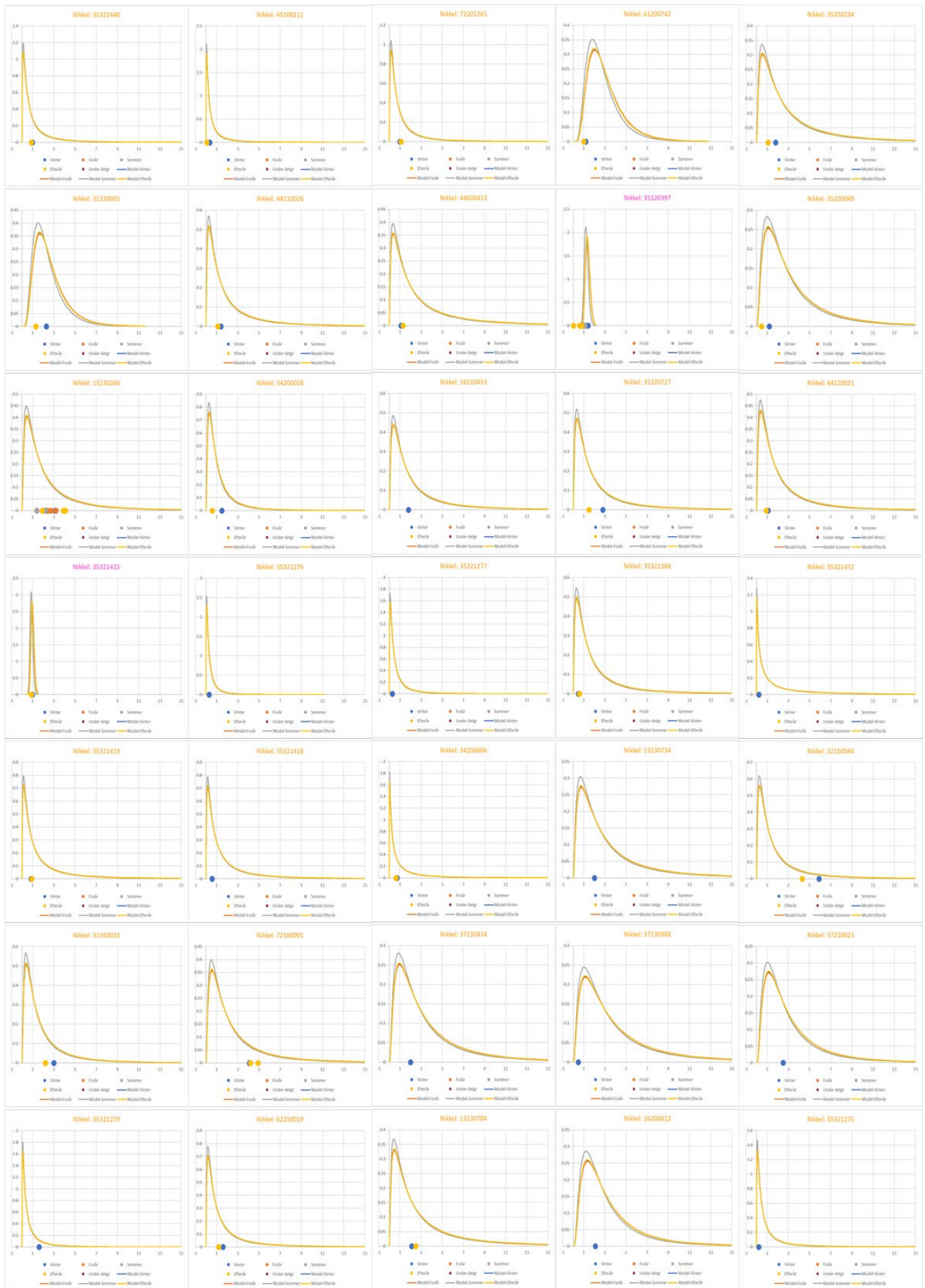


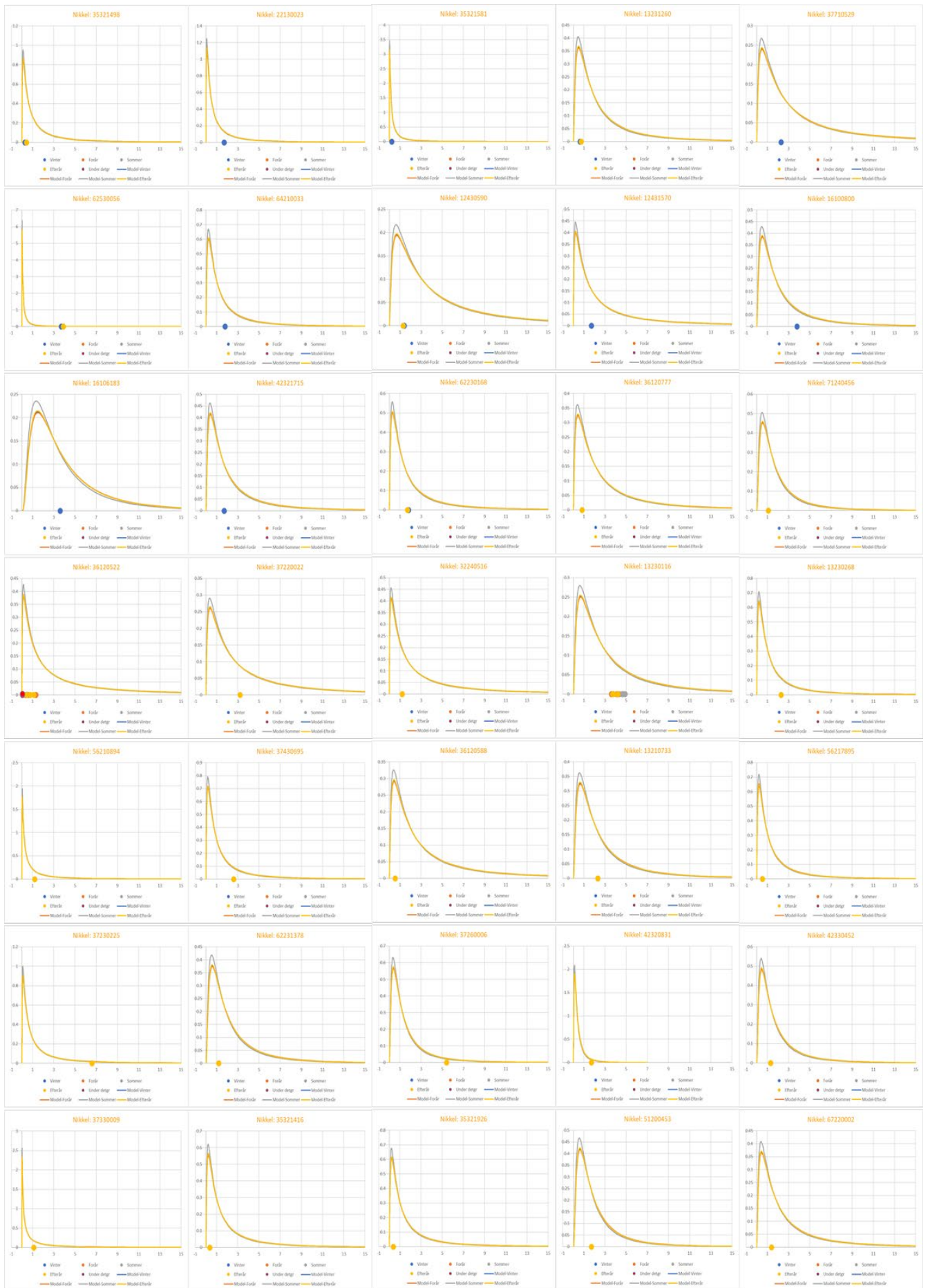


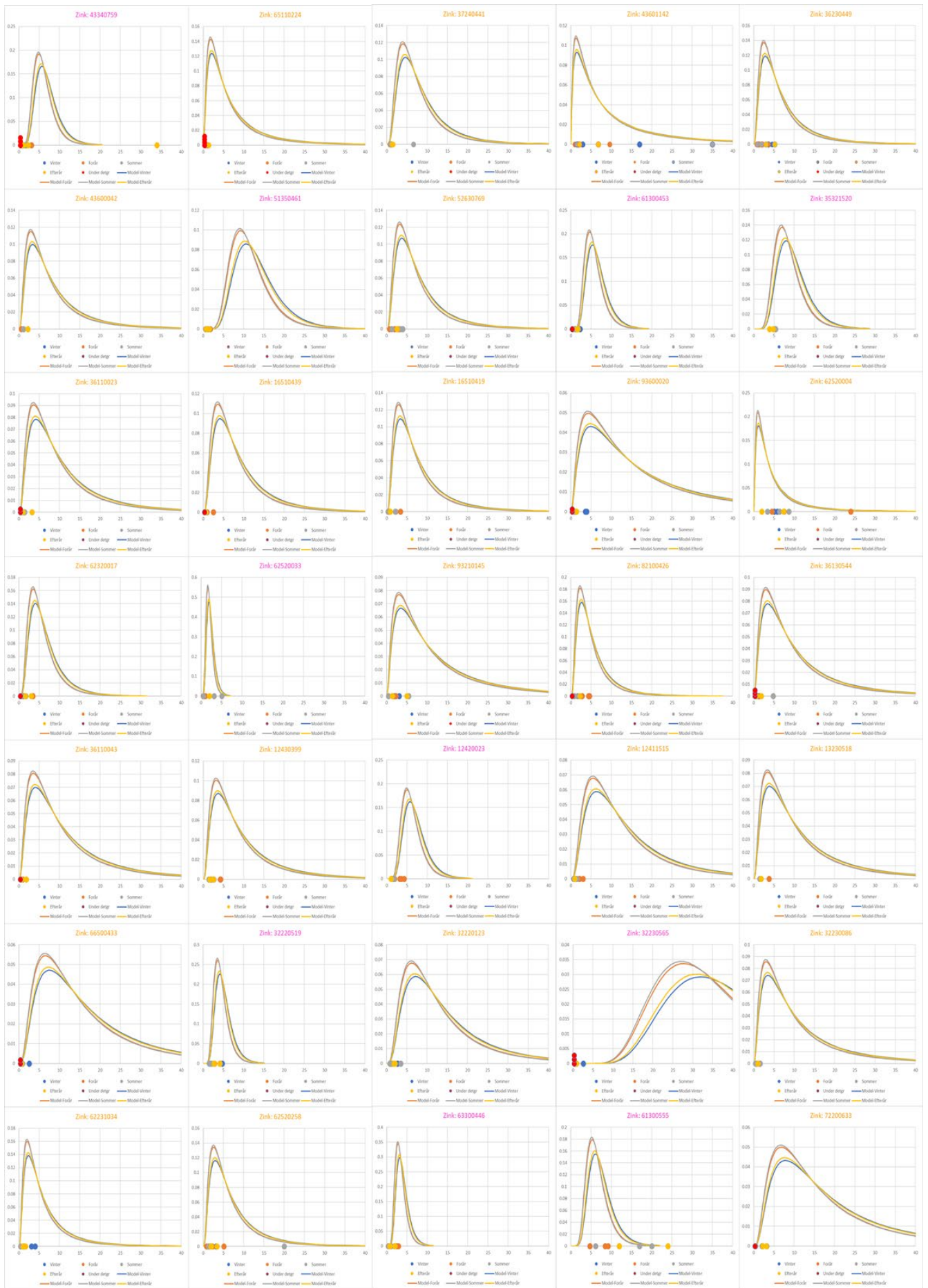


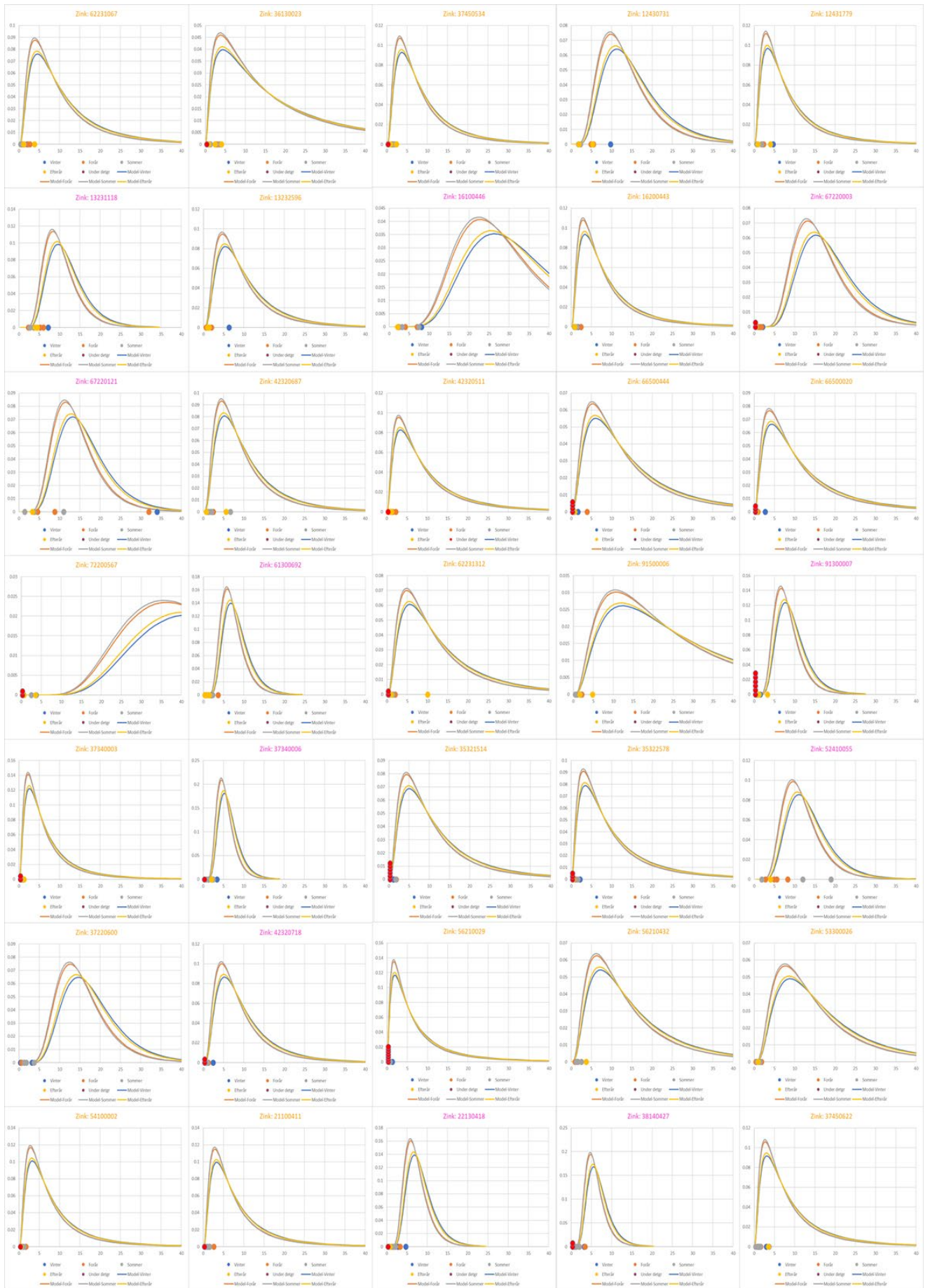


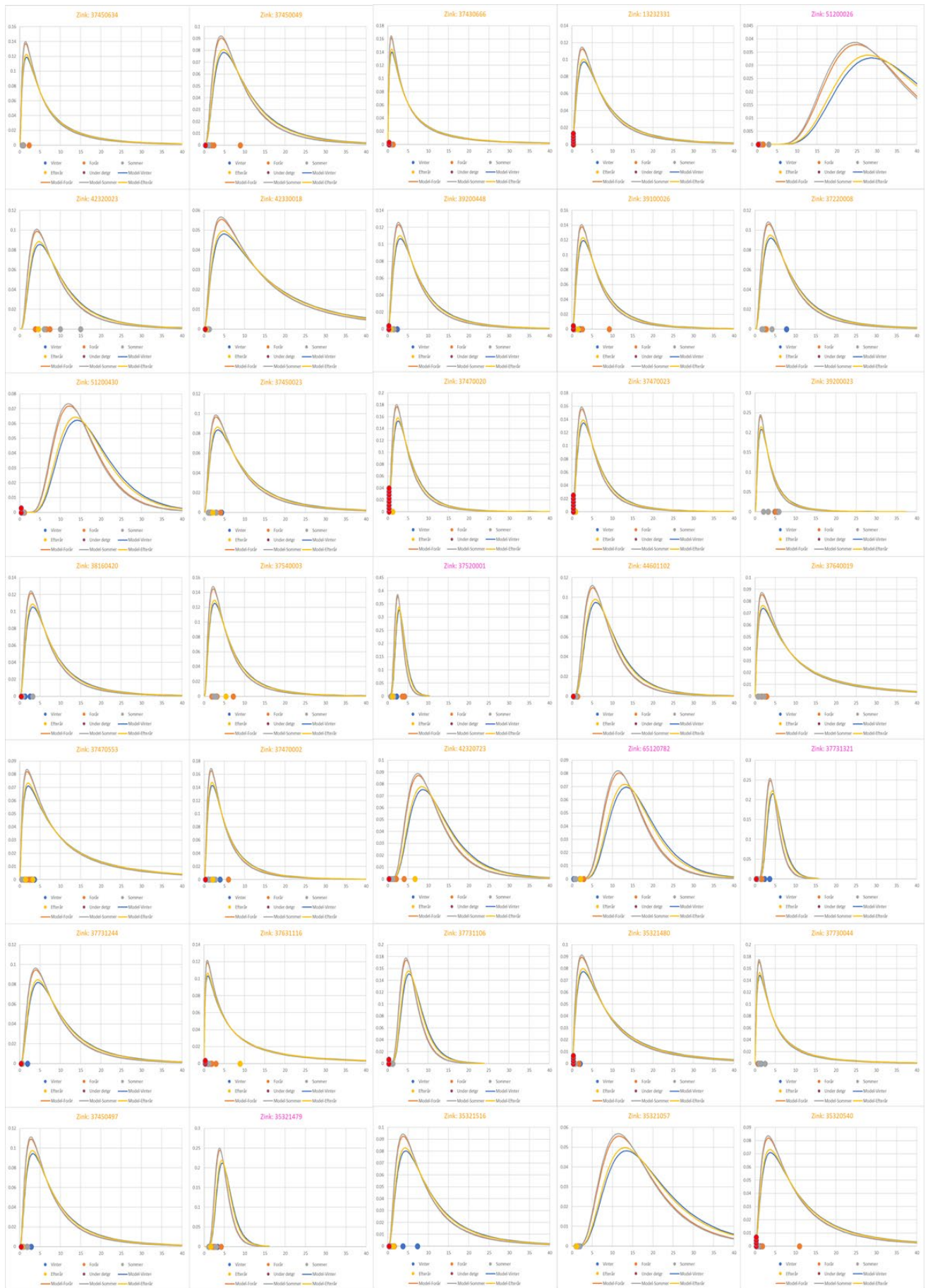


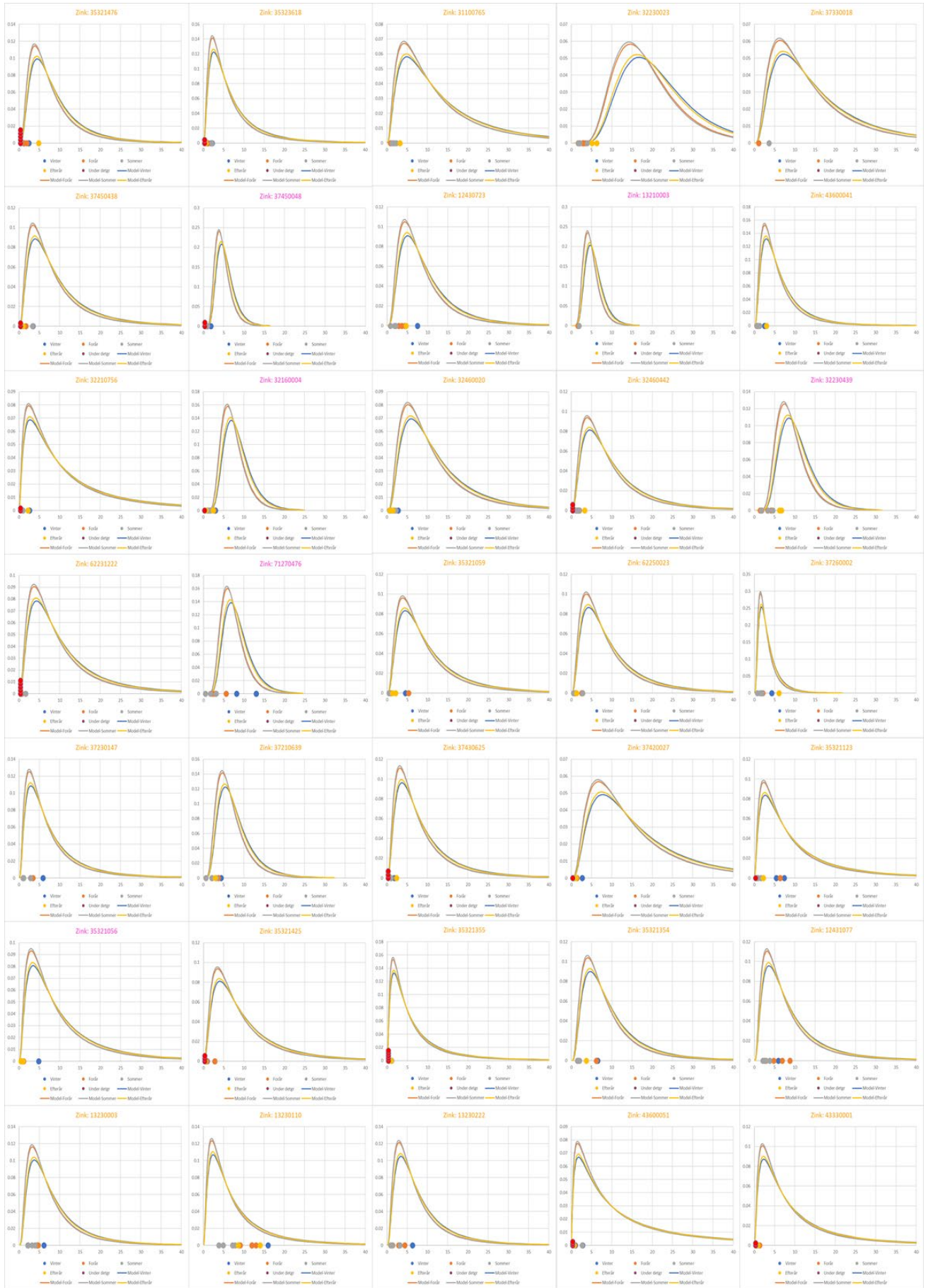


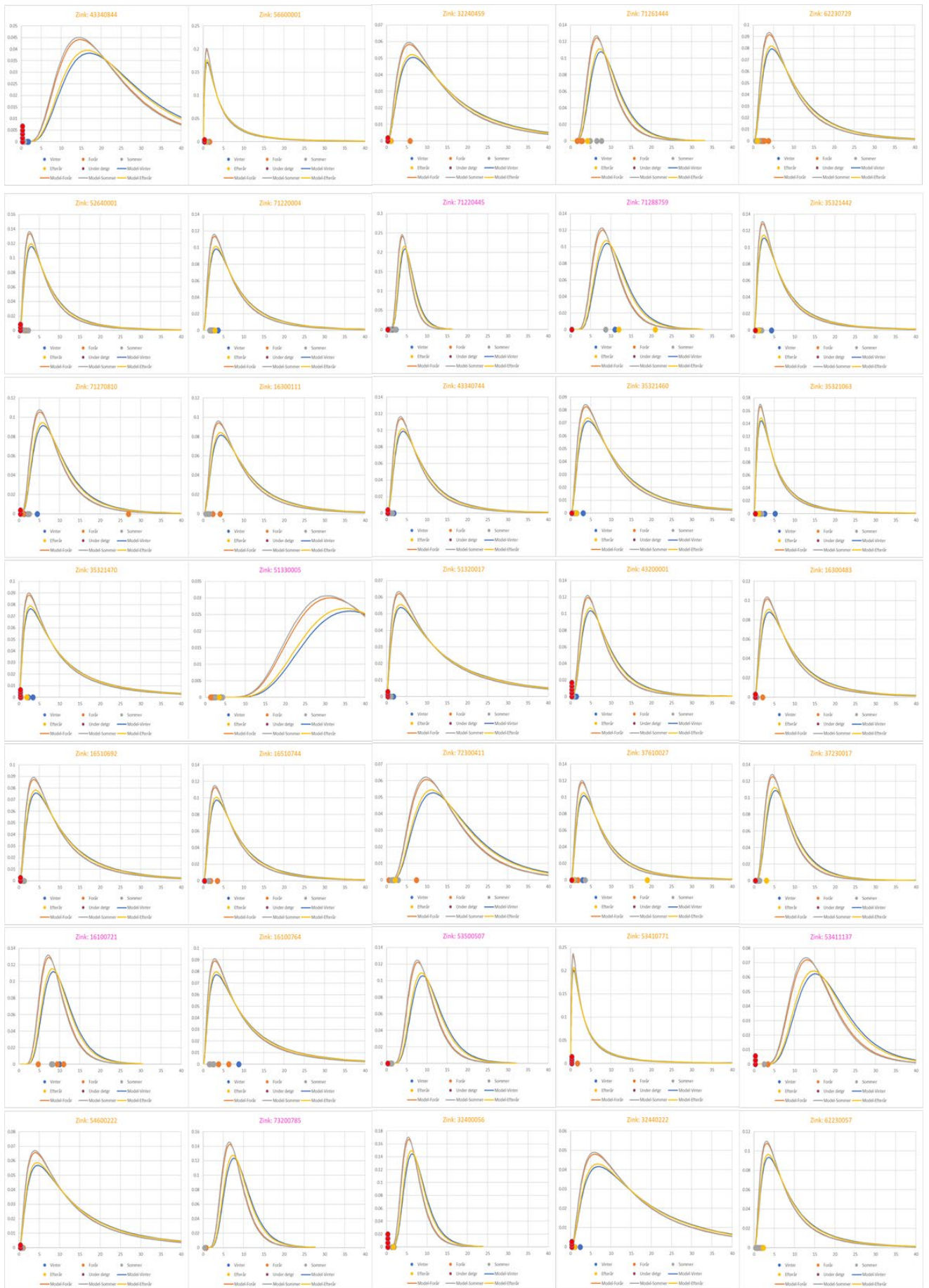


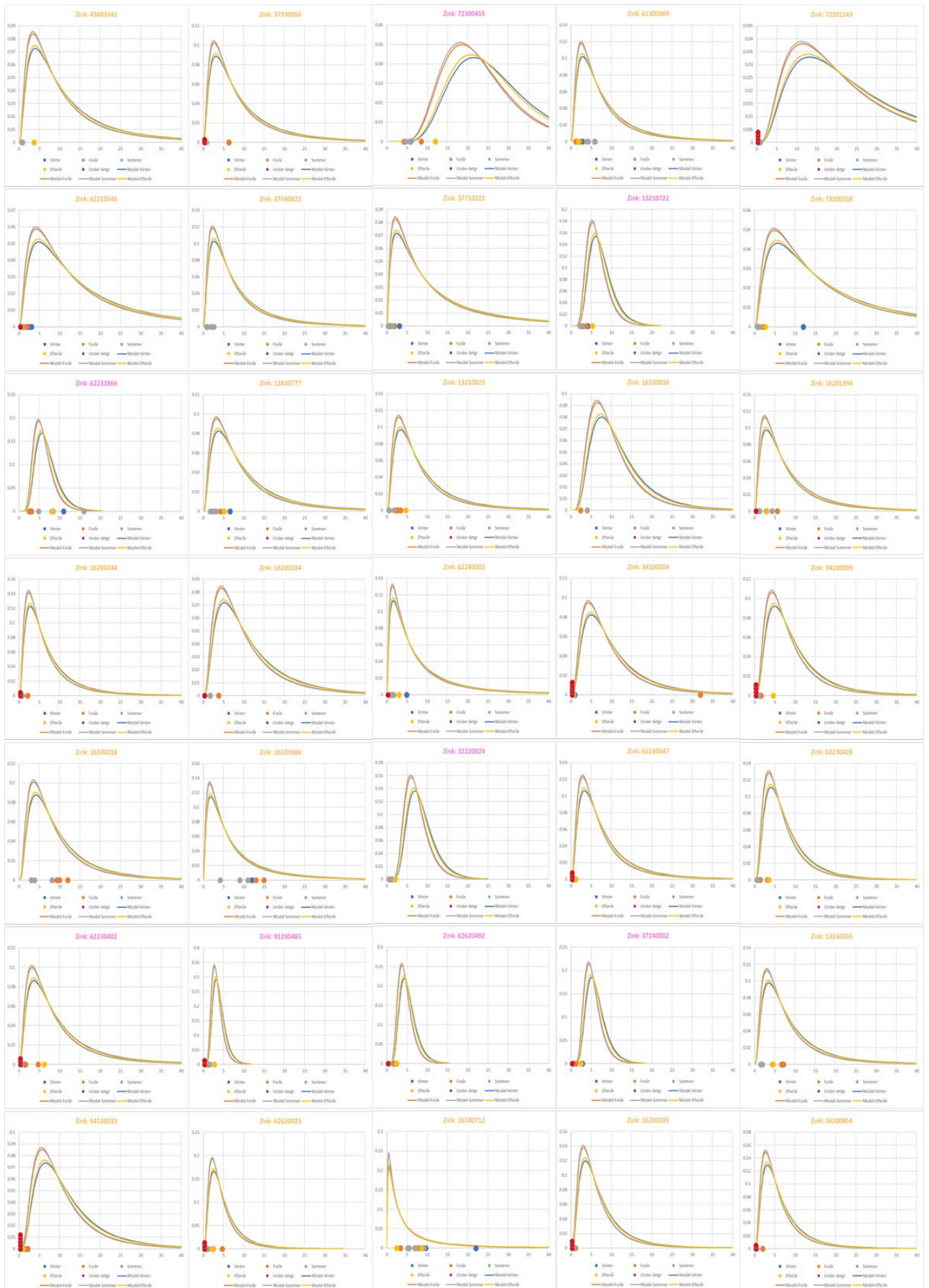


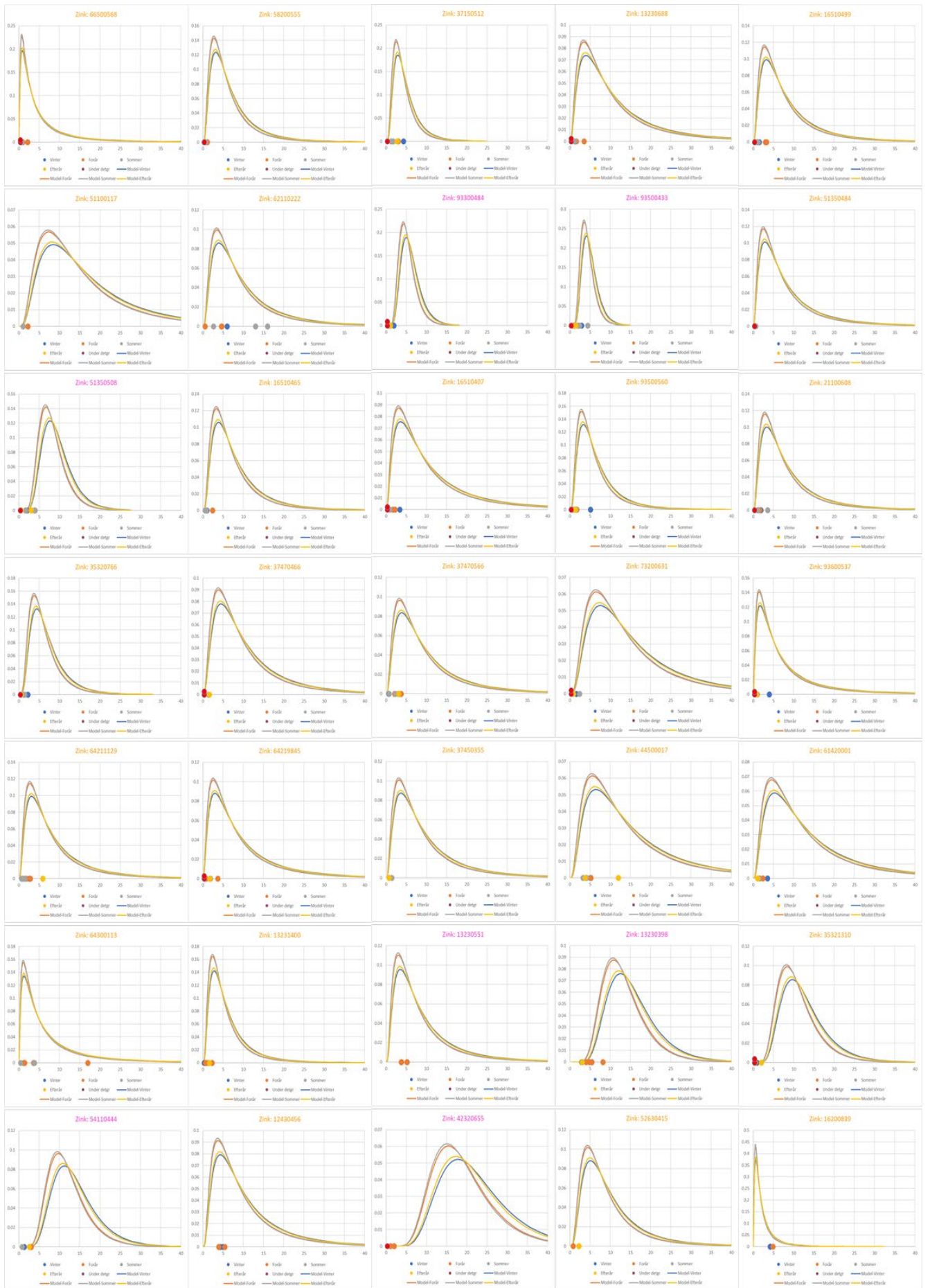


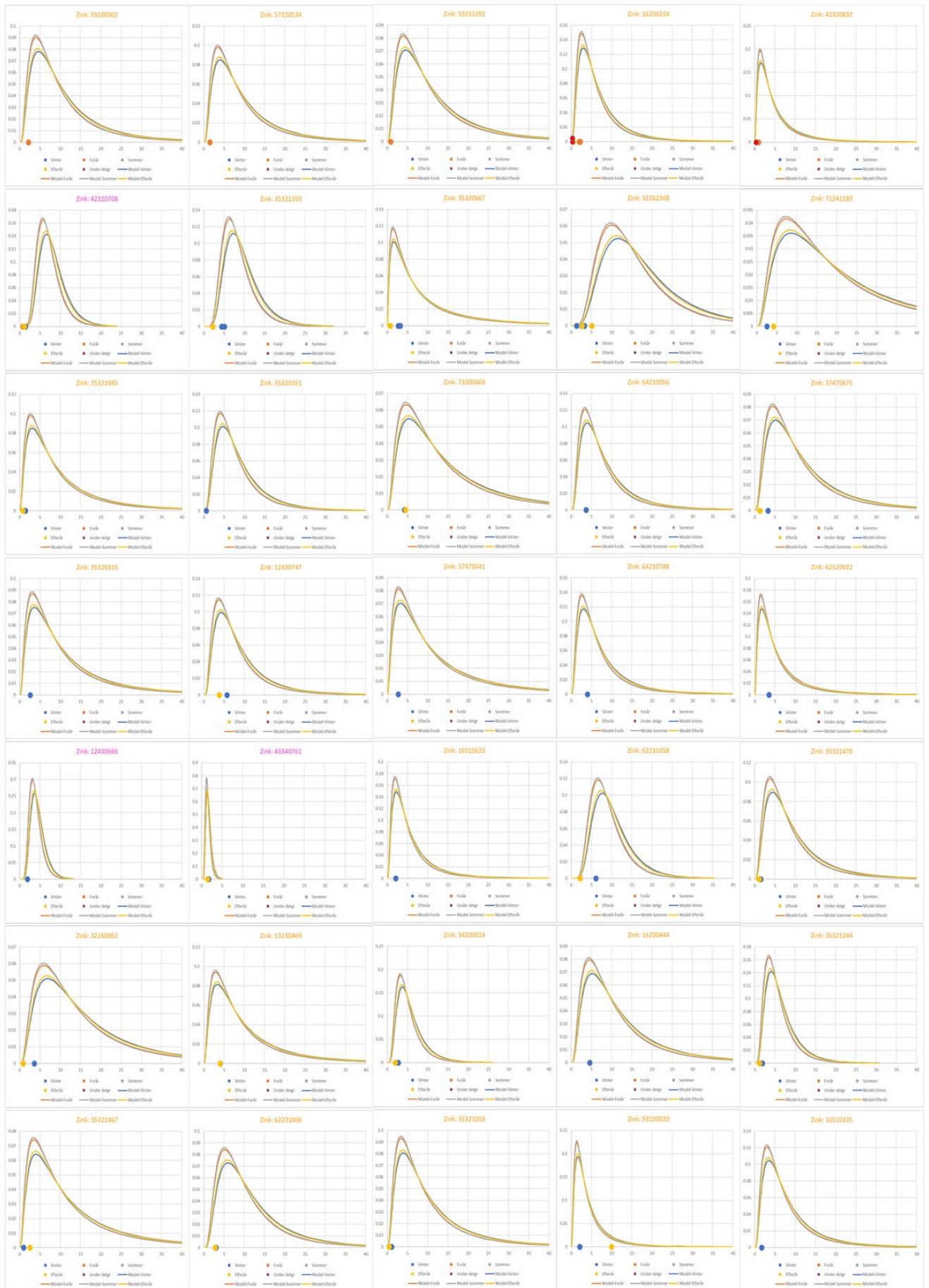


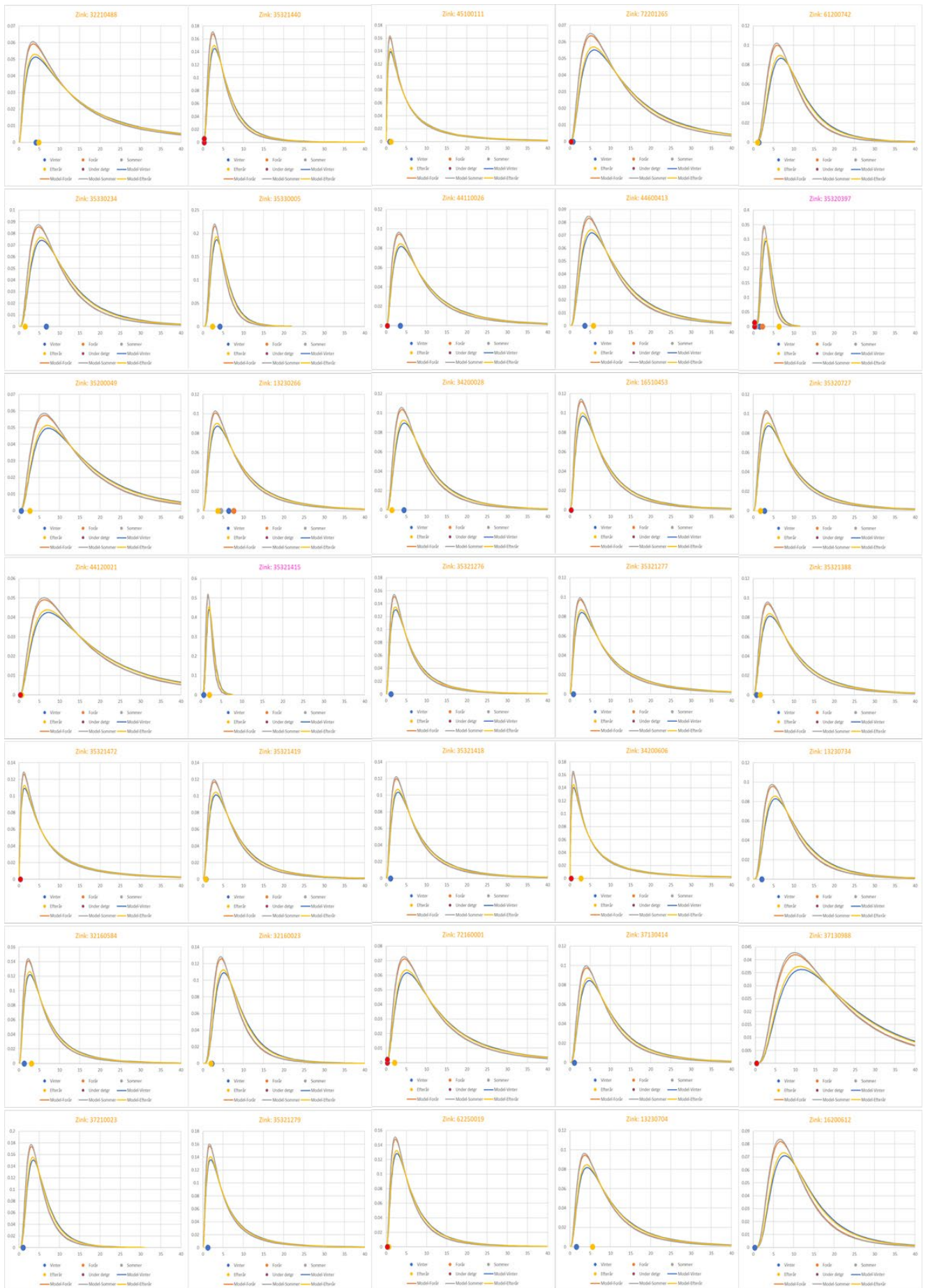


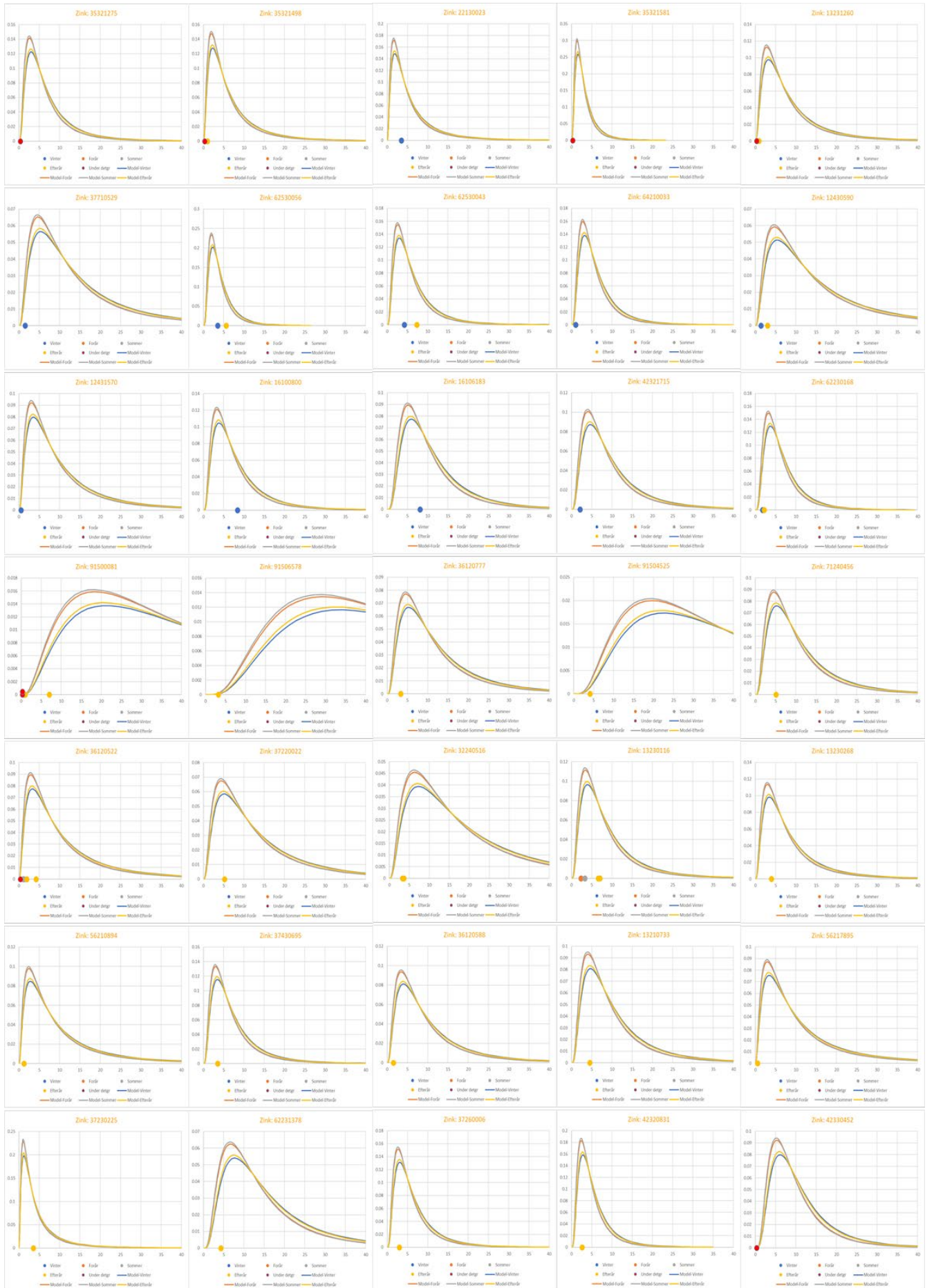


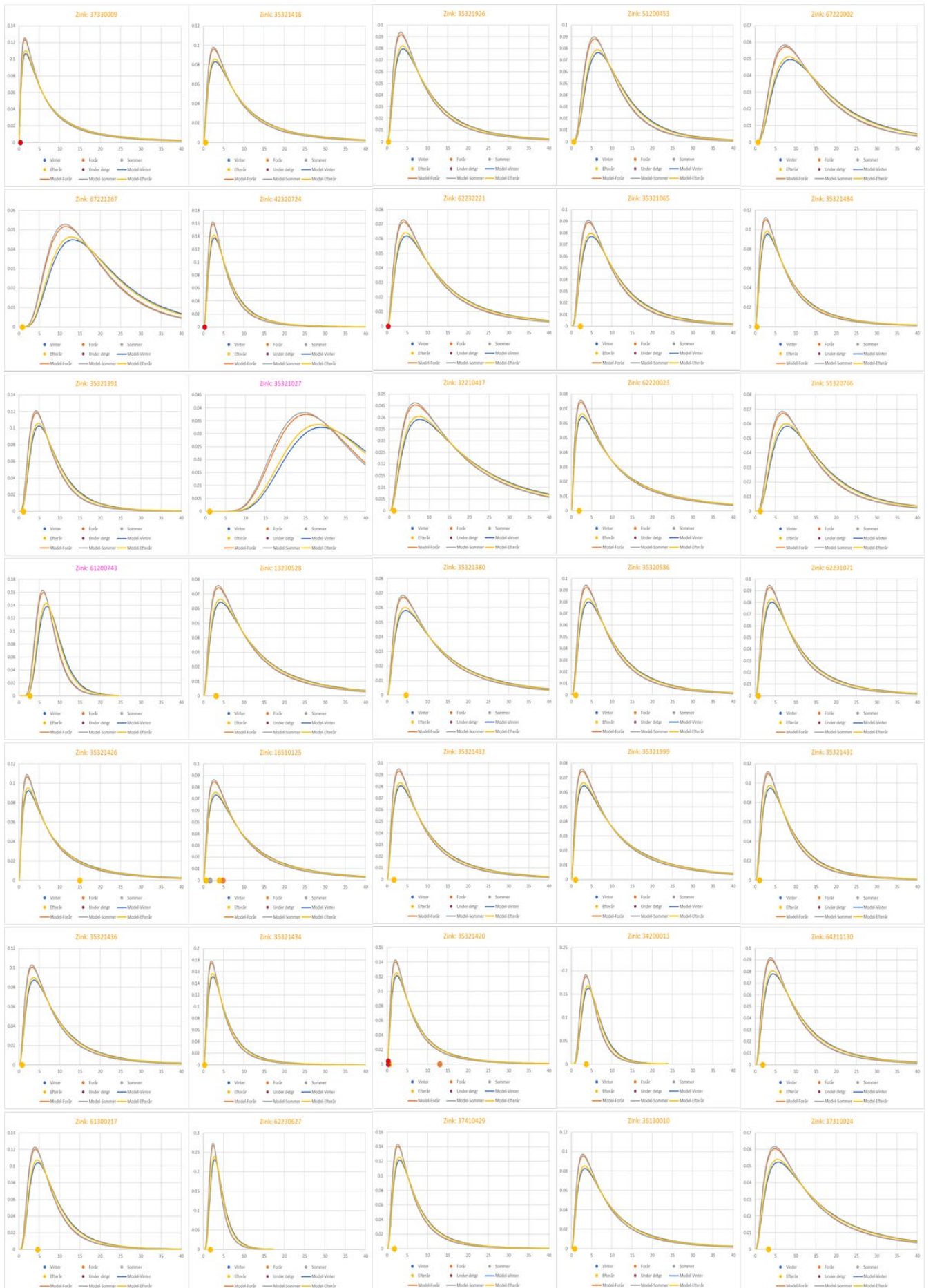












VIDEREUDVIKLING AF MODEL FOR METALLER I VANDLØBSVAND

MetalStat

Denne rapport indgår i et større arbejde med udvikling af statistiske modeller til estimering af koncentrationer af miljøfarlige forurenende stoffer i overfladevand for Miljøstyrelsen ved at videreudvikle på en pilotudgave af en statistisk model (MetalStat) for forekomsten af 5 metaller som opløst koncentration i vandløbsvand. Rapporten tager således udgangspunkt i afrapportering af MetalStats principper givet i Sørensen et al., 2022. Modellen er sat op, så metalkoncentrationen i vandløbet regnes for at være bestemt af det ID15 opland, som målestationen er placeret i, hvor et ID15 opland har en varierende udstrækning på typisk 10-15 km². Opgaven for 2022 er at få pilotudgaven af MetalStat valideret og modnet, så den kan indgå i Miljøstyrelsens forvaltning.